



# The Use of Biosimulation and Mathematical Modelling in Targeted Drug Discovery

a report by

**Seth Michelson**

*Principal Scientist and Biomathematician, Theranos, Inc.*

Discovering a new medicine is a multistep process that requires the investigator to identify:

- a cause–effect pathway (or pathways) inherent to a disease and its pathophysiology;
- those cells and molecular entities (e.g. receptors, cytokines, etc.) inherent to the control of those pathways (typically termed targets);
- an exogenous entity that can manipulate the molecular target to therapeutic advantage (typically termed a drug);
- with some level of specificity, how that manipulation modulates the disease effects (termed the mechanism of action of the drug); and
- the segment of the patient population most likely to respond to those manipulations (typically through the use of biomarkers).

Given these challenges, pharmaceutical drug discovery is an extremely complex and risky endeavour. Despite growing industry investment in research and development (R&D), only one in every 5,000 new drug candidates is likely to be approved for therapeutic use.<sup>1</sup> In fact, approximately 53% of compounds that progress to phase II trials are likely to fail, resulting in amortised costs of US\$800 million to US\$1.7 billion per approved drug.<sup>1–3</sup> Clearly, the crux of the problem is the high failure rate of compounds, especially those in late-stage clinical development. To solve this problem, one must clearly identify the most suitable compound for the most appropriate target in the subpopulation of patients it best pertains to, and then give them the optimum dose.

To adequately address these challenges, the pharmaceutical researcher must be able to understand and characterise the effects of diverse chemical entities on the pathways of interest in the context of the biology they are meant to affect. To make this possible, research scientists must have at their disposal the means to acquire the most pertinent and predictive information available. To do that, they need to answer the ultimate question: ‘What is the next best experiment I can do?’ Mathematical modelling and biosimulation are increasingly playing a key role in helping them answer this important question.



Seth Michelson is Principal Scientist and Biomathematician at Theranos, Inc., Menlo Park, California, a consumer healthcare systems company dedicated to building realtime, point-of-care monitoring systems. He has over 20 years of experience in modelling and simulating biological systems. Before moving to Theranos, Dr Michelson was Vice President of *In Silico* Research and Development at Entelos, Inc. Prior to joining Entelos, Dr Michelson was Director of Scientific Support Services at Roche Biosciences.

Dr Michelson received his PhD in 1987 and a BSc in 1984 from the University California, Los Angeles, and an MA in Applied Mathematics from the University of California, Berkeley.

## A Methodology for Idealising a System

Many people use the terms ‘modelling’ and ‘simulation’ synonymously. However, for the purposes of this editorial, I will define a model as a descriptive idealisation of a real system, while a simulation is only one method with which to explore that model and its dynamics. Simulations used to explore models of biological systems are termed biosimulations. Many researchers have used deductive logic and statistical reasoning to identify potential correlations in their data. Typically, these data are derived from bioassays meant to mimic the dynamics of the biological system described by the poster. As such, they suggest potential relationships between the biomolecular components in that system. As technologies have evolved, more sophisticated pattern recognition techniques have been used to infer pathway dependencies in these types of systems.<sup>4–6</sup> However, these deductive modelling efforts are wholly dependent on the sampling assumptions under which the data are acquired. For example, does the transient state of the cell cultures or tissue sections accurately represent the *in vivo* dynamics of the disease?

## Mathematical Modelling in Biology

As an alternative, researchers have constructed mathematical models of biological systems by rigorously characterising the relevant underlying interactions and dynamics into mathematical formalisms.<sup>7,8</sup> If one begins the modelling process at the bottom, using all the known parts of the system (typically thousands of genes and proteins), and assembling them piece by piece into a coherent whole, a mathematical model will eventually emerge.<sup>9,10</sup> However, whenever a new component or connection is discovered, the entire model must be reconfigured.<sup>7</sup>

