DDT Vol. 6, No. 24 December 2001

An information-flow model of the pharmaceutical industry

Charlie Hodgman, GlaxoSmithKline, Medicines Research Centre, Stevenage, UK SG1 2NY, tel: +44 1438 763244, fax: +44 1438 763082, e-mail: ch18380@glaxowellcome.co.uk

A model is presented of pharmaceutical research and development that is based on an information map. This is useful for understanding the relationship between different product classes, showing how efficiencies can be made, aiding management decision-making, and assisting the move towards integrated *in silico* research and development (e-R&D).

Probably all pharmaceutical companies have drug discovery process pipelines that are variations on a common theme. The study of disease states results in the selection of potential drug targets that are put into screens to identify compounds; these are then taken into development for assessment and formulation into the medicines eventually sold to relieve or cure illness. Although one set of experiments or trials leads on to another, there are points where decisions are made on which targets or compounds to progress through the process in preference to others, or when to abandon a particular compound because it will not be effective or profitable. Decisions are also made based on information from the clinic, which is used to decide which disease classes should be the target of further pharmaceutical R&D. Thus, from the perspective of information flow, we have a circular rather than linear process, with a large body of the information flowing out to the regulatory authorities for drug approval. Figure 1a shows this 'information space' model in its most general form.

By representing the drug discovery process, this model is useful in various ways. First, the position up the map equates to the physical size of entities. Thus, starting with compounds, the scale moves up through macromolecules, cells and people to end at the top with populations. This also means that the further up the map, the greater the physical complexity and variability, which in turn results in the greater unit cost of an experiment or trial. For example, consider the small costs associated with assaying one compound in a screen compared with the costs associated with a clinical trial or disease research programme (Fig. 1b). According to recent estimates[1], the cost of finding a validated target is more than ninefold higher than obtaining a lead compound for that target, and smaller companies need to be aware of the comparatively high cost of clinical trials[2]. The information also partitions into chemical, biological and medical information types, which are collated and analyzed by the appropriate type of informatics.

The left- and right-hand sides of the cycle in Figure 1 also have distinct characteristics. The right-hand half involves looking in increasing detail at fewer and fewer components (analysis and differentiation), whereas the left-hand half involves the reverse (synthesis and integration). It is, therefore, inevitable that the right-hand half (i.e. research) costs less than the left-hand side (development) in which the numbers of variables are increasing, thus resulting in the major attrition seen during drug development. Furthermore, the information generated in the development phase needs to be collated (over many years) for submission to the regulatory authorities who expect standardized protocols and procedures for information gathering.

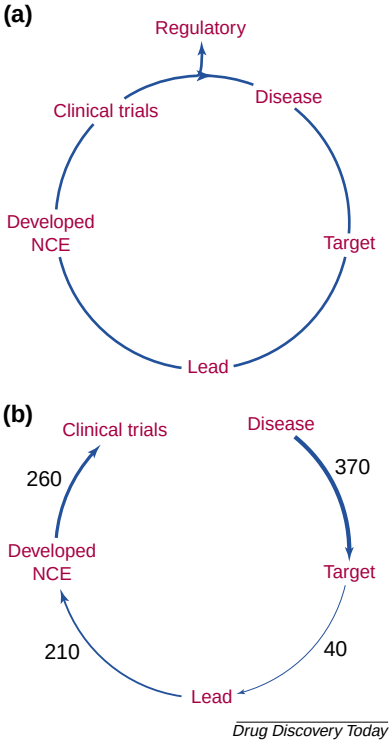

Figure 1. A pharmaceutical business information map. **(a)** shows the general flow through the information space for a small-molecule medicine; **(b)** shows the consequence of redrawing this figure with the thickness of the line corresponding to estimated costs for that stage of the process. The values are taken from Ref. 1 and the units are US$ millions. Abbreviation: NCE, new chemical entity.

Pharmaceutical product classes

In Figure 1, the path drawn through the 'information space' depicts the information flow leading to new small-molecule medicines. However, this is just one of five product classes that the pharmaceutical industry can choose. Other paths lead to biopharmaceuticals, diagnostics and surrogate markers, as shown in

  PII: S1359-6446(01)02019-0

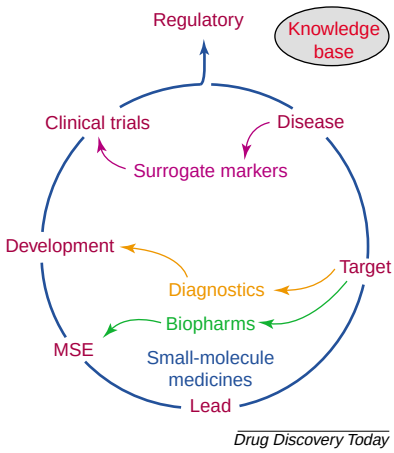

Figure 2. Pharmaceutical product classes. There are five classes: small-molecule medicines (blue), biopharmaceuticals (green), diagnostics (orange), surrogate markers (purple) and knowledge bases (red). The knowledge base is a static resource whereas the others involve information flow. The path for small-molecule medicine is shown in black for reference, whereas the other paths are shown in the colour of their product class. Abbreviation: MSE, medical safety and evaluation.

