Transforming the genome to drug discovery


Mark Murcko, Chief Technology Officer, and **Paul Caron**, Head of Informatics, Vertex Pharmaceuticals

'Capturing the value of the human genome requires us to revolutionize pharmaceutical R&D.'

Sequencing the human genome has provided us, as Eric Lander stated, with 'the parts list' necessary for understanding human biology [1]. Now, biotechnology and pharmaceutical companies are faced with the enormous challenge of translating genomic information into new, innovative medicines to improve the human condition. A thousand or more potential new drug targets have emerged as a result of sequencing the human genome, yet currently available drugs only target an estimated 500 different proteins [2]. If we are to effectively use the fruits of genomic research, we must re-engineer drug discovery and development. An ideal 'post-genomic' strategy would be inherently parallel, enabling researchers to pursue sets of targets simultaneously. Multiplexing across families of structurally similar proteins enables the reuse of chemical information and expertise [3]. These approaches, which leverage the way in which particular classes of chemical compounds interact with targets within the same family, could offer an efficient solution to the massive influx of potential new drug targets.

How far can we go?

Just as the fields of genomics and proteomics encompass the identification and classification of all the genes and proteins in a genome, the field of chemogenomics could be defined, at its limit, as the discovery and description of all possible drugs against all possible drug targets. Although this definition seems unattainable, it focuses attention on the appropriate organizational models for research and early clinical development.

Parallel drug design

Organizing targets in related gene families is not new. The use of both SAR data, to identify targets with similar patterns, and family-based screening strategies suggest that chemistry efforts can be focused toward related gene families [4,5]. However, incorporating traditional approaches to drug discovery in parallel for many targets across multiple biological and chemical disciplines represents a unique approach. Targets within a gene family often have similar *in vitro* assays and properties, enabling maximum reuse of expertise, as well as leveraging biology resources. In addition, a significant number of compounds that are designed and synthesized to target one gene family member often demonstrate activity against other family members, enabling medicinal chemistry on multiple targets to have a common, often potent, starting point. Potent chemical compounds set the stage for 'scaffold morphing' – literally, the metamorphosis of a chemical lead into multiple novel compound classes that inhibit each target within a gene family, based on a precise understanding of the points of interaction between the chemical lead and the target protein. The creation of numerous classes of chemically distinct compounds is driven by integrating structural biology and modeling approaches with combinatorial and medicinal chemistry. Coupled with 'target hopping' – in which the breadth and activity of each active chemical scaffold is explored against many members of the target family – multiple development candidates, against numerous protein targets, are designed synergistically with intellectual property that is transferable among related targets.

Avoid throwing the baby out with the bath water

In a traditional pharmaceutical organization, the reuse of knowledge across target families is typically limited by the division of discovery efforts into therapeutic areas that are confined to a single therapeutic indication. However, although this therapeutically based organizational structure could complicate gene family-based research, there remains a clear and compelling rationale for organizing late-stage development and commercial operations along such lines. Therefore, gene family-based research initiatives must be constructed in such a manner as to preserve the good and necessary features of therapeutic experience and expertise. Maximum efficiency is derived when gene family

1359-6446/02/$ – see front matter PII: S1359-6446(02)02267-5

initiatives simultaneously use specialized knowledge across multiple therapeutic areas.

Understanding key receptor–ligand interactions: drug specificity

The availability of complete gene sequence information for entire target families, together with representative protein structures, enables the construction of predictive three-dimensional models for entire protein families. These models enable interactions between a given substrate or inhibitor, and specific residues in the target can be mapped even when detailed structural data is unavailable. Characterizing the subset of residues that provide key inhibitor interactions facilitates the prediction of inhibitor specificity; for example, it has been demonstrated that single amino acid changes are sufficient to generate specificity in protein kinases [6–9]. Thus, genomic data can then be used to focus structure-based inhibitor design towards areas of the novel target molecules that enable engineering of specificity.

Target validation

One drawback to the wealth of genomic information is that most of the genes in a family cannot currently be linked to any therapeutic application. Although the shortage of target validation data can be addressed by many methods, including broad panels of gene knockouts, dominant negative mutants, antisense, expression profiling and proteomics, these processes can be slow and often provide only limited understanding of the target. Thus, potent and specific inhibitors arising from a chemogenomics effort are potentially the most effective and efficient resources for validating novel targets. These inhibitors can be used as biochemical probes to dissect the role of novel targets in both cellular experiments and animal models. A positive result not only links a specific target to an important biological process but also simultaneously provides an excellent chemical lead for further optimization.

Who knows what, when, and why does it matter?

In small groups, and as individuals, researchers can make breakthrough discoveries. However, the use and value of new insights are realized only when important knowledge produced in one discipline is disseminated to researchers across multiple fields in real-time, enabling scientists to continually evaluate and refine their hypotheses and make intelligent scientific decisions. Despite proclamations to the contrary by many companies, the degree of organizational integration required for parallel approaches to drug discovery still appears to be the exception rather than the rule in the biotechnology and pharmaceutical industry. For a parallel drug discovery strategy to be efficient, the research organization must be structured in a manner that enables intensive interactions across all disciplines.

Outlook

Biotechnology and pharmaceutical companies will not fully benefit from knowing the human genome without a major strategic overhaul. The chemogenomics approach enables a research team to consider all targets within large families in parallel. This approach capitalizes on the availability of genomic data, as well as established methods, such as combinatorial-, computational- and medicinal chemistry, HTS, structural biology, chemo- and bioinformatics; all of which, if properly organized, drive the efficient reuse of information, reagents, methods and expertise as research teams move from one group of targets to the next. Candidate molecules generated using this integrated approach hold the potential to transform medicine in the post-genomic era.

References

1 Lander, E.S. and Weinberg, R.A. (2000) Journey to the center of biology. *Science* 287, 1777–1782

2 Drews, J. and Ryser, S. (1997) The role of innovation in drug development. *Nat. Biotechnol.* 15, 1318–1319

3 Caron, P.R. *et al.* (2001) Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* 5, 464–470

4 Frye, S.V. (1999) Structure–activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* 6, R3–R7

5 Thorpe, D.S. (2000) Forecasting roles of combinatorial chemistry in the age of genomically derived drug discovery targets. *Comb. Chem. High Throughput Screen.* 3, 421–436

6 Wilson, K.P. *et al.* (1997) The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase. *Chem. Biol.* 4, 423–431

7 Lisnock, J. *et al.* (1998) Molecular basis for p38 protein kinase inhibitor specificity. *Biochemistry* 37, 16573–16581

8 Fox, T. *et al.* (1998) A single amino acid substitution makes ERK2 susceptible to pyridinyl imidazole inhibitors of p38 MAP kinase. *Protein Sci.* 7, 2249–2255

9 Gum, R.J. *et al.* (1998) Acquisition of sensitivity of stress-activiated protein kinases to the p38 inhibitor, SB203580, by alteration of one or more amino acids within the ATP binding pocket. *J. Biol. Chem.* 273, 15605–15610

Mark Murcko* and Paul Caron
Vertex Pharmaceuticals
130 Waverly Street
Cambridge, MA 02139, USA
**tel: +1 617 444 6100*
fax: +1 617 444 6680
e-mail: Mark_Murcko@vrtx.com