Expert Review

At the Biological Modeling and Simulation Frontier

C. Anthony Hunt,[1,3,4] **Glen E. P. Ropella,**[2] **Tai Ning Lam,**[1] **Jonathan Tang,**[3] **Sean H. J. Kim,**[3] **Jesse A. Engelberg,**[3] **and Shahab Sheikh-Bahaei**[3]

Abstract. We provide a rationale for and describe examples of synthetic modeling and simulation (M&S) of biological systems. We explain how synthetic methods are distinct from familiar inductive methods. Synthetic M&S is a means to better understand the mechanisms that generate normal and disease-related phenomena observed in research, and how compounds of interest interact with them to alter phenomena. An objective is to build better, working hypotheses of plausible mechanisms. A synthetic model is an extant hypothesis: execution produces an observable mechanism and phenomena. Mobile objects representing compounds carry information enabling components to distinguish between them and react accordingly when different compounds are studied simultaneously. We argue that the familiar inductive approaches contribute to the general inefficiencies being experienced by pharmaceutical R&D, and that use of synthetic approaches accelerates and improves R&D decision-making and thus the drug development process. A reason is that synthetic models encourage and facilitate abductive scientific reasoning, a primary means of knowledge creation and creative cognition. When synthetic models are executed, we observe different aspects of knowledge in action from different perspectives. These models can be tuned to reflect differences in experimental conditions and individuals, making translational research more concrete while moving us closer to personalized medicine.

KEY WORDS: agent-based; mechanistic; modeling; predict; simulation.

INTRODUCTION

The declining pace of new drug approvals, general inefficiencies, and significant rates of failures late in development are among the factors contributing to calls for reexamination of the methodological robustness of pharmaceutical and biotechnology research and development (R&D) (1–4). Among the inefficiencies is the fact that scientists and decision-makers translate results and conclusions from wet-lab experiments to patients using conceptual mappings where some assumptions about the wet-lab-to-patient mappings are intuitive and unknown, and thus unchallengeable. Those conceptual models exist in and rely on the minds of a changing cadre of domain experts. It has been argued that mathematical, systems biology models will help to address this issue, but we argue that primary reliance on the familiar inductive and deductive computational modeling methods will be inadequate: success requires that they be augmented with new classes of models and methods. Within the past decade, an expanding number of groups, working within different domains, have contributed to the development of a class of computational models that we argue can dramatically increase the efficiency, efficacy, reliability, and variety of plausible translational mappings, and thus facilitate and accelerate better decision-making.

To distinguish this class of models and its methods from the more familiar inductive mathematical modeling methods, we identify it as the synthetic (combining elements to form a whole) method of modeling and simulation (M&S). Our objective is to draw on recent examples to help make a case for the use of synthetic M&S within the critical decision-making stages of drug discovery and development. We explain 1) how synthetic M&S is made scientific 2) how synthetic M&S can augment mental models and system thinking with concrete virtual tissues, organs,

[1] Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California, USA.

[2] Tempus Dictum, Inc., Portland, Oregon, USA.

[3] The UCSF/UCB Joint Graduate Group in Bioengineering, University of California, Berkeley, California, USA.

[4] To whom correspondence should be addressed. (e-mail: a.hunt@ucsf.edu)

ABBREVIATIONS: ADMET, absorption, distribution, metabolism, elimination, and toxicity; ALC, alveolar-like cysts; AT II, alveolar type II; CA, cellular automata; CPM, cellular Potts models; CV, central vein; hACH, human adipose-derived stromal cell; ISL, In Silico Liver; MDCK, Madin-Darby canine kidney; M&S, modeling and simulation; MOF, multiple organ failure; OOP, object-oriented programming; PB, physiologically based; PD, pharmacodynamic; PK, pharmacokinetic; PCP, physicochemical property; PV, portal vein tracts; QSAR, quantitative structure-activity relationship; R&D, research and development; SM, similarity measure; SS, Sinusoidal Segment; SIRS, systemic inflammatory response syndrome; TNF, tumor necrosis factor; Z, zone.

and ultimately virtual patients; and 3) that synthetic M&S enables explorable, in silico wet-lab-to-referent mappings that are accessible to all members of an R&D organization as operating models of current knowledge and beliefs. More important is that synthetic M&S encourages and facilitates abductive reasoning (see Appendix for explanation): the primary means of knowledge creation and the primary source of creative leaps. The more mature descendants of these models may even begin capturing the gestalt of successful pharmaceutical R&D.

While it is true that no computational model can fully represent the complexity of biological systems, new model types are essential to achieving deeper insight into the causal, mechanistic networks responsible for disease and desired pharmacological phenotypes. We show how synthetic M&S can be used to discover, clarify, and challenge plausible linkages between biological mechanisms and phenotypes. Skeptics may also declare that models cannot mimic the complexity inherent within biology and thus cannot be correct. However, the advance of science depends on discovering and using better models. It will become clear how validation-supported synthetic models (defined in the Glossary) can expedite and improve R&D decision-making.

The text that follows is divided into five sections. In "Rationale for New Model Classes," we provide our rationale: we make the case for needing new classes of models and discuss how they are created. We follow that in "Analogues: From *In Vitro* Tissues to Interacting Organs" with descriptions of four examples of system-oriented, synthetic, biomimetic models that have provided new mechanistic insight into phenomena observed *in vitro* and *in vivo*. Those motivating examples provide context for "Reasoning Types and Their Different Roles in M&S," which discusses how the three types of reasoning—induction, deduction, and abduction (in Glossary) —are used in science and M&S. We describe how the two classes of models, inductive and synthetic, draw differently on inductive, deductive, and abductive reasoning to achieve their different objectives. Coupling the capabilities of well-established mathematical modeling methods with those of synthetic M&S will, for the first time, make the full power of the scientific method available to the M&S component of R&D. That discussion leads directly to "M&S and the Scientific Method," in which we develop the idea of scientific M&S (in Glossary). In the penultimate section, "Impact of M&S on Scientific Theory," we explain what it can accomplish. We argue that in order to achieve the above vision, we must expand computational M&S into scientific M&S. We provide a list of eight capabilities that synthetic models will need in order to achieve our vision. We then summarize in the Conclusions. A glossary of less familiar terms is included in the Appendix along with essential supporting information, including brief descriptions of inductive, deductive, and abductive reasoning. For convenience, selected key points made in "Rationale for New Model Classes," "Reasoning Types and Their Different Roles in M&S," "M&S and the Scientific Method," and "Impact of M&S on Scientific Theory" are provided as bulleted statements at the start of each subsection. A relatively comprehensive bibliography of primarily discrete event (in Glossary) biomedical models that combine synthetic and inductive methods is provided as Supplemental Material. See (5–11) for reviews of advances in and relevant biomedical applications of inductive mathematical M&S.

RATIONALE FOR NEW MODEL CLASSES

Envisioned New Model Classes

- Building an experimental apparatus is fundamentally different from "modeling the data."
- An objective is to build better working hypotheses about mechanisms.

What spatiotemporal mechanisms play roles in the emergence (in Glossary) of a pharmacological response? During drug discovery and development, current knowledge is often inadequate to answer that question. A research objective is to develop better working hypotheses about those mechanisms. Synthetic models can expedite that process. A dictum of the physicist Richard Feynman was "what I cannot create, I do not understand." It follows that to understand biological responses and their plausible generative mechanisms when uncertainty is large and data are chronically limited, we need to build extant (actually existing, observable), working mechanisms that exhibit some of those same phenomena. Building extant, plausible, analogue mechanisms is fundamentally different from the traditional approach of "modeling the data." In the latter case, the mechanisms are all conceptual. We cannot yet build hierarchical, modular, extant mechanisms out of biochemicals. However, as described below, we can build extant biomimetic mechanisms using object-oriented software tools.

Consider the following: A software engineer, given complete freedom, creates code that, when executed, produces mechanisms which give rise to multi-attribute phenomena that are strikingly similar to specified pharmacological phenomena. When the software engineer has limited biological knowledge, there may be no logical mapping from event execution in the simulation (in Glossary) to the biology during observation. However, biologically inspired requirements can be imposed to shrink and constrain the space of software mechanism and implementation options that successfully exhibit those same phenomena. A continuation of that process can lead to extant software mechanisms (and phenomena) that are increasingly analogous to their biological counterparts. In so doing, we are *not* building a model based exclusively on known biological facts and assumptions, because the facts are often insufficient to do so. Furthermore, keeping track of all the assumptions and assessing their compatibility can become an unwieldy, time-intensive task. Rather, we are exploring the space of reasonably realistic, biomimetic mechanisms that can cause the emergence of prespecified pharmacological phenomena. The focus is on inventing, building, exploring, challenging, and revising plausible biomimetic mechanisms. To distinguish the two modeling methods and help ensure a disciplined focus on methodology, we refer to models arrived at through the latter process as (biomimetic) *analogues* (in Glossary). To emphasize aspects of construction and method, specifically combining often varied and diverse elements, so as to form a coherent whole, we say *synthetic* analogues.

Bridging the Gap Between Wet-Lab and Traditional, Computational Models

- Gap-bridging computational models will be objects of experimentation, similar to wet-lab models.

To achieve our vision, we need in silico models that are different from the familiar inductive pathway, network, tissue transport, pharmacokinetic (PK), pharmacodynamic (PD), physiologically based (PB), and other related models. Synthetic models will have more in common with the wet-lab systems that they represent than with the current generation of equation-based models. Fig. 1 depicts the space of major model classes used today in pharmaceutical and biotechnology research. Model types on both sides have different uses. The ultimate referents for both are specified subsets of patients. Not represented are the conceptual mental models on which all scientists rely and the prosaic models (often supported by sketches of idealized mechanistic events) that describe these mental models. Moving to the right in Fig. 1, model aspects (in Glossary) become more realistic, relative to their referents. Moving up, similarities between model and referent attributes increase: the models become more biomimetic. The diagram excludes patients. Model organisms are in the upper right. *In vitro* cell and tissue models are next. Below them are cell-derived systems. Statistical and correlative models are to the lower left. Above them are the familiar, induced mathematical computer models. Included within the latter are network, pathway, PK, PD, and PB models; they are induced from data. The space above them includes inductive models based on discrete event formalisms, such as cellular automata (CA), cellular Potts models (CPM) (12–14), pi-calculi (15–17), etc.

In order to demonstrate that we understand how molecular level details interface with and exert influence at higher levels and emerge as features of a favorably altered patient phenotype, we need models and methods that can bridge the gap. We need models that are increasingly more like their referents—models that have extant mechanisms that generate emergent properties analogous to how phenomena emerge during wet-lab experiments. Those models will be synthetic, as are wet-lab models. An important use for such models will be testing hypotheses about mechanisms (rather than about patterns in data).

However, it is important to note that network, PK, PD, and other mathematical modeling methods do not need replacing—they do what they are intended to do very well, even though they will benefit from improvements. Nevertheless, to span the gap, we will need new model classes having new uses and capabilities (see Text Box below).

Analogue capabilities needed to bridge the gap in Figure 1

As engineered devices, what characteristics must a gap-spanning synthetic analogue exhibit to achieve the vision presented in the Introduction? They must exhibit the eight capabilities listed below, while also exhibiting characteristics toward the top, right of Fig. 11B. The challenge is to design, build, and validate such analogues.

1. *Transparency*: Simulation details, as they unfold, need to be visualizable, measurable, and comparable to those of referent wet-lab and animal systems.

2. *Experimental indistinguishability*: Measurements of analogue attributes made during execution are, to wet-lab domain experts (at a comparable level of abstraction), indistinguishable experimentally from measurements made on a referent wet-lab model during experimentation.

3. *Mappings*: Observables in silico, including components, are designed to be consistent with in vitro and in vivo observables. Doing so enables clear, iteratively concretizable mappings (Fig. 2; right side) between in vitro and in silico components and mechanisms.

4. *Local mechanisms*: As in wet-lab systems, the simulated physiological behaviors and measurements that emerge during execution (Fig. 2, mapping A) are the consequences of local mechanisms—local component interactions (mapping C). Components can be heterogeneous.

5. *Reusability*: Analogue components can be made quasi-autonomous, when needed, and thus can be easily reconfigured to represent different mechanistic hypotheses, aspects of the referent wet-lab system, and experimental conditions.

6. *Flexibility*: Because system aspects of interest will change, it must be relatively simple to increase or decrease the granularity of any analogue aspect in order to simulate an additional phenotypic attribute or change usage and assumptions. Achieving this capability will enable cycles of scientific M&S: testing, falsification, and analogue refinement, as new wet-lab observations become available or the space of targeted attributes expands.

7. *Adaptability*: Analogous to how object-oriented programming makes software objects more adaptable, the components and the system itself should be constructed so that they can be adapted easily to function as components in other analogues (e.g., a liver analogue can fit within different organism analogues; a hepatocyte analogue (validated against in vitro data) can fit within a liver analogue; components from a liver analogue can be used in a lung analogue, etc.).

8. *Components articulate*: To study alternate mechanistic hypotheses, and better understand normal-to-disease transitions and differences between individuals, it must be easy to construct—to *synthesize*—alternative systems and transform them during simulations (as in Fig. 8B). Doing so requires that it be easy to join, disconnect, and replace components, both vertically and horizontally, without having to significantly reengineer the system.
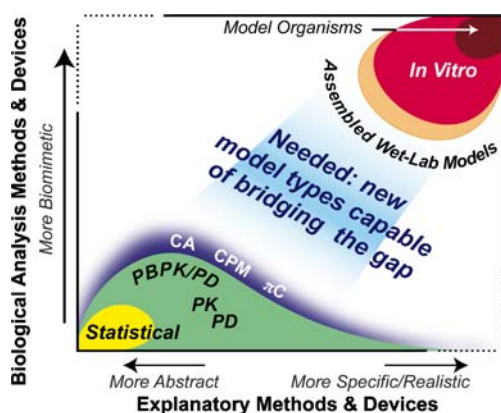
Fig. 1. Illustrated is the gap that exists between inductive, mathematical models and the wet-lab models used in biomedical research. For illustration purposes, model types, and the analytic and explanatory methods that use them, are arranged according to abstraction level *versus* biological character; in reality they are not independent. The arrangement of model types is discussed in the text. More abstract indicates a greater capability for simple and focused representation. More realistic indicates a greater capability for aggregating collections of facts. The biological axis (biomimetic) indicates the degree to which a model resembles and behaves, at some level of detail, like its wet-lab referent. An inductively defined, equation-based model, for example, can mimic time-course measures of an aspect of a biological system very well (high, aspect-specific biomimesis), but, as a complex algorithm implemented atop a numerical integrator, it is not at all realistic (yet the conceptual model to which it is tied may include some realistic features). An unvalidated agent-based model can implement detailed representations of almost any physiological process and yet be incapable of behaving like the referent in any particular context; hence, it exhibits high realism but little biomimicry. Models that can bridge the gap will be biomimetic analogues of their wet-lab counterparts; they can be used for evaluating explicit mechanistic hypotheses in the context of many aspects of the referent.

Instantiating a Mechanistic Hypothesis and Achieving Measurable Phenotype Overlap

- A synthetic analogue is an extant hypothesis: execution produces an observable mechanism.
- Analogues that are scientifically useful will have few 1:1 model-to-referent mappings.

When object-oriented, software engineering methods are used to implement a mechanism on the right side of Fig. 2, the product of the process is an extant hypothesis: these components (objects) will produce a mechanism upon execution. By so doing, we have instantiated (represented with a concrete instance) a mechanism in silico. A consequence of mechanism execution will be the emergence of phenomena that are similar (or not) to prespecified phenomena, such as a response following exposure to a xenobiotic. Execution produces a simulation with features that we can measure; those measurements enable testing the hypothesis. If phenomena similarities meet some prespecified criterion, then the simulation stands as a challengeable yet tested theory about abstract yet plausible mechanistic events that may have occurred during the wet-lab experiments.

There is an important lesson in the fact that *in vitro* models are useful, yet fundamentally different from their animal and patient referents, a lesson that can help provide
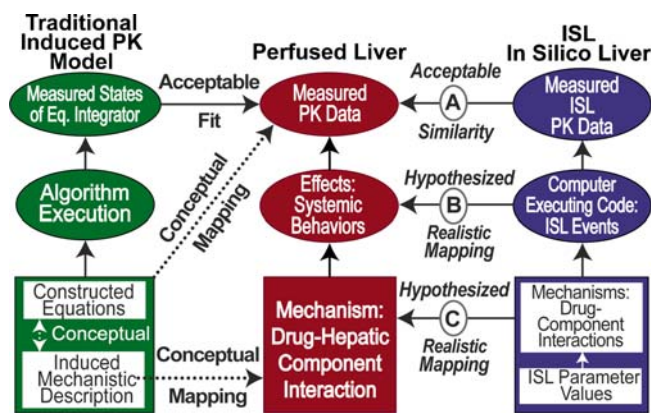

Fig. 2. Model-referent relationships (adapted from Fig. 1 in (65)). Shown are relationships between wet-lab, perfused liver experiments (*center*), traditional PK models (*left*), and In Silico Liver (ISL) analogues, which have begun bridging the gap in Fig. 1. *Center*: Rat livers in an experimental context (as in (89)) are the referent systems. During experiments, hepatic components interact with transiting drug molecules to cause changes in a drug's concentration-time profile. The system's behaviors during the experiment are reflected in the collected data. *Left*: The researcher identifies patterns in the data: drug (and possibly metabolites) levels in the hepatic out-flow profile. From those data and prior knowledge, an abstract, mechanistic description of what is thought to have occurred is offered, thus establishing abstract, conceptual mappings from that description to hepatic mechanisms. One or more equation-based PK models are selected, they are believed to be capable of describing the time course patterns identified in the data. The equations are known to be consistent with an idealized version of the mechanistic description. There is a conceptual mapping from that description to the equations. Software is executed to simulate parameterized equation output, enabling a quantitative mapping from simulated output to PK data. Metrics specify the goodness of fit. *Right*: A plausible, abstract mechanistic description is hypothesized and specified; it is similar but not identical to the one on the left side. Software components are designed, coded, verified, and assembled, and connected guided by the mechanistic specifications. The product of the process is a collection of micro-mechanisms rendered in software. A clear, concretizable mapping—c—exists between in silico components and how they plug together, and 1) hepatic physiological and micro-anatomical details, and 2) drug interactions with those components. Execution gives rise to a working analogue. Its dynamics are observable and intended to represent (mapping B) corresponding dynamics (believed to occur) within the liver during an experiment. Mapping B is also concretizable. Simulation measures provide time series data that are intended to mimic corresponding liver perfusion measurements. Quantitative measures establish the similarity between the two outflow profiles (mapping A).

guidelines for building scientifically useful synthetic analogues capable of bridging the gap. To present the lesson, we use a specific example, though the ideas are generalizable. The example wet-lab system is cultured Madin-Darby canine kidney (MDCK) cells, a well-established *in vitro* cell line. Epithelial cell cultures are widely used as model systems to support research in drug discovery and development. The Venn diagrams in Fig. 3 illustrate phenotypic relationships. There is overlap (measurable similarities and identified mappings) between phenotypic attributes of MDCK cell cultures and epithelial cells within a tissue, even though this phenotypic overlap is limited. By phenotype, we mean the variety of tissue attributes and cell behaviors associated with several aspects of each system, observed from particular,
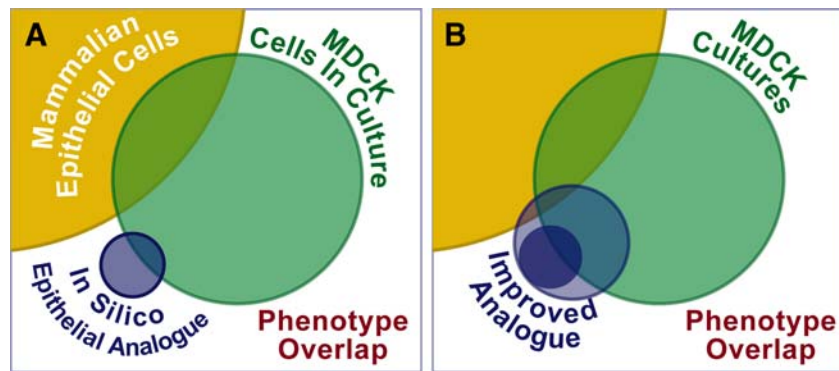
Fig. 3. Shown are examples of phenotype overlap. The shaded areas illustrate sets of phenotypic attributes. There is overlap (clear, direct similarities) of some systemic attributes of MDCK cultures and corresponding epithelial cell attributes in mammalian tissues. In the non-overlapping regions, the mapping between related attributes (and their generative mechanisms) is not straightforward. It is complex. An in silico, synthetic analogue of the class described on the right side in Fig. 2 can have a similar relationship to MDCK cultures. Grant *et al.* provide an example (86) in which cell components are quasi-autonomous. There is a set of operating principles along with component logic governing component interactions. Phenotypic attributes observed during execution are unique. Overlap (similarities) in phenotype between the analogue and MDCK cultures are intended to reflect similarities (but not precise matches) in components, mechanisms, and operating principles. **A** As in (86), the first analogue is simple and abstract. It validates when a set of its attributes are acceptably similar to a targeted set of MDCK culture attributes (area of overlap). As is the case with MDCK cultures relative to epithelial cells in mammalian tissues, the analogue will have attributes that have no MDCK counterparts (non-overlapping area). **B** Sequential, iterative refinement (see Fig. 12) of the first analogue leads to an improved analogue. Kim *et al.* provide an example (81). Its validation is achieved when an expanded set of its attributes are judged similar to an expanded target set of MDCK attributes.

comparable perspectives. Although phenotypic overlap is limited, *in vitro* models are extremely useful systems for understanding *in vivo* biology. Similarities do exist, and conceptual models that are broadly accepted provide descriptions of plausible relationships. Experimenting on MDCK cultures improves the researcher's insight into epithelial tissues. They have also proven useful because MDCK cultures are simpler (more abstract) and more easily controlled than *in vivo* systems, and thus more easily studied and understood.

