Beilstein-Institut The Chemical Theatre of Biological Systems, May 24th - 28th, 2004, Bozen, Italy

THE FINAL CURTAIN

KEVIN DAVIES

Editor-in-Chief, Bio-IT World, 1 Speen Street,
Framingham MA 01701, U.S.A.
E-Mail: Kevin_davies@bio-itworld.com

Received: 18th August 2004 / Published: 22nd July 2005

The 2004 Beilstein Bozen Workshop succeeded admirably in its stated goal of bringing together a diverse collection of interdisciplinary researchers, resulting in a stimulating and provocative conference. Despite their broad range of backgrounds - chemistry and biology, mathematics and biophysics, academia and big pharma - the speakers and attendees were united in their desire to understand more fully the dynamic interaction of genes, proteins and other cellular constituents, whether it be to model cellular pathways, or design and synthesize improved libraries of small-molecule drugs.

The preceding chapters in this monograph illustrate the remarkable quality of the presentations at the 2004 meeting. In this concluding chapter, I will simply aim to put some of these advances into a broader context.

THE PROMISE OF CHEMICAL GENOMCIS

In April 2003, the NIH marked the successful conclusion of the Human Genome Project - a massive international endeavour that was completed ahead of schedule, and under budget [1]. The full sequence is publicly available via several web sites, such as the Golden Path at UC Santa Cruz (genome.ucsc.edu). By May 2004, the month of the symposium, the sequence of approximately half of the human chromosomes had been published, along with the complete genomes of some 200 species, including mouse, rat, as well as preliminary drafts of the dog and chimpanzee genomes. Shortly after the completion of the human genome sequence, Francis Collins and his colleagues published a manifesto for the National Human Genome Research Institute, laying out a dozen or more key strategic goals over the next several years [2].

Interestingly, one of those goals encouraged academic groups to play a bigger role in drug discovery, using the tools of chemical genomics. Collins *et al.* wrote:

> "[Gleevec] offers promise that therapies based on genomic information will be particularly effective… A promising example of the gene-based approach to therapeutics is 'chemical genomics.' Providing such access more broadly… could lead to the discovery of a host of probes for biological pathways… Also needed are more powerful technologies for generating deep molecular libraries… A centralized database of screening results should lead to further important biological insights… Generating molecular probes for exploring the basic biology of health and disease in academic laboratories would not supplant the major role of biopharmaceutical companies in drug development, but could contribute to the start of the pipeline." [2].

Of course, academia has a rich tradition in drug development. Erythropoietin, tamoxifen, Taxol, AZT and many other important drugs were either purified or extracted by academic researchers, or required the tireless lobbying of physicians in order to gain the attention of drug companies [3]. The NIH has launched a Chemical Genomics Center, the flagstone of a ten-center nationwide Molecular Libraries Screening Centers Network.

The importance of chemical biology and improved synthetic techniques to revitalize the drug discovery process (and basic research) was one of the central themes of the symposium. Steve Ley reminded the audience that there is, by some estimates, a universe of 1060 small molecules. The Chemical Abstracts Service registry lists more than 23 million organic and inorganic compounds. And the famous Beilstein database, a compendium of organic chemicals going back to 1771, includes records of some 9.3 million compounds.

By contrast, the number of therapeutic targets is remarkably small. According to several investigators, the total number of targets in the 'druggable genome' amounts only to about 3,000 - that is, the number of genes in the human genome that encode biochemically tractable targets of medical/disease relevance [4](Fig. 1).