Figure 2. The only product that does not have a path in this model is a knowledge base, such as a proprietary database. However, few pharmaceutical companies are even contemplating selling their proprietary information and expertise. This is e-business and is still only the activity of biotechnology companies, for example, Incyte (Palo Alto, CA, USA), Curagen (Branford, CT, USA) and Inpharmatica (London, UK).

Improving business efficiencies

The major costs for a large pharmaceutical company arise from the failure rate of leads during development and clinical trials, target validation (although as a result of bioinformatics this cost is decreasing) and insufficient information sharing. A lack of information sharing results in duplication of effort, failure to become aware of problems early or a loss of knowledge when someone leaves the company. Information should feedback to earlier stages in the process to aid the prioritization of targets and compounds for screening – why screen compounds that others have found to be toxic? This type of information flow tends to happen on an informal basis or through a committee appointed for the purpose. However, with the increasing number of mergers, ever more high-throughput technologies and staff turnover, it is becoming harder for companies to keep track of such data. Therefore, information systems to cope with the increasing and varying data volumes, and to bring development and clinical data back into the research process are becoming increasingly important[3].

Information flow in other directions can also result in reduced costs. The structure of large organizations can result in comparatively poor communication between biologists in the research and the development halves of the company. However, the data from one arm can inform the work of the other. For example, the tissue specificity of a target is determined as part of its validation, but this also informs (pre)clinical staff which tissues to study when looking for potential unwanted side effects. Apart from the horizontal transfer of information, data can be reused in the vertical sense. For example, biochemical assay data can be used by chemists to decide which compound families should go into screens or which should be taken forward as leads, but can also be used by biologists to understand more about the structure and function of the target.

Novelty from crossing domains

As mentioned previously, the information map separates into three levels based on entity size: the chemical, biological and medical domains. The path to producing new medicines involves crossing these domain boundaries, and so any techniques that span them are likely to be especially useful. At the biological–medical interface, genetics uses population data to select targets in the discovery process

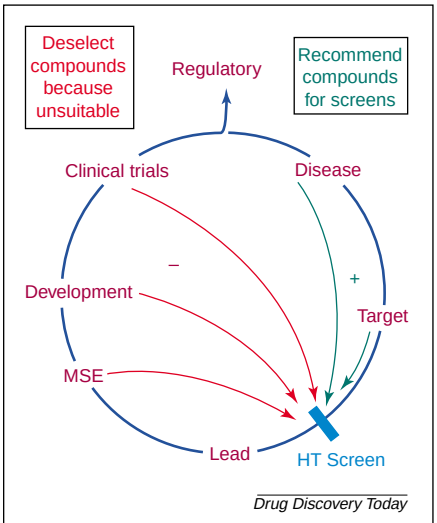

Figure 3. Information control mechanisms. To improve the efficiency and success rate of HTS, information should be fed forward from the disease and target analysis points about which compound classes are known to bind to the target (green arrows). Information should also feedback from development and the clinic about which compounds are known or strongly predicted to be unsuitable for reasons of safety, ability to be delivered or clinical effectiveness (red arrows). Abbreviation: MSE, medical safety and evaluation.

and for predicting therapeutic outcomes (pharmacogenomics) in the clinic. Given that these are at the higher end of the information map, this highlights the enormous value of genetics, genomics and informatics to the pharmaceutical industry, which has been affirmed in a recent report[1]. Biochemistry crosses the other interface and has been a mainstay of pharmaceutical research for decades.

Any computational technique that allows data from one domain to be used seamlessly by another offers great potential for *in silico* R&D. Perhaps the most sophisticated are the computer models of the heart. These contain an ensemble of equations corresponding to the kinetics of ion channels and biochemical reactions, in an assembly of *in silico* cells[4]. Hoffman-La Roche (Basel, Switzerland) used one version of this model in its submission of Posicor™ to the FDA. The Pharsight system (Mountain View, CA,