Although there are many similarities in measurable phenomena between systems, there are few precise, one-to-one behavioral mappings between structures in MDCK cultures and epithelial cell structures within mammalian tissues. True, there is a 1:1 correspondence between cells. However, because cell environments and genetics are different (cannot be precisely duplicated), mappings between *in vitro* and *in vivo*, which can be aspect- and perspective-dependent, may be nonlinear and in some cases complex (in Glossary). *That observation is instructive*. It suggests that an in silico analogue can become a scientifically useful representation of MDCK cultures (and eventually epithelial cells in tissues) without enforcing 1:1 mappings between its attributes and mechanisms and measures of cultured MDCK.

We can characterize wet-lab models that have extended lifetimes and are used in different experimental contexts with a variety of designs as being robust to context, even though its referent is specific. We need synthetic analogues that can be characterized similarly. The mappings from wet-lab model to referent are often somewhat different for each context because the aspect of interest will have changed. We can surmise that

an analogue built initially to have many 1:1 model-to-referent mappings may be solidly anchored to one referent aspect and attribute, and thus may have limited additional uses (without undergoing considerable reengineering).

Absent precise 1:1 mappings, scaling methods will be needed; their development can be separated from that of the analogue. The mechanisms responsible for generation of an MDCK culture phenomenon (e.g., stable cyst formation) are not grounded to any external measurement methods. Nor are they grounded directly to a tissue referent. The units, dimensions, and/or objects to which a variable or model constituent refers establish groundings (in Glossary and discussed further in "How a Model is Grounded Impacts How It Can and Should Be Used"). The components of parameterized PK models are typically grounded to metric space. The components of MDCK mechanisms are grounded to each other. The grounding of cells to each other and their environment is independent of any measures. From that fact, we can infer that analogues that bridge the gap will exhibit similar grounding. We measure wet-lab phenomena using metric devices. We cannot use those same devices to measure events during simulations.

Analogues That Bridge the Gap Will Be Executable Knowledge Embodiments Suitable for Experimentation

- Synthetic models present different aspects of knowledge in action, and do so from different perspectives.
- Separate, tuned copies of successful analogues can reflect differences in individual-specific attributes.

In synthetic analogues such as the In Silico Livers (ISLs) discussed below, components and their interactions represent micro-mechanistic features, including anatomical, physiological, and molecular details at different levels during execution. Because of such multi-level similarities, following several rounds of improvement, testing, and validation, descendant analogues of this class have the potential to evolve into executable representations of what we know (or think we know) about biological systems: *executable biological knowledge embodiments*. We expand on that idea below; see (18) for further discussion. Such embodiments are needed but are beyond the scope of current PK, PD, and related modeling methods. Knowledge embodiment is made feasible because synthetic analogues provide *concrete instances* of that knowledge rather than computational descriptions of conceptual representations. When an analogue is executed, it demonstrates when, how, and where our knowledge matches or fails to match details of the referent system. For that reason, Fisher and Henzinger (19) suggested referring to such simulation models as executable biology.

The envisioned synthetic analogues can facilitate the merger of knowledge and expertise contributed across organizational domains into executable and, therefore, observable and falsifiable systems of plausible mechanisms and hypotheses (20). Together, they will represent the current best theory for aspects of system function. It will be possible to observe different aspects of knowledge in action and do so from different perspectives, as we do with wet-lab systems. Adjusting (tuning) an ISL to represent (for example) a normal rat liver in one in silico experiment, a diseased rat liver in another (as in (21); see "In Silico Livers"), and a mouse or human liver in another will be relatively straightforward because uncertainty can be preserved and cross-validation of component functions can specify which features to tune and by how much. It will be feasible to take copies of the same analogue and tune each separately to reflect differences in measured, patient-specific attributes. The collective knowledge coupled with collective uncertainty can be made specific for groups of patients and even for individual patients.

Achieving the Vision Motivating Physiologically Based Pharmacokinetic (PBPK) Modeling and Simulation

- To discover and test plausible mechanistic details, we must experiment on (different) synthetic analogues.

A vision motivating research on synthetic analogues is identical to one that has motivated development of traditional PBPK models: by "accounting for the causal basis of the observed data, ... the possibility exists for efficient use of limited drug-specific data in order to make reasonably accurate predictions as to the pharmacokinetics of specific compounds, both within and between species, as well as under a variety of conditions" (22). However, PBPK model parameters necessarily conflate features and properties of the biology (aspects of histology, etc.) with drug physicochemical properties (PCPs) (23). In doing so, "the causal basis" becomes obscured due to the conflated biological features that were especially influential in causing some pattern in the data.

A purpose of conducting PK and other experiments that provide time course data is often to shed light on prevailing mechanistic hypotheses about drug dynamics, specifically to gain new knowledge regarding mechanistic details of disposition and metabolism. Most often, hypotheses about those details are induced from the data. Fitting inductive mathematical models to data is often used as evidence in support of particular hypotheses. To date, designing and conducting new wet-lab experiments has been the only practicable means to experimentally falsify those hypothesized, conceptual mechanisms. Experimenting on synthetic analogues provides a powerful new means of discovering and testing the plausibility of mechanistic details. A traditional, inductive, PK model hypothesizes an explanation of patterns in PK data (24). The mathematics of PBPK models describe data features predicted to arise from conceptualized mechanisms, which are typically described in sketches and prose. As illustrated on the left side of Fig. 2, there are unverifiable, conceptual mappings between equations and envisioned mechanisms. The methods used by synthetic analogues, as exemplified by the four cases described in "Analogues: From *In Vitro* Tissues to Interacting Organs," are different. They provide an independent, scientific means to challenge, explore, better understand, and improve any inductive mechanism and, importantly, the assumptions on which it rests.

Creating Synthetic Analogues and Defining Their Use

- Models that begin spanning the gap in Fig. 1 are generalized object- and agent-oriented constructions.
- Biomimetic agent-based analogues facilitate discovery and understanding of phenomena produced by systems of interacting components.
- Important use: better understand disease mechanisms and their interactions with interventions.
- By nesting agents and objects hierarchically, one can discover plausible upward and downward mechanistic linkages.

The biological mechanisms that generate system level phenomena are consequences of components at multiple levels interacting in parallel, primarily discretely, with other components in their local environment. Simulation of such behavior can and has been achieved by adopting discrete event M&S methods. Any interaction can be stochastic. So doing simulates uncertainties and is a means to preserve ignorance. See (25,26) for a generalized discussion of the advantages of using discrete event methods to model and simulate complex adaptive systems. Fisher and Henzinger (19) discuss how several formal, discrete event methods (Boolean networks, Petri nets (27), pi-calculi, interacting state machines, etc.) have been leveraged to gain mechanistic insight into biological phenomena. Advances in simulating complex biological phenomena have been accomplished using formal cellular automata (28–42) and cellular Potts models (13,43–52).

Most biological components are spatially organized, semi-modular, and quasi-autonomous: they include organs, tissue functional units, cells, subcellular systems, and macromolecular complexes. Synthetic analogues must be capable of exhibiting those same attributes. Greater component autonomy coupled with realistic yet abstract, spatially organized, biomimetic mechanisms have been achieved using agent-based and agent-oriented methods (53–55). Because all of the preceding methods are based ultimately on object-

oriented programming (in Glossary) methods, we suggest that analogues that begin spanning the gap in Fig. 1 will be considered generalized constructions in the object-oriented domain that uses agents (in Glossary). Consequently, we focus the following discussion on those methods. The four examples described in the sections that follow all use agents. In agent-based (in Glossary) modeling, quasi-autonomous, decision-making entities called agents are key components; see Appendix for available multi-agent M&S platforms. Other components, such as those representing specific compounds (biochemicals or xenobiotics), can be simple, reactive objects or properties of spaces. Reactive objects and agents follow sets of rules that govern their actions and interactions with other system components. In this context, an agent is a biomimetic object that can be quasi-autonomous; it has its own agenda and can schedule its own actions, much like we envision a cell or mitochondrion doing. When needed, an agent can change its operating logic. Agent-based modeling facilitates the production of systemic behaviors and attributes that arise from the purposeful interactions of changeable components. The resulting biomimetic analogues have advantages when attempting to understand and simulate phenomena produced by systems of interacting components, and that makes them prime candidates for bridging the gap in Fig. 1.

If we had analogues of the type just described, how would we use them in the context of drug discovery and development research? An important use would be to understand the mechanisms that generate disease-related phenomena and how compounds or formulations that interact with the mechanisms can alter those phenomena. Improved mechanistic knowledge will enable improved predictions, while helping to reduce requirements for new wet-lab experiments.

A feature of object-oriented analogues is that objects and agents can be either atomic or composite. Atomic components define the system's level of resolution—its granularity (in Glossary). Granularity is the extent to which a system is subdivided, with the smallest components being atomic. An atomic object has no internal structure and so cannot be subdivided—it simply uses its assigned logic. Granularity is also the level of specificity or detail with which system content is described: the more fine-grained, the more specific. Objects, both atomic and complex, are pluggable and can be replaced (as distinct from being subdivided) with more fine-grained, composite components that exhibit the same behaviors within the analogue under the same conditions. That replacement can even take place during a simulation. These components can exhibit hierarchical nesting, which makes it feasible to use analogues of this class to begin discovering plausible upward and downward linkages that are needed to enable instantiating (in Glossary) details of genotype-phenotype linkage. When the nested components are relationally grounded, one can avoid many of the multi-scale problems that plague metrically grounded, equation-based, inductive models. The phenomena emerging from mechanisms at one level can be used as input at another level. Greater nesting means more components, and that means more interactions and more simulation time to process, document, and record those interactions. In order to maintain parsimony, analogues should be designed with components that are just fine-grained enough to produce targeted phenomena and achieve the analogue's specified uses.

Representing Chemical Entity Attributes and Dynamics Within Biomimetic Analogues

- Components base their actions on information presented by the mobile objects (compounds) they encounter.
- Each type of component–compound interaction is a simple micro-mechanism.
- Components use simple logic to tailor their micro-mechanism to a subset of a compound's properties.

Recent reviews of in silico prediction of absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties and of quantitative structure-activity relationship (QSAR) methods chronicle advances in technical sophistication, plus increased use of consensus modeling. They also point out pitfalls of complicated correlation methods. Satisfaction with the quality of PCP predictions is counterbalanced by disappointment in predictions of organism-based properties for new compounds (56–62). Johnson reminds us (61) that many, nearly equivalent, correlational models are possible, making it easy to choose a wrong model for the next new compound. Furthermore, many of the models have no grounding in biological reality. A take-home message is that useful, reliable ADMET predictions of organism-based attributes of future compounds will need to be based on mechanistic insight. The representations of compounds within synthetic analogues, along with their interactions with system components, provide a new, potentially powerful approach to thinking about organism-based, structure-activity relationships.

Precise knowledge of the stoichiometry of biological component–compound interactions is rarely if ever available. Uncertainties at all levels are common. An advantage of discrete event methods is that both knowledge and ignorance (uncertainties) can be represented concurrently. Precise yet poorly informed assumptions can be avoided. The effective stoichiometry of influential, low level (fine-grained) interactions involving compounds can be represented at almost any convenient granularity level below that of the targeted phenomena, but the mappings from objects representing compounds to their referent molecules are not 1:1. Early in the analogue development process a scientifically justified level (moving down from targeted phenotypic attributes) at which to initially represent compounds is the first at which biological functional units are encountered, and that is typically a coarse-grained representation.

The presence of a compound can be represented as a property of a space or as mobile objects. We focus on the latter. Mobile objects representing chemical entities can map to an arbitrary number of molecules (see Appendix, Representing Compounds). During simulations such objects are typically passive. The agency (interaction-specific programming logic) to determine the outcome of a component–compound interaction typically resides with the biomimetic component or process; in some cases more than one component or process can be involved. An important feature of the synthetic approach from a biopharmaceutical sciences perspective is that each mobile object carries with it all the information that the empowered (active) component, a membrane transporter object for example, needs in order to

adjust and reparameterize its interaction logic. That information can include (or not) selected PCPs that domain experts understand intuitively (molecular weight, log*P*, measures of ionization state, etc.) along with bioactivity attributes (the chemical entity is a CYP 2C9 substrate, etc.). In that way, during a simulation an analogue system can accommodate objects representing any number of compounds, which of course is ideal for studying and exploring drug–drug interactions (63–65). A component empowered to interact can use compound identification information to adjust its engagement logic.

Early in an analogue's development, micro-mechanistic knowledge is insufficient to parameterize component–compound interactions *a priori* using compound-specific information. Micro-mechanism logic must be tuned individually for each of the first few compounds for which referent data are available. As the set of compounds enlarges, inductive modeling methods within the analogue's larger framework can be used, as in (66,67), to establish quantitative mappings from patterns in chemical entity information to patterns in parameter values of tuned component–compound interaction. Such a mapping will be the analogue's counterpart to a structure-activity relationship. In subsequent rounds of analogue use and refinement, the new knowledge contained in that relationship can be used, in some cases automatically, to provide an initial analogue parameterization for the next chemical entity to be studied. Simulations using those parameterizations will stand as crude predictions of the new compound's targeted attributes. The limited (artificial) intelligence available to analogue components at that stage can be improved systematically as the analogue is iteratively validated against wet-lab data for additional compounds. That primitive intelligence can be shared between different analogues and models within an organization's larger M&S framework.

Each type of component–compound interaction is a micro-mechanism. The micro-phenomenon that results is typically simple: transport occurred (or not), metabolism occurred (or not), spatial relocation occurred (or not), etc. When such phenomena are studied *in vitro* in simple systems, one often observes that just a few molecular descriptors account for the majority of the data variance for the compounds studied. Even when the more complex ADMET properties are analyzed collectively, simple, interpretable rules of thumb emerge (68). Given the simple and stochastic nature of most synthetic micro-mechanisms, a small change in the PCP space can correspond to a negligible or modest change in the micro-phenomenon, as well as the parameterization of its logic. Computationally simple methods are expected to suffice in predicting acceptable, micro-mechanism parameter values. The precision of estimated micro-mechanism parameterizations can be expected to vary randomly across the analogue. Nevertheless, the predicted, systemic target phenomena can still be accurate enough for R&D decision-making; see (67) for an example.

ANALOGUES: FROM *IN VITRO* TISSUES TO INTERACTING ORGANS

The evolution of object- and agent-oriented biological system models over the past decade is interesting but outside the scope of this article. Currently, there are no analogues that bridge the gap in Fig. 1. Analogues capable of beginning to bridge the gap began appearing only recently. Because model use has been different in each case, a straightforward comparison of what those different models do (or do not do) and how they do it would be misleading. Rather, the objective here is to provide examples of analogues that can evolve to become gap-spanning, scientifically useful in silico systems. We sought examples that would enable readers to envision analogues that could be useful within their own R&D domains. With that in mind, we limited examples to those that included drugs (one case) or to which objects representing drugs could be added (three cases) without requiring system reengineering. Because this field is relatively new, all four examples are early stage and still somewhat abstract. However, it is easy to imagine variants of each, in parallel, becoming incrementally more sophisticated and realistic.

Filling the Need for an Epithelial Cell Culture Analogue That Has Its Own Phenotype and Plausible Operating Principles

Even though their phenotypes are complex, *in vitro* cell cultures are among the simplest biological systems. Early examples of in silico explorations into mechanisms *in vitro* using CA and CPMs include (14,69,70). Thereafter, experimentation with agent-oriented methods increased. Within the past three years, considerable progress has been made improving *in vitro* mechanistic insight using primarily more sophisticated, cell-centered (71), and agent-oriented methods (53,64,72–77), including exploration of events in a crowded virtual cytoplasm (78). See Supplemental Material for a thorough listing of research progress that used variations of the synthetic method during the intervening years.

There are several agent-based, cell-centered, synthetic analogues to which drug objects could be added. Walker *et al.* (79) provide an early example representing cell–cell mechanisms of interaction using a synthetic, agent-based approach. Bindschadler and McGrath (72) achieved new insight into mechanisms of wound healing using an agent-based approach in which components were grounded purposefully to metric spaces, enabling direct comparison of simulated and wet-lab measurements. Zhang *et al.* (80) developed a sophisticated, 3D, multi-scale, agent-oriented analogue of solid tumor growth in which agent logic was controlled in part by conceptualizations of molecular details and gene-protein interaction profiles. Key features were grounded to metric spaces. The example that follows, even though drawn from our own work, was selected because the system uses relational grounding (discussed in "How a Model is Grounded Impacts How It Can and Should be Used").

The analogue is a simple synthetic analogue of human alveolar type II (AT II) epithelial cell cultures (81) similar to those used in drug discovery and development research (82–84), where selected system attributes are measured in the presence and absence of compounds of interest. Below, the discussion focuses on attributes in the absence of compounds. The targeted attributes are aspects of AT II cystogenesis in 3D matrix, which recapitulates several basic features of mammalian epithelial morphogenesis (85). To gain insight into the process, Kim *et al.* (81) created a concrete, stand-alone, in silico "cultured cell" system that had its own unique phenotype. The system was then refined iteratively so that its

phenotype had the relationship to *in vitro* cell phenotypes that is illustrated in Fig. 3. A case was made that when mappings from in silico components and interactions to their biological counterparts are intuitive and clear, even if abstract, then one can hypothesize that the causal events in silico have *in vitro* counterparts. Accepting the reasonableness of these mappings enabled making an important, additional claim: a mapping will also exist between in silico operating principles and cellular operating principles.

Representing an Epithelial Cell Culture as a Dynamic System of Coarse-Grained, Interacting Components

Detailed descriptions and methods are available in (81,86). Building blocks and their functions, along with assembly methods were proposed so that components and the assembled analogue mapped logically to wet-lab counterparts. Data accumulated during executions were compared against referent wet-lab data. When the analogue failed validation, it was revised and tested iteratively until pre-specified behaviors were achieved. The assembled components and their operating methods stood as a hypothesis: these mechanisms will produce targeted characteristics. Execution and analysis of results tested that hypothesis. Hereafter, to clearly distinguish in silico components and processes from corresponding wet-lab structures and processes, we use SMALL CAPS when referring to the in silico counterparts.

To produce the epithelial cell analogue, a cell culture was conceptually abstracted into four components. Cells, matrix (media containing matrix), free space (matrix-free media), and a space to contain them had in silico counterparts: CELL, MATRIX, FREE SPACE, and CULTURE. MATRIX and FREE SPACE were passive objects. MATRIX mapped to a cell-sized volume of extracellular matrix. A FREE SPACE mapped to a similarly sized volume that was essentially free of cells and matrix elements. Cells were quasi-autonomous agents, which mimicked specified behaviors of AT II cells in cultures. Each used a set of rules or decision logic to interact with their local environment. When two or more CELLS attached, they acted quasi-autonomously, independent of individual CELL activities. The CULTURE used a standard 2D hexagonal grid to provide the space in which its objects resided and moved about. The simulation was executed in discrete time steps, during which each CELL, in pseudo-random order, took actions based on its internal state and external environment. Having objects update pseudo-randomly simulated the parallel operation of cells in culture and the nondeterminism fundamental to living systems, while building in a controllable degree of uncertainty.

