Articles

Impact of Genomics and *in Silico* Related Technologies in the Drug Discovery Process

LÉAUTÉ, Jean-Baptiste *

Department of Valorization and Industrial Relationships, Paris 7 University, 2 Place Jussieu, 75005 Paris, France

In order to evaluate to what extent will genomics and *in silico* related technologies improve overall drug discovery process, we analyzed three studies comparing cost, time and attrition rate at each step of the drug discovery process, between standard pharmaceutical and genomics based approaches.

Keywords *in silico* drug discovery, bioinformatics, chemoinformatics, drug design, functional and structural genomics, virtual screening

Introduction

Today's pharmaceutical industry faces tremendous challenges. Intense product competition, patent expirations, reduced periods of exclusivity and price constraints are pressurizing pharmaceutical companies to lower costs, increase productivity and accelerate development. To meet investor expectations and achieve annual growth rates of 10%, each major pharmaceutical company must launch, on average, four new chemical entities (NCEs) per year, each with average annual sales of US $350 millions.

However, from 1996 to 2001, the industry launched, on average, less than one NCE per year per company. During the same period, pharmaceutical research and development (R&D) spending in the US increased by approximately 40%, whereas drug approvals declined by nearly 50%. Fig. 1 shows that average R&D expenditure as percent of sales has doubled since 1980, and is now 18% for US-based research companies.

And increase in R&D spending are not well correlated with approval of new products (Fig. 2), suggesting that simply increasing R&D spending does not appear to be the answer to the industry's pipeline challenges. On the other hand, the unexpected speed at which the human genome was sequenced, and the explosive development of related new technologies, such as: transcriptome, proteome, functional and structural genomics, pharmacogenomics, are felt to be the answer to the challenge.

The quantities and complexity of the data produced indicate that traditional methods and tools for accessing, analyzing and distributing these data are no longer viable. New computer information technology is used to handle and process all these data. Fortunately, informatics was fed by several coinciding factors: the ever-growing power of computers, more refined algorithms, platforms integration between data and technology, and the versatility of the Internet. Therefore, the bottleneck preventing the pharmaceutical industry from attaining enhanced levels of productivity includes the physical limitations required for wet-lab experiments. To accelerate and improve pharmaceutical drug

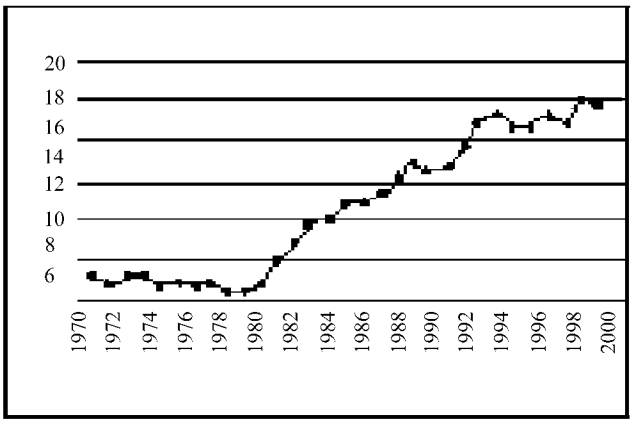

Fig. 1 Average R&D spending (US-based firms) as percent of sales (Source: Pharma Annual Survey 2000 Accenture analysis)

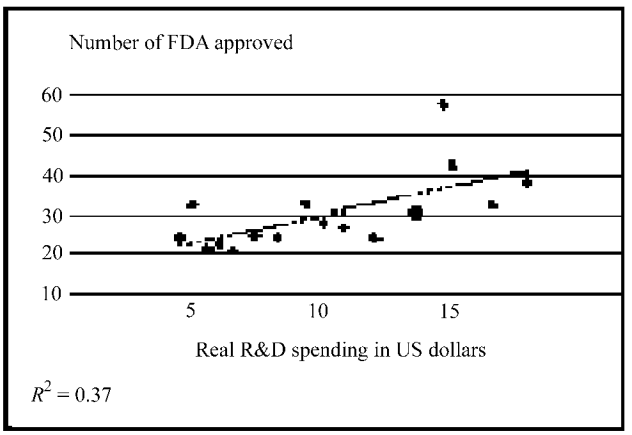

Fig. 2 Aggregate R&D spending vs. output (US-based firms) (Source: Pharma Annual Survey 2000 Accenture analysis)

* E-mail: leaute@paris7.jussieu.fr; Tel.: +33144278257; Fax: +33144278315
Received April 4, 2003; revised and accepted July 14, 2003.

discovery, wet-lab experimentation *in vivo* and *in vitro* has started to shift toward *in silico drug discovery* which analyzes and integrates both biological (bioinformatics) and chemical data (chemoinformatics).

