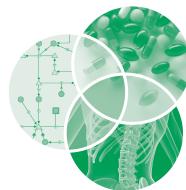


# Agent-based modeling: a systematic assessment of use cases and requirements for enhancing pharmaceutical research and development productivity



C. Anthony Hunt<sup>1\*</sup>, Ryan C. Kennedy<sup>1</sup>, Sean H. J. Kim<sup>2</sup>  
and Glen E. P. Ropella<sup>3</sup>

A crisis continues to brew within the pharmaceutical research and development (R&D) enterprise: productivity continues declining as costs rise, despite ongoing, often dramatic scientific and technical advances. To reverse this trend, we offer various suggestions for both the expansion and broader adoption of modeling and simulation (M&S) methods. We suggest strategies and scenarios intended to enable new M&S use cases that directly engage R&D knowledge generation and build actionable mechanistic insight, thereby opening the door to enhanced productivity. What M&S requirements must be satisfied to access and open the door, and begin reversing the productivity decline? Can current methods and tools fulfill the requirements, or are new methods necessary? We draw on the relevant, recent literature to provide and explore answers. In so doing, we identify essential, key roles for agent-based and other methods. We assemble a list of requirements necessary for M&S to meet the diverse needs distilled from a collection of research, review, and opinion articles. We argue that to realize its full potential, M&S should be actualized within a larger information technology framework—a dynamic knowledge repository—wherein models of various types execute, evolve, and increase in accuracy over time. We offer some details of the issues that must be addressed for such a repository to accrue the capabilities needed to reverse the productivity decline. © 2013 Wiley Periodicals, Inc.

How to cite this article:  
*WIREs Syst Biol Med* 2013. doi: 10.1002/wsbm.1222

## INTRODUCTION

Pharmaceutical research and development (R&D) is in the midst of a productivity decline. Unified, transdisciplinary, in silico modeling and simulation

(M&S) methods are viewed broadly as a promising countermeasure. Agent-based (AB) modeling is a phrase used currently to identify relatively young modeling methods that utilize software agents. We review and present evidence that AB methods will be essential contributors to successful M&S countermeasures.

\*Correspondence to: a.hunt@ucsf.edu

<sup>1</sup>Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA

<sup>2</sup>Department of Pharmacy and Therapeutics, University of Pittsburgh, Pittsburgh, PA, USA

<sup>3</sup>Tempus Dictum, Inc., Portland, OR, USA

Conflict of interest: The authors have declared no conflicts of interest in relation to this article.

Additional Supporting Information may be found in the online version of this article.

## TRI-FOCUS: AGENT-BASED M&S, R&D, AND EXPLANATORY MECHANISMS

The development of agent modeling tools and the availability of increasingly detailed, varied, and abundant data coupled with advances in computation

have made possible a growing number of agent-based modeling and simulation (AB M&S) applications across a variety of non-biomedical domains and disciplines. To illustrate, we identify early<sup>1–15</sup> and more recent<sup>16–32</sup> use cases. Macal and North<sup>33</sup> provide additional examples and a tutorial on AB M&S methods. Within the biomedical domain, AB M&S is a relatively new approach for studying systems composed of interacting components, some of which can be autonomous. The methods are used primarily to gain insight into mechanisms responsible for phenomena of living systems. We cite early<sup>34–40</sup> and more recent<sup>41–56</sup> examples of such applications. Amigoni and Schiaffonati<sup>57</sup>, An et al.<sup>58</sup>, and Edelman et al.<sup>59</sup> provide biomedically-focused reviews. Several applications exemplify the expanding variety of applications relevant to the pharmaceutical sciences.<sup>46,60–72</sup>

Pharmaceutical science stakeholders agree that there is a crisis within the broad pharmaceutical R&D domain (private and public): productivity continues to decline, even in the face of dramatic scientific and technical advances accompanied by a data deluge, especially at the molecular level. A flurry of recent reviews and commentaries<sup>73–102</sup> discuss the problems from several different perspectives and offer strategies and scenarios for how M&S methods can and are being used to enhance productivity. The different perspectives include pharmacometrics [including translational and physiologically based pharmacokinetics–pharmacodynamics (PBPK), and disease progression M&S], quantitative and systems pharmacology, and model-based drug development. Given the apparent fit between designed uses of AB M&S methods and the multifarious problems cited in the reviews that can benefit from computational M&S methods, it is surprising that only two of the citations mention AB M&S methods. No examples are presented within those citations of any current use of AB M&S methods within pharmaceutical companies. Does that situation signal ripe opportunity for AB M&S methods to add new value beyond that delivered using the established methods? That question, although provocative, is premature because one should not try to force-fit a particular M&S method to a particular set of problems. We should first evaluate where and how the cited domain experts envision M&S methods being used to reverse the productivity trend. We can then select those M&S methods that enable those uses.

The cited experts anticipate shifting focus from analysis of data to discovery and challenge of explanatory mechanisms. They see M&S becoming integral and essential to envisioned discovery and development

processes that are dramatically more productive in part because of improvements in use of M&S methods across all R&D activities. Yet they also make clear that the established practice is to select one or more modeling tools, drawn primarily from those already in use,<sup>75</sup> to address the particular problem at hand. The practice is the same independent of stage in the R&D process. Various scientists, often separated in time and space across R&D processes, integrate the derived information. Anticipating that having M&S integrated across R&D activities can improve productivity, what new demands are placed on M&S methods? We show why multifaceted, networked M&S use cases require drawing from an expanded computational modeling method and tool repertoire. Can current methods and tools fulfill the requirements? No. We justify that answer and discuss why AB and more advanced M&S methods are essential. It becomes clear that an analog (see Box 1) based knowledge repository will be needed. We identify M&S method requirements, and that brings into focus additional M&S use cases that we believe the repository's framework will likely need to enable.

## MODELING, SIMULATION AND THE PRODUCTIVITY CRISIS

Ideally, one should begin any M&S project without bias for any particular model types or methods.<sup>103</sup> Selecting M&S methods or tools in advance of specifying uses cases can constrain and even bias thinking about explanatory mechanisms. Hence, the first task is to clearly identify near and longer term expectations. Among the questions to be answered are these: what are the problems? What questions need to be answered? What new knowledge is sought and how will it be used? What decisions must be made? When are the deadlines? What resources are available? etc. Within pharmaceutical R&D, these questions are complicated by the nested, networked, multi-year, evolving nature of the overall enterprise. Scientific insight achieved at an early stage may be critical downstream (e.g., during dose ranging studies or clinical trials), and early planning for that prospect may influence selection of M&S methods in important ways. Yao et al.<sup>102</sup> illustrate the variety of relationships and common computational methods that support them (see Figure S1, Supporting Information).

Consider a model and method judged appropriate for a particular discovery or early development problem. The insight sought is intended to be useful for downstream as well as for current decision-making.

**BOX 1****DEFINITIONS**

**Actor:** an entity identifiable by an observer as a cause of an effect; an entity that participates in a process (plays a part); in computer science it is a mathematical model of computation that treats *actors* as the universal primitives: in response to a received message, an actor can make local decisions, create more actors, send messages, and determine how to respond to the next message received.

**Agent:** a software object that can schedule its own events (within an analog, it is quasi-autonomous); it senses and is part of its environment; it pursues and can revise an agenda within a larger script; some of 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 this context, it identifies a simulation model in which quasi-autonomous agents are key components. Terms that are often synonymous within the M&S literature include individual-based and multi-agent.

**Framework:** a carefully crafted assemblage of tools, devices (some software, some hardware, and occasionally even some wet-ware), usage protocols, good practices, etc., governed by a set of component interoperability standards. Upon satisfying use cases and listed requirements, we expect the framework to be an extensible, distributed, open, and loosely coupled (yet unified from the user's perspective).

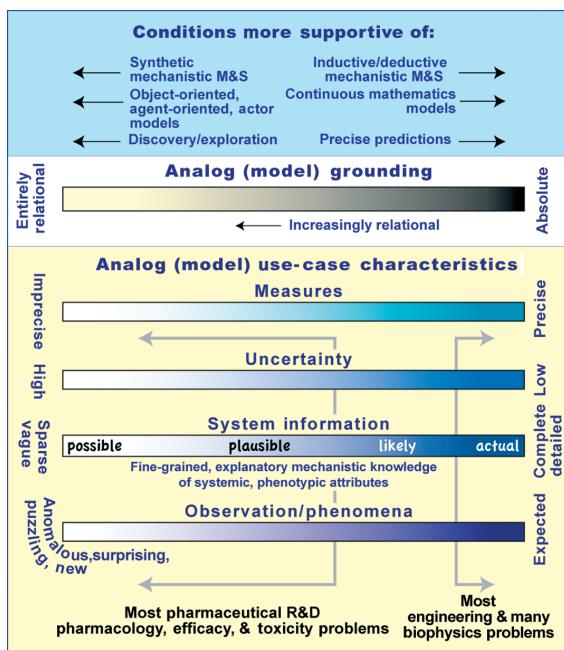
**Analog:** a software device that has (some) aspects and attributes that are similar to those of its R&D referent yet can exist and operate in isolation and in the absence of its R&D counterpart; in biomedical M&S, a model implemented in software that, when executed, produces phenomena that mimic one or more attributes measured or observed during referent wet-lab experiments. In this context, most analogs will be suitable for experimentation.

**In silico experiment:** it is precisely analogous to a wet-lab experiment. An analog is a hypothesis: the mechanism produced upon execution by interacting components will result in phenomena, often at different scales, that are similar, or not, to prespecified wet-lab phenomena. Measurement of features during execution enables testing the hypothesis. That activity is an *in silico* experiment.

**Analog based knowledge repository:** easily accessible, organized framework feature. Its content is an up-to-date instantiation of all relevant mechanistic knowledge in an accessible, easily understood, observable, and interactive form. It contains annotated records of analogs (current and falsified), their mechanisms, how they were composed plus the rationale, along with records of *in silico* experiments. All use cases (see 'Paramount use cases' subsection) can be achieved by employing the analog based knowledge repository. We envision domain experts making go/no-go decisions after interactive exploration of many scenarios (simulations) within the repository.