The initial list of targeted attributes was obtained from *in vitro* studies of primary human AT II cells (85). Observations determined that there was neither cell death nor proliferation. Achieving the targeted attributes (Fig. 5) required just three action options: migrate, attach to an adjacent CELL, or rearrange within a CLUSTER (Fig. 4B). Cells required three types of migration, separately or in pairs: random movement, CHEMOTAXIS, and CELL density-based migration. A CELL could switch its migration mode during execution. The CELL–CELL attachment action executed when two CELLS were in contact. CELL rearrangements within a CLUSTER were specified using axiomatic operating principles (81).
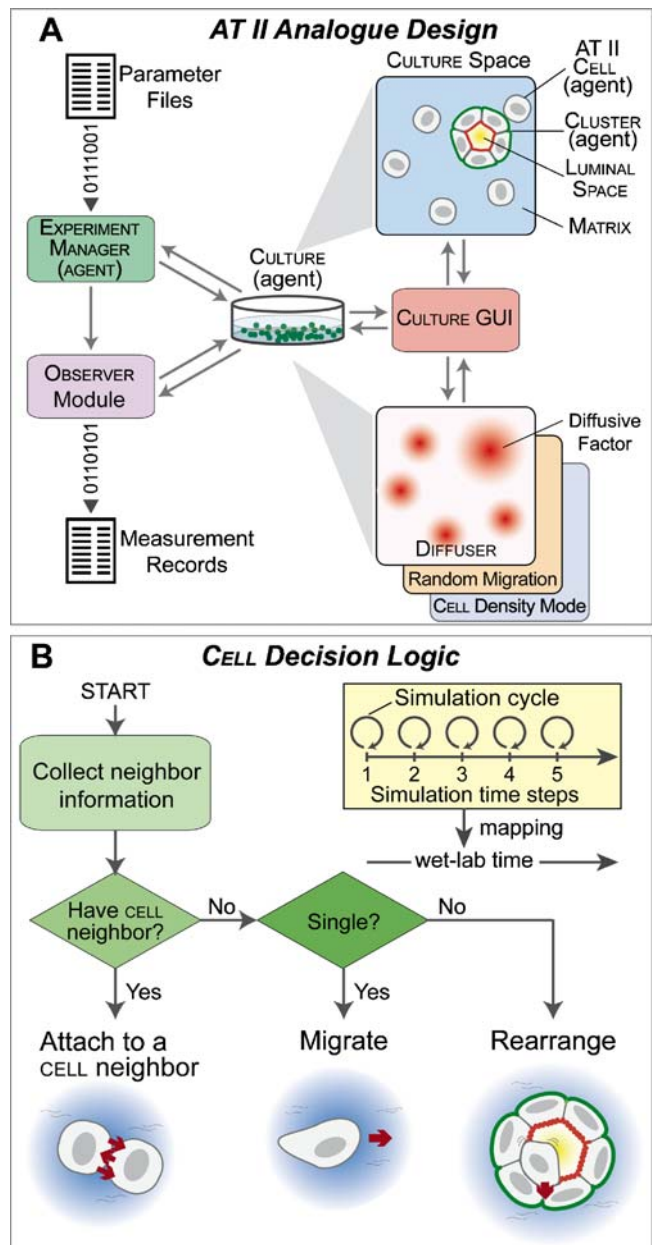

Fig. 4. AT II analogue design and CELL logic. **A** Shown is the AT II analogue design from (81). A hexagonal grid provides the space within which the four components interact. Cells are quasi-autonomous agents, which mimic AT II cell behaviors *in vitro*. Diffuser is a space to simulate diffusion of an abstract factor used to guide CHEMOTAXIS. The system-level components included EXPERIMENT MANAGER (the top-level system agent), OBSERVER (recorded measurements), and CULTURE graphical user interface (GUI). **B** Simulation time advances in steps corresponding to simulation cycles. Each simulation cycle maps to an identical interval of wet-lab time; during a cycle, every CULTURE component is given an opportunity to update. Every CELL, selected randomly, decides what action to take based on its internal state (clustered or single) and the composition of its adjacent neighborhood. Enabled CELL actions are CELL–CELL attachment, CELL migration, and rearrangement within a CLUSTER. A CELL within a CLUSTER can rearrange with other CELLS composing the CLUSTER, driven by a set of axiomatic operating principles (see (81) for specifics).
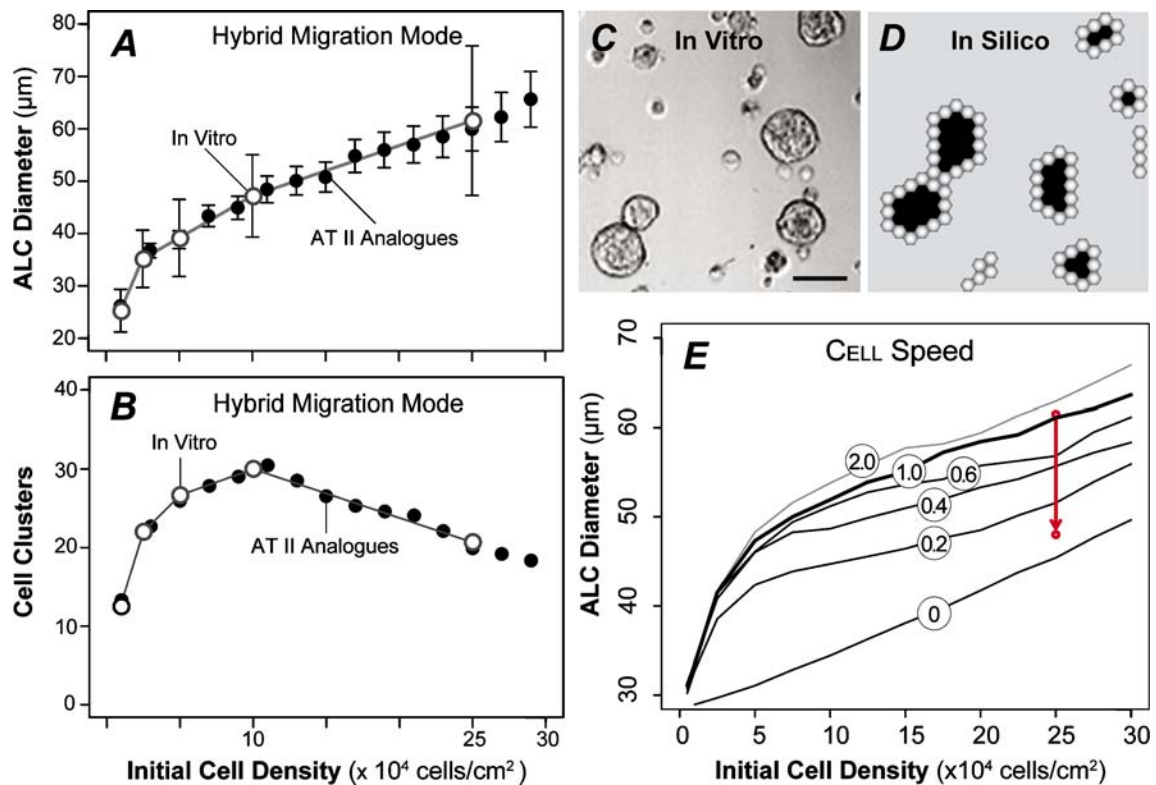
Fig. 5. AT II analogues and AT II cell cultures can exhibit quantitatively similar, phenotypic attributes. **A** Mean ALC diameters, both in silico and wet-lab, are graphed as a function of initial cell density. *Open circles*: mean *in vitro* diameter after 5.7 days; *vertical bars*: ±1 SD (*n*=25). *Filled circles*: mean analogue diameters after 100 simulation cycles (~6.1 days); *bars*: ±1 SD (*n*=100). The dominant migration mode was CELL density-based. At initial densities of ≤ 2,000 CELLS 10–15% of CELLS moved randomly; at higher initial densities, movement was CELL density-based. **B** *Open circles*: final, mean cluster count (averaged over three culture wells) for the *in vitro* experiments in **A**. *Filled circles*: final mean CLUSTER count in **A**. **C** Phase-contrast pictures after 4 d in 2% Matrigel. *Bar*: ~50 μm. **D** A sample image of simulated culture after 100 simulation cycles starting with 2,000 CELLS. Note that a hexagonal CYST within the discretized hexagonal space maps to a roundish cross-section through an ALC *in vitro*. Objects with *white centers* are CELLS. *Gray* and *black spaces* represent MATRIX and FREE (or LUMINAL) SPACE, respectively. **E** Shown are the consequences of changing CELL speed CELL in density-based mode. Speed (*circled*) is in grid units per simulation cycle; CELLS in **A–D** migrated 1 grid unit/cycle. Values are based on 100 Monte Carlo runs for 100 simulation cycles. The *arrow* pointing down shows the observed change in mean ALC diameter at the indicated initial cell density when Matrigel density was increased from 2% to 10%. Images adapted from (81).

Hypothesizing and Testing Plausible Mechanisms of AT II Cystogenesis

A sequence of increasingly stringent Similarity Measures (SMs) was used for validation and analogue refinement (81). So doing allowed discovery of constraints and mechanism changes (including logic adjustments or changes) that moved the analogue's behavior space (phenotype) toward targets and thereafter shrank it. Initially, every CELL was in a single, non-clustered state. As the simulation progressed, CELLS produced CULTURE level behaviors that qualitatively and quantitatively matched those observed *in vitro* (Fig. 5). Migrating single CELLS formed CELL–CELL attachments, which led to formation of small CLUSTERS. Some CLUSTERS migrated and merged with CELLS and other aggregates to form larger CLUSTERS. Cells within CLUSTERS rearranged themselves into configurations dictated by the axioms, causing adequately sized CLUSTERS to develop progressively into alveolar-like CYSTS (ALCS) having FREE SPACE surrounded by a CELL monolayer. ALCS maintained convexity and had no dimples; most remained stable until the simulation terminated. Note

that a structure having a regular hexagonal shape in hexagonal grid space maps to a circle in continuous space. For AT II CYSTOGENESIS, CELL activity patterns during simulations made clear how their mandates, the targeted attributes, are achieved. That clarity provided insight into and plausible explanations of AT II cystogenesis *in vitro*.

In vitro, the average ALC size increased monotonically with initial cell density. Similar patterns were observed during simulations: mean values and their standard deviations are graphed in Fig. 5A. In sparse CULTURES with < 1,000 CELLS, which mapped to ~1×10⁴ cells/cm² *in vitro*, CELLS formed small ALCS with diameters that were essentially the same as the referent mean diameter. In denser CULTURES, larger ALC diameters were observed. Changes in the number of clusters, as a function of initial cell density, were the same for both *in vitro* and in simulations (Fig. 5B). AT II cell migration speed was an important determinant of aggregation and ALC formation in 3D cultures. Intuitively, one would expect to achieve the production of larger ALCs by elevating cell speed. Doing so would increase the cell collision rate and thus accelerate aggregation. Conducting such experiments *in vitro*

is infeasible because a minimum level of extracellular matrix is required to sustain normal AT II cell behaviors. Testing that hypothesis for AT II analogues was straightforward and achieved by parametrically changing CELL migration speed. Slowing migration speed was expected to correlate to increased extracellular matrix densities. As shown in Fig. 5E, the results predicted a dramatic reduction in ALC formation when the extracellular matrix is stiffened. The prediction was tested and confirmed *in vitro* at a high initial cell density (85).

Simulating Tissue Responses *In Vivo*

Achieving Micro-mechanistic Insight to Ischemic Microvascular Injury

The work by Bailey *et al.* (87) demonstrates the potential of using synthetic models for hypothesis generation and knowledge discovery, and is also an excellent example of multiple iterative cycles of in silico experimentation coupled with wet-lab experimentation. They constructed a multi-cell, tissue-level, agent-oriented model of human adipose-derived stromal cell (hASC) trafficking through the microvasculature of skeletal muscle tissue after acute ischemia.

After ischemic injury to microvasculature, blood is re-routed to adjacent microvascular networks causing swelling and increases in wall shear stress and hydrostatic pressures at the adjacent site. The changes in wall shear stress and circumferential stress activate endothelial cells and perivascular cells, initiating the recruitment of circulating cells into the site of injury. Activated endothelial cells increase their surface expression of important cellular adhesion molecules that enable the circulating cells to home to the site of ischemic injury, adhere to the endothelium, extravasate, and incorporate into the injured tissue. In addition, activated endothelial cells and perivascular cells secrete a number of inflammatory chemokines, cytokines, and growth factors. These further activate endothelial and perivascular cells, as well as activating circulating cells in order to promote their recruitment into the site of ischemic injury. Intravenous delivery of hASC has been shown to help repair and regenerate injured tissue from ischemia. An analogue was used to gain mechanistic insight into that process.

Bailey *et al.* constructed an agent-based model to identify potential bottlenecks that may limit the efficiency of these therapeutic cells being recruited into the site of ischemic injury after intravenous injection. It was their hope that better clinical outcomes could be achieved through increasing the number of incorporated hASCs.

They used confocal immunohistochemistry images to manually reconstruct in silico the morphology of a characteristic microvascular network (Fig. 6). The endothelial cells lining the vessel surface, tissue resident macrophages, circulating monocytes, and therapeutically delivered stem cells were represented as individual software agents. The agent-based model was coupled with a network blood flow analysis program that calculated blood pressure, flow velocities, and shear stresses throughout the simulated microvascular network.

Individual in silico CELL behaviors were determined from a set of over 150 rules that were derived from independent literature. Each ENDOTHELIAL CELL, MONOCYTE, and HASC could, in a binary manner, either be positive or negative in their expression of each CELLULAR ADHESION MOLECULE. In a similar binary manner, each ENDOTHELIAL CELL, MONOCYTE, and TISSUE resident MACROPHAGE could either be in a positive or negative state of secretion for each of the CHEMOKINES and CYTOKINES. Thus, each CELL could be unique. A CELL's state of CELLULAR ADHESION MOLECULE expression and state of CHEMOKINE and CYTOKINE SECRETION were also dynamic, and depended upon the CHEMOKINE and CYTOKINE SECRETION states of neighboring CELLS (Fig. 6D).

Whether a circulating MONOCYTE or HASC rolled or adhered was dependent on a combination of factors. They had to experience the correct combination of CELLULAR ADHESION MOLECULE expression states and CHEMOKINE SECRETION states from a nearby ENDOTHELIAL CELL, and have also experienced a WALL SHEAR STRESS below a certain threshold level. If the CELL adhered for more than a specified number of time steps, it could then TRANSMIGRATE into the TISSUE space.

Bailey *et al.* simulated a microvascular network under normal conditions and after ischemic injury. The MICROVASCULAR network was representative of a microvascular network adjacent to the site of ischemic injury, where blood flow is increased due to its redistribution. When simulating ischemic injury, the PRESSURE at the feeding ARTERIOLE was increased by 25% and the resultant HEMODYNAMIC properties were re-calculated.

Simulation Experiments Implicated an Additional Cellular Adhesion Molecule

The analogue was verified without HASC by performing a series of in silico knockout experiments, comparing simulated and wet-lab phenomena where data were available, and by showing that the analogue could mimic three aspects of ischemic injury: (1) increase in wall shear stress and network rates; (2) up-regulation of specific cellular adhesion molecules expression by the endothelium; and (3) increased secretion levels of chemokines and cytokines. Additionally, two key monocyte properties had in silico counterparts. Following verification, they simulated hASC trafficking after INTRAVENOUS injection before and after ISCHEMIC injury. Lower than expected levels of HASC EXTRAVASATION were observed, and that led to a re-evaluation of their rule-set. hASCs do not express PSGL-1, however they hypothesized that there may exist an additional cellular adhesion molecule used for rolling, similar to PSGL-1, for which there was no counterpart in their analogue.

To explore that hypothesis, they included an additional CELLULAR ADHESION MOLECULE, termed SBM-X, with properties similar to PSGL-1 and determined whether so doing allowed the analogue to more closely mimic *in vivo* experimental results. Simulations showed that the inclusion of the new MOLECULE SBM-X was necessary to achieve targeted levels of HASC EXTRAVASATION. They subsequently tested their hypothesis *in vitro* and showed that small fractions of hASCs are able to roll on P-selectin even though they do not express PSGL-1. They proposed that the cellular adhesion molecule CD24 is a likely SBM-X candidate.
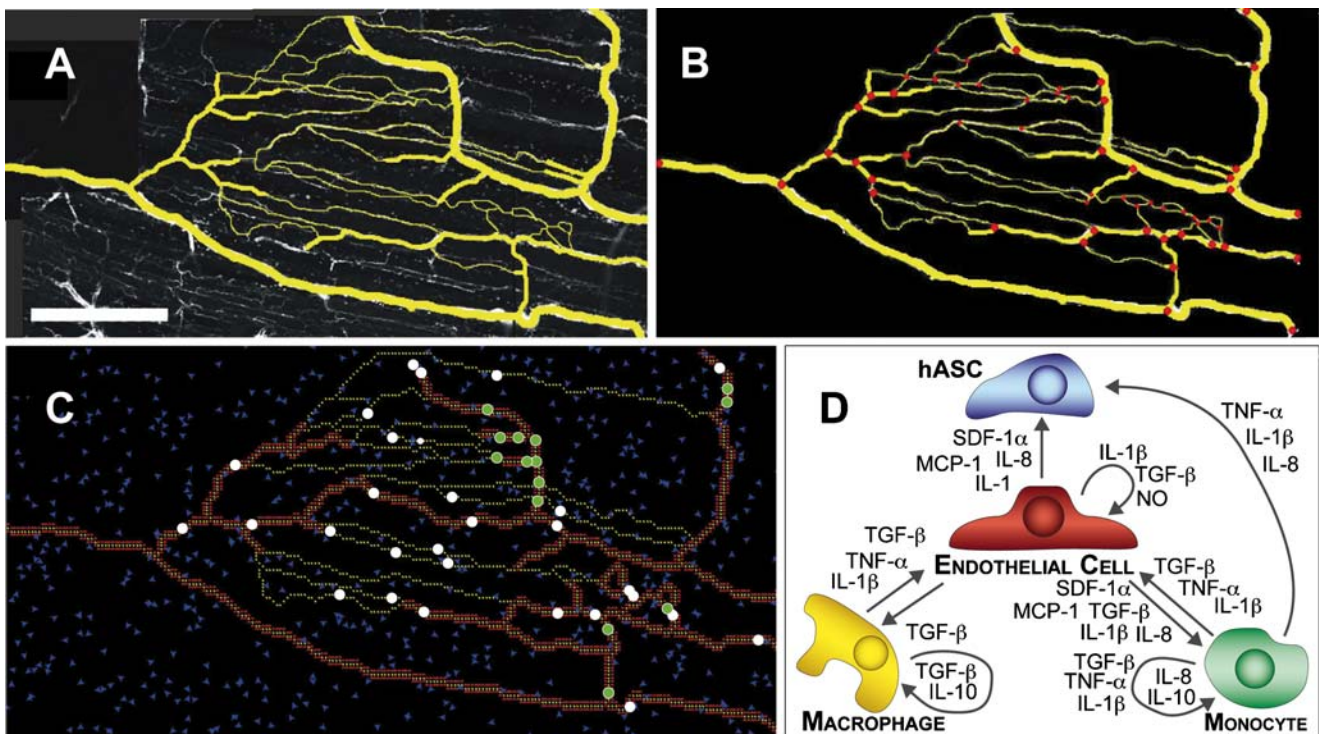
Fig. 6. Simulated human, adipose-derived, stromal cell (hASC) trafficking through the microvasculature during acute skeletal muscle ischemia (adapted from (87) with permission). The referent for the MICROVASCULAR NETWORK was skeletal muscle visualized using confocal microscopy following harvest, using a 20× objective. **A** Confocal microscopy image of mouse spinotrapezius muscle immuno-stained to visualize epithelial cells having BS1-lectin antibody (*white*). Vascular structures of interest were copied (*yellow*). Arterioles and venules were characterized based on vessel diameter. *Scale bar*: 1 mm. **B** The network in **A** was manually discretized into nodes (bifurcation points, marked *red*). Nodes were connected to form elements. **C** Screen-shot of simulation space. Nodes and elements were manually constructed within a NetLogo simulation space to mimic the referent network in A. *Red* SMOOTH MUSCLE CELLS line ARTERIOLES and VENULES. Simulated hASHs that have successfully EXTRAVASATED are *green*, otherwise they are *white*. ENDOTHELIAL CELLS are *yellow*; TISSUE MACROPHAGES present within the INTERSTITUM are *blue*. **D** Illustration of the complex and dynamic connections between the four CELL types. The listed, referent chemokines and cytokines have all been implicated in human ischemic injury. *Arrows* indicate connections between CELL populations and denote some combination of the following: induced SECRETION, changes in CELLULAR ADHESION MOLECULE expression, and/or INTEGRIN activation. All connections between nodes were based on relevant, independent, experimental literature. Images are adapted from (87).

In Silico Livers

An In Silico Liver is Constructed by Assembling Simple Components into a Larger, Multi-Level, Biomimetic Structure

In Silico Livers (ISLs) are advanced examples of biomimetic analogues designed specifically to help bridge the gap in Fig. 1. An ISL is not intended to be a model having a temporally stable structure. It is designed to be altered easily in order to explore many equally plausible mechanistic explanations for disposition-related observations. The ISL is an assembly of componentized mechanisms: purposefully separated and abstracted aspects of hepatic form, space, and organization interacting with COMPOUNDS. Each component mechanism has been unraveled from the complex whole of the hepatic-drug phenotype; it has its own unique phenotype, but that phenotype is much simpler than that of the entire lobule mechanism.

ISL-targeted attributes were divided into three classes (88): 1) compound-specific time course data, 2) microanatomical details (heterogeneous sinusoid structures, perivenous cross-connections between sinusoids, etc.), and 3)

experiment details (perfusion is single-pass or not, administered compounds pass though catheters and large vessels before entering lobular portal vein tracts, etc.). A primary use has been to use and reuse similar ISL structures to provide plausible micro-mechanistic explanations of hepatic disposition data for many drugs.

An ISL maps to a mammalian liver undergoing perfusion as in (89). It is physiologically based, yet abstract. An ISL represents a liver as a large, parallel collection of similar lobules (Fig. 7A). An ISL's functional unit is the same as the liver's: a lobule. Lobule structure is illustrated in Fig. 7B–D and detailed in (88,90,91). The following is an abridged description. Components mimic essential form and function features of a rat liver. Acinar flow patterns are represented by an interconnected, directed graph (Fig. 7B). Graph edges specify flow connections between objects called Sinusoidal Segments (SS). An SS, which maps to a unit of sinusoid function, is placed at each graph node. Multiple, different flow paths from portal vein tracts (PV) to central vein (CV) are present, as illustrated in Fig. 7B.