The aim of this paper is to discuss three studies previously published, in order to evaluate to what extent will genomic and *in silico* related technologies, improve productivity overall, and what will its effects be, when applied at various points of the value chain.

Methodology

The Boston Consulting Group, Accenture Group and Mc-Kinsey Group have conducted an extensive program of interviews with top scientists and research executives, in several major pharmaceutical companies and academic institutions, in an effort to compile accurate figures for all the main activities in the R&D process and to pin point critical success factors in drug discovery due to the emerging technologies. The result is a robust bottom-up model of R&D, based on the time, cost, and likely success rate, for each step of the value chain in pre and post genomic era. All numbers cited in here are for a relevant drug, that is one to which the technology under discussion could be applied. For the average drug, peak annual sales of 500 million $ and eleven years to patent expiration were assumed.

Results and discussion

Foreseeing the deficit of numerous leads, drug companies implemented a number of strategies in the late 1980s. Combinatorial chemistry was used to generate larger libraries of testing compounds, and high-throughput technology, including increasing miniaturization and automation, was deployed to screen these libraries more rapidly. Despite a tremendous advance in all aspects of the screening process, the improvements did not bring about the expected rise in productivity and the industry's drug pipelines still look decidedly thin. In this context, it is interesting to describe how genomic and related *in silico* technologies could be apply along the different steps of drug discovery process, in order to obtain a more positive outcome and determine the success of the whole operation.

Target identification

The sequencing of the human genome and numerous pathogen genomes resulted in an explosion of the number of potential drug targets. As recently as five years ago, the research community only knew 500 targets. High throughput genomics technologies opened up a vast new field of opportunities that will make it possible for researchers to find novel targets from a universe of as many as 10 000 of them. These targets represent both an unprecedented opportunity and a technological challenge for the pharmaceutical industry. The accompanying challenge is how to be the first to, identify these targets amongst thousands of candidate genes. At this stage, major tools are computer programs which search and scan genomes to add new members of target classes or new splice variants in well known members. This approach, called *in silico* gene hunting is an important part of bioinformatics and produces hundreds of target candidates. Not all of these potential gene targets will become drug targets and the big challenge is to select the most relevant ones for a given disease, a process called target validation.

Target validation

With genomic initiatives providing profusion of putative targets, some notion of gene function beyond what can be discerned from homology is crucial for the decision whether to continue or drop a target. The first step in target validation is the determination of the expression pattern of the gene of interest. Based on the expression patterns in healthy and diseased tissues and/or cells, a disease hypothesis for this gene can be generated. Among the most powerful and versatile genomic tools for target validation are high density arrays of oligonucleotides or complementary DNAs (*biochips*). Nucleic acid based arrays work by hybridisation of labelled RNA or DNA in solution to DNA molecules attached to specific locations on a chip. With this technology, the expression pattern of all genes in a given tissue, the so-called transcriptome, can be obtained in a very short time. It is essential to find genes which are specifically up or down regulated under the disease condition. These genes are "*a priori*" guilty by association, particularly when they are up-regulated during the disease process. In parallel or after a successful screening, further validation of the target will be necessary, using *functional genomics* which includes a plethora of technologies such as gene knockout and transgenic models. Already, computational methods have contributed to large-scale identification of proteins from two-dimensional gel electrophoresis, mass spectrometry and protein microarrays, a discipline called proteome. *Structural genomics* promises, via high-throughput structure determination, to produce a quantum leap in the number of available protein folds, making fold recognition and comparative protein modelling efforts much more effective.