Can the model, method, and information be easily utilized as needed within a tight time window at a downstream go/no-go decision juncture? If doing so proves problematic because of differences in models, methods, and/or tools, or perhaps because of inadequate records of analogs and (more notably) their *in silico* experiments, etc., uncertainties and delays increase unnecessarily along with the risk of a flawed decision. The situation is complicated further by a larger, pressing, overarching need: reverse the declining trend in productivity.

The cited experts make the case that M&S methods are essential to simultaneously increasing productivity of therapeutic drug development, and facilitating the recursive cycle of new knowledge buildup aimed at improved quality care. Accepting that position, we view the enterprise as a networked, experiment-intensive system that begins small (significantly left on the Systems Information spectrum in Figure 1). As R&D progresses, the system evolves and expands. It becomes a large, distributed, multiscale, multi-aspect M&S challenge. For a particular aspect of the system, its location on the Systems Information spectrum in Figure 1 inches to the right. In what variety of ways do the cited experts envision M&S methods being used? Answers are provided in Supplement–Use Cases; a sample of answers is provided in Box 3. We propose a set of seven paramount use cases that are derived from and subsume those particular use cases. We then present five requirements, which we argue must be satisfied to enable achieving those paramount use cases efficiently. We discuss M&S methods that will enable a unified M&S approach to satisfy those requirements. To our knowledge, such an analysis has not been done. We



**FIGURE 1** | A particular analog use case can be characterized by an approximate location on each of the four lower spectra. Spectrum location along with specifics about how and for what the simulation model will be used within the larger R&D context, determine which M&S methods are most appropriate for a given use case. Within a pharmaceutical R&D context, a use case includes details of the specific, biology-focused wet-lab experiment that analog execution is intended to model in some way. Depending on use case location on the spectra, an analog can be located anywhere along a spectrum of software devices (models) ranging from synthetic (all components designed to be plugged together and are thus replaceable) to purely inductive models. *System Information* includes current conceptual knowledge about the mechanisms on which wet-lab experiments focused. As the R&D process advances, evidence will shrink the space of possible mechanisms. The result will be a set of plausible analog mechanisms supported by validation evidence. Later, that set too will shrink. The result will be a smaller set of likely mechanisms (those that have survived several falsification experiments). An R&D project's product can be successful without being able to designate key mechanisms as either actual or likely. Actual mechanisms are typically known for engineered systems, but are typically lacking in therapeutics. Grounding is discussed in Box 2; additional information is provided as Supporting Information. Conditions on the far right of the bottom four spectra are supportive of models (typically, continuous equations) that rely exclusively on absolute grounding. Hunt et al.<sup>104</sup> make the case that when left of center on one or more of the bottom three spectra, models should rely more on relational grounding. When on the far left, early stage, purely qualitative M&S is still useful and productive: it facilitates goal-oriented research efforts by clarifying (unifying) current thinking about referent phenomena. Such models would typically be coarse grain and use relational grounding. Spectra colors were selected arbitrarily.

argue that such an examination is necessary and essential to discover strategies capable of restoring and enhancing productivity.

## INTEGRATIVE MANAGEMENT OF KNOWLEDGE AND UNCERTAINTY

Knowledge integration is an essential, paramount use case. For this discussion, we limit attention to knowledge relevant to the primary R&D project focus: the particular disease, morbidity, or health issue; the biologic or chemical entity treatment intervention; the pharmacological, toxicological, and clinical outcomes that are consequences of treatments; and the variety of mechanisms (even when vague) that are offered to explain those phenomena. We ignore other important categories of knowledge that can impact go/no-go decisions, but the approach and framework described below can be expanded to include them. Examples of those categories include regulatory science; chemistry, manufacturing, and control; human resources; accessible contract services; etc.

The cited reviews discuss roles of various computational models and methods in knowledge integration. However, we explain below that little knowledge actually resides in those models. Much of the mechanistic insight resides within mental models, and that presents a problem: mental model differences, similarities, and inconsistencies are difficult, and often impossible to ascertain. Mental models are subject to their own forms of error introduction and propagation. Increasing reliance on synthetic analogs (defined below under Mechanism-Based Approach) presents advantages because they can evolve into executable representations of what we know (or think we know) about biological systems. Those representations are called executable biological knowledge embodiments<sup>64,65,104</sup> and dynamic knowledge representations.<sup>34,58,60,61,105–109</sup> Such executions are suitable for knowledge discovery.<sup>58,108,109</sup> Knowledge embodiment is made feasible because synthetic analogs provide concrete instances of that knowledge rather than analytic descriptions of conceptual representations. When an analog is executed, it demonstrates when, how, and where our knowledge matches or fails to match referent system details, which enables and facilitates knowledge discovery.<sup>58,108,109</sup>

For current knowledge and beliefs to be useful (especially in a social context like shared model usage, validation, and falsification), it must be embedded in an analog and visible to the user (which is often not the case now). Analogs help build schemata for knowledge (and ignorance) representation, which can provide a mechanism for the curation and maintenance of the embedded knowledge. Users must be able to readily identify the knowledge and be able to discuss it, rely on it, dispute it, and falsify it (or not), all

while reflecting on how that knowledge impacts the project. To achieve these capabilities, most, if not all, modeling activities and all mechanism representations will need to take place within a framework of the type described in Box 1 and discussed below. The R&D project related content of that framework will become part of an interactive, analog based knowledge repository.

It should be noted that knowledge is not embedded in the any of the variety of pharmacometric or systems pharmacology models identified in the reviews.<sup>73–102</sup> Model refers specifically to the equations used to describe the referent aspects, dynamics, and features<sup>75,79,84–90,93,95</sup> as implemented in software. As documented by An<sup>105,107</sup> and Hunt et al.<sup>64,104</sup> embedding knowledge in pharmacometric and ordinary differential equation (ODE) models is challenging, if not problematic. Equations are typically used to describe patterns in data. The associated conceptual models do reflect knowledge, but it exists separate from the equations. Humans interpreting the I/O of the cited equations use prose and sketches to provide the semantic grounding for those models in terms of both the idealized, conceptual model and knowledge of the referent system. That semantic grounding is done manually, and is separate from the equations and their software implementations. Hence, neither knowledge nor semantics is *embedded* in those equations. However, the equation models are typically encapsulated within ‘ready to use’ software tools and packages. Bouzom et al.<sup>75</sup> lists and describes the more frequently used general and constrained tools and packages. The builders, especially of the constrained, domain-focused tools and packages, invest considerable effort in collecting, organizing, and enabling use of domain knowledge, information, and data related to a variety of model use cases. So doing considerably lowers the barrier to model parameterization and scenario exploration. The preceding points are expanded upon in Supporting Information using two concrete examples (Box 2).

## Acknowledging and Reducing Uncertainty

We focus on uncertainties directly associated with the mechanistic theories on which the project is based. At the start of a typical project, uncertainty is pervasive. Consequently, projects typically begin considerably left in Figure 1 Uncertainty and System Information spectra. Uncertainty sources are varied and plentiful. They need to be identified, annotated, and updated. As the project advances successfully, we inch to the right on both the System Information and

## BOX 2

### ANALOG GROUNDING ISSUES

Grounding issues must be addressed to satisfy the listed requirements. Key issues discussed in detail by Hunt et al.<sup>104</sup> are summarized in Supporting Information. The units, dimensions, and/or objects to which a variable or model constituent refers establish groundings. Absolute grounding: variables, parameters, and input–output (I/O) are in real-world units. Relational grounding: variables, parameters, and I/O are in units defined by other system components. Relational grounding requires a separate analog-to-referent mapping model. To satisfy requirements, a spectrum of model classes, methods, and groundings (illustrated in Figure 1) will be needed: absolute grounding occupies one extreme; relational grounding occupies the other.

Absolute grounding provides simple, interpretive mappings between simulation output, parameter values, and referent data. However, complex issues must be addressed each time one of the following occurs: expand the model to include additional phenomena; combining models; and/or model context changes. Expansions are challenging, even infeasible, when center-left in Figure 1. Reusability is hindered in part because the conflated semi-mechanistic, equation-based model and the model-to-referent mapping model have different use cases.

Analogs must evolve (become more complicated) as R&D advances and new mechanistic insight accumulate. That evolution will require changing, adding, and removing component linkages within analogs. We recommend keeping most component groundings relational. So doing facilitates component replacement, limiting any one component formulation solely to its coupling with the others. Any component can be replaced at will as long as the minimal I/O interface requirements are met. Starting with relationally grounded analog components allows the modeler to iterate progressively from qualitative to quantitative validation.

Uncertainty spectra. Productivity-enhancing M&S methods must facilitate and increase the pace of that movement in real time. We can infer that new M&S methods will be needed that are capable of generating new knowledge without requiring new wet-lab experiments. We can also infer that they must also be capable of exploration and advanced selection of

## BOX 3

## CLUSTERED EXAMPLES OF SPECIFIC M&amp;S USE CASES

These examples (selected randomly) were drawn from those listed in Supplement—Use Cases. They illustrate the diversity and scope of particular envisioned<sup>73–102</sup> M&S use cases.

