Industrialized molecular biology, information biotechnology, and the blockbuster drug model – alive and well at age 50

Jeffrey Augen
IBM Life Sciences
Route 100
Somers
NY 10589
USA
tel: +1 914 766 3657
fax: +1 914 766 8370

▼ The pharmaceutical industry is going through a period of dramatic change: we are witnessing the emergence of a new class of diagnostic tools based on complex patterns of gene expression; personalized therapeutics are appearing in the form of more precise patient stratifications; patent portfolios are shrinking along with the number of new FDA submissions [1]; and despite an 800% growth in R&D expenditures since the early 1980s, output of the drug discovery process has remained fairly constant [2]. Although these trends might seem troublesome at first, it is impossible to fully appreciate the therapeutic landscape outside the context of the dramatic advances taking place in gene sequencing, mRNA profiling, protein structure prediction and all supporting information technologies. Most significant among these has been the industrialization of molecular biology – a trend that will ultimately culminate in very rapid, inexpensive and freely available gene sequencing solutions for research and medicine.

From blockbuster to personalized medicine

The ultimate endpoint of this technological march is a molecular-level understanding of health and disease. This knowledge is at the core of a rapidly growing discipline known as 'personalized medicine'. The spectrum of business-model choices for drug discovery in this space range from pure 'blockbuster' at one extreme, to completely personalized at the other. At the blockbuster end of the spectrum, the development model is completely dedicated to strategies for treating large populations, whereas the personalized side implies customized treatments for very small populations and, potentially, single individuals.

Proponents of the personalized model believe that therapeutics will be customized according to complex patterns of gene expression and fine nuances of intermediary metabolism at the personal level. This model goes far beyond the prevention of adverse drug reactions; it includes specific selection and dosing based on a patient-specific understanding of absorption, distribution, metabolism and excretion (ADME) rates, and a specific understanding of potential side effects for a given individual. Importantly, however, this model does not imply the creation of custom compounds for each patient. Rather, the implication is that individuals will be treated with custom cocktails of existing drugs. Viewed in this way, the barriers to personalized medicine seem much lower. In fact, some elements of the concept present unique business opportunities for pharmaceutical, insurance and diagnostic companies, as well as for physicians.

The potential for better diagnostics

Companies will capitalize on the opportunity by marketing mRNA and protein expression profiling kits for specific diseases. For example, there are 7 million individuals in the USA suffering from chronic inflammation [3]. These individuals can be grouped into several different disease categories based on up- and down-regulation patterns of dozens of specific genes [4,5]. Different anti-inflammatory drugs are effective for treating the different profiles of chronic inflammation, and in some cases, the wrong drug can make the symptoms worse. The opportunity for diagnostic companies is even larger than one would expect because doctors must take advantage of these tools as they become available, or risk legal action from patients who received the wrong drug. Any legal analysis is beyond the scope of this discussion, but it is clear that

1359-6446/02/$ – see front matter PII: S1359-6446(02)02484-4

the responsibility for taking advantage of progressive and accurate diagnostic tools will lie with the physician. Furthermore, abdicating that responsibility could have legal ramifications – an additional force that will drive new business for the diagnostic companies. Similarly, Pharmaceutical companies will reap the benefits from more precise drug targeting and improved selection of patients for clinical trials. We already have examples of drugs that are genetically targeted, such as Herceptin for treating breast cancer, and a variety of reverse transcriptase inhibitors for treating AIDS.

In addition, the progress that has been made in dissecting the molecular biology of disease now means that compounds that failed clinical trial because patient subpopulations could not be characterized at a sufficiently detailed level can now be 'rescued' and re-evaluated. One recent and notable example is Thalidomide, it was abandoned in the early 1960s because of its mutagenic effects, but has now been resurrected as an anti-cancer drug. Finally, medical insurance companies could lower their costs by reducing the level of uncertainty associated with selecting pharmaceuticals. Ineffective or dangerous treatments increase costs, moerover, some estimates show that as many as 1/3 of all patients are currently treated with drugs that are genetically or biochemically inappropriate [6]. Clearly, eliminating inappropriate treatments would represent direct, bottom line earnings growth for health insurance companies.