Lead discovery

Chemistry, specifically *molecular modelling and screening*, supported by *chemoinformatics*, is being revolutionized by *in silico* technology. Computational approaches to drug design and screening can be either ligand or target-based. If, for a given therapeutic project, a set of active ligand molecules is known for the macromolecular target, but little or no structural information exists for the target, ligand-based computational methods can be used. Structure-based computational approaches require the 3D struc-

ture of the target. Quantitative structure activity relationship (QSAR) methods can be used, pharmacophore models developed and shape searches performed, based on the set of ligands.

Target-based virtual screening can be performed once a 3D structure of the target is available. Usually, *molecule docking* involves database search interrogation for compounds that fit into the binding site of the target structure in terms of shape complementarity and chemical matching. Structure-based computational docking used as a filter, can tremendously enrich the hit rate compared with random screening methods.

Virtual screening is the use of high-performance computing to analyse large database of chemical compounds in order to identify possible drug candidates, and is a technology that complements current advances in high-throughput chemical synthesis and biological assay. In experimental HTS (High Throughput Screening), robotics and automation allow several hundred thousand compounds to be screened in a month or two, assuming a suitable biological assay is available. With virtual screening, however, large databases or virtual libraries of compounds can be screened in a few days. Combinatorial libraries of analogs can be computationally designed and screened and subsets rapidly synthesized. Leads can be found through experimental high-throughput screening, as well as through virtual screening. Another emerging technology might eventually accelerate and increase the efficiency and productivity of drug development: chemical genomics. It is used for screening dozens of protein drug targets against thousands of small molecules, is a non-linear, parallel approach that can be used simultaneously to determine a novel target's function and identify a small molecule drug lead.

Lead optimisation

Eliminating poor candidates at earlier stages in the drug discovery process is of strategic importance, because it could reduce compound attrition rates during clinical trials. An increasingly popular approach to improve the quality of molecules is to design compound libraries that contain more "drug like" structures. "Drug likeliness" is used to indicate a broad range of physico-chemical properties, such as stability, solubility, lipophilicity, and pharmaco and toxico-properties. Based on experimental and computer methods, Lipinski's "rule-of-five" is a well-known rule that encodes a simple profile for absorption of an active compounds, basing the classification on a limit on molecular weight, lipophilicity, and hydrophilicity. *In silico* screening for drug likeness is then a central component of virtual screening.

Another emerging approach is the computational prediction of a compound's Absorption, Distribution, Metabolism, Excretion and Toxicity characteristics. It allows the elimination of compounds with poor ADME/T parameters before they advance to preclinical tests. These properties can be calculated quickly and can be easily applied to filter a large database. Likewise, filters can be applied on specific chemical substructures, *e.g.*, those associated with problems in chemical stability or toxicity.

Pre-clinical and clinical development

Pre-clinical characterization of new chemicals, up to human testing, will greatly benefit from predictors found in functional genomics, in vitro human cell tests, from mechanism based modelling and simulation derived from animal experimentation. At the pre-clinical phase, genomic technologies could be applied to toxicology, by using DNA microarray to explore the toxic effects of chemical agents on biological systems. Several studies indicated that gene expression profiles can produce a finger print associated to a specific drug's toxicity. To show the degree of relatedness between the toxic effects of compounds, large databases are constructed and computational analysis performed.

At entry into man, exploratory clinical research and indication findings based on minimal toxicology testing will enable economical clinical lead optimization. *In silico* modelling tools for predicting a compound's *in vivo* effects in humans, are still in their infancy but are predicted to facilitate early ranking a selection of the best lead compounds.

Physiological markers that correlate with elements of drug response (surrogate markers), applied in both preclinical and clinical trials, evaluate drug effects more efficiently than before, and are useful for fast identification of failing compounds. Computer modelling based on whole organ models could even provide tools to perform *in silico* clinical trials, and test for efficacy and side effects profiles. Beyond genomic technologies, "e-technologies," such as electronic patient recruitment and monitoring via the Internet, are expected to speed up the launch and completion of clinical trials. Pharmacogenomics, through its power to identify sub-groups of patients who respond differently to a drug under study, offers the promise of streamlining clinical trials. The three studies describe here, have analysed each steps of the drug discovery process evaluating: time, cost, and attrition rate, through pre and post genomic era.