*Knowledge integration:*

Quantitatively address questions concerning the functional relationship between prognostic factors, dosage, and outcomes.<sup>86,93,96</sup>

Integrate the basic components to describe and understand the complex interplay between the pharmacology of drug action and (patho-)physiological systems.<sup>85,93</sup>

*Management of uncertainty*

Account for uncertainty in the underlying assumptions and thus resulting prediction of drug effects.<sup>76,91</sup>

*Decision support*

Document decision-making from discovery through development to regulatory filing and improve answering postfiling questions and approval, as well as life cycle management of the asset.<sup>96</sup>

Provide a decision-making tool for selecting effective and safe doses, optimizing study sample sizes, evaluating alternative trial designs and making rational go/no-go decisions based on the probability of achieving predefined study goals.<sup>75,78,79,91,96,97</sup>

*Preclinical and clinical development*

Facilitate design and/or selection of lead compounds, selection of the first-in-human dose, early clinical trial design, and proof-of-concept studies of experimental drugs and drug combinations.<sup>74,75,82,83,85,102</sup>

Simulate outcomes of alternative study designs before the experimental investigation commences—incorporating different doses and/or different patients and computing the probability of a successful trial given the characterized patient population and proposed treatment regimens.<sup>75,78,86</sup>

Predict the toxic potential of chemicals and human adverse effect, and generate hypotheses about the putative molecular mechanisms of chemical-induced injury.<sup>94,99</sup>

*Drug-disease modeling*

Use to understand the relationship between drug effect and the natural progression of the underlying disease.<sup>76,85,91,93</sup>

Describe macrophysiological processes within a particular disease state and use to

understand likely modulation of those processes with specific interventions.<sup>73</sup>

*Drug-drug interaction, special populations*

Guide investigations in pediatric, the elderly and other special populations.<sup>76,84,88–91,95</sup>

Identify and assess complex drug-drug interactions early in drug development so that clinical studies could be planned or prioritized to assess the risk.<sup>79</sup>

*Prediction*

Allow for extrapolation of the PK properties across species and compounds.<sup>85,86,89,95</sup>

Ultimately predict human PK from *in silico*, *in vitro*, and physicochemical data.<sup>83</sup>

strategies to reduce uncertainties systematically in real time in parallel with advancing wet-lab experiments. Preference must be given to M&S methods that can do both.

Reliance on computational models that simply abstract away uncertainties is counterproductive: so doing does yield simple, easily managed models, which can be useful for particular use cases, but it also adds a new source of uncertainty. The core source of uncertainty is the experimental evidence on which the project's mechanistic theories are based (in the context of all other available knowledge and insight), and we recognize that current theory at a particular project stage is just one drawn from a space of possible or plausible mechanistic theories.

Studies conducted by Amgen<sup>110</sup> and Bayer HealthCare<sup>111</sup> found that published, preclinical, scientific findings that were important to their R&D efforts could be confirmed in only 11–25% of cases. A rule-of-thumb among early-stage venture capital firms is that at least 50% of published studies, even those in top-tier academic journals, cannot be repeated.<sup>112</sup> Consequently, there is significant risk that the evidence supporting the conceptual mechanistic models on which the project is based is flawed (even if the theory proves reasonably correct). Even when experiments are repeated, variability of results can be considerable. Rarely can measurements made on cell culture models be mapped quantitatively 1:1 to comparable measures made on animal models. Similarly, animal model phenomena rarely map 1:1 to human counterparts. Hence, there is considerable uncertainty about how results from different experiments can or should be mapped to—and thus influence—the conceptual mechanistic models on which project scientists are relying. Such uncertainty and mapping dilemmas present problems for those using conventional inductive, equation-based models (typically ODEs)

of the type discussed in the cited reviews. Modelers understand the issues: when using such methods, they would prefer to be on the far right of Figure 1. The conventional strategy is to request more data. So doing postpones mechanistic modeling. Unfortunately, it may be necessary for the project to advance beyond the next few go/no-go decision points before the requested data begins to come available. While waiting for the required data, the accepted strategy for dealing with uncertainty issues is to abstract them away: if we make particular simplifying assumptions, we can create an idealized, hypothetical mechanistic scenario that, if fully validated, will place us on the right side of Figure 1 where inductive, predictive models can be reliable. In doing so, model-grounding issues, discussed below, are typically ignored. As the project advances, there is no straightforward way to ‘add back’ the various uncertainties abstracted away, even if they could be measured, or undo the simplifying assumptions.

Analogs are particularly useful in managing uncertainty because they can be used to simulate possible and plausible mechanisms when center-left on the three lower Figure 1 spectra (although our focus here is on uncertainty, other elements that are often also abstracted away, such as biological or experimental details, may also be handled more concretely by analogs). An analog’s mechanism is a consequence of components interacting during execution. Monte Carlo variations in component specifications, rules specifying interactions, and parameterizations can cause phenomena generated during each execution to be unique. That process simulates the non-deterministic nature of biological phenomena. That variety can also represent uncertainty about generative mechanisms, experimental variability, and intra- and interindividual variability. The challenge then becomes to follow a disciplined protocol, as done in Sheikh-Bahaei and Hunt<sup>71</sup>, Hunt et al.<sup>109</sup>, and Lam and Hunt<sup>113</sup>, to identify and apply constraints in conjunction with *in silico* experimentation (see Box 1) to systematically shrink the space of acceptable Monte Carlo variations. Having multiple, equally satisfactory analogs of the same referent is an acknowledgment and representation of uncertainties that can shrink but not vanish.

## Avoiding Information Loss, ‘Warts and All’, is Essential

A product of modeling efforts cited in the reviews, irrespective of model type, is derived measures that are recorded for use by others. Examples include mean predicted ED50 ( $\pm$ SD), clearance, half-life,

volume of distribution, etc. Examples described in Supplement–Use Cases<sup>73–102</sup> illustrate that derived measures acquired during an early R&D stage can influence the perspective taken or focal aspect of interest at a later stage. However, derived measures are lossy (information is lost). Features present in the raw experimental observations are lost. Contextual information, including experiment method details, assay information, etc., may also be lost. Information loss is a concern: later, it may prove critical. Similar to issues discussed above, it becomes increasingly infeasible to recover lost information. Today, preventing such information loss can be challenging: proper measures are required to preserve, manage, and transfer information. During a project’s lifetime (and thereafter), it requires attention to those preservation details across different functional groups through each development stage, from preclinical to postmarketing. The risk of critical information loss increases for conceptual model descriptions enriched with quantitative, mechanistic, pharmacological, physiological, and systems biology details that do not readily reduce to simple, mathematical or statistical descriptions.

Despite the above concerns, we stress that derived measures along with highly abstracted models are essential: researchers need them. They provide a much needed bridge from particular and specific concrete modeling to the more powerful generalized models from which researchers will develop the theory for therapy controlled normal-to-morbid or diseased-to-normal transitions. For this reason, the framework must facilitate access to, and the analog based knowledge repository must house, curate, and facilitate access to analogs, metadata, and much of the experimental observations from which the metadata were derived.

Consider measures derived from multiple different models separated in time, combined and extended beyond their original model use cases, and how they may influence downstream conceptual models used for go/no-go decision-making. Conceptual models can be expected to push the go/no-go decision in one direction. A quite different direction may result if all the original models, including those that at an earlier time were judged ‘failures’, could be re-run together to directly inform decision makers when that decision is required. An analog based knowledge repository fulfilling the requirements presented below will guard against information loss and knowledge distortion. Having intuitive, easily understood analogs will reduce dependency on, and usefulness of, lossy, derived measures to inform domain experts’ conceptual models.

## EFFECTIVE DECISION-MAKING SUPPORT REQUIRES A UNIFIED FRAMEWORK

Today, domain experts rely primarily on their own conceptual models—*informed*, of course, by computational models—to make go/no-go decisions. Improved productivity can be achieved by changing that relationship: domain experts make go/no-go decisions after using simulations to explore many scenarios within a computational framework built and equipped specifically for scenario exploration in addition to enabling the multiple M&S methods needed to satisfy the requirements below. It will be the ‘framework’ within which all model use cases occur. We envision the framework becoming an **analog based knowledge repository**: an up-to-date instantiation of all accumulated, new, and proprietary mechanistic knowledge in an accessible, easily understood, observable, and interactive form. As a consequence, R&D project team members will have reduced dependency on the difficult to challenge conceptual mechanistic models of current, past, and no longer available domain experts.

Current experiment records and/or protocols capture and preserve some current, underlying, mechanistic conceptualizations, however, that contextual information is typically decoupled from data and can become unavailable in subsequent phases of development. A good practice will be that information is documented and provided as part of annotation within the framework. It will not be ancillary, as in OpenABM.<sup>115</sup> In order to effectively incorporate data, especially wet-lab data across discovery and development stages for future use, automated capabilities will be needed that enable and facilitate metadata annotations, which may include biological, anatomical, and physiological details across biological scales. Of course, the framework will also provide methods for storage, curation, composition, and execution of analogs representing what is known about domain-specific referent systems. Its core constituents (specific items) will be data, semantic relationships, workflow actions, and computational components. There will also be a variety of derived constituents including data sets, semantic networks (e.g., XML ontologies), workflows, plus analogs and their components. Under Resources, we list currently available tools that can be used to organize and complete all of the preceding tasks.

As an example of a knowledge repository decision support use case, consider a task to estimate the likelihood of success for a clinical trial given a compound and cohort sample specifications. The

following may seem futuristic, yet tools for enabling all capabilities (listed under Web Resources) are in use today within different domains. The project’s inter-disciplinary team, including regulatory scientists, statisticians, programmers, regulatory staff, etc., populate the knowledge repository with experimental protocols (Workflows under Resources), experimental data sets, analog mechanisms (Executable models under Resources) representing the cohort and compounds, and maps between the terms (objects, methods, variables, parameters, graphs, etc.) used in all these elements. Software agents and actors within the repository simplify the process. A user specifies a set of objectives or criteria for a successful (or unacceptable) clinical trial outcome and assembles at least one analog that might plausibly generate data satisfying the objectives. Repository agents facilitate the composition by periodically checking the consistency and completeness of the evolving analog. The user then executes the analog according to study design workflow(s). Execution results will exhibit the systemic causation and variability generated by fine-grain, networked events embedded in analog components. Repository agents will compare and contrast the results to the objectives and present the user with similarity scores validating or falsifying analogs. Repository agents may also present a suite of alternatively composed analogs consistent with the data, ontologies, and workflows previously installed. Similarity scores provide a rudimentary estimate of the likelihood of success for the clinical trial. Further execution and similarity scores of alternative analogs provide refinement of and confidence (or the lack thereof) in those estimates. Note that because we are left of center in Figure 1, the approach avoids assuming the existence of a perfectly accurate analog. Any analog that achieves all face validation targets and satisfies the various similarity measures will be considered valid (until falsified). We envision all of the preceding being completed within hours.