The SS structure shown in Fig. 7C maps to a unit of sinusoid function that includes spatial features. An SS is a
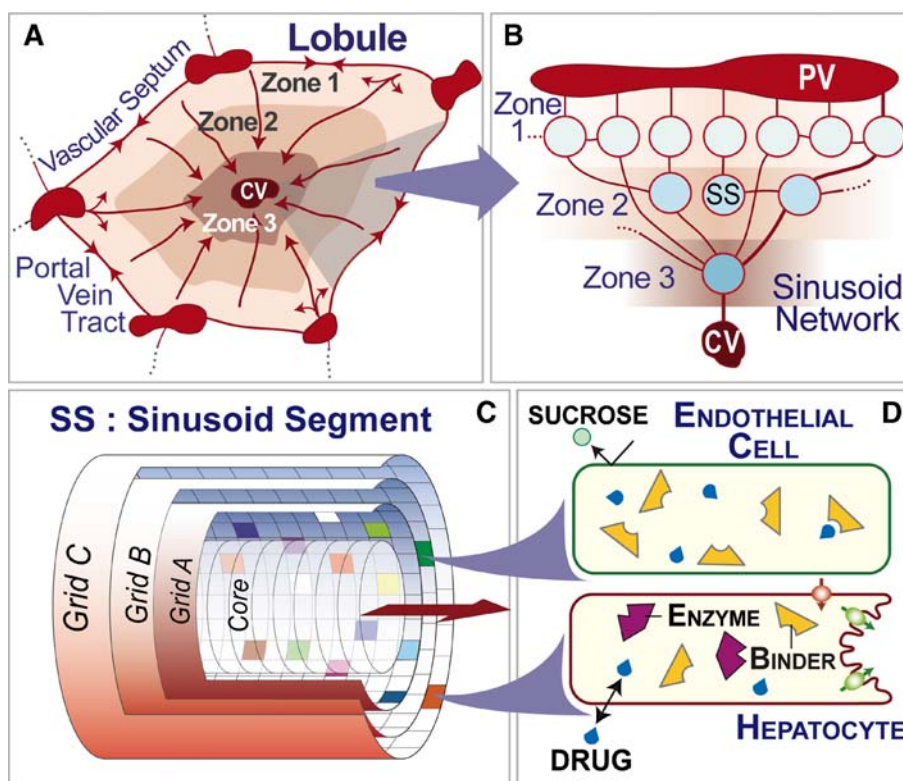
Fig. 7. Illustrated are hepatic lobular structures and their ISL counterparts. **A** A schematic of a cross-section of a hepatic lobule showing the direction of flow from the terminal protal vein tracts (PV) through sinusoids in three concentric zones to the central hepatic vein (CV). Different zones can have quantitative differences in structural and functional characteristics. **B** A portion of the SINUSOID NETWORK is shown. It is an interconnected, three-zone, directed graph (lines connecting shown as *circles*). It maps to a portion of a lobular sinusoid network. Data from the literature are used to constrain the graph size and structure. *Circles*: sinusoidal segments (SS). **C** A schematic of a sinusoidal segment (SS): one SS occupies each node specified by the directed graph (in Fig. 7B). Grids map to hepatocyte spaces; they contain objects that map to intracellular functionality. From Grid A, they can access the other spaces. Grid locations have properties and that govern their interaction with mobile COMPOUNDS. Different shadings of Grid A illustrate the potential for representing heterogeneous properties. Objects functioning as containers (for other objects) map to cells, and can be assigned to any grid location. **D** Shown are ENDOTHELIAL CELL and HEPATOCYTE. Objects representing all needed intracellular features can be placed within. Two types of intracellular BINDERS recognize COMPOUNDS: those that simply bind (BINDER) and those that also map to enzymes and can METABOLIZE. Bile attributes can be represented easily when needed.

discretized, tube-like structure comprised of a BLOOD "core" surrounded by three identically sized 2D grids, which together simulate a 3D structure. It can be replaced by more realistic 3D grid when it is required to do so to achieve some targeted attribute. Two SS classes, S_1 and S_2, are specified to provide sufficient variety of COMPOUND travel paths. Compared to an S_2, an S_1 on average has a shorter internal path length and a smaller surface-to-volume ratio.

ISL parameters are grouped into three categories: 1) those that control LOBULE graph and 2) SS structures, and 3) those that control LOBULAR component interactions with COMPOUNDS. Compounds are represented using objects that move through the LOBULE and interact with encountered SS features. A typical COMPOUND maps to many drug molecules (see Representing Compounds in the Appendix). A COMPOUND'S behavior is determined by the PCPs of its referent compound, along with the LOBULE and SS features encountered during its unique trek from PV to CV. During a simulation cycle, an encountered component "reads" the information carried by a COMPOUND and then uses it to customize its response, in compliance with its parameter values, following some pre-specified logic. That feature enables multiple, different COMPOUNDS to be percolating through SS features during the same experiment.

Objects called CELLS (Fig. 7D) map to an unspecified number of cells. They function as containers for other objects. A grid location and its container are the current limit of spatial resolution. Cells contain a stochastic parameter-controlled number of BINDERS in a well-stirred space. Binders map to transporters, enzymes, lysosomes, and other cellular material that binds or sequesters drug molecules. In the cited work, a BINDER within an ENDOTHELIAL CELL only bound and later released a COMPOUND. A BINDER within a HEPATOCYTE is called an ENZYME because it can bind a substrate COMPOUND and either release or METABOLIZE it. Additional objects can be added as needed, as in (65), to represent uptake and efflux transporters, specialized enzymes, and pharmacological targets, without compromising function of objects already

present. Because of the stochastic nature of ISL simulations, each in silico experiment generates a slightly different outflow profile.

Experimenting on ISLs to Better Understand Micro-mechanisms and Predict Hepatic Drug Disposition

In (90), a single parameterized ISL structure was arrived at iteratively, and held constant for antipyrine, atenolol, labetalol, diltiazem, and sucrose. Parameters sensitive to compound-specific PCPs were tuned so that ISL outflow profiles were validated separately and together against rat perfused liver outflow profiles. Each ISL component interacted uniquely with each of the five compounds. The consequences of ISL parameter changes on outflow profiles were explored. Selected changes altered outflow profiles in ways consistent with knowledge of hepatic anatomy and physiology and compound PCPs. That level of validation enabled the authors to posit that static and dynamic ISL micro-mechanistic details, although abstract, mapped realistically to hepatic mechanistic details. Because ISL mechanisms are built from finer-grained components, there is precise control over conflation. The causal basis is present in the component–COMPOUND interaction logic (axioms, rules). An expectation has been that at some level of granularity, the complexity will be sufficiently unraveled so that the logic for a given micro-mechanism (its "causal basis") will rely heavily on only a few easily specified COMPOUND and biological attributes.

Subsequently, in (91), quantitative mappings were established between drug PCPs and ISL parameter values for the above four sets of drug PCPs and the corresponding sets of PCP-sensitive, ISL parameter values. Those relationships were then used to predict PCP-sensitive, ISL parameter values for PRAZOSIN and PROPRANOLOL given only their PCPs. Relationships were established using three different methods: 1) a simple linear correlation method, 2) the Fuzzy c-Means algorithm, and 3) a simple artificial neural network. Each relationship was used separately to predict ISL parameter values for PRAZOSIN and PROPRANOLOL given their PCPs. Those values were then used to predict disposition details for the two drugs. All predicted disposition profiles were judged reasonable (well within a factor of two of referent profile data). The parameter values predicted using the artificial neural network gave the most precise results. More noteworthy, however, was that the simple linear correlation method did surprisingly well. That is because the ISL is an assembly of micro-mechanisms where each is influenced most by a small subset of PCPs. The results suggest that when using the synthetic method of assembling separated micro-mechanism, a parameter estimation method, which reasonably quantifies the relative differences between compound-specific behaviors at the level of detail represented by those micro-mechanisms, will provide useful, ballpark estimates of hepatic disposition. That bodes well for using synthetic analogues for predicting PK properties, given only molecular structure information.

Starting with the ISLs described above, Park *et al.* (21) discovered that they could alter a small subset of ISL parameter values, tuned previously to match diltiazem's outflow profile, and match diltiazem's outflow profile from a diseased rather than a normal rat liver (Fig. 8). Dynamic tracing features enabled spatiotemporal tracing of differences in dispositional micro-mechanisms. Differences in ISL attributes mapped to measures of histopathology. By measuring disease-caused differences in local, INTRALOBULAR and within-zone effects (Fig. 9), they obtained heretofore-unavailable views of how and where hepatic drug disposition *may* differ in normal and diseased rat livers from diltiazem's perspective. The approach and technology represent an important step toward unraveling the complex changes from normal to disease states and their influences on drug disposition.

Because the causal, mechanistic differences occur at the micro-mechanism level, it is easy to morph—transform—a NORMAL ISL into a DISEASED ISL (Fig. 8B). The morphing stands as a hypothesis for how and where disease may have altered hepatic micro-architectural features and processes.
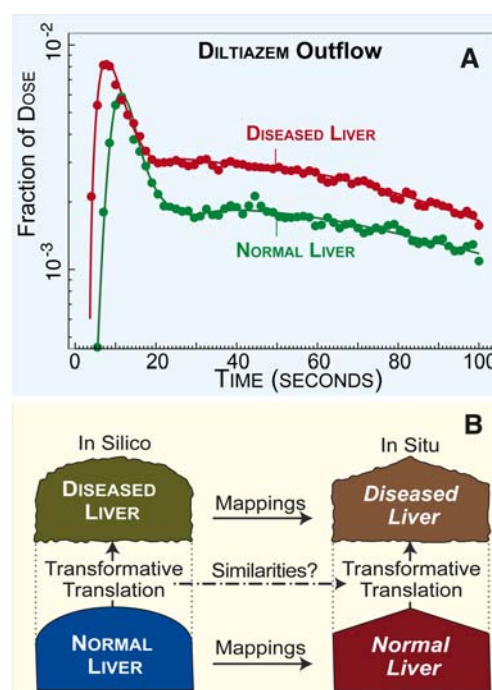

Fig. 8. ISL properties. **A** Outflow profiles of NORMAL and DISEASED ISLs are compared. Values are smoothed, mean DILTIAZEM levels (fraction of dose per collection interval) from the NORMAL and DISEASED ISL that achieved the most stringent, pre-specified Similarity Measure: >90% of simulated outflow were within a factor of 0.33 of corresponding wet-lab values. **B** Illustrated is a simple example of the model-to-model translational mapping mentioned in the "Introduction." Eleven of 25 key ISL parameters' values that were tuned to create the DILTIAZEM outflow profile from a NORMAL LIVER were altered to obtain the validated DILTIAZEM outflow profile from a DISEASED LIVER. Three of the 11 were LIVER structure parameters. Their change mapped to disease-caused changes in referent liver micro-anatomical characteristics. Nine of the eleven were parameters governing movement and interaction of DILTIAZEM with LIVER components, such as moving between spaces and the probability of METABOLISM after being bound to an ENZYME. For the attributes targeted, intermediate ISL parameterizations (of those 11) can be used to document the incremental transformation of a NORMAL to a DISEASED LIVER. The details of such a transformation provide a working, abstract hypothesis for the mechanisms of actual disease progression; they specify what must be changed (morphed) to translate results from one wet-lab model to another.
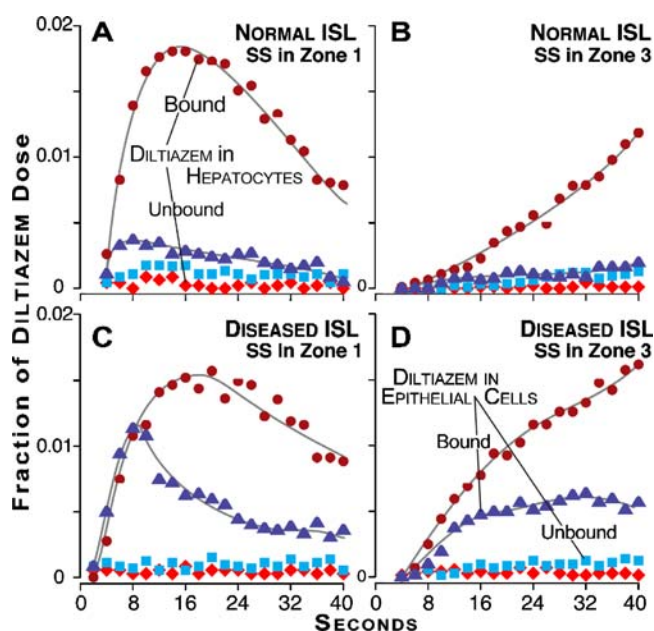
Fig. 9. Temporal changes within comparable ISL micro-mechanisms. Dispositional events within an ISL from a variety of perspectives can be measured following dosing. Doing so gives an unprecedented view of plausible dispositional detail. Examples included measuring BOUND and UNBOUND DILTIAZEM within HEPATOCYTES, number of METABOLIC events occurring within any one of the three Zones, or the fraction of DOSE within a particular SINUSOIDAL SEGMENT (SS). The values graphed include the latter. Values graphed in **A–D** are the amounts of DILTIAZEM in four different states: BOUND and UNBOUND in HEPATOCYTES, and BOUND and UNBOUND in ENDOTHELIAL CELLS. **A** and **B**: a NORMAL LIVER; **C** and **D**: a DISEASED LIVER; **A** and **C**: the focus is SS #14 in Zone 1; **B** and **D**: the focus is SS #33 in Zone 3.

The transformation methods are generalizable. For example, a validated analogue of one *in vitro* cell culture system can be morphed into a different analogue representing a second *in vitro* cell culture system (and vice versa). The process will present a dynamic hypothesis of where and how COMPOUND interaction properties differ between analogues.

It is impractical to obtain liver perfusion data for large numbers of compounds. However, *in vitro* disposition properties can be measured using cultured hepatocytes and other cell types. An advantage of the componentized analogue approach is that an ISL validated for several compounds can be re-used to obtain ballpark estimates of hepatic disposition properties of other compounds by using *in vitro* data and taking advantage of ISL component replacement capabilities. Sheikh-Bahaei *et al.* (66) validated HEPATOCYTE monolayers for four different compounds. The referent system was hepatocytes in a sandwich-culture system that enabled estimating biliary excretion. Their HEPATOCYTES were based on the same container object concept used to create ISL HEPATOCYTES. That similarity opens the door to unplugging the HEPATOCYTES from an ISL that has been validated for several compounds and replacing them with HEPATOCYTES that have been tuned and validated *in vitro* for one or more other compounds for which no liver perfusion data is (or will be) available. Following adjustments based on cross-model validation studies, the outflow profile from such an ISL, given the new compound's PCPs, will stand as a ballpark prediction of that drug's hepatic disposition.

An Analogue of Interacting Organs

Unraveling the Mechanisms of Systemic Inflammatory Response Syndrome and Multiple Organ Failure

Gary An (18) engineered a multi-level, two-organ analogue (gut and lung) to explore plausible causal mechanisms responsible for the clinical manifestation of multi-scale disordered acute inflammation, termed systemic inflammatory response syndrome (SIRS) and multiple organ failure (MOF), and how they may respond to therapeutic interventions (92). Following iterative refinement and parameter tuning, An discovered course-grained, multi-level analogue mechanisms that achieved several targeted attributes related to SIRS and MOF. Simulations used abstract and discrete analogues of gut and lung, each comprising fixed cell-mimetic agents forming ENDOTHELIAL and EPITHELIAL TISSUES, mobile cell-mimetic agents corresponding to inflammatory cells, and mobile objects that mapped to pro-inflammatory mediators. Although exploration of the consequences of drug interventions was not part of the latest study, system design enables adding drug-mimetic objects when the need arises.

Component interactions have biological counterparts that extend from intracellular mechanisms to clinically observed phenomena in the intensive care setting. Different types of CELL agents encapsulate specific mechanistic knowledge extracted from *in vitro* experiments. The model was used to explore the likelihood of the two prevailing hypotheses about the nature of disordered systemic inflammation: that it is a disease of the endothelium or that it is a disease of epithelial barrier function. The former paradigm points to the endothelial surface as the primary communication and interaction surface between the body's tissues and the blood, which carries inflammatory cells and mediators. However, there is also compelling evidence that organ dysfunction related to inflammation is primarily manifest in a failure of epithelial barrier function. An's multi-level analogue of interacting gut and lung enabled exploration of plausible mechanisms that unify those two hypotheses. Simulations produced qualitative phenomena that mimicked attributes of multi-organ failure: severe inflammatory insult to one organ led to both organs failing together.

Abstract, Coarse-Grained Representation of Epithelial and Endothelial Tissues Responding to INFLAMMATION

The 3D TWO-ORGAN system was comprised of the six layered, 2D spaces (square grids) shown in Fig. 10. Greater detail was not needed for the attributes targeted. Together they formed the GUT–LUNG axis. From bottom they are Z=0, 1, 2... 5. Different CELL types populated each 2D grid. The bottom three layers formed the GUT; the top three formed the LUNG. The two ORGANS shared the same design (92,93): a layer of EPITHELIAL CELLS on the top (Z=2 & 5) and a layer of ENDOTHELIAL CELLS in the middle (Z=1 & 4). The bottom layer of each ORGAN (Z=0 & 3) was a space for mobile INFLAMMATORY CELLS, CYTOKINES, etc. Cells in each layer were able to influence the state of CELLS in the layers above and below.

In deference to the parsimony guideline, each ORGAN was composed of an EPITHELIAL surface, which determined
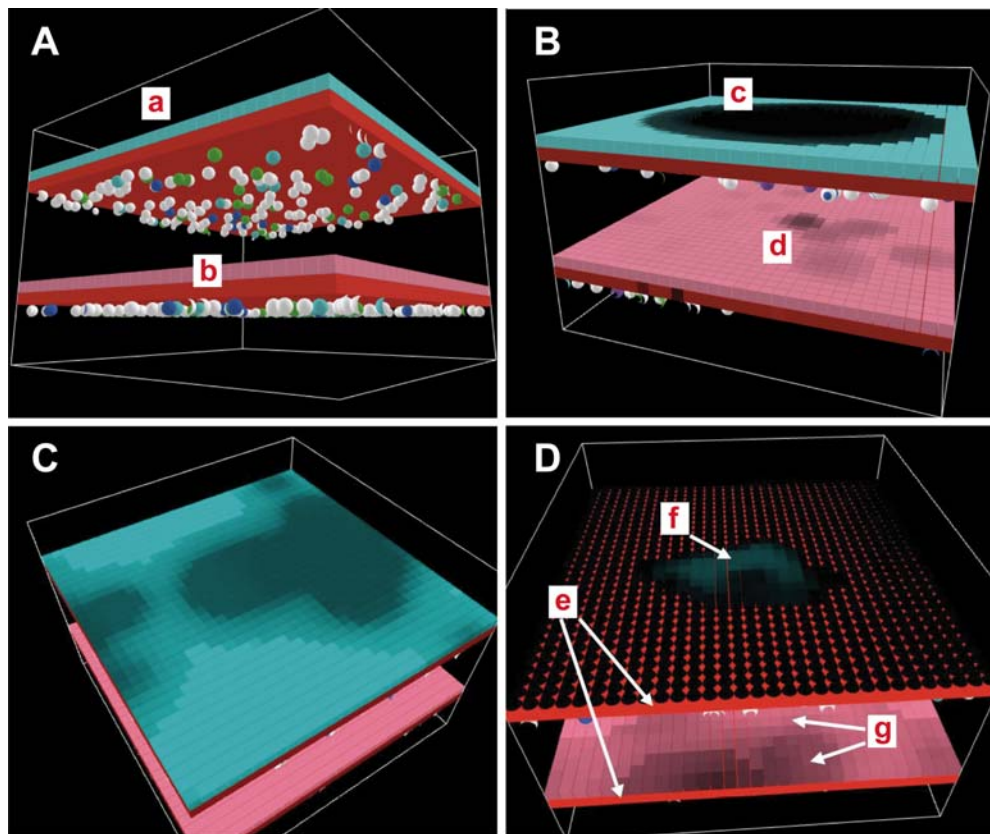
Fig. 10. Screenshots of Multi-Bilayer GUT–LUNG Axis (adapted from (18)). All illustrations are from (18) with permission. **A** Illustrated is the multiple bilayer topology of the interacting GUT–LUNG system. *Letter a*: the PULMONARY bilayer; on top (*aqua*) is the layer of PULMONARY EPITHELIAL CELLS. Each CELL in this and the other layers is a separate agent. Below that (*red*) is the layer of PULMONARY ENDOTHELIAL CELLS. Below that are spherical INFLAMMATORY CELLS. *Letter b*: the GUT bilayer; the three, similarly configured layers are comprised of the same cell types as above. Top (*pink*): GUT EPITHELIAL CELLS; middle: (*red*) GUT ENDOTHELIAL CELLS; bottom: INFLAMMATORY CELLS. Circulating INFLAMMATORY CELLS move between the GUT–LUNG bilayers. **B** Shown is an example of GUT barrier dysfunction. The GUT–LUNG system was run starting with PNEUMONIA as the initial perturbation. *Letter c*: the localized INJURY to the PULMONARY bilayer; *letter d*: the shaded areas demonstrate areas of the GUT EPITHELIAL layer experiencing impaired TIGHT JUNCTION PROTEIN METABOLISM due to GUT ISCHEMIA from decreased SYSTEMIC OXYGENATION arising from PULMONARY EDEMA. **C** Shown are the effects of GUT ISCHEMIA on PULMONARY OCCLUDIN levels (serving as a proxy for PULMONARY barrier dysfunction) after 72 h of an experiment that started with sub-lethal ISCHEMIA. Levels of both CYTOPLASMIC and CELL WALL OCCLUDIN levels reached a nadir at ~24 h (not shown). Thereafter recovery progressed as INFLAMMATION subsided. Shades of *gray*: partially recovered PULMONARY EPITHELIAL CELLS. **D** Shown are the PULMONARY effects 72 h into an experiment that started with a lethal level of ISCHEMIA but also with supplementary OXYGEN to 50% (up from normal level of 21%), which added to the OXYGEN levels that could be generated by the damaged LUNG. CYTOPLASM and CELL WALL OCCLUDIN levels dropped to minimal levels by 12 h (not shown). The supplementary OXYGEN blunted the effects of PULMONARY EDEMA by keeping OXYGEN levels above the ISCHEMIC threshold for ENDOTHELIAL CELL activation. Consequently, ENDOTHELIAL CELLS survived the interval of most intense INFLAMMATION and that allowed EPITHELIAL CELLS to begin recovering their TIGHT JUNCTIONS. *Letter e*: intact ENDOTHELIAL CELL layer. *Letter f*: recovering PULMONARY EPITHELIAL CELLS. *Letter g*: intact and recovering GUT EPITHELIAL CELLS.