The Boston Consulting Group Study[1] draws on more than 100 discussions at nearly 50 companies and academic institutions in 2001.

Before genomics technology, developing a new drug has cost companies on average $880 million, and has taken about 15 years from start to finish, that is, from target identification to regulatory approval (Fig. 3). Of this cost, about 75 percent can be attributed to failures along the way. Fig. 4 shows cost and time to market at each step of the drug discovery process using genomics technology

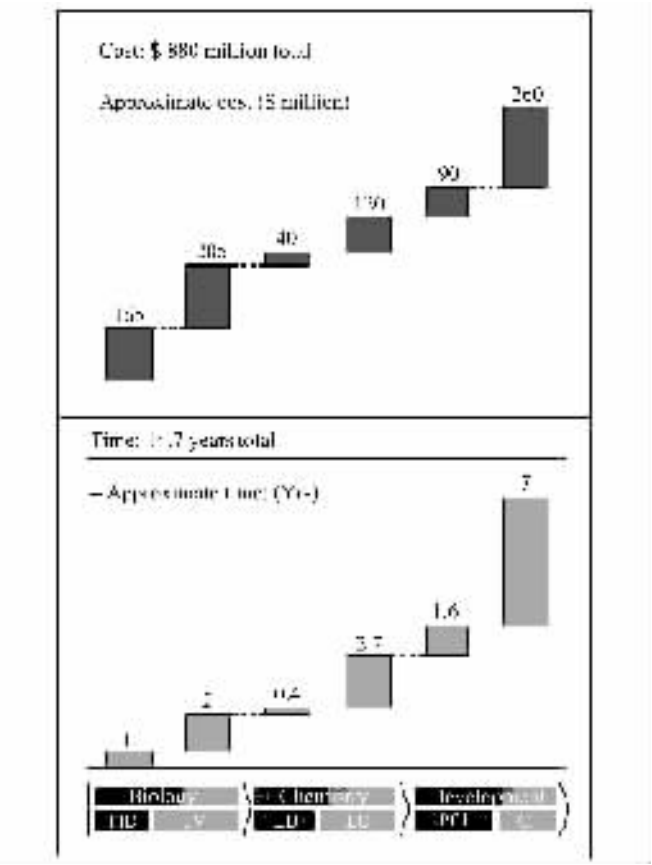

Fig. 3 Drug discovery process traditional approach. TID : Target Identification, TV : Target Validation, LD : Lead Discovery, LO : Lead Optimization, PCI : Pre Clinical, CI : Clinical.

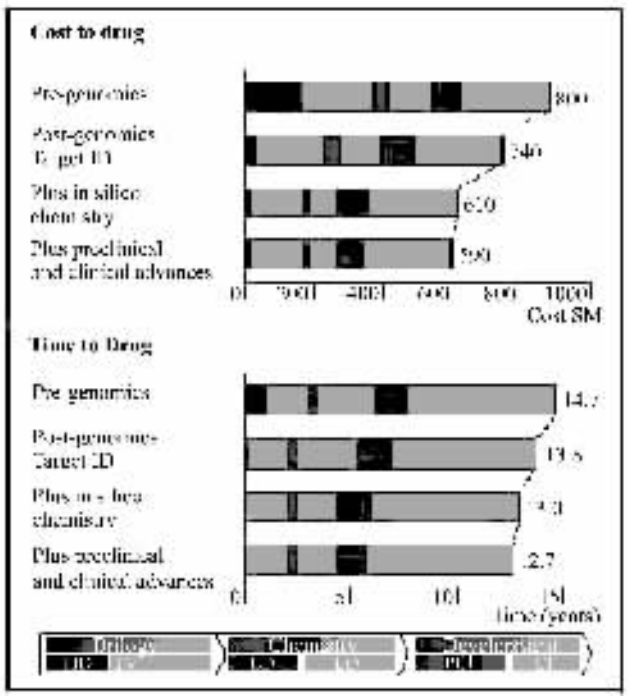

Fig. 4 Cost and time for different R&D models. TID : Target Identification, TV : Target Validation, LD : Lead Discovery, LO : Lead Optimization, PCI : Pre Clinique, CI : Clinique (Source : BCG analysis).