As a technical note, an analog may contain components based on different formalisms. Some may be graph theoretic. Some may use ODEs. Others will be (or will use) agents. The preceding discussions illustrate that referring to the modeling activities as being AB is misleading. Agent-directed or agent-oriented are more accurate descriptions.<sup>114</sup>

## FROM PARTICULAR TO PARAMOUNT M&S USE CASES

We worked through each of the cited reviews and commentaries<sup>73–102</sup> and identified more than ninety general and specific M&S use cases. Although each

article presented its assessment from a particular perspective, there was, as might be expected, considerable overlap in specific use cases. That overlap guided and facilitated clustering them into seven categories that span all pharmaceutical R&D activities. Given use cases, we sought general requirement statements that would subsume two or more uses within each category. They are listed in the subsequent section.

## Paramount Use Cases

The following are called paramount M&S use cases because they subsume the particular use cases in Supplement—Use Cases clustered under the following seven categories.

### *Knowledge Integration*

Given multiple organizational perspectives (disease state, prognostic factors, drug characteristics, cohort variability, disease progression, drug effects, timelines, budgets, etc.) and multiple data sets (qualitative and quantitative; some possibly incommensurate) from various experiments and experimental models, solve for a collection of alternative, composable, explanatory mechanisms parameterized by, and validated against, available data. Here, ‘solve’ is defined as applying constraints within an iterative protocol (see e.g., Tang and Hunt<sup>116</sup>, Park et al.<sup>117</sup>, Engelberg et al.<sup>118</sup>) to shrink a set of possible mechanisms into a much smaller set of plausible (supported early by qualitative validations) and incrementally more likely mechanisms (supported by quantitative validation).

Use collected mechanisms to assemble concrete, consistent, comprehensive, clinically relevant ‘stories’ about the co-evolution, during and following treatment, of subject (multiscale), the condition being treated, and treatment as well as (absorption, distribution, metabolism, and elimination) ADME plus response monitoring in the case of a drug. In this context, a story is the narrative created during simulation, and a simulation results from executing an *in silico* experiment. There are several reasons why a good simulation story is important.<sup>119,120</sup> Use assembled stories to discover and make explicit potential conflicts, voids, and ambiguities within an analog based knowledge repository, thereby providing immediate predictive and analytical use, while bringing into focus paths for repository improvement.

### *Management of Uncertainty*

Present for *in silico* experimentation, plausible mechanisms exhibiting variability, both composite and

singular, that are qualitatively and quantitatively similar to that seen (or expected) in laboratory experiments or clinical trials, including matching changes in variability across different cohorts and study designs. Use simulations to estimate probability of achieving clinical target efficacy outcomes, and when feasible, estimate probability of specified, undesired effects.

### *Decision Support*

Present contextualized assemblies of plausible mechanisms parameterized to provide an evidence-based justification for componentized estimates of likelihood of success at critical stages within the development process. Use the assemblies and their stories to estimate efficacy and safety windows, including confidence intervals, and to demonstrate sample sizes within appropriate population cohorts needed for constraining study outcomes to within those windows.

### *Preclinical and Clinical Development*

Preclinical and clinical development along with postmarketing uses are covered by requirements listed above and below.

### *Drug-disease Modeling*

Solve for plausible mechanism composites satisfying multiple long-term aspects of diseases, drug effects and fates within cohorts, placebo effect, and disease modifying interventions. Use those mechanisms to predict compound behavior (PK) within contexts of interest, including hypothetical mechanisms for lack of adherence, dropout, and multiple drug interaction.

### *Drug-drug Interaction, Special Populations*

Solve for plausible mechanism composites of compound interactions of interest, including time-dependent inhibition, induction, and competition between a parent compound and its metabolites. Solve for plausible mechanism composites for various cohorts (animals, children, and adults) based on classifications of compound behavior in each cohort. Hypothesize and build mechanism translation maps between analog cohorts (e.g., between adults and children). Use plausible translation maps to design clinical trials for human cohorts.

### *Prediction*

Solve for plausible mechanism composites by multi-objective search within constraints defined by fine-grained classifications of cohorts (particular attributes at multiple levels and scales) and compounds (e.g., particular physicochemical properties), including their formulation, across model types (*in vitro* to *in vivo*, animal to human). Use the plausible solutions

to predict the outcomes of precise and particular regimens on individually characterized cohorts, both across (translation) and within model types.

## REQUIREMENTS

To enable all paramount use cases, the framework must enable, and analog systems must meet, these five requirements.

1. An analog's components and spaces will be concrete (enabling knowledge embodiment), wherein its details will be directly defined by its use cases. Analog components will be somewhat modular, in schedule as well as state. So doing helps accomplish the following framework activities.
  - a. It enables defining and annotating component- and module-to-biological counterpart mappings, making them explicit, intuitive, and easily understood. It enables experimentation on concrete analogs. It is essential for wet-lab R&D scientists and decision makers to easily follow, interpret, and comment, unassisted, on simulation details. For that, the embodied knowledge needs to be easily accessible, which requires transparency in representation and execution—form and function.
  - b. It enables making modules quasi-autonomous and thus more biomimetic.<sup>65,104</sup> Analog components composing virtual cells, organs, animal models, and ultimately individuals, must exhibit some level of autonomous behavior in order to improve similarity with biological counterpart phenotypes, while increasing their explicative and predictive utility.
  - c. Components can be adapted easily to represent different past and future experiment designs and protocols.
  - d. It is straightforward to change mechanistic detail (granularity, resolution) to simulate additional attributes or experiments. It facilitates scaling (translation, morphing) among *in silico* experimental systems to represent transitions from *in vitro* to animal models, from animal model to human cohorts, and from normal health to morbid.
  - e. It enables reusing analogs and components along with their embedded knowledge, for study of new chemical entities (or biologics) and new intervention scenarios under similar or different morbidity constraints or epigenetic influences.
  - f. It facilitates verification through unit testing, where each component can be tested in isolation as well as in the composed analog context.
  - g. It facilitates versioning, where each component can evolve independent of other components.
  - h. It enables building trust in surviving analogs by accumulating direct *in silico*-to-wet-lab validation evidence, where measures taken during *in silico* experiments are mapped quantitatively to counterpart measures taken during wet-lab experiments. So doing reduces reliance on derived measures, which can mask information loss, assumptions, and uncertainty removal. Many of those validation exercises will rely extensively on pharmacometrics and conventional modeling methods.
  - i. It facilitates archiving analog and mechanism evolution along with *in silico* experiment successes and failures within the framework. The latter is particularly important. When an analog or *in silico* experiment fails in some way, we acquire new knowledge, e.g., a feature of an analog mechanism thought to have a particular *in vitro* biological counterpart, does not. However, in a different context, that mechanism or some variant may prove useful.
2. Components and spaces can be assembled easily to simulate current, past, and future laboratory or clinical experiments. Generating many alternative, plausible, testable (through *in silico* experimentation) components for each function/structure and then selecting against those that fail, is needed to insure generation of new knowledge. So doing helps accomplish two framework activities.
  - a. It becomes increasingly easy to construct (plug together) and explore alternative mechanistic hypotheses and intervention scenarios. It facilitates contrasting their predictions during simulation. So doing can help avoid dictating 'best methods' prematurely. The latter keeps the door open for yet-to-be-stated M&S use cases while

- increasing opportunities for serendipitous insight and/or discovery.
- b. It becomes increasingly easy to construct multiscale, multiresolution, multi-attribute analogs (eventually individualized virtual patients) composed of heterogeneous (form, function, methods, formalisms, etc.) components.
  - 3. Simulation experiments are feasible in the presence and absence of chemical entity objects (hereafter, CE-objects). They are also feasible in the presence of multiple CE-objects. Components within analogs can recognize different CE-objects and adjust their response accordingly.
  - 4. Coarse grain (from the perspective of biological organization) phenomena will derive mostly from local component interactions at a finer grain (local includes a living entity's immediate environment). When required, finer grain mechanisms can respond to coarser grain phenomena.
  - 5. Semi-automated modeling methods are needed to more rapidly complete three critical activities:
    - a. Conduct *in silico* experimentation to explore and shrink spaces of competing mechanistic hypotheses plus alternative mechanism instantiations and parameterizations.
    - b. Use cross-model validation methods (based on quantitative similarity measures) to discover parsimonious options to increase and decrease component and analog granularity when milestones change and when new questions require new use cases, which necessitates changing targeted attributes and/or shifting attention to new aspects and phenomena.
    - c. Discover testable hypotheses about how changes in clinical, field, or laboratory measures may be linked mechanistically to observed or observable changes in particular biological level phenomena, especially molecular-level phenomena.

## MECHANISM-BASED APPROACH

New projects are typically initiated based on evidence (even if limited) supporting conceptual mechanistic models (morbidity or disease progression, cause-effect, pharmacology, clinical outcomes, etc.)

that often include hypothesized molecular targets. There are large gaps in the mechanistic knowledge landscape that must be filled strategically to facilitate making 'correct' go/no-go decisions. The mechanistic landscape spans cell cultures, model organisms, and humans. It also spans pharmacology, toxicology, disposition, metabolism, and more. Today, no one member of the R&D enterprise has a comprehensive 'view' of that landscape, yet it is clear from the cited experts that full knowledge of the current state of that landscape and what can be predicted from it is needed at each go/no-go decision juncture. Knowledge of—or use of one or more features on—that mechanism landscape is common to all identified model use cases.

The ability to navigate, use, and leverage current mechanistic insight is a common feature of all paramount use cases. Enabling those use cases requires transitioning from models that are separate and distinct computational (or diagrammatic) descriptions of conceptual representations (e.g., the systems biology, PBPK, and pharmacodynamic models discussed in the cited reviews) to concrete instances of that knowledge. Doing so is feasible using synthetic analogs of mechanisms.<sup>65</sup> We create analogs by combining (plugging together) specific elements, often varied and diverse, so as to form a coherent biomimetic whole. The analog is synthetic because it is a software mechanism constructed from extant, autonomous components (in this case, executable software components) whose existence and purpose are independent of the model or mechanistic landscape that they comprise. The expectation is that, upon execution, the interacting elements and components—the mechanism undergoing simulation—will exhibit event sequences and outcomes that are measurably similar to counterpart mechanisms within the corresponding wet-lab experiments that will be used to validate the analog. See Yan et al.<sup>119</sup>, Lam and Hunt<sup>113</sup>, Tang and Hunt<sup>116</sup>, Engelberg et al.<sup>118</sup>, and Park et al.<sup>118</sup>, for validated examples of synthetic analogs and their mechanisms, all designed for use center-left on the three lower Figure 1 spectra.