ORGAN integrity, and an ENDOTHELIAL/BLOOD interface, which provided for initiation and propagation of INFLAMMATION. The EPITHELIAL CELL layer was validated separately against data from *in vitro* cell monolayer models used to study epithelial barrier permeability. The EPITHELIAL CELL layer was concatenated with the ENDOTHELIAL/INFLAMMATORY CELL layers to produce an abstract, coarse-grained gut analogue. It was separately validated against observations made on *in vivo* wet-lab models of the inflammatory response of the gut to ischemia. Finally, the GUT ORGAN and a similarly

constructed PULMONARY ORGAN were combined to create a gut-pulmonary axis analogue, the behavior of which was expected to map to *in vivo* and clinical observations on the crosstalk between these two organ systems.

The interaction of a layer of ENDOTHELIAL CELLS with a population of different INFLAMMATORY CELLS was described in (92,93). The latter included NEUTROPHILS, MONOCYTES, T-CELLS, etc. All CELLS were agents. The analogue used objects that mapped to specific mediators, including endotoxin, tumor necrosis factor (TNF), and IL-1. Other objects mapped

to receptors, including L-selectin, ICAM, TNF receptors, and IL-1 receptors. The interaction of those objects mapped to signals being transferred between inflammatory cells and endothelial cells. The rules and operating principles used by CELLS were abstract yet strove to reflect current knowledge: positive and negative feedback relationships were implemented using simple arithmetic relationships. Receptor status was expressed as either on or off. The system enabled simulating the dynamics of the innate immune response and exploring plausible mechanisms of systemic inflammatory response syndrome due to a disease of endothelial cells.

EPITHELIAL CELL agents are the smallest functional unit of the GUT–LUNG analogue. Each agent maps to a single epithelial cell in the context of its response to inflammatory mediators, including nitric oxide, and pro-inflammatory cytokines, including TNF and IL-1. Any EPITHELIAL CELL can form a TIGHT JUNCTION with any of its eight EPITHELIAL CELL neighbors. The process required FUNCTIONAL and LOCALIZED TIGHT JUNCTION PROTEINS and could be impaired by INFLAMMATION. That analogue mechanism mapped to the production and localization of tight junction proteins being impaired in a pro-inflammatory cytokine milieu. The EPITHELIAL component was first validated against *in vitro* data. That validated component was used directly as its *in vivo* counterpart. Variables within each CELL controlled levels of TIGHT JUNCTION components along with INTRACELLULAR, pro-INFLAMMATORY signals. The phenomena of pro-inflammatory signals impairing tight junction function, and tight junction dysfunction leading to epithelial barrier failure were targeted attributes. Simple rules specified factor creation and how those factors interacted. Component levels and rules were tuned to achieve a satisfactory degree of similarity between analogue phenomena during simulated treatments and corresponding reported observations. When TIGHT JUNCTION formation was impaired (or its components inhibited), a factor permeated the EPITHELIAL CELL layer in a process that mapped to epithelial barrier failure. The same analogue mechanism was used with LUNG EPITHELIAL CELLS in a process that mapped to pulmonary edema, impaired oxygenation, and further injury.

Linking GUT and LUNG Analogues into a Higher Level, Interacting System of Organs

An inherent property of synthetic system models is composability (the linkage or establishment of component, inter-relationships) downward, by nesting components within components, and laterally, by linking components at a similar level (linkage of the SS in the ISL is an example). By linking GUT and LUNG, An demonstrates the important point that upward composability can yield significant benefits—in the form of improved insight—and was necessary for achieving the research objective.

For the GUT ORGAN, two coupled phenomena were targeted: 1) the gut can fail in the presence of severe ischemia, but 2) it can recover from less severe ischemia. In the GUT, ISCHEMIA interfered with formation of TIGHT JUNCTION components and thus EPITHELIAL barrier function. Each GUT ENDOTHELIAL CELL had an ISCHEMIC INJURY parameter that enabled the control of the proportion of ENDOTHELIAL CELLS having ISCHEMIC INJURY. The latter led to production of a pro-INFLAM-MATORY signal (called CELL-damage-byproduct). The ENDOTHE-LIAL COMPONENT for both GUT and LUNG also needed to recover from perturbations simulating both infectious and non-infectious insult, where the infectious insult replicated and actively damaged the system. It needed to do so while mimicking recognized component mechanisms. GUT EPITHELIAL CELLS responded to that byproduct. The process simulated epithelial barrier dysfunction. Simulation results are consistent with the hypothesis that the CELL-damage-byproduct was responsible for activation of circulating INFLAMMATORY CELLS and that, in turn, led to ORGAN injury. If we accept the analogue-to-referent mappings as being reasonable, then there may be an *in vivo* counterpart to the CELL-damage-byproduct. The LUNG employed essentially the same mechanisms. For the LUNG ORGAN, two coupled phenomena were targeted: epithelial barrier dysfunction results in lung edema and impaired systemic oxygenation, and supplemental oxygen can increase the sublethal threshold of hypoxia. Pulmonary EPITHELIAL barrier dysfunction led to impaired OXYGENATION, which in turn affected systemic ENDOTHELIAL CELL OXYGENATION status.

For the two\ coupled ORGANS, two attributes were that a severe inflammatory pulmonary insult (such as pneumonia) can lead to gut failure (Fig. 10B), and gut ischemia can lead to pulmonary failure (Fig. 10C). Given those, three modes of ORGAN crosstalk were specified: 1) INFLAMMATORY CELLS moved between Z=0 and Z=3 carrying INFLAMMATORY signals; 2) the CELL-damage-product produced by ISCHEMIC ENDOTHELIAL CELLS in GUT moved to LUNG where it could activate LUNG INFLAMMATORY and ENDOTHELIAL CELLS. That process had a negative impact on ENDOTHELIAL function and EPITHELIAL barrier function, which, in turn, impacted systemic OXYGENATION; and 3) all ENDOTHELIAL CELLS were dependent on a baseline OXYGENATION level. It decreased as LUNG dysfunction increased. Simulations with the two ORGANS coupled together showed that a severe insult to one organ (which, for example, may map to pneumonia) led to MOF: LUNG INFLAMMATION led to impaired systemic OXYGENATION and GUT ischemia, which, in turn, fed back to the LUNG, potentiating PULMONARY dysfunction and lowering the analogue's sublethal ischemic threshold.

REASONING TYPES AND THEIR DIFFERENT ROLES IN M&S

Three Types

- Deduction, induction, and abduction play different, essential roles in M&S.
- Methodical use and documentation of all three reasoning types is required.

The three methods of reasoning are induction, deduction, and abduction. For completeness, they are described and discussed in the Appendix. Conditions supportive of all three reasoning methods are sketched in Fig. 11A. Traditional mathematical M&S involves little if any abduction. However, each of the four example systems in "Analogues: From *In Vitro* Tissues to Interacting Organs" was realized following cycles of abduction, induction, and deduction. The same iterative, scientific cycles characterize the discovery and knowledge generation processes during the early stages of
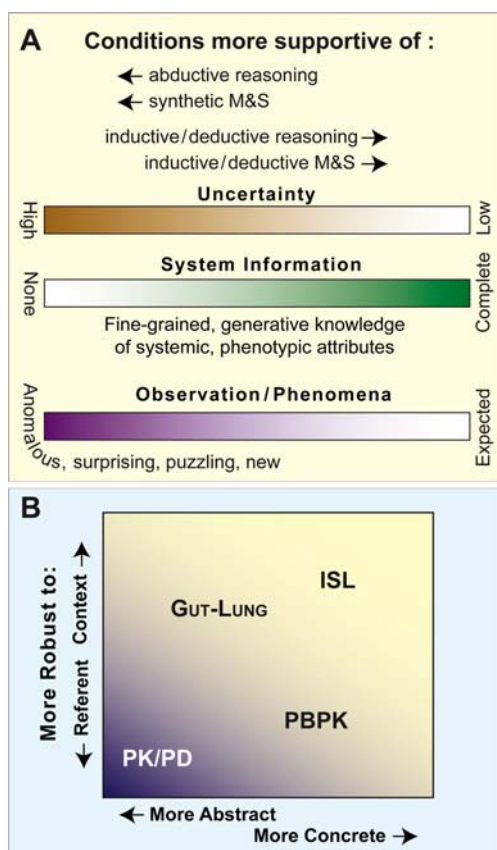
Fig. 11. Analogue characteristics. **A** Conditions supportive of all three reasoning methods are sketched. Obviously, everyone associated with a pharmaceutical or biotechnology R&D effort would like knowledge about all wet-lab research systems to be rich and detailed, and for uncertainties to be limited. Such conditions (*toward the far right side*), which are common in non-biological, physical systems, favor developing inductive models that are increasingly precise and predictive. However, the reality is that we are most often on the *left side*, where frequent abduction is needed and synthetic M&S methods can be most useful. **B** Four different model types are characterized in terms of robustness to context or referent, as discussed in the text. In terms of components and variables (input/output), PK/PD models (like many inductive, equation-based models) and the GUT–LUNG analogue are abstract enough to represent different families of referents, whereas the ISL and most PBPK models are more concrete and so less flexible.

biopharmaceutical research; we return to that idea below. The three types of reasoning contribute differently to the creation of new knowledge. Consequently, some review is warranted along with a discussion of where and how best to use the three reasoning methods to enable scientific M&S, as discussed in "M&S and the Scientific Method."

A model is a physical, mathematical, or logical representation of a referent system. The word *model* shares its etymological ancestors with the word *measure*. A model is, fundamentally, a measurement device or method for some referent. We use *analogue* to refer to models that are physical, such as an *in vitro* tissue model, and to distinguish products of the synthetic method from inductive models. An analogue's existence, operation, mechanism, etc. are entirely independent of the modeler and the referent. In those cases where the analogue consists of a computer running a program, we call it a simulation. Note that a simulation must be executing—prior to

and after the computer executes the program it is merely a set of instructions (program) and an instruction interpreter/executor (computer). Computation is a form of inference or reasoning: it is deduction. It is noteworthy that biological processes are also largely deductive, but the challenge is that biology is a young science; we do not know much about the language of biological processes or axioms.

In order to understand how computational models are and should be used for mechanism and knowledge discovery, one needs to understand how induction, deduction, and abduction relate to computational M&S. Those relationships are at the core of both synthetic and scientific M&S.

Synthetic Analogues Encourage Abductive, Scientific M&S

- Abductive inference dominates upstream discovery and development.
- Experimenting on synthetic analogues encourages abductive inference in exactly the same way as wet-lab experimentation.
- Abduction, induction, and deduction are necessary for discovery and development decision-making.
- Scientific M&S requires designing and conducting experiments on analogues designed to qualify as objects of experimentation.
- Multiple competing hypothetical mechanisms (models) are required.

Models are used throughout the drug development pipeline from discovery to post-marketing surveillance and from laboratory production to manufacturing. However, model purpose and usage vary within that pipeline. Downstream models focus on achieving some specific objective, like documenting that disease progression can be halted or improving a drug's supply chain. Upstream, scientific models focus on adding new domain knowledge and reducing uncertainties. We argue that abductive inference is most important for upstream M&S, and we will show that experimenting on synthetic analogues encourages abductive inference in exactly the same way as wet-lab experimentation.

None of the three methods of reasoning adds new knowledge on its own. In particular, deduction, being purely syntactic, is incapable of adding new knowledge. Because most current computational analogues are, independent of some larger, descriptive context, deductive devices, it is justifiable to doubt the extent to which such an analogue can be scientific. Such an analogue is a statement about what is currently known or believed. The means for making computational analogues scientific lies in model usage and how that usage fits into the larger research enterprise. The same is true of a wet-lab model.

New knowledge comes about by seeking and confronting contrast, anomaly, and surprising or unexpected observations. Our models evolve fastest when they fail to capture the world around us. When that occurs, we respond by constructing explanatory hypotheses—often relying on abduction—which are usually manifold and typically wrong at first. The collection of initial hypotheses is refined iteratively through rational analysis, including experimentation and deduction in both the minds of the

researchers as well as in computer simulations. Those that survive are further refined in the face of these newly induced models. At the end of this iterative process, the most robust explanatory and predictive hypotheses can be integrated into larger bodies of theory.

Granted, the above is a caricature of the actual process, which is extremely complex and social. But from the perspective presented, it is clear that abduction is very important to the scientific method. Scientific models—in silico or in the mind of a domain expert—are primarily abductive. This should be common sense, since science is about capturing, refining, and ultimately reducing our ignorance of a given system. As scientists, we deal more with what we do not know or do not understand than with what we do know or understand.

Computational analogues differ from other model types (e.g. *in vitro*, *in situ*, *in vivo*) because they rely entirely on machinery that has been explicitly and aggressively designed so that variation, anomaly, and surprise are minimized to the point where they are vanishingly small. For the most part, computational analogues are deterministic, well-controlled, and predictable devices. Because of this special status, the overwhelmingly popular uses of computational analogues do not involve experimentation. By contrast, consider an *in vitro* model. It is also explicitly and aggressively designed to minimize variation. However, there are always system component aspects about which the experimenter is fundamentally ignorant. Obviously, that is the component of interest (cells, tissue explant, etc.). We still treat artificial machines as experimental systems, even though well-understood theories are known to govern their behavior. In contrast, we often believe we fully understand and can validate computational programs. Most researchers do not treat a computational analogue as an experimental apparatus.

Instead, we inscribe into them what we expect to conclude from them. As stated above, that is deduction. The conclusions are the same as the premises, just transformed by a formal system grammar. Hence, if we maintain that computational analogues are completely verified (we know precisely what they do) and they are purely deductive (truth-preserving), then we cannot rely on them as rhetorical devices in and of themselves without committing the fallacy of *petitio principii*—assuming the conclusion. This situation makes it clear that in order to avoid fallacy, any scientific rhetoric of which a computational model is a part must include the other two types of inference (abduction and induction), and to do that must draw on *additional models*, especially those in the mind of the researcher. That realization implies that scientific research involving computational analogues—scientific M&S—is characterized by testing *multiple* similarly plausible models, just as abduction requires testing multiple hypotheses and induction requires multiple observations. Note that abduction and induction occur at a level above the computational analogues.

The described framework for the scientific use of computational models requires designing and conducting experiments on the analogue and constructing analogues that merit being objects of experimentation. Each of the four biomimetic systems in "Analogues: From *In Vitro* Tissues to Interacting Organs" had multiple predecessors (which, at the time were considered plausible models about some aspect of

phenotype) that were challenged experimentally and found in some way wanting.

M&S AND THE SCIENTIFIC METHOD

Two Major Categories of Robustness: You Cannot Have Both

- Inductive models tend to be robust to changes in referent; synthetic models tend to be robust to changes in context.
- Synthetic models are more useful early on, while inductive models are more useful later.
- As they mature, synthetic models will also be useful later.

For the computational models in Fig. 1, including those that will bridge the gap, there are two main categories (types I and II) of robustness, each with a subcategory. Short- and long-term uses determine which category should be selected to achieve a given M&S objective, and that in turn impacts which model type, inductive or synthetic, may be best. A model cannot be robust in all ways. Analogues in the first category (type I) are robust to context changes, yet fragile to changes in referent. Some of these analogues can still be abstract enough to work for families of referents. Analogues in the second category (type II) are robust to referent changes, yet fragile to changes in context. Some analogues within this category can still be abstract enough to work for families of context. Wet-lab models can be similarly classified. For example, MDCK cells can be cultured under many different conditions with a variety of additives; the cells are robust to context changes. However, as epithelial cells, they cannot mimic cardiac myocytes and in that way they are fragile to changes in referent. Embryonic stem cells can be prodded to transform into representations of many different cell types, but to be maintained as stem cells their environment must be tightly controlled; they are robust to referent changes, yet fragile to changes in context. Synthetic and inductive models often identify more strongly with one of these categories, as illustrated in Fig. 11B. A fully synthetic and concrete analogue, such as the ISL, is robust to changes in context, yet fragile to changes in referent. It can be used to represent a liver in almost any context, but it cannot be used to represent a lung (however, some of its parts could be reused in a lung analogue). The generic two-layer subsystems used by An in the fourth example are robust to changes in context, and they have contextual patterns suitable for multiple referents. The ORGAN subsystems are synthetic, yet abstract enough to represent other organs or tissues. A fully inductive, detailed, PBPK model is expected to be robust to changes in referent, yet it is fragile to changes in context. The same model can be used to represent any number of individuals and even different mammals, but only under similar conditions. Traditional PK and PD models are sufficiently abstract and general to be robust to changes in referent within referent patterns across contexts. Consequently, such models can be used to characterize data from many different referents. The patterns in the data can arise during different experimental contexts.

Models grounded to a metric space or hyperspace (discussed in "How a Model is Grounded Impacts How It

Can and Should be Used") are robust to changes in referent while being fragile to changes in usage or experimental protocol. Inductive models tend to be grounded absolutely because they model relations between variables or quantities, not qualities or mechanisms. However, a generalized inductive model (e.g., exponential decay, saturation, or a sigmoid) can easily show qualitative relationships between quantities. Such models can be robust to variations in the ratios between the quantities even though they depend fundamentally on the quantities they relate.

Relational models are robust to changes in use or experimental protocol, yet they are fragile to changes in referent. It is natural for synthetic models to be internally grounded because they model relations between identified and hypothesized components. A more generalized and abstract synthetic model, such as cellular automata can show organizational patterns between constituents. Consequently, it can be robust to changes in referent, when the various referents have similar organization.

In general, because inductive models tend to be robust to changes in referent, and synthetic models tend to be robust to changes in context, we recommend synthetic analogues for explanation and hypothesis generation earlier in R&D and inductive models for prediction and late-stage hypothesis falsification. When synthetic analogues are relational, they are best for knowledge embodiment.

The vision presented in the Introduction requires analogues with long lifecycles. To have long lifecycles, analogues must be capable of adjusting easily to incorporate new knowledge. This process requires using all three forms of inference (94). It is noteworthy that many inductive models have short lifecycles; some are never used again following their initial application. Synthetic and inductive models have different adjustment capabilities depending on the type and source of the data. When new knowledge comes from changes to context, then a synthetic model will be most appropriate. An example of change in context would be many different types of wet-lab experiments using the same cell line. The latter is often the case for wet-lab models used early in support of R&D. When the new knowledge comes from well-studied changes to the referent, then an inductive model is most appropriate. An example of the latter would be an expanded, phase four clinical trial.

How a Model is Grounded Impacts How it Can and Should be Used

- Knowledge embodiment requires synthetic analogues that are relational.
- Inductive models are typically grounded to metric spaces.
- Metric grounding complicates combining models to form larger ones.
- Relational grounding enables flexible, adaptable analogues, but requires a separate analogue-to-referent mapping model.
- Biomimetic analogues designed to support drug discovery and development must have long lifecycles.

As stated earlier, the units, dimensions, and/or objects to which a variable or model constituent refers establish

groundings. Inductive models are typically grounded to metric spaces. So doing provides simple, interpretive mappings between output and parameter values and referent data. Because phenomena and generators (in Glossary) are tightly coupled in such models, the distinction between phenomenon and generator is often small. Metric grounding creates issues that must be addressed each time one needs to expand the model to include additional phenomena and when combining models to form a larger system. Adding a term to an equation, for example, requires defining its variables *and premises* to be quantitatively commensurate with everything else in the model. Such expansions can be challenging and even infeasible when knowledge is limited and uncertainty is high, as on the left side of Fig. 11A. A model synthesized from components all grounded to the same metric spaces—a PBPK model for example—is itself grounded to the Cartesian composite of all those metric spaces. The reusability of such a model is limited under different experimental conditions or when an assumption made is brought into question.

Grounding to hyperspaces increases flexibility. A hyperspace is a composite of multiple metric spaces (and possibly non-spatial sets). Grounding to a hyperspace provides an intuitive and somewhat simple interpretive map (see (95)). Phenomena and generators are more distinct, because derived measures will often have hyperspace domains and co-domains, making them more complex as interpretive functions. Hyperspaces are often intuitively discrete, so they do not require discretization. They thus handle heterogeneity better than does a model grounded to a metric space. The High Level Architecture (IEEE standard 1516-2000) and federated systems for distributed computer simulation systems are examples of hyperspace grounding. Their focus is to define interfaces (boundary conditions) explicitly so that components adhere to a standard for such interfaces.

Dimensionless, relational grounding is another option. In equation-based models, dimensionless grounding is achieved by replacing a dimensioned variable with itself multiplied by a constant having the reciprocal of that dimension. That transformation creates a new variable that is purely relational. It relies on the constant part of a particular context. The components and processes in synthetic models need not have assigned units; see (65,96,97) for examples. The first, third, and fourth of the above examples use relational grounding: each constituent is grounded to a proper subset of other constituents. Relational grounding enables synthesizing flexible, easily adaptable analogues. However, a separate mapping model is needed to relate analogue to referent phenotypic attributes.