* For target identification, The potential savings would be about 140 million dollars and one year per drug for time to market.

* With *in silico* chemistry, the potential saving would be about 130 million dollars and nearly one year per drug for time to market.

* During development phase, the potential savings would be of 20 million dollars and 0.3 year per drug for time to market.

By using genomics technology, companies could, on average, saves nearly $ 300 million and two years per drug, largely as a result of efficiency gains. This represents a 35 percent cost and 15 percent time-savings.

Mc Kinsey study[2] begins in the year 2000, and it assumes current costs, performance levels, and technology.

A comparison of a low-throughput, low-novelty approach, which was used in the industry pre-1990s, with a high-throughput, high-novelty approach in 2000s, is shown in Table 1. The numbers shown are for the success rate at each stage of development, and the risk-adjusted cost per New Chemical Entity (NCE).

Table 1 Attrition rates and cost for different R&D models

	Low throughput, low novelty	High throughput, high novelty
Key drivers	Success rate	Success rate
Input[a]	50 targets (30% novel)	200 targets (70% novel)
Target validation	30	35
Hit generation	90	90
Lead optimization	90	90
Biological validation	75	50
Preclinical	50	50
Phase I	70	70
Phase II	50	30
Phase III	70	70
FDA filing	90	90
Output[b] (NCEs)	2	3.6
Risk adjusted cost/NCE (US $ B)	0.7	1.0

[a] Number of targets evaluated annually ; [b] number of new drugs ; NCE : new chemical entity ; B : billion (from McKinsey Quartely report 2001).

Under this model, a typical pharmaceutical company could increase its yearly R&D output to 3.6 new drugs, up from 2, but would also have a higher attrition rate. As it can be seen in that table, biological validation and Phase II testing will probably be the key drivers for the relative success of the genomic approach. Without substantial improvements, the total risk-adjusted costs of research and development for the genomic approach could be 40% higher than the traditional approach per NCE generated. More-

over, the higher risk that is associated with high-throughput, high-novelty approach did not generate higher returns.

In that study, analyse of product launched between 1991 and 2000 for the top 15 pharmaceutical companies, shows a lower average present value of sales for these products than product 's sales from traditional approach (US $2.8 billion compared with US $3.6 billion).

In a scenario with only a minimal improvement in technology, the average pharmaceuticals company 's annual R&D budget for its output of new drugs should double, from $800 million in 2000 to $1.6 billion in 2005, and then decrease to $1.2 billion by 2010.

In a scenario with a moderate improvement in technology, the average pharmaceutical company 's annual R&D budget for new drugs output would increase to $1.3 billion by 2005 and decrease to $700 million by 2010 (Table 2).

The accenture study[3] shows that, emerging genomic technologies such as functional genomics, proteomics, gene chips, pathway predictive tools, high-throughput expression systems, bioinformatics, virtual screening, chemoinformatics and modelling, have the potential to fundamentally change the drug discovery process, by allowing a shift from the traditional linear approach to target creation, to an evolving model in which target validation is conducted in parallel to lead discovery and optimization.

Table 2 Annual R&D budget for output of new drugs ($ billion)

Type of Scenario	1995	2000	2005	2010
Minimal improvement in technology	0.8	1.6		1.2
Moderate improvement In technology			1.3	0.7

(from McKinsey Quartely report 2001)

Applying these new technologies at multiple points during the drug discovery process, could potentially improve target identification, and attrition rate, enhance lead optimization and at last, improve clinical trial designs to speed approval.

Figs. 5a and 5b compare the timeframes and sequences of the traditional process with a more dynamic, multi-functional model incorporating these new technologies. The time to market could be divided by two, from 18 years with the traditional approach, down to 8 years with the new technologies.