## Exploratory Iteration

One final, critical capability is needed in order to enable the remaining paramount use cases. It must be straightforward to repeat any simulation experiment using a different compound, set of compounds, or no compound. So doing can be accomplished most easily by building analogs using object-oriented software

methods and using different mobile objects to map to different compounds of interest. We require that objects representing drugs can be added (or not) and, when ‘inactive’, their presence will not interfere with any already validated mechanism. Examples of that approach are provided in Sheikh-Bahaei and Hunt<sup>71</sup>, Lam and Hunt<sup>113</sup>, Park et al.<sup>118</sup>, Yan et al.<sup>121</sup>, and Sheikh-Bahaei et al.<sup>122</sup>. In those reports, one CE-object maps to a very small amount of referent compound in a small aliquot of measurable material taken from a referent wet-lab experiment, such as culture media, blood, tissue, etc. In time, these CE-objects will carry information that distinguishes one from another, such as structure specifications, particular physicochemical properties, affinities for biological substrates, etc. It follows that any analog component that might feasibly interact with a CE-object must be able to ‘read’ structure specifications and particular physicochemical properties, and use that information to adjust (or not) its interactions with that object.

To enable the latter, analog components will need to be ‘intelligent’ (able to use artificial intelligence methods): they will be programmed with the results of many earlier validation and falsification experiments and can—absent user intervention—arrive at a customized parameterization that determines how they will interact with a new CE-object. To illustrate, imagine that analogs of each of a battery of *in vitro* model systems have achieved degrees of validation for ten compounds. One analog’s referent is an *in vitro* system used to characterize metabolic profiles and predict human metabolic clearance. Focus on that analog: upon validation, for each mechanistic event, it retains its parameterizations for all ten CE-objects. The project team needs an answer to this question. Given three, competing, new chemical entities, are their expected metabolic clearance values within a target range? Each analog component, for each unique mechanistic event, can use available framework tools, follow a provided protocol, and construct a predictive map from the space of selected structure specifications and particular physicochemical properties to the space of validated parameterizations. We extend the mapping to the three new compounds and arrive at unique analog parameterizations for each CE-object counterpart. We then conduct analog clearance experiments using each of the three CE-objects. We use the results to predict wet-lab clearance measures. Sheikh-Bahaei and Hunt<sup>71</sup> described a prototype example of such a protocol. With a knowledge base of only ten compounds, we would have only limited trust in those predictions. However, we would expect to have more trust after doing the same using a

knowledge base of fifty compounds. At that stage we could begin characterizing families of compounds as well as particular molecules. As more information is embedded, the knowledge repository increasingly facilitates the work of the domain expert.

The same basic approach can be used to explore expected outcomes of *in silico* toxicology experiments, experiments in analogs of animal disease models, and even analogs of normal and special human cohorts. There would be no technical barriers to simulation experiments that explore possible drug–drug interactions.

## Beyond Paramount Use Cases

Figure 1 helps to bring into focus important challenges faced by any new, high risk, high gain pharmaceutical R&D project. It can be characterized as being left of center on the bottom three spectra. A task is to acquire just enough new information and knowledge (move right in Figure 1) to make better-informed, ‘good’ go/no-go decisions sooner within budget constraints and in the face of considerable uncertainty. The expectation is that simulations will use available knowledge to provide essential information and/or guidance moving forward. To do so, M&S methods must be engaged at the start of the project. Otherwise there will be considerable risk that M&S efforts will lag behind wet-lab efforts. The logical beginning scenario would be to pull together components from models successfully supporting more mature projects, modify them as appropriate, and assemble them to begin being synergistic with wet-lab experiments. Even when the new effort is far left in Figure 1 spectra, speculative and qualitative analogs can be constructed to help researchers think clearly, explicitly, and concretely about the referent and design specific, well-focused wet-lab experiments. That scenario is an example of model use cases beyond those that directly support product development and approval. It also speaks to requirements, covered by those above, which have apparently not yet been considered.

A related, also essential use case will occur once a new product has been approved: revisit the process *in silico* (from start to finish) to explore alternative R&D, knowledge acquisition paths that would have been more time, cost, and/or knowledge effective. To do so will require simulating phases of the R&D process, including various wet-lab experiments performed or not. Such exploratory simulation may seem futuristic, but it is easy to believe that, with such capabilities, the field will have achieved dramatic productivity gains and that

is the objective. These additional M&S use cases illustrate that future development of simulation models must take place within a common framework, if the simulations are expected to make a lasting contribution to the knowledge base. Simulated R&D activities will be analogs of past or future, real or considered R&D activities. None of the analogs used, however, will be fully detailed or fully validated. Because we will be center-left on the bottom three Figure 1 spectra, we know that they will (always) be flawed in ways yet to be determined. Those flaws will be hidden by built-in uncertainties.

Simulations that support and add plausible detail to early conceptual models, such as hybrids of systems biology, PBPK, and pharmacodynamic models, will be useful. Simulation models that can also falsify aspects of those conceptual models, earlier rather than later, will be especially valuable because hypothesis falsification, not validation, generates the new knowledge that will be needed for making go/no-go decisions.<sup>123,124</sup> Enabling mechanism falsification in addition to mechanism validation is another, important use case not specifically identified in the cited reviews. Falsification of an analog requires that its mechanism be concrete and particular.<sup>65</sup> Plausible conceptual mechanisms, especially those that are more complicated, can in theory be falsified using wet-lab experimentation, but doing so can be challenging and time- and resource-intensive.

## Analog Based Knowledge Repository and AB Models

Representing within a knowledge repository the variety of data structures, experimental observations, and documented biological phenomena that may influence go/no-go decisions during the R&D projects is beyond the scope of any one model of computation (MoC). A successful analog based knowledge repository will require the co-existence of multiple, occasionally inconsistent, yet equally valid, models of the same referent(s), as has been done within other domains.<sup>125</sup> Inconsistency robustness<sup>126</sup> is important to the aspect-oriented nature of scientific M&S.<sup>65,104,127</sup> The management of inconsistent models is part of the strategy for integrating MoCs into a coherent knowledge repository. The choice of underlying MoC for any particular activity is driven by the focal aspects or model (analog or component) use cases. Variety in aspects is better realized by variety in MoCs. For example, a dataflow MoC can co-exist with a discrete event analog of a mouse cancer model,

the former representing a functional relationship between high volume population-oriented variables (like concentration in a PBPK model) and the latter representing mechanistic relationships between unique, biomimetic components. For the foreseeable future, such a repository will not, itself, be AB, but it will be partially agent-directed.<sup>115</sup> Agents such as Experiment Agents used in our work<sup>118,128–130</sup> follow protocols and perform tasks exactly analogous to those human modelers perform: set up and execute various models (analog, modules, etc.) and use pharmacometric methods. Drawing on work from Lee and colleagues<sup>125,131,132</sup>, we anticipate no restrictions being placed on MoCs, except where required for managed composition, execution, and analysis as has been done with Ptolemy II<sup>125</sup> for M&S distributed computing scenarios. Depending on use case requirements, analogs may be realized by any given MoC, including continuous-time systems, ABMs, ODEs, systems and process networks, Stream X-Machines, etc.

## Conclusion

An eruption of recent articles, including those cited,<sup>73–102</sup> is focusing attention on the declining productivity crisis stressing pharmaceutical R&D. The hundred plus authors of the more than thirty cited reviews and commentaries argue from different perspectives (pharmacometrics, quantitative and systems pharmacology, model-based drug development, etc.) for expanded use of M&S methods as the primary means to reverse the downward productivity trend. They describe more than ninety particular M&S use cases. The Box 3 includes particular examples. They were merged into paramount use cases listed above. We made the case that facile realization of those use cases will require use of an analog based knowledge repository that fulfills five requirements. AB M&S methods will play important roles along with those methods already in use, but additional capabilities and methods are required. Importantly, there are no technology gaps. The five requirements can be satisfied by the tools listed under Resources, which are in use within other domains.

Reversing the productivity decline requires rethinking how M&S methods can and should be used within the larger R&D process. It requires shifting focus from analysis of data to discovery of explanatory mechanisms. The latter, conditioned paramount use cases, requires implementing the requirements listed above. As illustrated in Figure 1, and as explained in Hunt et al.<sup>104</sup> and Supporting Information, so doing

requires increased reliance on relational grounding and diminished reliance on absolute grounding. It also requires being clear about how computational

models are being used and continuously exploring how barriers to participation in M&S activities can be lowered.

## ACKNOWLEDGMENTS

This work was supported in part by the CDH Research Foundation, the Alternatives Research and Development Foundation, and the EPA (G10C20235). The funding bodies had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication. We gratefully acknowledge commentary and suggestions provided by Brenden Peterson, Rada Savic, and Andre J. Jackson.