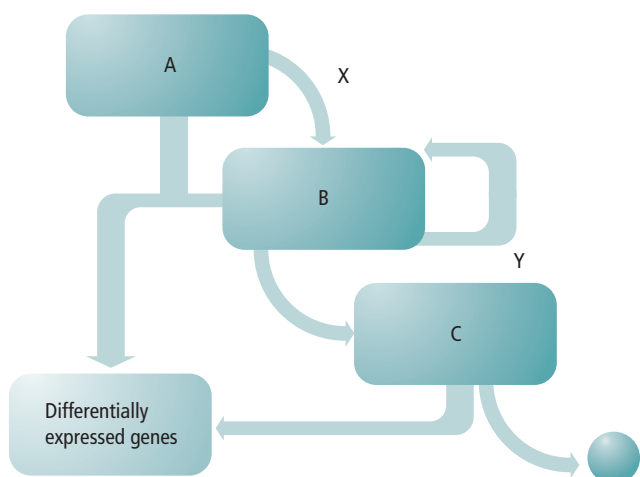
In contrast, one can begin with clinical observations about the overall behaviour of a system and its disease. The model is then built piecemeal from the top down by identifying only those physiological subsystems required to reproduce clinically relevant outcomes. The systems, organs, tissues, cells, proteins, biomolecules and pathways that comprise each of these subsystems are added in progressively greater detail, subsystem by subsystem, until the model sufficiently reproduces the observed clinical behaviours.<sup>7</sup> In this way, each subsystem is constrained by the overall activity of its parent system and, ultimately, the model as a whole.

## Biosimulation – A Means of Characterising the Solution Set of a Model

As the models of human disease are typically much too large and complex to derive simple closed-form solutions or to perform simple steady-state analyses, the best way to understand the model and the biology it represents is by simulating its dynamics in a computer. Numerically solving the equations that make up a model of human disease and projecting these solutions forward in time is termed predictive biosimulation.<sup>11,12</sup>



**Figure 1: A Hypothetical Cascade of Disease Pathophysiology**



Three cell types (A, B and C) under the control of two cytokines (X and Y). Inappropriate signals emerge upstream and deleterious effects are exerted downstream.

However, biosimulation is not the answer. Rather, it is the means to answering the next question. To make optimum use of this technology in the discovery and development of new drugs, one must marry the model, its biosimulation and its experimental design into an integrated knowledge-gathering process. By way of an example, consider a simple model of a hypothetical disease cascade (see *Figure 1*).

Suppose one believes that three separate cell types are intrinsic to the progression and pathophysiology of a particular disease. Suppose further that cell A is inappropriately activated by an endogenous signal, which upregulates the secretion of cytokine X. Cytokine X, in turn, stimulates cell B to upregulate the expression of cytokine Y, as well as its autocrine receptor. Finally, cytokine Y signals cell C to alter its physiology in a deleterious way. While a genomic snapshot of this process may yield a set of differentially expressed genes (e.g. cytokine X, cytokine Y and the cytokine Y autocrine receptor), which correlate strongly with the disease state, do these genes necessarily identify optimal therapeutic targets? Or, is it possible that the most influential target actually lies upstream of cytokine X (e.g. the receptor on cell A that transmits the initial signal) or downstream (e.g. the cytokine Y receptor on cell C)? A pharmaceutical researcher may then formulate these questions in the form of hypotheses, and explicitly characterise each within the model. Then he/she can simulate the impact of each on the progression of the disease. It is then incumbent on the experimentalist to design and run the bioassay that tests the most pertinent of these hypotheses.

### The Target

Why model and simulate a biological system in the first place? If we already know everything there is to know about the disease and its dynamics, there would be no need to model them, and one's ability to ultimately predict clinical success based on pre-clinical data would be significantly greater. However, past performance has shown this is not the case. Therefore, it is inevitable that during the modelling exercise, researchers will uncover a critical gap in their knowledge. Once these knowledge gaps have been identified, their impact on the system and its dynamics can be explored *in silico* by formulating alternative hypotheses to span them. This process allows the researcher to evaluate the impact of a particular physiological hypothesis on the ultimate course of the disease, and then plan a course of

experiments accordingly. Therefore, data collection efforts can be more clearly focused, assays can be better designed and the resultant data can be more effectively interpreted. By iteratively focusing modelling and data collection on those biological subsystems that exert the greatest impact on the vital aspects of a disease and its phenotypes, predictive biosimulation helps us gather the most pertinent information available.

The challenge then facing the pharmaceutical scientist is how to most efficiently interpret his/her data in the context of the dynamic human disease being simulated in the computer. How much does a particular physiological pathway contribute to the overall physiology of the disease? Is there a feedback mechanism in place that amplifies or dampens this signal? Are there redundant or back-up pathways that will mitigate attempts to inhibit the pathway? By systematically categorising what one knows and, more importantly what one does not know, and building testable hypotheses to explain each gap, one can use these data to systematically test the impact of each hypothesis (guess) on the decision at hand.