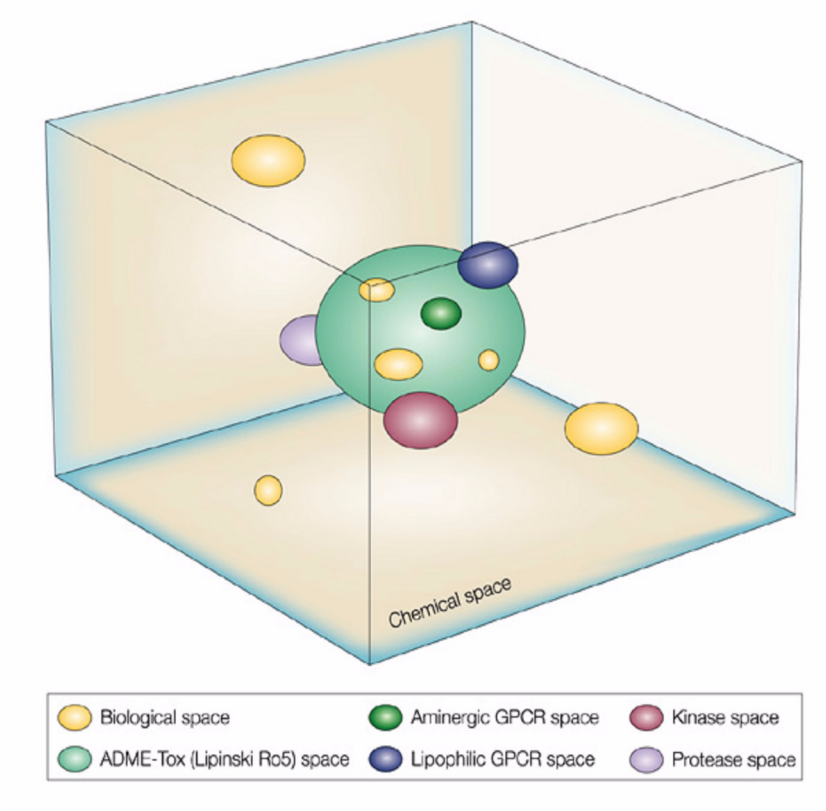

Figure 1. Chemical space and the druggable genome. The box depicts the complete set of all possible molecular structures, or chemical space - 1060 or more. The coloured spheres represent subsets of potentially therapeutic molecules. (Published with permission by Nature Publishing Group. Peti-Zeman, S. (2004) Exploring biological space. *Nature Reviews Drug Discovery*, Horizon Symposia; May).

Given the vast excess of small molecules compared to therapeutic targets, why has the production of new medical entities declined so dramatically in recent years? There are many factors (see below), but part of the problem is at the lead identification stage. Many mediocre small-molecule libraries have been created during the past 10-20 years by decorating basic core structures, rather than building diversity into the basic core structure. (Steve Ley's description of this depressing trend was "absolutely pathetic.") Ley's group takes a target-oriented synthesis approach to facilitate the synthesis of important compounds (Fig. 2).

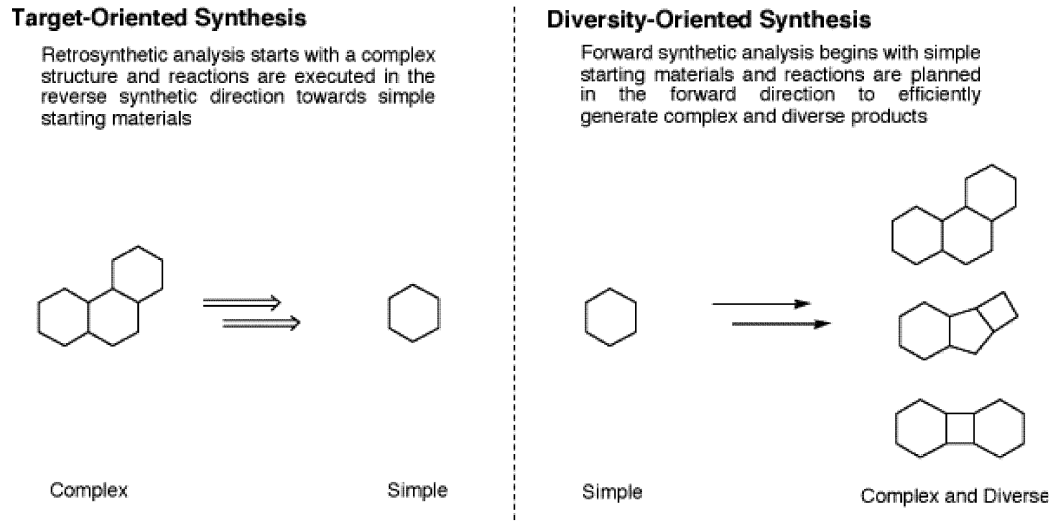

Figure 2. Target-oriented synthesis and diversity-oriented synthesis. (Published with permission from Cell Press. Burk, M.D., Lalic, G. (2002) *Chemistry & Biology* **9**:535-541).

A similar theme was sounded by Keith Russell (AstraZeneca). It is not uncommon for drug screens employing in-house or commercial libraries of a million compounds or more to fail, at least initially, because of lack of diversity, quality and stability of the molecules. Both presenters underscored the importance of new software and robotic devices to improve the synthesis of new compound libraries.

The process of diversity-oriented synthesis (Fig. 2), championed by chemists such as Harvard's Stuart Schreiber, shows how chemical diversity can be injected into libraries using just a handful of combinatorial steps [5]. Such methods not only bode well for lead identification in the future, but also provide an arsenal of promising new tools for studies of chemical genetics, whether it be the incorporation of novel amino acids into proteins, the combinatorial synthesis of novel antibiotics, or the development of *in vivo* biological assays (Fig. 3) [6]. The explosion of public funding in this area, and new database resources such as PubChem and Chembank, auger well for the future development of this field.