## REFERENCES

- Arthur WB, Durlauf SN, Lane DA, eds. *The Economy As An Evolving Complex System II*, Santa Fe Institute Series. Santa Fe, NM: Westview Press; 1997.
- Axelrod R. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press; 1997.
- Bankes SC. Agent-based modeling: a revolution? *Proc Natl Acad Sci USA* 2002, 99:7199–7200. doi:10.1073/pnas.072081299.
- Bonabeau E. Agent-based modeling: methods and techniques for simulating human systems. *Proc Natl Acad Sci USA* 2002, 99(suppl 3):7280–7287. doi:10.1073/pnas.082080899.
- Davidsson P. Multi agent based simulation: beyond social simulation. In: Moss S, Davidsson P, eds. *Multi Agent Based Simulation. Lecture Notes in Computer Science*. Berlin: Springer; 2000 1979:97–107. doi:10.1007/3-540-44561-7\_7.
- DeAngelis DL, Cox DK, Coutant CC. Cannibalism and size dispersal in young-of-the-year largemouth bass: experiment and model. *Ecol Model* 1979, 8:133–148.
- DeAngelis DL, Mooij WM. Individual-based modeling of ecological and evolutionary processes. *Annu Rev Ecol Evol Syst* 2005, 36:147–168. doi:10.1146/annurev.ecolsys.36.102003.152644.
- Epstein JM, Axtell R.. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, DC: Brookings Institution Press and The MIT Press; 1996.
- Grimm V. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecol Model* 1999, 115:129–148. doi:10.1016/S0304-3800(98)00188-4.
- Judson OP. The rise of the individual-based model in ecology. *Trends Ecol Evol* 1994, 9:9–14. doi:10.1016/0169-5347(94)90225-9.
- Kephart JO. Software agents and the route to the information economy. *Proc Natl Acad Sci USA* 2002, 99(suppl 3):7207–7213. doi:10.1073/pnas.082080499.
- LeBaron B. Short-memory traders and their impact on group learning in financial markets. *Proc Natl Acad Sci USA* 2002, 99(suppl 3):7201–7206. doi:10.1073/pnas.072079699.
- Parunak HVD, Savitt R, Riolo RL. Agent-based modeling vs. equation-based modeling: a case study and users' guide. In Sichman JS, Conte R, Gilbert N, eds. *Multi-Agent Systems and Agent Based Simulation Lecture Notes in Computer Science*. Berlin: Springer; 1998 1534:10–25. doi:10.1007/10692956\_2.
- Rose KA, Cowan JH. Predicting fish population dynamics: compensation and the importance of site-specific considerations. *Environ Sci Policy* 2000, 3(suppl 1):433–443. doi:10.1016/S1462-9011(00)00054-X.
- Schelling TC. *Micromotives and Macrobbehavior*. New York, NY: W.W. Norton & Company, Inc ; 1978.
- Auchincloss AH, Riolo RL, Brown DG, Cook J, Diez Roux AV. An agent-based model of income inequalities in diet in the context of residential segregation. *Am J Prev Med* 2011, 40:303–311. doi:10.1016/j.amepre.2010.10.033.
- Bobashev GV, Goedecke DM, Yu F, Epstein JM. A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches. In: Henderson SG, Biller B, Hsieh MH, Shortle J, Tew JD, Barton RR, eds. *Proceedings of the 2007 Winter Simulation Conference IEEE, New York*; 2007, 1511–1516. doi:10.1109/WSC.2007.4419767.
- Bousquet F, Le Page C. Multi-agent simulations and ecosystem management: a review. *Ecol Model* 2004, 176:313–332. doi:10.1016/j.ecolmodel.2004.01.011.
- Caplat P, Anand M, Bauch C. Symmetric competition causes population oscillations in an individual-based model of forest dynamics. *Ecol Model* 2008, 211:491–500. doi:10.1016/j.ecolmodel.2007.10.002.

20. Castella J-C, Verburg PH. Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam. *Ecol Model* 2007, 202:410–420. doi:10.1016/j.ecolmodel.2006.11.011.
21. Charles S, Subtil F, Kielbassa J, Pont D. An individual-based model to describe a bullhead population dynamics including temperature variations. *Ecol Model* 2008, 215:377–392. doi:10.1016/j.ecolmodel.2008.04.005.
22. Faugeras B, Maury O. Modeling fish population movements: from an individual-based representation to an advection–diffusion equation. *J Theor Biol* 2007, 247:837–848. doi:10.1016/j.jtbi.2007.04.012.
23. Garcia R. Uses of agent-based modeling in innovation/new product development research. *J Prod Innovat Manage* 2005, 22:380–398. doi:10.1111/j.1540-5885.2005.00136.x.
24. Garcia R, Jager W. From the special issue editors: agent-based modeling of innovation diffusion. *J Prod Innovat Manage* 2011, 28:148–151. doi:10.1111/j.1540-5885.2011.00788.x.
25. Grimm V, Railsback SF. *Individual-based Modeling and Ecology*. Princeton, NJ: Princeton University Press; 2005.
26. Grosman PD, Jaeger JAG, Biron PM, Dussault C, Oullet J-P. Trade-off between road avoidance and attraction by roadside salt pools in moose: An agent-based model to assess measures for reducing moose-vehicle collisions. *Ecol Model* 2011, 222:1423–1435. doi:10.1016/j.ecolmodel.2011.01.022.
27. López-Sánchez M, Noria X, Rodríguez JA, Gilbert N. Multi-agent based simulation of news digital markets. *Int Jf Comput Sci Appl* 2005, 2:7–14.
28. Kohler TA, Gumerman GJ, Reynolds RG. Simulating ancient societies. *Sci Am* 2005, 293:76–82, 84. doi:10.1038/scientificamerican0705-76.
29. Pan X, Han CS, Dauber K, Law KH. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI Soc* 2007, 22:113–132. doi:10.1186/1471-2105-6-228.
30. Sato H. Simulation of the vegetation structure and function in a Malaysian tropical rain forest using the individual-based dynamic vegetation model SEIB-DGVM. *Forest Ecol Manage* 2009, 257:2277–2286. doi:10.1016/j.foreco.2009.03.002.
31. Tang W, Bennett DA. Agent-based modeling of animal movement: a review. *Geography Compass* 2010, 4:682–700. doi:10.1111/j.1749-8198.2010.00337.x..
32. Rand R, Rust RT. Agent-based modeling in marketing: Guidelines for rigor. *Int J Res Market* 2011, 28:181–193. doi:10.1016/j.ijresmar.2011.04.002.
33. Macal CM, North MJ. Tutorial on agent-based modelling and simulation. *J Simul* 2010, 4:151–162. doi:10.1057/jos.2010.3.
34. An G. Agent-based computer simulation and sirs: building a bridge between basic science and clinical trials. *Shock* 2001, 16:266–273.
35. Casal A, Sumen C, Reddy TE, Alber MS, Lee PP. Agent-based modeling of the context dependency in T cell recognition. *J Theor Biol* 2005, 236:376–391. doi:10.1016/j.jtbi.2005.03.019.
36. Kreft JU, Booth G, Wimpenny JWT. Bacsim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* 1998, 144:3275–3287. doi:10.1099/00221287-144-12-3275.
37. Mansury Y, Kimura M, Lobo J, Deisboeck TS. Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model. *J Theor Biol* 2002, 219:343–370. doi:10.1006/jtbi.2002.3131.
38. Segovia-Juarez JL, Ganguli S, Kirschner D. Identifying control mechanisms of granuloma formation during M. tuberculosis infection using an agent-based model. *J Theor Biol* 2004, 231:357–376. doi:10.1016/j.jtbi.2004.06.031.
39. Thorne BC, Bailey AM, Peirce SM. Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Brief Bioinform* 2007, 8:245–257. doi:10.1093/bib/bbm024.
40. Walker DC, Southgate J, Hill G, Holcombe M, Hose DR, Wood SM, Mac Neil S, Smallwood RH. The epitheliome: agent-based modelling of the social behaviour of cells. *Biosystems* 2004, 76:89–100. doi:10.1016/j.biosystems.2004.05.025.
41. Artzy-Randrup Y, Rorick MM, Day K, Chen D, Dobson AP, Pascual M. Population structuring of multi-copy, antigen-encoding genes in *Plasmodium falciparum* eLife 2012, 1:e00093. doi:10.7554/eLife.00093.
42. 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:e1000294. doi:10.1371/journal.pcbi.1000294.
43. Bogle G, Dunbar PR. Agent-based simulation of T-cell activation and proliferation within a lymph node. *Immunol Cell Biol* 2010, 88:172–179. doi:10.1038/icb.2009.78.
44. Fallahi-Sichani M, El-KebirM, Marino S, Kirschner DE, Linderman JJ. Multiscale computational modeling reveals a critical role for TNF- $\alpha$  receptor 1 dynamics in tuberculosis granuloma formation. *J Immunol* 2011, 186:3472–3483. doi:10.4049/jimmunol.1003299.
45. Joyce KE, Laurienti PJ, Hayasaka S. Complexity in a brain-inspired agent-based model. *Neural Networks* 2012, 33:275–290. doi:10.1016/j.neunet.2012.05.012.