Hybrids of the above grounding methods are also possible. Some models can be synthesized by plugging together components that are simpler models. For example, in (80) output of metrically grounded, equation-based models of subcellular molecular and cell cycle details contribute to rules used by cell level agents. Such coupling makes them somewhat relational because not every component must be connected to every other component (or adhere to a standard adhered to by all other components, as with the High Level Architecture). However, their synthesis will depend in a fundamental way on their grounding, sometimes to a metric space, as in (98,99). The High Level Architecture (and similar) standards can be considered as hybrids, because they

provide openness and extensibility that allow some subsystems to integrate based on one standard and others to integrate based on another standard.

Biomimetic analogues designed to support drug discovery and development research are expected to evolve and become more realistic and useful. Consequently, those that do will have long lifecycles. Some will mature to become virtual tissues and organs: components in virtual patients. For these analogues, we suggest they begin as relational analogues and remain so to the degree feasible, and that separate mapping models be developed in parallel.

Synthetic analogue development to date by different groups has managed the grounding issue differently. Twelve examples are discussed briefly in the Appendix.

Experimenting on Synthetic and Inductive Analogues

- Synthetic models are relatively specific, particular, and concrete; inductive models are relatively general, representative, and more abstract.
- Experiments on inductive models will discover characteristics of the data from which the model was induced; experiments on synthetic models will discover characteristics of the composition, the mechanisms, and the model's systemic phenotype.
- When trying to explain a system about which we are ignorant or uncertain, use abduction: it is ignorance-preserving.
- Synthesis facilitates knowledge discovery by helping to specifically falsify hypotheses.
- Inductive models preserve the truth about patterns in data. Synthetic models exercise abduction while representing knowledge, uncertainty, and ignorance.

We argue for conducting experiments on computational analogues as if they were naturally occurring organisms or tissues. That is precisely what is done with very complex software and hardware systems developed purely for engineering purposes (e.g. flight code for an automatic pilot). However, in these engineering contexts, the purpose behind such testing is to clamp down on the exhibited variation and ensure that it stays within specified, controlled tolerances. When designing and planning wet-lab experiments, we include engineering tasks such as clamping down on the variation of those parts that are not objects of the experiment, such as temperature, pH, pO_2, etc. The difference is that those wet-lab experiments have a different objective: to explore the living component of the system. The tightly controlled, well-understood, predictable parts of the supporting laboratory equipment are ancillary to the primary purpose, which is to refine and increase our understanding of the biological material being studied.

If we replace the biological material with a computational analogue in a supporting framework like the one described in Fig. 4, does it still make sense to use the *whole apparatus* to refine and increase our understanding of a smaller component of the system, such as the simulated cells in Fig. 4? The answer depends on the nature of that model and the model's current location in model space. If it is a straightforward implementation of, for example, a simple mathematical equation that is well understood, then the answer is "no." Of the models in the references cited in Supplemental Material, the vast majority developed within the past 15 years required considerable experimentation. That is because the phenotype (as in Fig. 3) resulting from the initially conceived mechanism was too far removed from targeted phenotypic attributes. Experimentation (several cycles of a protocol such as the one in the next section) was needed to locate a region of model space (mechanism space) for which the phenotype was more biomimetic.

Both inductive and synthetic methods depend differently on the means and measures used to gather the data and related information that becomes the focus of the model engineering effort. Inductive models contain an inherent commensurability amongst the measures, because induction finds and reproduces connected patterns in whole data sets. Synthesis, however, combines heterogeneous data with information from disparate sources and discovers ways to compose them; some information sources are ad hoc, whereas others are highly methodical. Even though the methods are fundamentally different, both inductive and synthetic models of biopharmaceutical interest will often be appropriate for experimentation.

Experiments on any model—mental, wet-lab, inductive, or synthetic—can help the scientist think about and discover plausible characteristics of a referent system's mechanisms. Experiments on inductive models do so by exploring characteristics of the data from which the model was induced. That is because the mappings are among data, conceptualized mechanisms and referent, as illustrated on the left side of Fig. 2. Experiments on synthetic models do this by exploring model organization (how components co-operate/interact), which is hypothesized to map to referent organization, as illustrated by mapping C on the right side of Fig. 2.

Both modeling types have their strengths and weaknesses. Inductive models, because they rely directly on the measures used to take the data, are susceptible to the fallacy of inscription error (the logical fallacy of assuming the conclusion and programming in aspects of the result you expect to see). This weakness is a natural result of the combination of the extrapolative properties of induction and the truth-preserving properties of deduction. By contrast, as discussed in (100), synthetic models, like the examples described in "Analogues: From *In Vitro* Tissues to Interacting Organs," can contain abiotic and arbitrary artifacts, assumptions, and simplifications made for the convenience of the builders. Note the partial overlap of phenotypes in Fig. 3. These properties provide direction for when one style should be preferred over the other. When trying to clearly specify the parts of a referent about which we are ignorant or uncertain, an ignorance-preserving technique like abduction combined with synthetic M&S should be the center of attention. When trying to specify the parts of a referent about which we have deep knowledge, a truth-preserving technique like deduction should be the focus. Where we possess enough reliable knowledge to warrant extrapolation and precise prediction, induction should be the focus. Hence, synthesis can be most useful as an upstream modeling method focused on discovering and falsifying hypotheses during knowledge discovery and synthesis, and while honing down and selecting hypotheses that are most believable. When we become confident of the generative mechanisms, inductive

counterparts of synthetic analogues can kick in, allowing us to approach engineering or clinical degrees of understanding, intervention, and prediction. Inductive models are best for preserving the truth about patterns in data. Abductive synthetic models are best for exercising abduction and representing current knowledge and beliefs clearly, as well as areas of ignorance and uncertainty.

When R&D goals require the capabilities of both modeling methods, both model types can be implemented for parallel simulations within a common framework (101). In the sections that follow, we identify several important issues and discuss them in context of both synthetic and inductive M&S. Background information on computational models in support of scientific discovery is provided in the Appendix.

The Scientific Method in Iterative Analogue Refinement

- Conceptual mechanisms can be flawed in ways that only become obvious after they are implemented synthetically.
- Following a rigorous protocol facilitates generating multiple mechanistic hypotheses and then eliminating the least plausible through experimentation.
- An iterative model refinement protocol is the heart of abductive, mechanism-focused, exploratory modeling.

The scientific method provides a procedure for investigation, the objective of which is knowledge discovery (or questioning and integrating prior knowledge). The method begins with phenomena in need of explanation or investigation. We pose hypotheses and then strive to falsify their predictions through experimentation. The traditional inductive modeling approach illustrated on the left side of Fig. 2 is often part of a larger scientific method that includes wet-lab experiments. On its own, however, inductive modeling is not scientific because new knowledge (about the referent) is not generated.

The stages in scientific M&S are illustrated on the right side of Fig. 2. The assembly of micro-mechanisms in each of the four examples in "Analogues: From *In Vitro* Tissues to Interacting Organs" was a hypothesis. Each execution was an in silico experiment. Measures of phenomena during execution provided data. When that data failed to achieve a pre-specified measure of similarity with referent wet-lab data, the mechanism was rejected as a plausible representation of its wet-lab counterpart (for a detailed example, see (65)). In all four examples discussed in "Analogues: From *In Vitro* Tissues to Interacting Organs," many mechanisms were tested and rejected en route to the mechanisms discussed in the cited papers. Multiple rounds of iterative refinement followed by mechanistic failure illustrate the fact that complex conceptual mechanisms can be flawed in ways that are not readily apparent to the researcher. The flaws only become obvious after we actually invest in the effort to implement and test the mechanism synthetically.

The iterative analogue refinement process benefits from following a protocol. We have used the protocol presented in Fig. 12 successfully (65,75,81,96,97). It strives to adhere to the guideline of parsimony, which is important when building agent-oriented analogues that are expected to become increasingly complex. The protocol facilitates generating multiple mechanistic hypotheses and then eliminating the least plausible through experimentation.
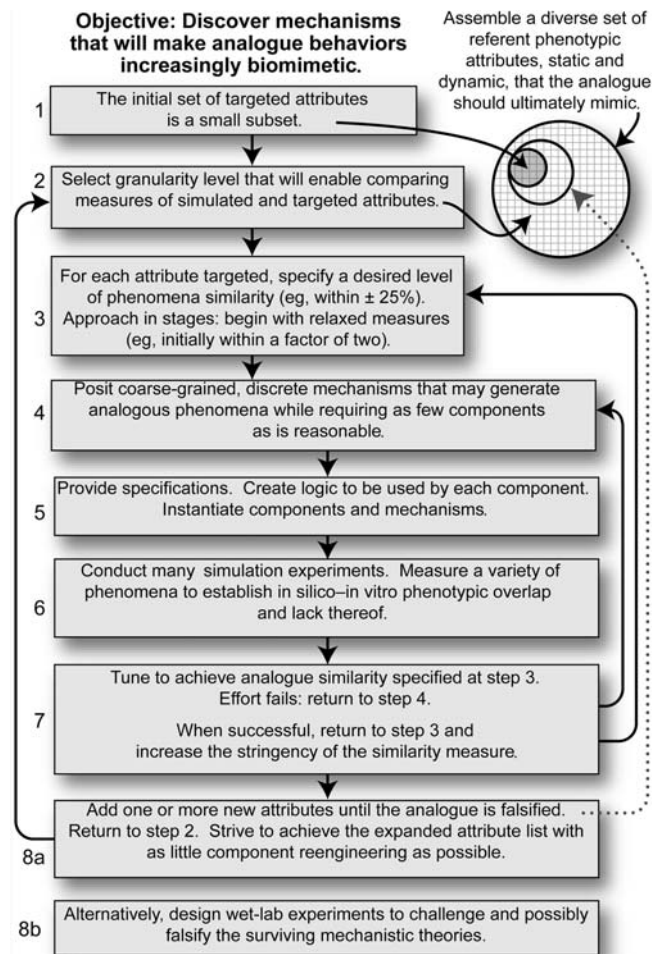

Fig. 12. An iterative protocol for refining and improving synthetic analogues. Abductive reasoning may be required at steps 4–8. Induction and deduction occur during steps 5–7.

The iterative model refinement protocol is the heart of abductive, mechanism-focused, exploratory modeling. When faced with the task of building a scientifically relevant, multi-attribute analogue in the face of significant gaps in the body of knowledge used to guide the process, parameterizations and model components must strike a flexible balance between too many and too few. Doing so can be complicated by the fact that a validated, parsimonious, multi-attribute analogue will be over-mechanized ("over-parameterized") for any one attribute. Too many components and parameters can imply redundancy or a lack of generality; too few can make the model useless for researching multi-attribute phenomena.

Scientific progress can be measured in scientific M&S protocol cycles completed. It follows that we want to make the process as easy as possible. Decisions made at the beginning and during the first protocol cycle can dramatically impact the level of effort required to complete subsequent cycles. Strategies that can work well for inductive M&S may not be appropriate for synthetic M&S.

Making Predictions Using Synthetic Analogues

- Synthetic models make predictions about component relations; inductive models make predictions about variable relations and patterns in data.

Synthetic analogues are ideal for discovering plausible mechanisms, relations between components, and mechanism-phenotype relationships. They are best for exercising abduction and representing current knowledge. They are good at explanation. Because of the uncertainties reflected in stochastic parameters and mappings to the referent, they are not as good as inductive, equation-based models at precise, quantitative prediction; their predictions will be "soft." However, they can make effective relational predictions. Synthetic models make predictions about component relations (quantitative or qualitative), and inductive models make predictions about variable relations (quantitative or qualitative) and patterns in data. An ISL, for example, can be used to make predictions about where and how within the system two DRUGS administered together may effectively interact. However, when such predictions are absolutely grounded (via the grounding map), then an important distinction between synthetic and inductive model predictions lies in the error estimate for the prediction.

IMPACT OF M&S ON SCIENTIFIC THEORY

Instantiating and Exploring Theories of Translation

- Theories of translation will arise from contrasting analogues.

A precondition for understanding if and when observations on wet-lab research models can translate to patients (and vice versa) is to have a method to anticipate how each system will respond to the same or similar new intervention at the mechanism level. The ISLs (21) and the other analogues described above enable developing that method. Building an analogue of each system within a common framework allows exploration of how one analogue might undergo (automated) metamorphosis to become the other. When successful, a concrete mapping is achieved. Such a mapping is a hypothesis and an analogue of a corresponding mapping between the two referent systems, as in Fig. 8B. The analogue mapping can help establish how targeted aspects of the two referent systems are similar and different both at the mechanistic level and, importantly, at the systemic, emergent property level. The vision is that the analogues along with the metamorphosis method can be improved iteratively as part of a rational approach to translational research.

Abductive Reasoning and Synthetic M&S Can Help Manage the Information and Data Glut

- Synthetic analogues help alleviate the information and data glut.
- Combining synthetic with inductive models better preserves and progressively enhances knowledge.
- Exclusive reliance on inductive and deductive methods starves R&D of abductive opportunities.

We concur with An's observations regarding dynamic knowledge representation and ontology instantiation (18,102), and argue that the data and information glut impacting pharmaceutical R&D is caused by our knowledge schemes and knowledge bases (into which that data and information should fit) being incomplete or unnecessarily

abstract. The available schemes are largely represented in the formalized, prosaic Methods sections of published scientific papers, and to some extent in Discussion sections; they are also represented within similar documents within organizations. The majority of that information consists of relationships between quantities, yet the schemata available for cataloging relationships between quantities are ambiguous. Examples of elements of schemata are $\log P$, clearance, level of gene expression, media composition, response, etc. The schemata are designed purposefully to lose, forget, abstract away, and/or ignore some concrete details of experiments and cases. On one hand, such abstraction is good because it facilitates the extraction of fundamentals and major trends: the take-home messages from specific experiments. On the other hand, there are many concrete details which, were they captured by the schemata, would permit a more complete cataloging of experiments and observations. In some cases, such improvements could enable semi-automatic extraction, hypothesis generation, evaluation, and hypothesis selection. We posit that complementing the current schemata with synthetic analogues, advanced progeny of the four examples in "Analogues: From *In Vitro* Tissues to Interacting Organs," would be a significant step toward more satisfactory schemata. The process would begin alleviating the information and data glut, and allow semi-automated hypothesis generation/testing and theory development.

A synthetic analogue is a schema for biomimetic constituents. An inductive model is a schema for quantities. Either, alone, is inadequate. Together, knowledge generated can be preserved and its value progressively enhanced as the process advances.

A natural consequence of the rapid advances in –omic technologies, which are quantitative, coincident with the rise of molecular biology, coupled with advances in computational methods, has been a heavy focus on forward mappings and inductive models (discussed further under Generator–phenomenon relationships in the Appendix). Methods to support abduction and synthesis—scientific M&S—have not advanced as quickly. A consequence is the current information and data glut. Methods (knowledge schema) for rapid hypothesis generation and refinement are sorely needed (102).

We suggest that a contributor to the inefficiencies responsible for significant rates of failures late in development is the fact that scientists and decision-makers currently translate results and conclusions from wet-lab experiments to patients using conceptual mechanisms and mappings where some assumptions are intuitive and unknown, and thus unchallengeable. It has been argued that mathematical, systems biology models will help address this issue, but we believe the situation is actually made worse by over-reliance on inductive and deductive computational modeling methods that reduce opportunities for and ignore the importance of abductive reasoning. Scientific M&S provides the means to concretely challenge those concepts and preserve those that survive.

Discovery and Development Research Needs Explanatory Models to Complement Inductive Models

- Synthetic analogues are best at explanation; inductive models are best at precise prediction.
- Strong theory and good science depend on having both heuristic value and predictive value.

- When in pursuit of new mechanistic insight, the emphasis should be on generation and exploration of multiple hypotheses.
- We risk being stuck on the current scientific plateau until we implement complementary methods to generate and select from competing, explanatory hypotheses.

The explanatory power and heuristic value of a hypothesis come from its ability to make specific statements about the network of micro-mechanisms (lower level generator–phenomenon relationships; see Appendix for further discussion) that produce a rich phenotype. Synthetic analogues are best at explanation. On the other hand, the predictive power of a hypothesis comes from its ability to make specific statements about the end conditions (the context, state, situation, etc. it will obtain) given some initial conditions. Inductive models are best at prediction. Strong theory and good science depend on having both heuristic value and predictive value. Recall Richard Feynman's dictum: "what I cannot create, I do not understand." Correlations are devoid of heuristic value, yet they can provide predictive value.

An attractive property of inductive models is that, because they relate quantities, they are relatively easy to validate. However, because they do not relate specific components, the mechanisms and the generator–phenomenon map they are intended to represent remain conceptual. All the hypothesized generators for the phenomena modeled are embedded in the prosaic and pictorial descriptions of the models, not in the mathematics. The descriptions are not actually part of the induced model. There is only the conceptual linkage illustrated on the left side of Fig. 2. That situation makes generator falsification very difficult because many generator configurations and different sets of generators can produce the same phenomena, which, when measured, contains patterns specified by the validated mathematical, inductive model. Within the biomedical, pharmaceutical, and biotechnology domains, that difficulty has resulted in too little focus on falsification and an overzealous focus on data validation (of patterns), as distinct from the more heuristic forms of validation (103). When in pursuit of new mechanistic insight, the emphasis should be on generation and exploration of multiple hypotheses. The validation process should involve repeated attempts to falsify a population of hypotheses and select the survivors. As demonstrated herein, the technology and methods are available to complement any mechanistically focused inductive model with synthetic explanatory modeling methods that offer concrete analogues of the conceptualized mechanisms. Without complementary methods to generate and select from competing, explanatory hypotheses, we risk remaining stuck on the current scientific plateau. Information and data will continue to grow, further overwhelming the individual scientist's ability to reason scientifically and to decide which therapeutic candidates to select and which to eliminate.

Scientific Multi-modeling

- Modeling a wet-lab biological system is multifaceted, requiring all three reasoning methods along with multiple models and model types.
- The phases of multi-modeling are construction, evaluation and selection, and refinement.

Scientific modeling, like wet-lab science, requires use of all three modes of inference. It also requires development of inductive and synthetic models to provide heuristic as well as predictive value. Use of the three inference modes requires developing and using multiple models. Multiple models are necessary because discovering plausible forward maps requires targeting multiple phenomena (and multiple measures). Discovering plausible inverse maps requires exploring multiple generators (and multiple measures).

Any satisfactory synthetic analogue can be falsified by placing additional demands on its phenotype, by selectively expanding the set of targeted attributes to which it is expected to be similar (Step 8a in Fig. 12). A beauty of synthetic analogues of the types described in "Analogues: From *In Vitro* Tissues to Interacting Organs" is that we can observe the networked micro-mechanisms and "see" what mechanistic features were most likely responsible for the analogue's inability to survive falsification. The data in Fig. 9 are an example. This knowledge improves insight into the referent mechanism and is an example of the analogue's heuristic value. If all existing analogues have been falsified, then one must step back and invent new analogues containing new micro-mechanistic features that may survive falsification. Following the preferred (but more resource-intensive) approach, two or more somewhat different, yet equally plausible analogues are created, and one or more survives falsification. The process of analogue falsification and survival provides valuable new knowledge about the analogue and about the referent micro-mechanisms.

Falsification of a synthetic analogue requires a precise criterion, one that requires the use of inductive models. The quantitative comparison of comparable analogue and wet-lab phenomena typically focuses on data features, a task for which inductive models are ideally suited. Statistical models are also useful. Because multiple attributes are always being targeted, the falsification decision will be based on multi-attribute comparisons. Following the repeated cycles of refinement in Fig. 12, analogues become more resistant to incremental falsification, and begin earning trust. Trustable synthetic analogues must be robust to both context variance and constituent variance. To build trust, the analogue must be groundable using absolute units (clock time, ml, moles, etc.), which requires concurrent development of mapping models. It becomes clear that modeling a wet-lab biological system to expedite research progress is a multifaceted undertaking requiring exercising all three reasoning methods and the development and use of multiple models and multiple model types. These tasks will be facilitated and made easier by insisting on analogues that strive to exhibit the capabilities in the Text Box. Further, multi-modeling has three different phases: construction, followed by evaluation and selection (for or against), and refinement.

Complement, Not Replace; Evolution, Not Revolution

- Synthetic methods complement inductive methods.
- Synthetic methods can help ferret out bad information.
- Inclusion of synthetic M&S early in R&D will accelerate the process.

One might surmise that because synthetic M&S methods draw extensively on recent computer science advances and

thus are relatively new, that they are intended to replace the "old" inductive methods. That is not the case: M&S efforts complement wet-lab efforts. As illustrated in Fig. 11, synthetic methods complement the well-established inductive methods. Synthetic M&S simply provides a dramatic expansion in options for using M&S to advance science and facilitate discovery, especially where mechanistic insight is needed, toward the left side of Fig. 11A. Synthetic M&S is a product of the evolutionary process that is driving computational M&S methods to become more biomimetic, to bridge the gap in Fig. 1.