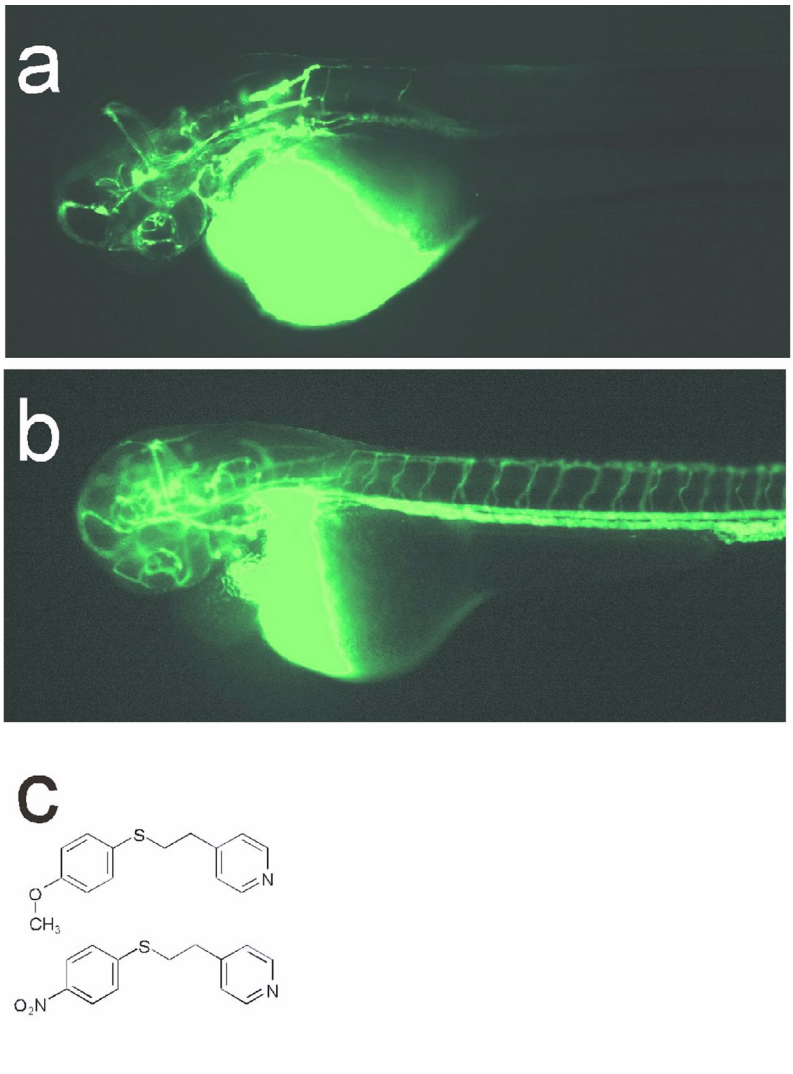

Figure 3. Chemical rescue of zebrafish genetic cardiovascular defect. Microangiograms of a gridlock zebrafish embryo before **(a)** and after **(b)** treatment with the small molecule GS4021. **(c)** Shows the chemical structures of two small molecules gridlock suppressors identified by *in vivo* screening, with GS4021 on the top. (Published with permission from Nature Publishing Group. Peterson, R.T. *et al.* (2004) *Nature Biotechnology* **22**:595-599).

Several speakers addressed the importance of virtual screening and biosimulation for modelling drug-protein interactions and 'reducing the haystack' of potential drug candidates, and the development of new software algorithms and databases to facilitate research advancement. Sometimes, however, the answer is to make the haystack bigger. As described in the 2002 Beilstein workshop report, Graham Richards and his colleagues' distributed computing project produced a virtual library of 3.5 billion molecules, by concocting 100 modifications for each of 35 million compounds, with considerable success [7].

Taken together, these initiatives will help investigators tackle one of the most challenging problems faced in medicine today - the crippling cost and time required for drug discovery.

THE COST OF DRUG DISCOVERY

The importance of boosting the quality and speed of discovery of new drugs cannot be overestimated. The net cost of drug development is put at $800 million, according to an oft-quoted 2001 study from the Tufts Center for Drug Development. (This includes the 'cost of capital,' the result of treating R&D costs as an investment rather than an expense.)

In his book *The $800 Million Pill*, author Merrill Goozner points out that this $800 million estimate also includes the cost of developing me-too drugs; the cost of developing stereoisomers of approved drugs to extend a drug patent; and the practice of conducting clinical trials not for the purpose of seeking FDA approval, but simply to gather data to persuade doctors to prescribe a particular brand of medication. "In short," writes Goozner, "if the industry funded academic economists at Tufts had factored out the half of industry research that is more properly categorized as corporate waste, their number would have been similar to that of the Global Alliance." [3].