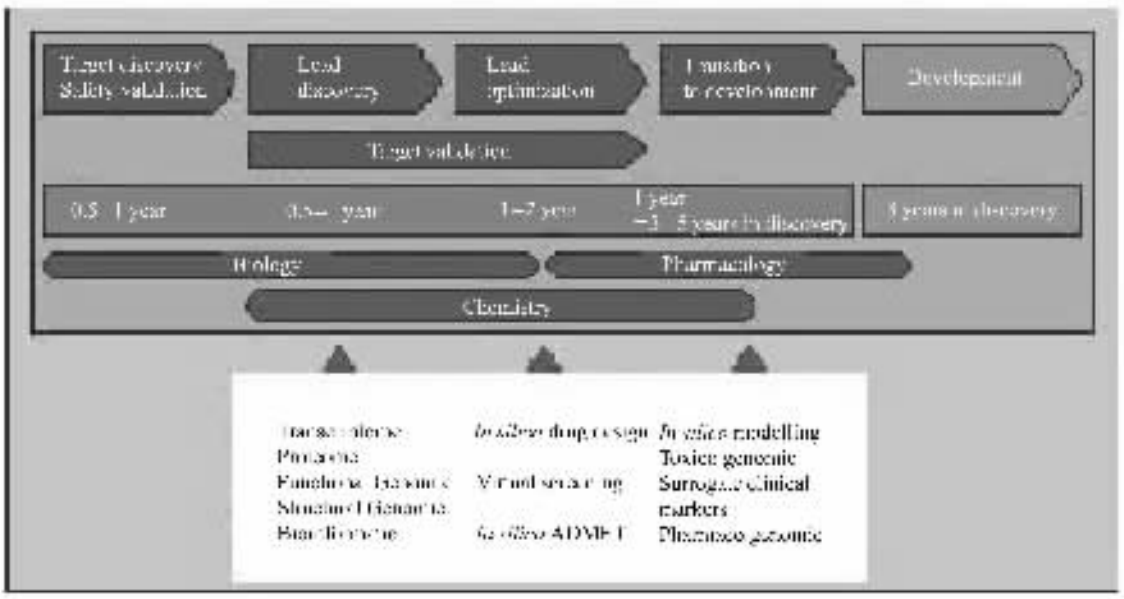

Fig. 5a Drug discovery process: evolving approach.

Fig. 5b Drug discovery process: traditional approach.

Conclusion

Technologies integrating high throughput techniques with powerful new computing capabilities are the main advance of the genomic era. Most likely, they will enhance productivity by boosting efficiency. These three studies demonstrate clearly that genomics technologies and *in silico* R&D have the potential to rapidly and fundamentally change the drug discovery process by transforming numerous aspects of the traditional approach. According to Accenture and BCG studies, duration of drug discovery process could be cut by 50 percent, and 15 percent. Applying genomics technologies, the number of new target would increase the in put by 4 and then the out put by 2. However, the Mc Kinsey study predicts with new technologies, an increase in attrition rate and new drug development costs until 2005 and then decrease by 2010.

This contradicts the BCG 's conclusions stating that genomics could yield significant savings in cost (more than 50 percent) and in time (two years per drug). But that study did not say in which delay these savings would be realised.

It appears then, that pharmaceutical companies need to make significant progress in the discovery process for novel targets. Significant improvements will be needed particularly at the early stages of target validation, and lead optimization. These improvements should include changes in various processes, including decision processes whether compounds should or not go through. Investments in the upcoming technologies such as functional genomics, structural proteomics, bioinformatics, *in silico* drug design and screening and computational prediction tools, should also be considered.

The simultaneous optimization of several biological and chemical parameters and the discovery of rules on how molecular characteristics can be stored and modified are the main challenges in data mining for drug discovery. The challenge for pharmaceutical companies is towards a better integration of these new emerging technologies within the drug discovery process. However, investments in new technologies will require parallel investments in top-rate bench science from biologists, chemists, pharmacologists and computer scientists, to truly deliver meaningful improvements.

References

1 Tollman, P.; Guy, P. A *Revolution in R&D, How Genomics and Genetics are Transforming the Biopharmaceutical Industry, Andersen Consulting, BCG Report*, Nov. **2001**.

2 Barnerjee, P.; Myers, S. *High Performance Drug Discovery, an Operating Model for a New Era, Executive Briefing, Accenture Report*, **2001**.

3 Ma, P.; Zemmel, R. *Splicing a Cost Squeeze into the Genomics Revolution, McKinsey Quartely Report*, No. 2, **2001**.

(E0304041 PAN, B. F.; DONG, H. Z.)