46. Kim PS, Lee PP. Modeling protective anti-tumor immunity via preventative cancer vaccines using a hybrid agent-based and delay differential equation approach. *PLoS Comput Biol* 2012, 8:e1002742. doi:10.1371/journal.pcbi.1002742.
47. Klann MT, Lapin A, Reuss M. Agent-based simulation of reactions in the crowded and structured intracellular environment: Influence of mobility and location of the reactants. *BMC Syst Biol* 2011, 5:71. doi:10.1186/1752-0509-5-71.
48. Li NY, Verdolini K, Clermont G, Mi Q, Rubinstein EN, Hebda PA, Vodovotz Y. A patient-specific *in silico* model of inflammation and healing tested in acute vocal fold injury. *PLoS ONE* 2008, 3:e2789. doi:10.1371/journal.pone.0002789.
49. Marino S, El-Kebir M, Kirschner D. A hybrid multi-compartment model of granuloma formation and T cell priming in Tuberculosis. *J Theor Biol* 2011, 280:50–62. doi:10.1016/j.jtbi.2011.03.022.
50. Martínez IV, Gómez EJ, Hernando ME, Villares R, Mellado M. Agent-based model of macrophage action on endocrine pancreas. *Int J Data Mining Bioinform* 2012, 6:355–368. doi:10.1504/IJDMB.2012.049293.
51. Mina P, di Bernardo M, Savery NJ, Tsaneva-Atanasova K. Modelling emergence of oscillations in communicating bacteria: a structured approach from one to many cells. *J R Soc Interface* 2013, 10:20120612. doi:10.1098/rsif.2012.0612.
52. Pogson M, Holcombe M, Smallwood R, Qwarnstrom E. Introducing spatial information into predictive NF- $\kappa$ B modeling—an agent-based approach. *PLoS ONE* 2008, 3:e2367. doi:10.1371/journal.pone.0002367.
53. Smallwood R. Computational modeling of epithelial tissues. *Wiley Interdiscip Rev Syst Biol Med*, 2009, 1:191–201. doi:10.1002/wsbm.18.
54. Stern JR, Christley S, Zaborina O, Alverdy JC, An G. Integration of TGF- $\beta$ - and EGFR-based signaling pathways using an agent-based model of epithelial restitution. *Wound Repair Regen* 2012, 20:862–871. doi:10.1111/j.1524-475X.2012.00852.x.
55. Tang J, Enderling H, Becker-Weimann S, Pham C, Polyzos A, Chen CY, Costes SV. Phenotypic transition maps of 3D breast acini obtained by imaging-guided agent-based modeling. *Integr Biol* 2011, 3:408–421. doi:10.1039/C0IB00092B.
56. Zhang H, Vaksman Z, Litwin DB, Shi P, Kaplan HB, Igoshin OA. The mechanistic basis of Myxococcus xanthus rippling behavior and its physiological role during predation. *PLoS Comput Biol* 2012, 8:e1002715. doi:10.1371/journal.pcbi.1002715.
57. Amigoni F, Schiaffonati V. Multiagent-based simulation in biology—a critical analysis. *Model-Based Reason Sci Technol Med Stud Comput Intell* 2007, 64:179–191.
58. An G, Mi Q, Dutta-Moscato J, Vodovotz Y. Agent-based models in translational systems biology. *Wiley Interdiscip Rev Syst Biol Med* 2009, 1:159–171. doi:10.1002/wsbm.45.
59. Edelman LB, Eddy JA, Price ND. *In silico* models of cancer. *Wiley Interdiscip Rev Syst Biol Med* 2010, 2:438–459. doi:10.1002/wsbm.75.
60. An G. *In silico* experiments of existing and hypothetical cytokine-directed clinical trials using agent-based modeling. *Crit Care Med* 2004, 32:2050–2060. doi:10.1097/01.CCM.0000139707.13729.7D.
61. An G, Bartels J, Vodovotz Y. *In silico* augmentation of the drug development pipeline: examples from the study of acute inflammation. *Drug Dev Res* 2011, 72:187–200. doi:10.1002/ddr.20415.
62. Alberts S, Keenan MK, D’Souza RM, An G. Data-parallel techniques for simulating a mega-scale agent-based model of systemic inflammatory response syndrome on graphics processing units. *Simulation* 2012, 88:895–907. doi:10.1177/0037549711425180.
63. Fallahi-Sichani M, Flynn JL, Linderman JJ, Kirschner DE. Differential risk of tuberculosis reactivation among anti-TNF therapies is due to drug binding kinetics and permeability. *J Immunol* 2012, 188:3169–3178. doi:10.4049/jimmunol.1103298.
64. Hunt CA, Ropella GEP. Moving beyond *in silico* tools to *in silico* science in support of drug development research. *Drug Dev Res* 2011, 72:153–161. doi:10.1002/ddr.20412.
65. Hunt CA, Ropella GE, Lam TN, Tang J, Kim SH, Engelberg JA, Sheikh-Bahaei S. At the biological modeling and simulation frontier. *Pharmaceut Res* 2009, 26:2369–2400. doi:10.1007/s11095-009-9958-3.
66. Kim SH, Jackson AJ, Hur R, Hunt CA. Individualized, discrete event, simulations provide insight into inter- and intra-subject variability of extended-release, drug products. *Theor Biol Med Model* 2012, 9:39. doi:10.1186/1742-4682-9-39.
67. Lam TN, Hunt CA. Discovering plausible mechanistic details of hepatic drug interactions. *Drug Metab Dispos* 2009, 37:237–246. doi:10.1124/dmd.108.023820.
68. Marino S, Linderman JJ, Kirschner DE. A multi-faceted approach to modeling the immune response in tuberculosis. *Wiley Interdiscip Rev Syst Biol Med* 2011, 3:479–489. doi:10.1002/wsbm.131.
69. 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, 328:294–305. doi:10.1124/jpet.108.142497.
70. Pepper JW. Drugs that target pathogen public goods are robust against evolved drug resistance. *Evol Appl* 2012, 5:757–761. doi:10.1111/j.1752-4571.2012.00254.x.

71. Sheikh-Bahaei S, Hunt CA. Enabling clearance predictions to emerge from *in silico* actions of quasi-autonomous hepatocyte components. *Drug Metab Dispos* 2011, 39:1910–1920. doi:10.1124/dmd.111.038703.
72. Sun X, Zhang L, Tan H, Bao J, Strouthos C, Zhou X. Multi-scale agent-based brain cancer modeling and prediction of TKI treatment response: incorporating EGFR signaling pathway and angiogenesis. *BMC Bioinform* 2012, 13:218. doi:10.1186/1471-2105-13-218.
73. Allerheiligen SR. Next-generation model-based drug discovery and development: quantitative and systems pharmacology. *Clin Pharmacol Ther* 2010, 88:135–137. doi:10.1038/clpt.2010.81.
74. Baneyx G, Fukushima Y, Parrott N. Use of physiologically based pharmacokinetic modeling for assessment of drug-drug interactions. *Future Med Chem* 2012, 4:681–693. doi:10.4155/fmc.12.13.
75. Bouzom F, Ball K, Perdaems N, Walther B. Physiologically based pharmacokinetic (PBPK) modelling tools: how to fit with our needs? *Biopharm Drug Dispos* 2012, 33:55–71. doi:10.1002/bdd.1767.
76. Edginton AN, Joshi G. Have physiologically-based pharmacokinetic models delivered? *Expert Opin Drug Metab Toxicol* 2011, 7:929–934. doi:10.1517/17425255.2011.585968.
77. Gobburu JV, Lesko LJ. Quantitative disease, drug, and trial models. *Annu Rev Pharmacol Toxicol* 2009, 49:291–301. doi:10.1146/annurev.pharmtox.011008.145613.
78. Grasela TH, Slusser R. Improving productivity with model-based drug development: an enterprise perspective. *Clin Pharmacol Ther* 2010, 88:263–268. doi:10.1038/clpt.2010.117.
79. Grillo JA, Zhao P, Bullock J, Booth BP, Lu M, Robie-Suh K, Berglund EG, Pang KS, Rahman A, Zhang L, et al. Utility of a physiologically-based pharmacokinetic (PBPK) modeling approach to quantitatively predict a complex drug-drug-disease interaction scenario for rivaroxaban during the drug review process: implications for clinical practice. *Biopharm Drug Dispos* 2012, 33:99–110. doi:10.1002/bdd.1771.
80. Huisenga W, Telgmann R, Wulkow M. The virtual laboratory approach to pharmacokinetics: design principles and concepts. *Drug Discov Today* 2006, 11:800–805.
81. Iyengar R, Zhao S, Chung SW, Mager DE, Gallo JM. Merging systems biology with pharmacodynamics. *Sci Transl Med* 2012, 4:126ps7. doi:10.1126/scitranslmed.3003563.
82. Jiang W, Kim S, Zhang X, Lionberger RA, Davit BM, Conner DP, Yu LX. The role of predictive biopharmaceutical modeling and simulation in drug development and regulatory evaluation. *Int J Pharm* 2011, 418:151–160. doi:10.1016/j.ijpharm.2011.07.024.
83. Jones HM, Dickins M, Youdim K, Gosset JR, Atkins NJ, Hay TL, Gurrell IK, Logan YR, Bungay PJ, Jones BC, Gardner IB. Application of PBPK modelling in drug discovery and development at Pfizer. *Xenobiotica* 2012, 42:94–106. doi:10.3109/00498254.2011.627477.
84. Khalil F, Läer S. Physiologically based pharmacokinetic modeling: methodology, applications, and limitations with a focus on its role in pediatric drug development. *J Biomed Biotechnol* 2011, 2011:907461. doi:10.1155/2011/907461.
85. Mager DE, Jusko WJ. Development of translational pharmacokinetic-pharmacodynamic models. *Clin Pharmacol Ther* 2008, 83:909–912. doi:10.1038/clpt.2008.52.
86. Nucci G, Gomeni R, Poggesi I. Model-based approaches to increase efficiency of drug development in schizophrenia: a can't miss opportunity. *Expert Opin Drug Discovery* 2009, 4:837–856. doi:10.1517/17460440903036073.
87. Prokop A, Michelson S. Development: multiscale CSB—simulation tools. In: *Systems Biology in Biotech and Pharma. Springer Briefs in Pharmaceutical Science and Drug Development*, Springer, New York; vol 2; 2012, 87–102. doi:10.1007/978-94-007-2849-3\_8.
88. Rostami-Hodjegan A. Physiologically based pharmacokinetics joined with *in vitro-in vivo* extrapolation of ADME: a marriage under the arch of systems pharmacology. *Clin Pharmacol Ther* 2012, 92:50–61. doi:10.1038/clpt.2012.65.
89. Rowland M, Peck C, Tucker G. Physiologically-based pharmacokinetics in drug development and regulatory science. *Ann Rev Pharmacol Toxicol* 2011, 51:45–73. doi:10.1146/annurev-pharmtox-010510-100540.
90. Sinha V, Kimko HHC. Recent developments in physiologically based pharmacokinetic modeling. *Clin Trial Simul AAPS Adv Pharm Sci Series* 2011, 1:483–499. doi:10.1007/978-1-4419-7415-0\_21.
91. Suryawanshi S, Zhang L, Pfister M, Meibohm B. The current role of model-based drug development. *Expert Opin Drug Discovery* 2010, 5:311–321. doi:10.1517/17460441003713470.
92. Zhang L, Pfister M, Meibohm B. Concepts and challenges in quantitative pharmacology and model-based drug development. *AAPS J* 2008, 10:552–559. doi:10.1208/s12248-008-9062-3.
93. Zhang L, Sinha V, Forgue ST, Callies S, Ni L, Peck R, Allerheiligen SR. Model-based drug development: the road to quantitative pharmacology. *J Pharmacokinet Pharmacodyn* 2006, 33:369–393.
94. Jack J, Wambaugh J, Shah I. Systems toxicology from genes to organs. *Methods Mol Biol (Clifton, NJ)* 2013, 930:375–397. doi:10.1007/978-1-62703-059-5\_17.