Information Technology is at the core of this discussion, and it can be argued that one of the most important tools for modern day drug discovery is the relational database; equally as important as combinatorial chemistry and gene sequencing. Furthermore, data mining and pattern discovery are core technologies that will be required to drive the next generation of blockbuster drug development. As a result, it is highly probable that information technology platforms for drug discovery will become key differentiators for the pharmaceutical industry over the next several years. The blockbuster model might be changing, but it is certainly alive and well and will probably remain so.

Disease reclassification

One of the most substantial changes that will affect blockbuster drug development is related to disease classification. Today, diseases are identified by phenotype, which typically includes clinically identifiable symptoms and results from specific chemical and physical tests. For example, diabetes is identified by abnormalities in sugar metabolism, coronary artery disease involves restricted blood flow through clogged arteries, and most cancers are identified by tumor classification, often using a microscope.

Over the next several years a huge amount of data will be gathered through expression-array data analysis. These data will be used to reclassify illnesses based on the specific patterns of up- and down-regulation of genes. Large databases will be populated with these patterns, which will become the subject of data mining and pattern discovery experiments. Eventually, the patterns will be sub-classified and separated into distinct groups for further analysis. Once separated into groups, individual gene products that emerge as most significant within the patterns will be studied at the biochemical level, with the ultimate goal of finding new drug targets. Once validated, these targets will become the basis for new lead compound optimization experiments. Areas of differentiation will include algorithms for pattern discovery, infrastructure for linking internal and external (outside the firewall) databases and software for molecular modeling.

Thus, in the future, disease names and classifications will change dramatically. Even more important will be the realization that, at the molecular level, many apparently dissimilar diseases are actually biochemically related. Such relatedness has already been seen. For example, statin drugs, which were designed for treating high cholesterol, and COX-2 inhibitors, which were designed to treat inflammation, are both effective treatments for certain types of lung cancer because they also inhibit angiogenesis.

In addition to these surprising relationships, new discoveries have been made regarding the biochemical nature of many diseases that, previously, were thought to be fully understood. Often, these discoveries arise from statistical analysis of large amounts of clinical and metabolic data. Again, a central role for information technology is apparent because the initial discovery will probably be made entirely *in silico*. One outstanding example is coronary artery disease; our understanding has been revolutionized during the past few years as a result of statistical studies across large patient populations. One shocking conclusion is that the most important risk factor for coronary artery disease is not elevated cholesterol but elevated levels of C-reactive protein (CRP) [7]. High circulating levels of CRP appears to cause plaque rupture and blood clot formation, which cause heart attacks and strokes [8,9]. Furthermore, individuals with high levels of CRP are at risk even if they have low cholesterol levels [10]. However, statin drugs, which were originally designed to lower cholesterol levels by inhibiting the enzyme HMG-CoA reductase, are also very effective at reducing high levels of CRP. Therefore, statin drugs are helpful to individuals with normal cholesterol levels if they also exhibit elevated CRP [11]. As many disease states raise the level of CRP – inflammation and diabetes to name just two – statin drugs might have a much broader appeal than previously anticipated. They attest to the power of the blockbuster model.

Two interesting conclusions emerge from this discussion: The blockbuster drug model is alive and well; and the marriage of information technology and biochemistry is an essential for the next generation of drug discovery.

Concluding remarks

The migration from drugs that target large populations, to drugs that target individuals will continue to be driven by the industrialization of molecular biology and increasingly rapid target identification processes. For each illness and treatment, the migration will terminate at a specific point on the continuum. The outcomes range from completely customized therapeutics (custom drugs and mixtures for each individual) to drugs that are targeted more precisely to smaller populations (but not individuals). Migration has already begun, and there

is no obvious force to slow the progression. By contrast, patient demand, economic forces and the desire of physicians to provide more precise and effective medical treatments will probably act in concert to hasten the rate of migration.