A few years ago, however, the Global Alliance for Tuberculosis Drug Development convened a panel of industry experts to calculate the cost of developing a new drug against tuberculosis. The report, published in October 2001 [8], concluded that the total costs to discover and develop a new anti-TB drug - including the costs of failures -ranges from $115 million to $240 million. These costs could be divided into $40-125 million for discovery efforts, and $76-115 for preclinical development through phase III trials.

Even drugs that are approved in record time, such as Novartis' Gleevec, which was approved by the FDA in 2001 in a mere 10 weeks, took a decade to reach the market. The molecule formerly known as STI-571 was originally synthesized by Jurg Zimmerman in 1992. Much of the increase in drug development time and cost during the past two decades is in the area of clinical trials - difficulty recruiting patients, meeting FDA guidelines, and the more complex nature of the diseases being treated all conspire to lengthen the process and raise the cost. Clearly, greater insight earlier in the process as to the likely efficacy and toxicity of potential new drugs could reap profound dividends later on.

The recent saga of Iressa is instructive. The lung cancer drug was approved after a costly trial involving some 12,000 patients, which caused the company something of a dilemma, as indicated by Sir Tom McKillop, CEO of AstraZeneca: "There are 50% of patients with lung cancer who do not get any benefit from Iressa … but then there is a small group, maybe 10-20% of patients [who] get an almost a miraculous response, and their whole life can be transformed and extended for years." [9].

The reason for these contrasting results has now been clarified. In a pair of papers published in 2004, researchers in Boston found that patients who respond to Iressa have acquired mutations in the target protein that increase the binding of the drug. A new diagnostic test is currently in preparation [10,11].

The examples of Gleevec and Iressa, two of the most high-profile cancer drugs approved in recent years, demonstrate how critical it is to marry rational drug design with improved understanding of the target patient population. Most, perhaps all, drugs do not produce the same results in all patients. Using pharmacogenomic and other tools to stratify the patient population so that the positive effects of a drug are not shrouded by non-responders or worse, adverse events, will pay dividends in the long run.

IN CONCLUSION

The convergence of technologies and disciplines showcased at the 2004 Beilstein symposium invited several speakers to talk about the new holy grail of modern biology - systems biology. In 2003, researchers at Curagen published a proteome map of some 7,000 proteins in *Drosophila* that they dubbed "The dawn of systems biology" [12]. Be that as it may, the explosion of interest in integrative biology - the interplay of genome-wide expression, proteomic, metabolomic and *in silico* studies - lends credence to the view that we are on the verge of describing in more holistic terms the functioning of cellular systems.

Several speakers at the symposium discussed areas in which this could be achieved, including work at Hoffman-LaRoche, in collaboration with Entelos Inc., to produce 'virtual patients' for diseases such as asthma, obesity and arthritis [13]. In a related vein, researchers at Gene Network Sciences, in Ithaca New York, are building a computational model of a colon cancer cell.

It is to be hoped that the rich potential inherent in these models acquires sufficient validation from the academic community that they might form an integral part of the discussion at the 2006 Beilstein Bozen Workshop.

REFERENCES

[1] Davies, K. (2001) *Cracking the Genome*. The Free Press, New York.

[2] Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S. (2003) A vision for the future of genomics research. *Nature* **422**:835-847.

[3] Goozner, M. (2004) *The $800 Million Pill*. UC Berkeley Press, Berkeley.

[4] Hopkins, A.L., Groom, C.R. (2002) The druggable genome. *Nature Rev. Drug Discov.* **1**:727-730.

[5] Burke, M.D., Berger, E.M., Schreiber, S.L. (2003) Generating diverse skeletons of small molecules combinatorially. *Science* **302**:613-618.

[6] Peterson, R.T., Shaw, S.Y., Peterson, T.A. *et al.* (2004) Chemical suppression of a genetic mutation in a zebrafish model of aortic coarctation. *Nature Biotech.* **22**:595-599.

[7] Richards, W.G. (2003) Pattern recognition and distributed computing in drug design. In: *Molecular Informatics: Confronting Complexity.* (Hicks, M. G., Kettner, C. Eds), Proceedings of the Beilstein-Institut Workshop, May 13-16, 2002, Bozen, Italy.

[8] Global Alliance for Tuberculosis. The economics of TB drug development. (2001). http://www.tballiance.org/

[9] Humphreys, A. (2003) Ability to manage. *Med. Ad. News* **22**:36-49.

[10] Lynch, T.J., Bell, D.W., Sordella, R. *et al.* (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *New Engl. J. Med.* **350**:2129-2139.

[11] Paez, J.G., Janne, P.A., Lee, J.C. *et al.* (2004) EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* **304**:1497-1500.

[13] Giot, L., Bader, J.S., Brouwer, C. *et al.* (2003) A protein interactive map of *Drosophila melanogaster. Science* 302:1727-1736.

[14] Uehling, M.D. (2003) Model patient. *Bio-IT World* **2**:30-36.