Integration of various published results into a synthetic analogue accomplishes two things: 1) it helps build schemata for knowledge (and ignorance) representation, and 2) it also provides a mechanism for the curation and maintenance of the embedded knowledge. Both are possible because components and mechanisms in synthetic models are concrete, whereas those in inductive models are conceptual. Because synthetic models are heuristic, falsification during the evaluation of the Fig. 12 protocol will spawn abductive explanations for the hypothesis's failure. The weakest part of the failed analogue will most often be some new feature of the hypothesized mechanisms. However, in some cases, the weakest part may be some previously added component or logic drawn from badly designed experiments or invalid or false conclusions. The heuristic nature of synthetic models can highlight and facilitate the correction or revisitation of those previously accepted components; see (65) for an example. Synthetic M&S in conjunction with wet-lab experimentation and inductive M&S methods curates and maintains the integrity of current knowledge but also acknowledges our current state of ignorance.

One can argue that progress in a pharmaceutical R&D effort is tied to opportunities for mission-centered abductive reasoning. Circumstances requiring abductive reasoning associated with results of wet-lab experiments are common during discovery and early development. Opportunities to exercise abductive reasoning have traditionally been tied to the number of experiments and their duration. A complementary synthetic M&S effort can be added incrementally. With that effort comes a dramatic increase in opportunities for goal-directed abductive reasoning. If progress and better decision-making are positively correlated with knowledge and insight gained through goal-directed abductive reasoning, then we can anticipate that the early inclusion of scientific synthetic M&S will accelerate the process.

CONCLUSIONS

We presented a case that the means are currently available to speed the progression from discovery to clinical use and production of new therapeutic products. We showed that the process requires multi-modeling (multi-scale, multi-formalism, multi-technology, multi-person, multi-referent, multi-context, etc.) within a common framework, and that it, in turn, requires taking full advantage of computers during R&D. Multi-modeling that can be semi-automated is the M&S frontier. For multi-modeling to be successful, it must become methodically scientific. Scientific M&S will accelerate the above progression by facilitating fast-paced cycles of

hypothesis (about mechanisms) generation, selection, and falsification. Each cycle requires synthetic modeling and simulation coupled with inductive methods; during such a cycle, abduction drives the creation of mechanistic hypotheses. Those mechanistic hypotheses that meet criteria are selected for experimentation (wet-lab and in silico) designed to ensure that only those with explanatory, heuristic value survive falsification. The cyclic process exercises and leverages the mental models of domain experts in new ways. However, the new knowledge created can be instantiated and retained within the analogues and their framework making it immediately accessible to all members of an R&D organization. We explained that making M&S scientific requires extension of rigorous methodology to include planned use of all three reasoning methods: abduction, induction, and deduction. So doing will make M&S a methodologically sound, increasingly productive tool similar to wet-lab models, laboratory equipment, and large-scale experiments.

APPENDIX

Glossary

Abduction: arrival at a conjecture based on a pattern observed in one or a few particular cases; construction of hypothetical speculation (consistent with current knowledge) about the process by which an outcome (phenomenon) came to be, where the hypotheses are all equally reasonable as long as they lead to the outcome; arrival at a conjecture (hypothesis) that would, if true, explain the relevant evidence (abductive reasoning is discussed following the Glossary)

Agent: [technical] an object within an object-oriented program that can schedule its own events [within an analogue: it is quasi-autonomous; it senses and is part of its environment; it pursues and can revise an agenda within a larger script; it is identifiable by an observer as a cause of an effect; its attributes and actions may be designed to represent biological counterparts, whereas others will deal with issues of software execution]

Agent-based: something formulated with or built up from agents; [in agent-based modeling] a model designed for simulation in which quasi-autonomous agents are key components

Analogue: anything that is analogous or similar to something else, and that exists and operates in isolation even in the absence of

a referent; a system that has aspects and attributes that are similar to those of a referent system; a model implemented in software that, when executed, produces phenomena that mimic those of the model's referent

Aspect: the perspective taken when a system is observed; one of many functional effects that results when a system executes

Complex: consisting of interconnected or interwoven parts; involved or intricate [a complex system may have many cause–effect relationships or one in which it is difficult to tease out cause–effect relationships]

Deduction: automatic and/or mechanical transformation of a set of statements; purely mechanical (syntactic) transformation of the premises to a conclusion; transformation of assumptions into conclusions within a formal system [where if the assumptions are true, the conclusions and every intermediate step in the transformation are also guaranteed to be true]

Discrete event: a measurable, individual transaction or status change with specific start and end timing that can be attributed as a source of requirement

Discrete Event Simulation: a simulation in which the model evolves over time by representing changes as separate events [in a continuous simulation, the system evolves as a continuous function (differential)]

Emergence: a descriptive term for systems that have complex maps between mechanisms and phenomena, where it is not obvious how the properties and characteristics of the mechanisms generate the observed phenomena

Generator: a process or event that causes, produces, or contributes to a phenomena (generator–phenomenon relationships are discussed in a following section)

Granularity: the extent to which a system is subdivided; a measure of the size (degree of discretization) of components that make up a system

Grounding: units, dimensions, and/or objects to which a variable or model constituent refers

Induction: arrival at a conjecture (universal conclusion) based on a pattern observed in many particular cases; generalization: reasoning from detailed facts to general principles; generalization drawn from patterns in observed data

Instantiate: to construct a particular, concrete instance or individual member of a class

Object-oriented programming (OOP): at the risk of over-generalization, an effort to make software more tacit and intuitive; classes are defined so as to package logic/code and state variables inside a wrapper, thereby separating and abstracting an object's *interface* (behavior) from its *implementation* (mechanism); this separation facilitates the synthetic construction of many systems of objects without re-engineering the internals of any object used; this property makes OOP software more easily adapted, disassembled, and reassembled into other system organizations; it makes an ideal method for constructing software for biomimetic analogues

Scientific modeling and simulation: a particular use of models within a larger context in which 1) the model is abductively or inductively inferred and in which a deductive analogue is designed and constructed to be the object of experimentation, and 2) where experimentation is characterized by cycles of abductive, inductive and deductive reasoning focused on generating, testing, and falsifying hypotheses about mechanisms and their relationships to phenomena

Simulation: a real, implemented, or instantiated model during execution; a process of generating behavior of a model using a simulator; execution of a model by a simulator

Synthetic model: a model system constructed from extant, autonomous components whose existence and purpose are independent of the model they comprise; one formed specifically by combining elements, often varied and diverse, so as to form a coherent whole

Further Explanations

Representing Compounds: One COMPOUND within one SS grid space (in Fig. 7C) can be viewed as mapping to a wet-lab lower limit of detection. For example, it may be viewed as the limit of detection of referent compound in a biopsy sample that has a volume $1/5,000^{th}$ that of an average lobule. At that limit, some biopsy samples will test negative for drug, even though we are confident some is present. By analogy, an empty space within the ISL during simulation cycle maps to "no detectable drug." Even though we can trace the change in location of a specific COMPOUND during execution, there is no mapping to corresponding changes in location for specific drug molecules. Drug molecules comprising the amount to which a COMPOUND maps during one simulation cycle are not identical to those to which that COMPOUND maps during the next simulation cycle.

Inductive, deductive, and abductive reasoning: Conditions supportive of all three reasoning methods are sketched in Fig. 11A. *Induction* is arrival at a conjecture (universal conclusion) based on a pattern observed in many particular cases. We start with induction because of its pervasiveness in pharmaceutical and biological modeling domains. Because induction begins with the measurement of a collection of objects, it is fundamentally and foremost a method for studying phenomena, not mechanism. That realization is important. The measure that is defined in order to take measurements on the set of objects is a model. The data from which a pattern is induced is a statement *solely* about the phenomena for which the measure was designed. Hence, every data set, and subsequently every inductive model, has embedded in it the aspects or usage protocols plus premises commensurate with the measure used to take the data. Naturally, using more data or other information during induction allows for a more fine-grained model, including induction over multiple dimensions and (though difficult) heterogeneous data types. In general, a model induced from large sets of the same type data will be more precise, more specific, and more suitable for prediction. On the other hand, a model induced from lots of variant data (measures of different phenotypic attributes) will be less precise but more resilient and more general; its predictions will reflect greater uncertainty. In the former context, one has to worry most about over-fitting data, which makes the model too specific to a single data set to be useful. At the same time, one has to worry about inducing a model that is too general and misses crucial patterns in the data. Those two issues are broadly recognized and appreciated in the ADMET, QSAR, and PK literature.

Often, the modeler is also interested in the generative mechanisms responsible for the data. In those cases, features of idealized, conceptual constructs (a feedback circuit, a two-compartment model, etc.) can provide mechanistic insight. This is particularly true when the mathematical description of an idealized measure of a construct feature (given a set of premises) is known to have a form that is the same (or nearly the same) as the induced mathematical description (a sum of exponentials, for example). How well the parameterized features of the conceptual construct map to components and features of the referent, along with the acceptability of the premises, are separate issues outside the scope of this discussion. The initially induced mathematical description is prosaically expanded to include the conceptual features, but remains a hypothesis about patterns in the data. The process just described corresponds to the conceptual mapping arrows on the left side of Fig. 2.

Deduction is automatic and/or mechanical transformation of a set of statements. It is the purely mechanical (syntactic) transformation of the premises to a conclusion. As such, no meaning (semantics) need exist for deductive systems. All executing computer programs (absent human or real-world interactions) are deductive systems. Likewise, mathematical transformations within domains like algebra, calculus, or CA are also deductive systems. The most fundamental element of deduction is the engine that actually makes the transformation. In the case of a computer program, the instruction pointer provides the impetus for transformation. In mathematics, the engine is the human manipulating the symbols. A simulation is an operating, deductive system designed to mimic the behavior of some referent. Its alphabet and grammar are specified in part by the language in which the program is written and in part by the constructs the programmer creates. The premises are statements about the initial conditions of the program. The conclusions are statements about the final conditions of the program. The conclusions (of interest) become the outputs or "behaviors" of the simulation, but no new knowledge can be created. Any meaning applied to the premises, grammar, or conclusions are inferred by those examining the program and simulation, and so remain conceptual and hypothetical.

What about interactive programs? Are they deductive, and if not, what are they? In a non-interactive simulation, what is true in the beginning remains true at the end and in between. The truth-preserving property of non-interactive programs holds even in the cases where, for example, variables take on values that oscillate, are stochastic, or produce a chaotic time series. Thus, an inductive model implemented as a deductive simulation is truth-preserving. However, an interactive program (for example, one that controls periodic blood sampling during an experiment, and causes injection of an amount of insulin based on the reading and its internal logic) can be written such that sentences are added, modified, or removed over time. Such systems are termed non-monotonic (104) because new premises ("facts") can change a conclusion that was formerly true to false, or vice versa. As we argue in the text, scientific M&S involves more than purely deductive (and monotonic) systems. Given that argument, it is clear that, for example, a non-interactive algorithmic implementation of a system of ordinary differential equations (even a very large system) is not a complete

model of a system; the supporting prose, diagrams, and artwork typically provide the necessary context and mechanism descriptions.

Abduction (105–109) is arrival at a conjecture based on a pattern observed in one or a few particular cases. Abduction is conceptualizing (multiple) mechanisms: explanations that, if true, could account for or generate a similar anomalous, interesting, or surprising observation. Abductive inference involves hypothesis generation and selection, and is an important occurrence during wet-lab research. Abduction is most likely and appropriate for ambiguous systems (the left side of Fig. 11A). The task of resolving the ambiguity is best approached through multiple aspects and with multiple mechanistic hypotheses measured with multiple measures.

To illustrate abductive inference, consider the following situation: measurements of an experimental treatment group exhibit unexpected values when compared to those of control groups and data from past experiments. Further, the results do not fit well with known categories of similar phenomena. In such a situation, researchers offer many speculative, candidate explanations: were this condition or circumstance true, it could explain the anomalous or new observations. Some explanations may focus on material used in the experiment (a possible bad batch of reagent, etc.). Other explanations may focus on the conduct of the experiment (the samples may have been mishandled, etc). Others are ideas about mechanistic explanations. Generation of varied explanatory hypotheses, some highly speculative, following the observation of the anomalous behavior is part of abduction. The next phase involves a process that narrows the competing ideas to those deemed most plausible. Following abduction, the consequences of these hypotheses are logically or experimentally deduced, and then evaluated using induction. After testing, when the set of plausible hypotheses is dramatically reduced, those remaining represent the current best explanation(s) until some new observation falsifies one or more of them. At that stage, the entire scientific reasoning cycle may repeat itself.

The above cycle (Fig. 12) occurs frequently when engineering and refining a synthetic analogue. The behaviors of the first implementation often fall short, frequently far short of expectations, even though it is the modeler's best hypothesis about how components should plug together to obtain the targeted phenomena. That shortfall falsifies the best hypothesis. The modeler has learned that the selected region of mechanism space is too abiotic. The modeler must rethink plausible micro-mechanisms. The solution is to jump to another region of mechanism space and experiment to determine if the new micro-mechanisms produce phenomena that are more similar to targeted phenomena. So doing exercises creativity. When improvements are seen, the modeler can conjecture that the new micro-mechanisms are more biotic. Each failed cycle exercises thinking creatively about plausible mechanisms. Each improvement in the similarity of the analogue's behavior to that of the referent adds new knowledge and improves insight into referent mechanisms. Both failed and fruitful cycles are often characterized by abductive reasoning.

Like induction, abduction starts with a measure selected by the researcher. As such, the hypothetical mechanisms

inferred (current beliefs) are inherently and irrevocably dependent upon the measure through which the phenomena are defined and revealed. They are aspect- and perspective-dependent. Change aspect and/or perspective and the hypothetical mechanisms inferred may change. Unlike induction, however, abduction does not necessarily produce (overly) precise or (overly) general mechanistic explanations. More often, they are specific to the observation or experiment and its context. That is because the focus of abduction is on the current few interesting cases. The hypotheses (explanatory models) must be elaborated through deduction followed by validation through induction in order to learn how precise or general each hypothetical mechanism actually is. For that reason, abduction preserves ignorance, in contrast to the truth preservation of deduction. The researcher is just as ignorant after abducing an explanation as before. There is a spectrum for abduction analogous to the precision-generic spectrum for induction. It is heterogeneity–homogeneity. The number of distinct aspects (model usage protocols) the measure tolerates dictates the heterogeneity of the abduced model. For example, an anomalous set of drug plasma levels at intervals after dosing will produce a population of similar, explanatory hypotheses. They can be described as being relatively homogeneous. On the other hand, an anomalous set of measures that also consider levels of a surrogate marker, blood pressure, muscle tone, etc. will produce an array of very different hypotheses. Because of that variety, they can be described as being heterogeneous. Granted, an overly enthusiastic modeler can construct a complex explanation for a simple phenomenon, but there are philosophical, professional, and practical motivations to minimize the complexity of the abduced model so that it is only as complex as it must be to explain the heterogeneous set of measured phenomena.

Grounding Examples: The following are twelve examples of biomimetic, synthetic analogues (all agent-oriented) that manage grounding differently, even though uses are broadly similar. From first to last, the analogues range from being thoroughly grounded to metric space (and thus requiring reengineering to reuse the analogue for referents having different contexts) to being more relational (and thus requiring a separate mapping model to quantitatively relate analogue behaviors to referent behaviors).

Odell *et al.* (76) present a molecular mechanism for positioning during cell division of the cleavage furrow during cytokinesis. The analogue contrasts opposite effects of dynamic and stable microtubules on cleavage furrow positioning. It uses a few hundred thousand differential equations to keep track of how, via Newton's and Fick's laws, myriad individual cytoskeletal elements (agents, e.g., MT "segments" and individual MKLP1 motors) move, continuously with time, through a spatial continuum of viscous cytoplasm in response to forces these parts exchange after collisions between them. As such, the analogue is thoroughly grounded in metric space.

Zhang *et al.* (80) present a lattice-based model of tumor growth, which uses a set of equations to represent epidermal growth factor receptor signaling network in the extracellular space, cell membrane, cytoplasm and nucleus. The epidermal growth factor receptor signaling network regulates CELL DIVISION, APOPTOSIS, and migration processes that collectively give rise to tumor-like phenotypes. The authors parameterized their model using measurements taken from the literature, and the parameter values were grounded to metric space.

Bindschadler *et al.* (72) present a confluent monolayer of CELLS that is INJURED by forcibly removing a strip of CELLS. The remaining monolayer HEALS through a combination of CELL migration, spreading and PROLIFERATION. The authors measured parameters of L1 fibroblasts and used them to construct their analogue. Components and rules used equations that had units of the *in vitro* measurements. The model is agent-based, so, in theory, mechanisms could be dissociated from specific units.

Christley *et al.* (53) present an analogue that was used to mimic patterns of mesenchymal condensation in which cells of the embryonic vertebrate limb in high-density culture undergo chondrogenic pattern formation. They used wet-lab values of diffusion coefficients to specify other parameter values, thus grounding the analogue to metric space. They used parameter sweeping to discover which parameter settings produced results most in agreement with *in vitro* data. The model design was such that the mapping model, rather than being merged with the analogue, could be made separate.

Walker *et al.* (79) developed an agent-based analogue of epithelial cells, and used it to simulate epithelial monolayer growth in culture. A set of rules governed CELL cycle actions, and was coupled with a physical model (equations) to account for physicochemical determinants of CELL bonding and movement. Separate mappings were used to draw quantitative relationships between in silico and *in vitro* features.

Galle *et al.* (77) developed an individual, agent- and physics-based representation of cells in a monolayer of epithelial-derived cancer cells. TUMOR CELL shape change was dependent on physical constants and forces, which provided grounding to metric space. Cell parameters were grounded to wet-lab dimensions. The applications discussed were somewhat limited by the precise grounding.

Ridgway *et al.* (78) simulate diffusion and reaction kinetics in a crowded virtual cytoplasm relying on metrically grounded reaction rates. The authors invested considerable effort in quantitative validation. This analogue has the flexibility to simulate different types of particles, but context is limited by the precise grounding.

Garmire *et al.* (64) describe an in silico Transwell device for the study of drug transport and drug–drug interactions. Being agent-based, the analogue has a general form that allows for its use with a wide range of different parameter settings. However, the particular version described was carefully parameterized so that simulation results were directly comparable to wet-lab results, and most values were grounded in metric space.

Cai *et al.* (73) present an analogue that was used to explore population-scale and CELL-scale behaviors in a WOUND-HEALING CELL migration assay. The analogue uses equation-based modeling, but the authors created and used generalized dimensionless forms of the equations. The individual-based analogue components were not grounded to metric space, although the analogue was parameterized using experimental observations.

Robertson *et al.* (74) present a multiscale analogue to explore issues of morphogenesis, specifically mesendoderm

migration in the Xenopus laevis explant model. Equation-based, mass action kinetic models specify intracellular processes. Events at the cell level used an agent-based model of mesendoderm migration across a fibronectin extracellular matrix substrate. CELLS were grounded to time and the dimensions of their grid space, which in turn were grounded to metric space. Given the flexible design, it should be straightforward to uncouple the system and mapping models.

Engelberg *et al.* (75) represent cells using atomic agents to better understand cell level principles of operation enabling formation of stable, semi-spherical tumor spheroids *in vitro*. The design was fully relational. Comparing simulated with wet-lab observations used a separate mapping model. However, because analogues were parameterized using diffusion coefficients and cell sizes, that grounding locked in certain other parameter values.

Lam *et al.* (65) used a simplified perfused liver analogue with some features similar to the ISL. The goal was discovering plausible mechanistic details of hepatic drug interactions. The analogue was fully relational. All quantitative mapping from analogue to referent measures used separate, simple mapping models.

Multi-agent platforms: A number of multi-agent M&S platforms are available. Swarm (http://www.swarm.org) is an established software framework supporting multiple programming languages (110). RePast (http://repast.sourceforge.net) is an integrated modeling environment that provides visual modeling and machine learning libraries (111). MASON (http://cs.gmu.edu/~eclab/projects/mason) is a lightweight Java-based framework with a modular visualization package (112). NetLogo (113) and StarLogo (114) are easy-to-use, integrated modeling environments based on a custom programming language, Logo. All provide requisite features for agent-based, discrete event simulation. These and other M&S toolkits are reviewed in (115,116). As demonstrated in the next section, agent-oriented analogues can develop into complicated devices. Thorne *et al.* (55) have observed that, as a consequence of that and the fact that building and working with biomimetic, synthetic analogues is a relatively new domain, it can be logistically difficult to communicate all of the relevant details. During a simulation, hundreds to millions of different components can be in play. That makes many simulation experiments computationally intensive, an issue that can be addressed using advanced technologies such as *cloud computing*.

Generator–phenomenon relationships within wet-lab models and synthetic analogues: Because generator–phenomenon relationships are at the heart of most synthetic modeling efforts, definitions and clarifications are needed. A generator is a tangible object that exists in the real world. A purely mathematical object is not a generator. A concept is not a generator. However, a computer executing instructions compiled from the description of a mathematical object (in a programming language) is a generator. A phenomenon is a pattern or event that, in principle, is observable, experienceable or perceivable by some other thing (no consciousness or intention is required). Phenomena remain *potentials* until or unless they are measured or used. Measures are a type of model, with a referent and an analogue. Generators are

distinct from the phenomena they (help to) generate. Within wet-lab systems there are many generator–phenomenon layers, many of which are still unknown and unobserved. A phenomenon in one generator–phenomenon layer can be a generator in another generator–phenomenon layer. Forward and reverse maps and generator–phenomenon relationships are related. Deduction is a forward mapping from generators to phenomena. Deduction and thus forward mappings preserve knowledge. Knowledge preserved in a simulation is not scientifically interesting. However, the *measurement* of the phenomena generated can be interesting when there are wet-lab counterparts measured similarly. All interesting simulation efforts *begin* by defining their measures, their usage context, their experimental protocol, etc.

