Network and systems biology: essential steps in virtualising drug discovery and development

Christoph Wierling[1,2], Thomas Kessler[1,2], Lesley A. Ogilvie[1], Bodo M.H. Lange[1], Marie-Laure Yaspo[2], Hans Lehrach[2,3,*]

[1]Alacris Theranostics GmbH, Berlin, Germany
[2]Department of Vertebrate Genomics, Max Planck Institute for Molecular Genetics, Berlin, Germany
[3]Dahlem Centre for Genome Research and Medical Systems Biology, Berlin, Germany

The biological processes that keep us healthy or cause disease, as well as the mechanisms of action of possible drugs are inherently complex. In the face of this complexity, attempts at discovering new drugs to treat diseases have alternated between trial-and-error (typically on experimental systems) and grand simplification, usually based on much too little information. We now have the chance to combine these strategies through establishment of 'virtual patient' models, centred on a detailed molecular characterisation of thousands or even, in the future, millions of patients. In doing so, we lay the foundations for truly personalised therapy, as well as a far-reaching virtualisation of drug discovery and development in oncology and other areas of medicine.

Section editor:
Hans V. Westerhoff, University of Amsterdam, NL.

Introduction

Biological systems and disturbances in them, for example, those leading to diseases, are enormously complex. Not only are diseases and drug actions complex, they differ from patient-to-patient; and in a tumour often from cell-to-cell, with every cell potentially reacting differently to the drugs the patient receives [1,2]. In the field of cancer, numerous avenues of complexity are presented, including intra-tumour heterogeneity, clonal development and various mechanisms of developing resistance to drug treatment. The complexity of the biological processes we are trying to affect by a drug is often mirrored by the complexity of drug actions. Many drugs that have been designed to bind to a single target, on detailed analysis show much more complicated mechanisms of action [3]; an overall complexity that presents a fundamental challenge to current approaches to drug discovery and development [4]. There is therefore growing recognition that new concepts and strategies are required to address these challenges.

Here we focus on how technological developments in tandem with the ever-growing molecular knowledge base on complex disease processes and individuals world-wide are pushing forward the development of systems-level computational 'virtual patient models'; models that will enable evaluation of the likely outcome of all possible therapies on every individual patient and provide the possibility of carrying out every step of the drug discovery and development process on collections of virtual patients or a virtual clinical (or preclinical) trial.

Corresponding author:. H. Lehrach (lehrach@molgen.mpg.de)

How can we handle this complexity in the process of drug development?

Any complex artefact we generate (e.g. a modern airplane or a complex computer chip) has usually been designed logically, with major effort being spent to make sure that the overall complexity is subdivided into well isolated subsystems, for example, modules. Biological systems are, however, designed by evolution through trial-and-error not by logic. Discovering drugs to correct mistakes in these systems will therefore, more likely than not, require trial-and-error on a massive scale.

Many early drugs have been the result of trial-and-error over many generations and by many individuals, leading to the identification of, for example, plants with pharmacological effects, serving as the basis for further drug development. Similarly, trial-and-error strategies have been used for large-scale screens of many different compounds in biological systems that, hopefully, mirror the disease process (e.g. the NCI60 dataset [5]).

As more and more molecular mechanisms have been identified, this has been replaced/complemented by much more mechanism-based, focussed strategies. Approaches that have often ignored the enormous complexity of biological systems, the complexity of many drug mechanisms, and the large differences between superficially similar individual patients and disease states; it is because of this that many drugs, particularly in the case of cancer, only help a small fraction of the patients receiving them [6,7].

Major new developments to address this complexity include the increased use of high-throughput screening techniques or virtual screening approaches, allowing analysis of very large and even virtual compound libraries [8–11]; complemented by an enormous increase in biological information through projects such as the human genome project [12] and an increasing set of high throughput-omics techniques [13], which have generated invaluable data on human development, physiology and evolution, as well as identifying molecular factors driving disease states [14]. The enormous progress in molecular techniques in the past decade, and in particular 'Next Generation Sequencing' (NGS), now allows us to generate more information on single patients than was previously available on the whole of human biology.

Drinking from a firehose: what to do with the data?

Several computational methods are available to integrate and analyse large datasets, with a range of statistical methods being used to identify a set of genes or gene expression signatures (biomarkers) that, potentially, enable discrimination of distinct molecular subgroups within a patient cohort, or that are associated with certain treatment responses [17–21]. Such statistical classification of samples may clearly make sense when the susceptibility of a given patient to specific drugs is assessed, for example, for improving the treatment regime [22–25]. However, whether the different gene signatures obtained by different statistical studies actually reflect the underlying biology of the disease in an individual patient is questionable [26–28]. To obtain further insight, the analysis has been conceptually advanced by taking into account information on disease-related general biology, for example, by classical GO term or pathway enrichment methods as implemented in R packages such as Bioconductor [29], or web tools including DAVID [30,31], pathway commons [32] and ConsensusPathDB [33,34].

In addition, the enormous complexity of the combination of biological systems and drug action, and the flood of data that we can now generate, require, but also make possible, new analysis concepts and approaches [15,16], which have been categorised under the headings 'network biology' or 'systems biology'.

Network biology

In comparison to these classical approaches towards sample stratification, more recent work has taken information about the biology of the underlying cellular networks into account. Network biology approaches [35–37], in general aim more at topological, qualitative analyses, with several statistical or graph theoretical concepts being explored for applications in drug discovery and personalised medicine [38–42]. The aim of these studies is mainly to generate deeper insights into biological causes of a disease or to allow more efficient stratification of individual samples in the disease context. Indeed, integration of network information and gene expression data has generated higher predictive power than gene expression-based methods alone [43]. Nevertheless, network biology may be seen simply as an improved approach towards understanding of molecular interactions and their perturbations based on statistical analysis of large groups of samples, while molecular events within a single sample remain inaccessible by these methods.

Implementation of such methods are facilitating identification of drug targets, helping to optimise drug efficiency, minimising toxicity and drug side effects (reviewed in [44]); however, despite the significant insights gained, we are still faced with the challenge of delineating the extensive and dynamic changes occurring in the diverse and molecular components and biological pathways within diseased (as well as healthy) individuals [45,46].

Systems biology

Systems biology, in general, tends to describe more quantitative, model-based analysis. We focus here on the use of the term *systems biology* as the establishment and use of computer models replicating, as far as possible, known biological mechanisms, not in the statistical context of many clinical samples, but in the context of actual biological events occurring in one distinct sample at a time.