### The Compound

Once a drug target has been identified and validated, the focus of the drug discovery effort shifts to building a compound that will modulate that target in a therapeutically beneficial way. This relationship between chemical character and biological activity fundamentally depends on how well the medicinal chemist understands the biological impact of his/her compound when he/she is constructing a structure-activity relationship (SAR). To do this, the medicinal chemist must understand his/her compound from the biologist's perspective. This requires the team to optimally sample the biological activity space. This process is best depicted in *Figure 2*. The more informative the sampling data obtained from the bioassays, the higher the probability of finding an optimal compound. As always, the question is 'What is the next best experiment to do?' By explicitly representing each variation of a compound's activity under each hypothetical assumption, the impact of a compound's pharmacodynamics (PD) versus hypothesised exposure levels (pharmacokinetics or PK) can be tested *in silico*, and those hypotheses yielding the greatest impact on the decision-making of the chemist can focus the biologist on designing the most informative assay possible to increase the power of the sampling scheme.

### The Patients

To be a medicine, a drug must be safe, effective and biologically accessible – i.e. it must remain at the target site in sufficient concentration for a sufficient period of time to exert its therapeutic effects. Patient-to-patient variability affects these aspects of medicinal character and must be accounted for when selecting and optimising a candidate compound and its dosing regimen.

The goal of the clinical trial is to prove that the effects researchers observe in the laboratory can, in fact, be generalised to a reasonably large segment of the human clinical population. However, since laboratory animals are typically derived from inbred strains, the effects of genetic variation are relatively well controlled in the laboratory. This control of experimental variance improves a scientific study by separating signal from noise. However, our ability to apply these results to humans and their inherent variability – e.g. variable absorption, distribution, metabolism and excretion (ADME) characteristics – is severely hampered. Additionally, the influence of environmental variability on test animals is further suppressed by

housing them in controlled animal care facilities. The fact that the human population is not composed of genetically controlled, environmentally sequestered subjects but is, instead, an outbred collection of highly variable populations whose behaviour is constantly varying is a constant source of difficulty for clinical research scientists.

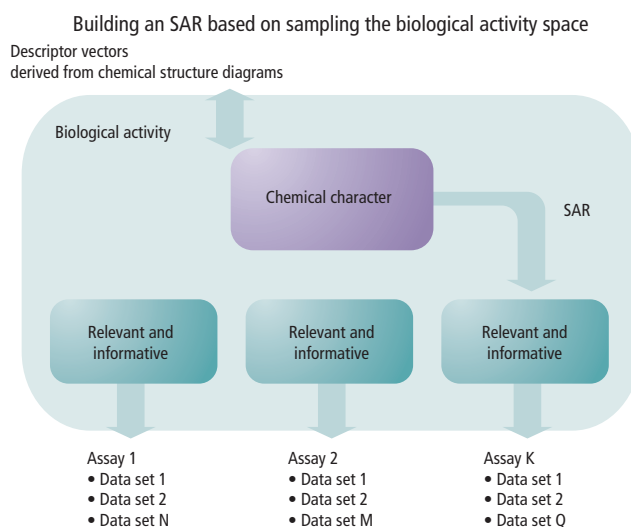
How can pharmaceutical researchers account for patient variability during the drug discovery process when the bulk of research and development is performed with cellular systems and test animals instead of human subjects? To fully comprehend the impact of patient variability on the medicinal characteristics of candidate compounds, researchers must capture, in explicit quantifiable form, the effects of genetic and environmental variations on the underlying disease physiology. In the context of the mathematical model and its biosimulation, the impact of each of these variable influences can be represented as specific hypotheses, and then mathematically represented in the model as explicit expressions. Michelson et al.<sup>13</sup> detailed how one can use this strategy to explicitly represent patient variability in a predictive biosimulation to explore the impact of that variability on clinical trial design and execution.

### The Data – Populating the Model

To develop new drugs, we need to realise that biology is telling us a story, and that it is speaking to us in the only language it knows. To understand that story we must, therefore, converse with it on its own terms. To do that, we need to become as fluent in the language of biology as possible. The conversation we undertake with biology can take many forms. It can be purely observational, à la Darwin, or in an exploratory sense it can be fairly broad but lacking in depth. However, drug discovery and development require that we formulate our conversation as an in-depth interview, diving into and thus characterising those pieces of the story most relevant to the pathology at hand. The way we conduct that in-depth interview is to design the most informative assays possible, and from them acquire the most important data available.