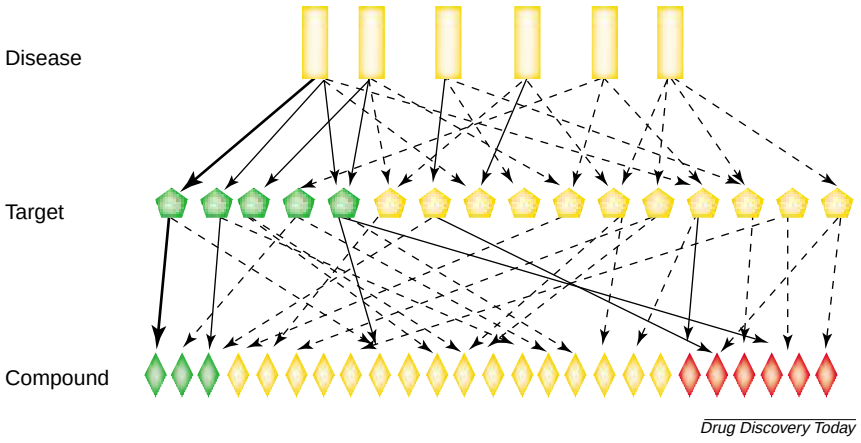

Figure 4. Interconnected portfolios. The different disease areas, targets and compounds are depicted as rectangles, pentagons and diamonds, respectively. The green shaded figures are validated targets or leads and the red diamonds represent failed leads. The arrows show how the three portfolios connect: dashed arrows show the potential or predicted relationships, solid arrows represent experimentally observed associations, and the thick arrows show successful associations that have resulted in the compound going into development.

USA)[5] takes basic PK and PD parameters and returns clinical-trial scenarios that are most likely to succeed. At the biological–chemical interface, tools of computational chemistry are being used in large and in specialist (e.g. De Novo Pharmaceuticals, Cambridge, UK) companies in conjunction with information about target structure to conduct virtual compound-screening and predict orphan compound classes of high intellectual property (IP) value.

e-R&D

Internet technology has already advanced to the point where e-business is becoming a reality. The industry model presented here can contribute in various ways to e-R&D, particularly for smaller companies. For example, it can show where the information generated by a company resides in the pharmaceutical-industry information space. A smaller company can then begin to see how it can contribute to the discovery and development of new pharmaceutical company products, by knowing which of the product-class circuits (Fig. 3) are involved. The small company should establish connections with other companies to complete the circuit needed to produce products for the class in question. A simple example of this is the partnership of Gemini Genomics (Cambridge, UK) and Axis-Shield (Tayside, UK) to develop diagnostic tests for osteoporosis.

If a company wishes to extend the market for its information, it should look to equivalent levels on the map, and then focus on companies using information in that area. This is new use of the same data, as described previously. For companies seeking to expand into new product classes, then collaborations, licencing deals or mergers are needed with companies having expertise in the areas and on the paths of that product class. Perhaps an example of this was the merger of Celltech (Slough, Berkshire, UK), which was a strong company in biopharmaceuticals and drug development, with Chiroscience (Cambridge, UK), which had more small-molecule expertise.

Companies have portfolios of disease interests and/or targets and/or compounds. Such a portfolio represents a third dimension to the model, where each slice corresponds to an item in the portfolio. To complicate the issue, there are many relationships between these three portfolios (Fig. 4), and so the challenge for discovery is to find the appropriate path through these different portfolios. Their content also changes over time.

Finally, this model demonstrates the distribution of information resources in a fully integrated pharmaceutical company, and it provides a first-level approximation of what a company 'knowledge space' looks like. The detailed construction of a knowledge space comes from making the data available to staff who work at equivalent levels on the information map, although they will require different query tools because they will be addressing different questions. The company will gain the efficiencies described previously. A recent report suggests that the efficiency gains will correspond to US$300 million and two years of development time per medicine[1]. This same report also outlines the importance of integrating the biological and chemical information for further reducing times and costs.

Caveats

The model presented here, of using an information map to view the processes taking place within pharmaceutical companies, is a tool to improve efficiency and save time in the R&D of new pharmaceutical products. Like any model, it is likely to have some shortcomings when particular aspects are examined in detail. For example, the chemical and biological information produced during the development phase is often generated concurrently rather than sequentially. However, despite this, the model has proved useful in various types of discussions, from management issues to project prioritization.

References

1 Flanagan, A. *et al.* (2001) A revolution in R&D: the impact of genomics. Boston Consulting Group (http://www.bcg.com/publications/files/genomics.pdf)
2 Fox, A.W. (2001) Clinical development – look before you leap. *Nat. Biotechnol.* 19, BE27–BE29
3 Augen, J. (2001) Information technology to the rescue. *Nat. Biotechnol.* 19, BE39–BE40
4 Noble, D. *et al.* (1999) Biological simulations in drug discovery. *Drug Discov. Today* 4, 10–16
5 Lehmann, H.P. (1999) Pharsight trial designer. *J. Am. Med. Assoc.* 281, 955–957