The exploration of an inverse map from phenomena to generators (via abduction or induction) requires the researcher to hypothesize generators that could result in the observable phenomena. The question posed is this: given phenomena, what hypothetical generators (and measures) might generate them? When we find a hypothetical generator in the form of biomimetic mechanisms we are just as ignorant of the wet-lab system as we were before, but we now have a concrete instance of a plausible mechanism where before we only had vague, unchallenged concepts. In the absence of other concrete, competing theories, it can stand as the current best explanation for the phenomena. A full study of any inverse map requires abduction (several hypotheses), deduction (simulation), and induction (falsification), and that can only be done with concretizable hypotheses: synthetic analogues. It cannot be done with hypotheses that remain conceptual.

A generator–phenomenon map is not one-to-one. Many generator compositions (mechanisms) can produce phenomena, which when measured are indistinguishable. Also, different phenomena can result from the same generator composition. The same ISL micro-mechanisms and COMPOUNDS can produce almost an endless variety of outflow profiles. It follows that studying an inverse mapping requires multiple, seemingly plausible hypotheses, which then compete against each other during simulation experiments. Those that survive spawn additional, more refined hypotheses, following an iterative refinement protocol like the one in "The Scientific Method in Iterative Analogue Refinement." Inductively defined models implicitly handle multiple generators for the same phenomena. That is because they relate quantities and measurements, and are therefore inherently phenomenal. As stated in the text, they are robust to changes in composition. Synthetically defined analogues implicitly handle multiple phenomena for the same generators. They are inherently generative because they relate constituents and objects: they are robust to changes in context.

REFERENCES

1. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. Nat Biotechnol. 2004;22:1253–9.
2. Butcher EC. Can cell systems biology rescue drug discovery? Nat Rev Drug Discov. 2005;4:461–7.
3. van der Greef J, McBurney RN. Innovation: rescuing drug discovery: *in vivo* systems pathology and systems pharmacology. Nat Rev Drug Discov. 2005;4:961–7.

4. Kola I. The state of innovation in drug development. Clin Pharmacol Ther. 2008;83:227–30.

5. Chien JY, Friedrich S, Heathman MA, de Alwis DP, Sinha V. Pharmacokinetics/Pharmacodynamics and the stages of drug development: role of modeling and simulation. AAPS J. 2005;7: E544–59.

6. Byrne HM, Alarcon T, Owen MR, Webb SD, Maini PK. Modelling aspects of cancer dynamics: a review. Phil Trans R Soc A. 2006;364:1563–78.

7. Sumpter DJT. The principles of collective animal behaviour. Phil Trans R Soc B. 2006;361:5–22.

8. Boccalettia S, Latorab V, Morenod Y, Chavez M, Hwanga D-U. Complex networks: structure and dynamics. Phys Rep. 2006;424:175–308.

9. Goutelle S, Maurin M, Rougier F, Barbaut X, Bourguignon L, Ducher M, et al. The Hill equation: a review of its capabilities in pharmacological modelling. Fundam Clin Pharmacol. 2008;22:633–48.

10. Daun S, Rubin J, Vodovotz Y, Clermont G. Equation-based models of dynamic biological systems. J Crit Care. 2008;23:585–94.

11. Demin I, Goryanin. Kinetic modelling in systems biology. BocaRaton: Chapman & Hall; 2008.

12. Glazier JA, Balter A, Poplawski NJ. Magnetization to morphogenesis: a brief history of the Glazier-Graner-Hogeweg model. In: Anderson ARA, Chaplain MAJ, Rejniak KA, editors. Single-cell-based models in biology and medicine. Basel: Birkhäuser; 2007. p. 79–107.

13. Graner F, Glazier JA. Simulation of biological cell sorting using a two-dimensional extended Potts model. Phys Rev Lett. 1992;69:2013–6.

14. Glazier JA, Graner F. Simulation of the differential driven rearrangement of biological cells. Phys Rev E. 1993;47:2128–54.

15. Cardelli L. Brane calculi. Lecture notes in computer science (Computational Methods in Systems Biology) v3082. Springer; 2005. p. 257–78.

16. Priami C, Quaglia P. Beta binders for biological interactions. Lecture notes in computer science (Computational Methods in Systems Biology) v3082. 2005; p. 20–33.

17. Phillips A, Cardelli L, Castagna G. A graphical representation for biological processes in the stochastic pi-calculus. Transactions On Computational Systems Biology VII, Lc Notes Compu Sc. 2006;4230:123–52.

18. An G. Introduction of an agent-based multi-scale modular architecture for dynamic knowledge representation of acute inflammation. Theor Biol Med Model. 2008;5:11.

19. Fisher J, Henzinger TA. Executable cell biology. Nat Biotechnol. 2007;25(11):1239–49.

20. Hunt CA, Ropella GE, Park S, Engelberg J. Dichotomies between computational and mathematical models. Nat Biotechnol. 2008;26:737–8.

21. Park S, Ropella GE, Kim SH, Roberts MS, Hunt CA. Computational strategies unravel and trace how liver disease changes hepatic drug disposition. J Pharmacol Exp Ther. 2009;328(1):294–305.

22. Rowland M, Balant L, Peck C. Physiologically based pharmacokinetics in drug development and regulatory science: a workshop report (Georgetown University, Washington, DC, May 29–30, 2002). AAPS PharmSci. 2004;6:6.

23. Rescigno A. Foundations of pharmacokinetics. New York: Kluwer; 2003. p. 17–21.

24. Rescigno A. On the use of pharmacokinetic models. Phys Med Biol. 2001;49:4657–76.

25. Zeigler BP. Multifacetted modelling and discrete event simulation. San Diego: Academic; 1984.

26. Zeigler BP, Praehofer H, Kim TG. Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. San Diego: Academic; 2000.

27. Chaouiya C. Petri net modelling of biological networks. Brief Bioinform. 2007;8(4):210–9.

28. Von Neumann J, Burks AW. Theory of self-reproducing automata. Urbana: University of Illinois Press; 1966.

29. Düchting W, Dehl G. Spread of cancer cells in tissues: modelling and simulation. Int J Bio-Med Comput. 1980;11: 175–95.

30. Toffoli T. Cellular automata as an alternative to differential equations in modelling physics. Physica D. 1984;10:117–27.

31. Bodenstein L. A dynamic simulation model of tissue growth and cell patterning. Cell Differ. 1986;19:19–33.

32. Cocho G, Pérez-Pascual R, Rius JL. Discrete systems, cell–cell interactions and color pattern of animals. I. Conflicting dynamics and pattern formation. J Theor Biol. 1987;125:419–35.

33. Cocho G, Pérez-Pascual R, Rius JL, Soto F. Discrete systems, cell–cell interactions and color pattern of animals. II. Clonal theory and cellular automata. J Theor Biol. 1987;125:437–47.

34. Chowdhury D, Sahimi M, Stauffer D. A discrete model for immune surveillance, tumor immunity and cancer. J Theor Biol. 1991;152(2):263–70.

35. Celeda F, Seiden PE. A computer model of cellular interactions in the immune system. Immunol Today. 1992;13(2):56–62.

36. Ermentrout GB, Edelstein-Keshet L. Cellular automata approaches to biological modeling. J Theor Biol. 1993;160:97–133.

37. Smolle J, Stettner H. Computer simulation of tumour cell invasion by a stochastic growth model. J Theor Biol. 1993;160:63–72.

38. Kansal AR, Torquato S, Harsh GR, Chiocca EA, Deisboeck TS. Simulated brain tumour growth dynamics using a three-dimensional cellular automaton. J Theor Biol. 2000;203:367–82.

39. Dormann S, Deutsch A. Modelling of self-organized avascular tumour growth with a hybrid cellular automaton. In Silico Biol. 2002;2:35.

40. Alber MS, Kiskowski MA, Glazier JA, Jiang Y. On cellular automaton approaches to modeling biological cells, IMA Vol Math Appl. 2003;134:1–39.

41. Deutsch A, Dormann S. Cellular automaton modeling of biological pattern formation: characterization, applications, and analysis. Birkhäuser: Boston; 2005. p. 334.

42. Hatzikirou H, Deutsch A. Cellular automata as microscopic models of cell migration in heterogeneous environments. Curr Top Dev Biol. 2008;81:401–34.

43. Hogeweg P. Computing an organism: on the interface between informatic and dynamic processes. Biosystems. 2002;64:97–109.

44. Turner S, Sherratt JA. Intercellular adhesion and cancer invasion: a discrete simulation using the extended Potts model. J Theor Biol. 2002;216:85–100.

45. Izaguirre JA, Chaturvedi R, Huang C, Cickovski T, Coffland J, et al. CompuCell, a multi-model framework for simulation of morphogenesis. Bioinformatics. 2004;20:1129–37.

46. Longo D, Peirce SM, Skalak TC, Davidson L, Marsden M, Dzamba B, et al. Multicellular computer simulation of morphogenesis: blastocoel roof thinning and matrix assembly in Xenopus laevis. Dev Biol. 2004;271:210–22.

47. Peirce SM, Gleson EJV, Skalak TC. Multicellular simulation predicts microvascular patterning and in silico tissue assembly. FASEB J. 2004;18(6):731–3.

48. Chaturvedi R, Huang C, Kazmierczak B, Schneider T, Izaguirre JA, et al. On multiscale approaches to three-dimensional modelling of morphogenesis. J R Soc Interface. 2005;2(3):237–53.

49. Merks RMH, Glazier JA. A cell-centered approach to developmental biology. Phys Stat Mech Appl. 2005;352(1):113–30.

50. Anderson ARA, Chaplain MAJ, Rejniak KA, editors. Single-cell-based models in biology and medicine. Basel: Birkhäuser; 2007. p. 349.

51. Chen N, Glazier JA, Izaguirre JA, Alber MS. A parallel implementation of the Cellular Potts Model for simulation of cell-based morphogenesis. Comput Phys Commun. 2007;176:670–81.

52. Newman SA, Christley S, Glimm T, Hentschel HG, Kazmierczak B, Zhang YT, et al. Multiscale models for vertebrate limb development. Curr Top Dev Biol. 2008;81:311–40.

53. Christley S, Alber MS, Newman SA. Patterns of mesenchymal condensation in a multiscale, discrete stochastic model. PLoS Comput Biol. 2007;3(4):e76.

54. Thorne BC, Bailey AM, DeSimone DW, Peirce SM. Agent-based modeling of multicell morphogenic processes during development. Birth Defects Res C Embryo Today. 2007;81(4):344–53.

55. Thorne BC, Bailey AM, Peirce SM. Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. Brief Bioinform. 2007;8(4):245–57.

56. Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A, Li Y. In silico ADME/Tox: why models fail. J Comput Aided Mol Des. 2003;17:83–92.

57. Norinder U, Bergström CA. Prediction of ADMET properties. Chem Med Chem. 2006;1:920–37.

58. Dudek AZ, Arodz T, Gálvez J. Computational methods in developing quantitative structure-activity relationships (QSAR): a review. Comb Chem High Throughput Screen. 2006;9:213–28.

59. Dearden JC. *In silico* prediction of ADMET properties: how far have we come? Expert Opin Drug Metab Toxicol. 2007;3:635–9.

60. Ekins S, Mestres J, Testa B. *In silico* pharmacology for drug discovery: applications to targets and beyond. Br J Pharmacol. 2007;152:21–37.

61. Johnson SR. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). J Chem Inf Model. 2008;48:25–6.

62. Hou T, Wang J. Structure-ADME relationship: still a long way to go? Expert Opin Drug Metab Toxicol. 2008;4:759–70.

63. Liu Y, Hunt CA. Mechanistic study of the cellular interplay of transport and metabolism using the synthetic modeling method. Pharm Res. 2006;23(3):493–505.

64. Garmire LX, Garmire DG, Hunt CA. An *in silico* transwell device for the study of drug transport and drug–drug interactions. Pharm Res. 2007;24(12):2171–86.

65. Lam TN, Hunt CA. Discovering plausible mechanistic details of hepatic drug interactions. Drug Metab Dispos. 2009;37(1):237–46.

66. Sheikh-Bahaei S, Hunt CA. Prediction of *in vitro* hepatic biliary excretion using stochastic agent-based modeling and fuzzy clustering. In: Perrone LF *et al*, editors. Proceedings of the 37th Conference on Winter Simulation, Monterey, CA, Dec 03–6. 2006. p. 1617–24.

67. Yan L, Sheihk-Bahaei S, Park S, Ropella GE, Hunt CA. Predictions of hepatic disposition properties using a mechanistically realistic, physiologically based model. Drug Metab Dispos. 2008;36(4):759–68.

68. Gleeson MP. Generation of a set of simple, interpretable ADMET rules of thumb. J Med Chem. 2008;51:817–34.

69. Düchting W, Vogelsaenger T. Aspects of modelling and simulating tumor growth and treatment. J Cancer Res Clin Oncol. 1983;105:1–12.

70. Stevens A. Simulations of the aggregation and gliding behavior of myxobacteria. In Biological motion, lecture notes in biomathematics, vol 89. New York: Springer. 1990. p. 548–55.

71. Andersen T, Newman R, Otter T. Shape homeostasis in virtual embryos. Artif Life. 2009;15:161–83.

72. Bindschadler M, McGrath JL. Sheet migration by wounded monolayers as an emergent property of single-cell dynamics. J Cell Sci. 2007;120:876–84.

73. Cai AQ, Landman KA, Hughes BD. Multi-scale modeling of a wound-healing cell migration assay. J Theor Biol. 2007;245(3):576–94.

74. Robertson SH, Smith CK, Langhans AL, McLinden SE, Oberhardt MA, Jakab KR, *et al*. Multiscale computational analysis of Xenopus laevis morphogenesis reveals key insights of systems-level behavior. BMC Syst Biol. 2007;1:46.

75. Engelberg JA, Ropella GE, Hunt CA. Essential operating principles for tumor spheroid growth. BMC Syst Biol. 2008;2:110.

76. Odell GM, Foe VE. An agent-based model contrasts opposite effects of dynamic and stable microtubules on cleavage furrow positioning. J Cell Biol. 2008;183(3):471–83.

77. Galle J, Hoffmann M, Aust G. From single cells to tissue architecture-a bottom-up approach to modelling the spatiotemporal organisation of complex multi-cellular systems. J Math Biol. 2009;58(1–2):261–83.

78. Ridgway D, Broderick G, Lopez-Campistrous A, Ru'aini M, Winter P, Hamilton M, *et al*. Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. Biophys J. 2008;94(10):3748–59.

79. Walker DC, Southgate J, Hill G, Holcombe M, Hosea DR, Wood SM, *et al*. The epitheliome: agent-based modelling of the social behavior of cells. Biosystems. 2004;76:89–100.

80. Zhang L, Athale CA, Deisboeck TS. Development of a three-dimensional multiscale agent-based tumor model: simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. J Theor Biol. 2007;244(1):96–107.

81. Kim SH, Yu W, Mostov K, Matthay MA, Hunt CA. A computational approach to understand *in vitro* alveolar morphogenesis. PLoS ONE. 2009;4(3):e4819.

82. Forbes B, Ehrhardt C. Human respiratory epithelial cell culture for drug delivery applications. Eur J Pharm Biopharm. 2005;60:193–205.

83. Sakagami M. *In vivo*, *in vitro* and *ex vivo* models to assess pulmonary absorption and disposition of inhaled therapeutics for systemic delivery. Adv Drug Deliv Rev. 2006;58:1030–60.

84. Sporty JL, Horálková L, Ehrhardt C. *In vitro* cell culture models for the assessment of pulmonary drug disposition. Expert Opin Drug Metab Toxicol. 2008;4:333–45.

85. Yu W, Fang X, Ewald A, Wong K, Hunt CA, *et al*. Formation of cysts by alveolar type II cells in three-dimensional culture reveals a novel mechanism for epithelial morphogenesis. Mol Biol Cell. 2007;18:1693–700.

86. Grant MR, Hunt CA. An *in silico* analogue of *in vitro* systems used to study epithelial cell morphogenesis. Computational Methods In Systems Biology, Proceedings, Lc Notes Comput Sc. 2006;4210:285–97.

87. Bailey AM, Lawrence MB, Shang H, Katz AJ, Peirce SM. Agent-based model of therapeutic adipose-derived stromal cell trafficking during ischemia predicts ability to roll on P-selectin. PLoS Comput Biol. 2009;5(2):e1000294.

88. Hunt CA, Ropella GE, Yan L, Hung DY, Roberts MS. Physiologically based synthetic models of hepatic disposition. J Pharmacokinet Pharmacodyn. 2006;33(6):737–72.

89. Hung DY, Chang P, Weiss M, Roberts MS. Structure-hepatic disposition relationships for cationic drugs in isolated physiological models. JPET. 2001;297:780–9.

90. Yan L, Ropella GE, Park S, Roberts MS, Hunt CA. Modeling and simulation of hepatic drug disposition using a physiologically based, multi-agent *in silico* liver. Pharm Res. 2008;25(5):1023–36.

91. Yan L, Park S, Sheikh-Bahaei S, Ropella GEP, Hunt CA. Predicting hepatic disposition properties of cationic drugs using a physiologically based, agent-oriented *in silico* liver. In: Rajaei H, Wainer GA, Chinni MJ, editors. Proceedings of the 2008 Spring Simulation Multiconference, SpringSim 2008, Ottawa, Canada, April 14–17, 2008. SCS/ACM 2008. 2008a; p. 162–6.

92. An G. *In silico* experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. Crit Care Med. 2004;32:2050–60.

93. An G. Agent-based computer simulation and sirs: building a bridge between basic science and clinical trials. Shock. 2001;16(4):266–73.

94. Addis TR, Gooding DC. Simulation methods for an abductive system in science. Found Sci. 2008;13:37–52.

95. Fages F. From syntax to semantics in systems biology towards automated reasoning tools. In: Priami C *et al*., editors. Trans Comput Syst Biol IV, LNBI. 2006;3939:68–70.

96. Tang J, Ley KF, Hunt CA. Dynamics of *in silico* leukocyte rolling, activation, and adhesion. BMC Syst Biol. 2007;1:14.

97. Kim SHJ, Park S, Mostov K, Debnath J, Hunt CA. Computational investigation of epithelial cell dynamic phenotype *in vitro*. Theo Biol Med Model. 2009;6:8.

98. Gennari JH, Neal ML, Carlson BE, Cook DL. Integration of multi-scale biosimulation models via lightweight semantics. Pac Sym Biocomput. 2008;2008:414–25.

99. Sun Z, Finkelstein A, Ashmore J. Using ontology with semantic web services to support modeling in systems biology. In: Weske M, Hacid M-S, Godart C, editors. WISE 2007 Workshops, LNCS. 2007; 4832:41–51.

100. Hunt CA, Ropella GE, Yan L, Hung DY, Roberts MS. Physiologically based synthetic models of hepatic disposition. J Pharmacokinet Pharmacodyn. 2006;33(6):737–72.

101. FURM: A Functional Unit Represenmtation Method, http://furm.org/ (accessed 6/6/09).

102. An G. Dynamic knowledge representation using agent-based modeling: ontology instantiation and verification of conceptual models Ch. 15. In: Maly IV, editor. Methods in molecular biology: systems biology 500 Humana Press (Springer Science); 2009. doi:10.1007/978-1-59745-525-1_15.

103. Xiang X, Kennedy R, Madey G, Cabaniss S. Verification and validation of agent-based scientific simulation models. In: Yilmaz L., editor. Proceedings of the 2005 Agent-Directed Simulation Symposium, April 2005. The Society for Modeling and Simulation International 2005;37:47–55.

104. Moore RC. Semantical considerations on nonmonotonic logic. Art Intel. 1985;25:75–94.

105. Peirce CS. How to make our ideas clear. Pop Sci Monthly. 1878;12:286–302.

106. Peirce CS. Deduction, induction, and hypothesis. Pop Sci Monthly. 1878;13:470–82.

107. Yu CH. Abduction? Deduction? Induction? Is there a logic of exploratory data analysis?. 1994. http://www.creative-wisdom. com/pub/Peirce/Logic_of_EDA.html (accessed 5/24/09).

108. Magnani L. Abduction, reason and science—processes of discovery and explanation. New York: Kluwer; 2000.

109. Gabbay DM, Woods J. A practical logic of cognitive systems, volume 2: the reach of abduction: insight and trial. Elsevier. 2005.

110. Minar N, Burkhart R, Langton C, Askenazi M. The Swarm simulation system: a toolkit for building multi-agent simulations. Working Paper 96-06-042, Santa Fe Institute, Santa Fe, NM. 1996.

111. North MJ, Collier NT, Vos JR. Experiences creating three implementations of the Repast agent modeling toolkit. ACM TOMACS. 2006;16:1–25.

112. Luke S, Cioffi-Revilla C, Panait L, Sullivan K, Balan G. MASON: a multiagent simulation environment. Simulation. 2005;81:517–27.

113. Sklar E. NetLogo, a multi-agent simulation environment. Artif Life. 2007;13:303–11.

114. Klopfer E. Technologies to support the creation of complex systems models—using StarLogo software with students. Biosystems. 2003;71:111–22.

115. Gilbert N, Troitzsch K. Simulation for the social scientist. Maidenhead: Open University Press; 2005.

116. Railsback SF, Lytinen SL, Jackson SK. Agent-based simulation platforms: review and development recommendations. Simulation. 2006;82:609–23.