It is equally important to realise during this interview that biology is alive and, as such, dynamic. That means that for us to derive the most benefit from our interview, we must formulate our questions so that they adequately query those dynamics. The way to do that is to design bioassays that include time as a variable, and thus sample the time course of the hypothetical processes we suspect underlie the pathophysiology of the disease. The data extracted from these experiments may be based on cohort studies, or may be acquired longitudinally within individual subjects. The latter design is from a statistical point of view much more informative, since the biostatistician can now derive estimates for the impact of both between- and within-subject variances. These data help the researcher account for the types of patient variability highlighted above. How often, exactly what data to measure and what they mean in

**Figure 2: The Conversation Between Medicinal Chemist and Biologist During the Development of a Structure–Activity Relationship**



SAR = structure–activity relationship.

the context of the disease and its evolution should be guided by the same principles outlined above.

### Conclusions

The pharmaceutical industry is experiencing a failure rate in drug development that is, in the long term, unsustainable. It is unlikely that a brute force approach – launching multiple clinical trials to test diverse target hypotheses – will lead to the needed reductions in the cost and risk of development programmes. Rather, the industry needs to focus its efforts on acquiring, as early as possible, the most pertinent information available. To do that, one must explicitly identify what is known and not known, and then systematically develop a set of biologically testable hypotheses to span those gaps. One way to do that is to formalise the biology at hand in a mathematically rigorous way, and use the biosimulation infrastructure to drive the development of the next best experiments. Perhaps the greatest challenge posed by these technologies will be the impulse to base critical decision-making primarily on model predictions. The most important thing to remember is that the model and its simulation do not provide the answers one seeks; they merely help one develop one's questions more completely and focus the conversation with biology in the most productive way. When synthesising data from *in vitro*, *in vivo* and, increasingly, *in silico* sources, critical human thought will always play the pre-eminent role. Bioassays are represented as sampling processes of the biological activity space. These data are used by the chemist to develop the SAR. ■

- Pharmaceutical Research and Manufacturers of America (PhRMA), Pharmaceutical Industry Profile 2006. <http://www.phrma.org/files/2006%20Industry%20Profile.pdf>
- Gilbert J, Hense P, Singh A, Rebuilding big pharma's business model, *In Vivo, The Business and Medicine Report*, Windhover Information, Inc., 2003;21(10):1–10. [http://www.bain.com/bainweb/PDFs/Marketing/rebuilding\\_big\\_pharma.pdf](http://www.bain.com/bainweb/PDFs/Marketing/rebuilding_big_pharma.pdf)
- Dimasi J, Hansen R, Grabowski, H, The price of innovation: new estimates of drug development costs, *J Health Econ*, 2003;22:151–85.
- Kightley DA, Chandra N, Elliston K, Inferring gene regulatory networks from raw data: a molecular epistemics approach, *PSB*, 2004;9:498–509.
- King RD, Whelan KE, Jones FM, et al., Functional genomic hypothesis generation and experimentation by a robot scientist, *Nature*, 2004;427:247–52.
- Chakraborty AK, Dustin ML, Shaw AS, *In silico* models for cellular and molecular immunology: successes, promises and challenges, *Nat Immunol*, 2003;4:933–6.
- Friedrich CM, Paterson TS, *In silico* predictions of target clinical efficacy, *Drug Discov Today*, 2004;3:216–22.
- Ho RL, Biosystems modelling for *in silico* target validation: challenges to implementation, *Emerg Therap Targets*, 2000;4:699–714.
- Noble D, Levin J, Scott W, Biological simulations in drug discovery, *Drug Discov Today*, 1999;4:10–16.
- Noble D, Modeling the heart – from genes to cells to the whole organ, *Science*, 2002;295:1678–82.
- Hall K, Baillie R, Michelson S, Biosimulation: dynamic modeling of biological systems, *Annu Rep Med Chem*, 2002;37:279–88.
- Musante CJ, Lewis AK, Hall K, Small- and large-scale biosimulation applied to drug discovery and development, *Drug Discov Today*, 2002;7(20)(Suppl.):S192–S196.
- Michelson S, Sehgal A, Friedrich CM, *In silico* prediction of clinical efficacy, *Curr Opin Biotech*, 2006;17:1–5.