This migration is also driving significant change in the pharmaceutical and diagnostic industries. Most significant among these changes, is the rapid adoption of cutting edge information technologies. The percentage of revenue spent on IT is steadily increasing at almost every major pharmaceutical company and some of the smaller pharmaceutical and biotechnology companies already have the look and feel of an IT company. Complete platforms include infrastructure for heterogeneous database management, data mining and pattern discovery algorithms, large-scale database infrastructure and applications for *in silico* molecular docking.

Thus, there are two complimentary, but separate, potential areas for infrastructure growth. The first is based on high performance integer arithmetic and facilitates pattern matching of gene sequences. Such infrastructure usually includes Linux clusters, a perfect design for rapid and efficient division of sequence homology problems. The second is based on high performance, floating-point intensive computation (i.e. large powerful computers). Such infrastructure requires very large machines in the teraflop or greater range, and it is the driver for *in silico* molecular modeling, protein folding, chemical kinetics, and systems modeling. Finally, raw compute power is rapidly becoming a differentiator in the drug discovery business. Grid computing, based on a shared virtual system that can provide this level of compute horsepower, is becoming an important dynamic in the high performance computing marketplace. In the future, it will be very difficult for pharmaceutical companies to compete effectively without access to such a system. The alternative would be to acquire and maintain multiple supercomputer sites, which is not a cost-effective solution, regardless of the size of the company.

So, personalized medicine is not a challenge, but an opportunity. The winners will be companies that can leverage this opportunity, and the differentiator will almost certainly be efficient use of information technology.

References

1 Office of Planning (HFP-1) US food and Drug Administration (2001) Performance report to Congress for the prescription Drug User Free Act of 1992. http://www.fda.gov

2 Tufts Center for the Study of Drug Development. Outlook 2001. http://www.csdd.tufts.edu/InfoServices/ImpactReports.asp

3 Blackwell, D.L. *et al.* (2002) Summary health statistics for US adults: national health interview survey, 1997. *Vital Health Stat*. 10 (205) http://www.cdc.gov

4 Joyce, D.E. *et al*. (2001) Gene expression profile of antithrombotic protein c defines new mechanisms modulating inflammation and apoptosis. *J. Biol. Chem*. 276, 11199–11203

5 Chen, B.P. *et al*. (2001) DNA microarray analysis of gene expression in endothelial cells in response to 24-h shear stress. *Physiol. Genomics* 7, 55–63

6 Kohn, L.T. and Corrigan, J., eds (1999) To err is human: building a safer healthcare system, National Academy Press

7 Blake, G.J. and Ridker, P.M. (2001) Novel clinical markers of vascular wall inflammation. *Circ. Res.* 89, 763–771

8 Libby, P. (2001) Current concepts of the pathogenesis of the acute coronary syndromes. *Circulation* 104, 365–372

9 Freeman, W.E. Jr (2002) Risk factors versus inflammation in atherothrombotic disease. *Circulation* 106, discussion e31

10 Steinberg, D. (1997) Low density lipoprotein oxidation and its pathobiological significance. *J. Biol. Chem*. 272, 20963–20966

11 Albert, M. (2001) Effect of statin therapy on c-reactive protein levels: the pravastatin inflammation/CRP evaluation (PRINCE): a randomized trial and cohort study. *JAMA* 286, 64–70

Information Biotechnology: A supplement to Drug Discovery Today

How to cite an article in this supplement:

Huang, S. (2002) Rational drug discovery: what can we learn from regulatory networks? *Drug Discov. Today* 7 (Suppl.), S163–S169

How to cite the whole supplement:

Lawrence, R.N. and Owen, M. eds (2002) *Information Biotechnology: A Supplement to Drug Discovery Today* Vol. 7, No. 20, Elsevier Science

To purchase additional copies of this supplement, please contact:

Stacey Sheeky, Commercial Sales Department, Current Trends, Elsevier Science, 84 Theobald's Road, London, UK WC1X 8RR

tel: +44 20 7611 4449, fax: +44 20 7611 4463, e-mail: stacey.sheekey@bmn.com