95. Kuepfer L, Lippert J, Eissing T. Multiscale mechanistic modeling in pharmaceutical research and development. *Adv Exp Med Biol* 2012, 736:543–561. doi:10.1007/978-1-4419-7210-1\_32.
96. Wetherington JD, Pfister M, Banfield C, Stone JA, Krishna R, Allerheiligen S, Grasela DM. Model-based drug development: strengths, weaknesses, opportunities, and threats for broad application of pharmacometrics in drug development. *J Clin Pharmacol* 2010, 50(9 suppl):S1–S46S. doi:10.1177/0091270010377629.
97. van der Graaf PH, Benson N. Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharmaceut Res* 2011, 28:1460–1464. doi:10.1007/s11095-011-0467-9.
98. Morgan P, Van Der Graaf PH, Arrowsmith J, Feltner DE, Drummond KS, Wegner CD, Street SD. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov Today* 2012, 17:419–424. doi:10.1016/j.drudis.2011.12.020.
99. Valerio LG. *In silico* toxicology for the pharmaceutical sciences. *Toxicol Appl Pharmacol* 2009, 241:356–370. doi:10.1016/j.taap.2009.08.022.
100. Vicini P. Multiscale modeling in drug discovery and development: future opportunities and present challenges. *Clin Pharmacol Ther* 2010, 88:126–129. doi:10.1038/clpt.2010.87.
101. van der Graaf PH. *CPT: Pharmacometrics Syst Pharmacol* 2012, 1:e8. doi:10.1038/psp.2012.8.
102. Yao L, Evans JA, Rzhetsky A. Novel opportunities for computational biology and sociology in drug discovery. *Trends Biotechnol* 2010, 28:161–170. doi:10.1016/j.tibtech.2009.06.003.
103. Shannon, RE. Introduction to the art and science of simulation. In Simulation Conference Proceedings, 1998. Winter; 1998 1:7–14.
104. Hunt CA, Ropella GEP, Lam TN, Gewitz AD. Relational grounding facilitates development of scientifically useful multiscale models. *Theor Biol Med Model* 2011, 8:35. doi:10.1186/1742-4682-8-35.
105. 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. doi:10.1186/1742-4682-5-11.
106. An G. Dynamic knowledge representation using agent-based modeling: ontology instantiation and verification of conceptual models. *Methods Mol Biol (Clifton, NJ)* 2009, 500:445–468. doi:10.1007/978-1-59745-525-1\_15.
107. An GC. Translational systems biology using an agent-based approach for dynamic knowledge representation: an evolutionary paradigm for biomedical research. *Wound Repair Regen* 2010, 18:8–12. doi:10.1111/j.1524-475X.2009.00568.x.
108. Ropella GEP, Hunt CA. Foundations for Extended Life Cycle Biological Models. CiteSeerX10.1.1.1.2664 2003. Available at: [http://biosystems.ucsf.edu/publications/Ropella\\_UCITR\\_2003.pdf](http://biosystems.ucsf.edu/publications/Ropella_UCITR_2003.pdf) (accessed January 3, 2013).
109. Hunt CA, Ropella GE, Yan L, Hung DY, Roberts MS. Physiologically based synthetic models of hepatic disposition. *J Pharmacokinet Pharmacodyn* 2006, 33:737–772.
110. Begley CG, Ellis LM. Drug development: raise standards for preclinical cancer research. *Nature* 2012, 483:531–533. doi:10.1038/483531a.
111. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov* 2011, 10:712. doi:10.1038/nrd3439-c1.
112. Osherovich L. Hedging against academic risk. *Sci Business eXchange* 2011, 4(15). doi:10.1038/scibx.2011.416.
113. Lam TN, Hunt CA. Mechanistic insight from *in silico* pharmacokinetic experiments: roles of P-glycoprotein, Cyp3A4 enzymes, and microenvironments. *J Pharmacol Exp Ther* 2010, 332(2):398–412. doi:10.1124/jpet.109.160739.
114. OpenABM. Available at: <http://www.openabm.org/standards>. (Accessed January 5, 2013).
115. Yilmaz L, Ören T, eds. *Agent-directed Simulation and Systems Engineering*, Wiley Series in Systems Engineering and Management. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, ISBN: 978-3-527-40781-1; 2009.
116. Tang J, Hunt CA. Identifying the rules of engagement enabling leukocyte rolling, activation, and adhesion. *PLoS Comput Biol* 2010, 6:e1000681. doi:10.1371/journal.pcbi.1000681.
117. Park S, Kim SH, Ropella GEP, Roberts MS, Hunt CA. Tracing multiscale mechanisms of drug disposition in normal and diseased livers. *J Pharmacol Exp Ther* 2010, 334:124–136. doi:10.1124/jpet.110.168526.
118. Engelberg JA, Datta A, Mostov KE, Hunt CA. MDCK cystogenesis driven by cell stabilization within computational analogues. *PLoS Comput Biol* 2011, 7:e1002030. doi:10.1371/journal.pcbi.1002030.
119. Bigelow JH, Davis PK. (2002). Implications of multi-resolution modeling (MRM) and exploratory analysis for validation. Proceedings of Foundations, 2002 V&V Workshop. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.118.9430&rep=rep1&type=pdf>. (Accessed February 17, 2013).
120. Danos V, Feret J, Fontana W, Harmer R, Krivine J. Rule-based modelling of cellular signalling. CONCUR 2007—concurrency theory. *Lect Notes Comput Sci* 2007, 4703:17–41. doi:10.1007/978-3-540-74407-8\_3.

121. Yan L, Sheikh-Bahaei S, Park S, Ropella GEP, Hunt CA. Predictions of hepatic disposition properties using a mechanistically realistic, physiologically based model. *Drug Metab Dispos* 2008, 36:759–768. doi:10.1124/dmd.107.019067.
122. Sheikh-Bahaei S, Maher JJ, Hunt CA. Computational experiments reveal plausible mechanisms for changing patterns of hepatic zonation of xenobiotic clearance and hepatotoxicity. *J Theor Biol* 2010, 265:718–733.
123. Popper K. *Conjectures and refutations: The Growth of Scientific Knowledge*. London, UK: Routledge; 2002.
124. Popper K. *The Logic of Scientific Discovery*. The Logic of Scientific Discovery. London, UK: Routledge; 2002 ISBN 0-203-99462-0.
125. Hylands C, Lee EA, Liu J, Liu X, Neuendorffer S, Xiong Y, Zhao Y, Zheng H. Overview of the Ptolemy project, Technical Report UCB/ERL M01/11, University of California, Berkeley, 2001. Available at: <http://ptolemy.berkeley.edu/publications/papers/03/overview/overview03.pdf>. (Accessed January 5, 2013).
126. Hewitt C. Formalizing common sense for scalable inconsistency-robust information integration using Direct Logic (TM) reasoning and the Actor Model. arXiv:0812.4852 v82., 2008.
127. Yilmaz Y, Hunt CA. Advanced concepts and generative simulation formalisms for creative discovery systems engineering. In: A. Tolk, and L.C. Jain, eds. *Intelligent-Based Systems Engineering ISRL 10*. Springer Berlin, Germany; 2011, 233–258. doi:10.1007/978-3-642-17931-0\_9 (Chapter 9).
128. Kim SH, Park S, Ropella GEP, Hunt CA. Agent-directed tracing of multi-scale drug disposition events within normal and diseased *In Silico* Livers. *Int J Agent Technol Syst* 2010, 2:1–17. doi:10.4018/jats.2010070101.
129. Ropella GEP, Hunt CA. Cloud computing and validation of expandable *in silico* livers. *BMC Syst Biol* 2010, 4:168. doi:10.1186/1752-0509-4-168.
130. Ropella GEP, Kennedy RC, Hunt CA. Falsifying an enzyme induction mechanism within a validated, multiscale liver model. *Int J Agent Technol Syst* 2012, 4, 1–14. doi:10.4018/jats.2012070101.
131. Goderis A, Brooks C, Altintas I, Lee EA, Goble C. Heterogeneous composition of models of computation. *Future Gener Comput Syst* 2009, 25:552–560, 2009. doi:10.1016/j.future.2008.06.014.
132. Eidson JE, Lee EA, Matic S, Seshia SA, Zou J. Distributed real-time software for cyber-physical systems. *Proc IEEE* 2011, 100: 45–59. <http://dx.doi.org/10.1109/JPROC.2011.2161237>. (Accessed January 5, 2013).
133. Hewitt, C. Actor model of computation: scalable robust information systems. arXiv:1008.1459 v28, 2010.

## WEB RESOURCES

Listed below are available, open-source tools of the type that can be used now to begin satisfying the listed interactive Knowledge Repository requirements.

### Databases

PostgreSQL (relational): <http://www.postgresql.org/>  
 ZOBD: <http://zodb.org/>  
 NoSQL: <http://nosql-database.org/>  
 RDF/SPARQL: <http://www.w3.org/TR/rdf-sparql-query/>  
**Data [de]composition**  
 R: <http://www.r-project.org/>  
 HDF: <http://www.hdfgroup.org/HDF5/>

### Semantics

Protege: <http://protege.stanford.edu/>  
 Ontopia: <http://www.ontopia.net/>  
 Jena: <https://jena.apache.org/>  
 Pellet: [http://clarkparsia.com/pellet/~\(complete\)](http://clarkparsia.com/pellet/~(complete))  
 Racer: <http://www.sts.tu-harburg.de/~r.f.moeller/racer/> (complete)  
 Hermit: <http://hermit-reasoner.com/>

### Executable models

**Code**  
 Eclipse: <http://eclipse.org/>  
 Netbeans: <http://netbeans.org/>  
**Graphical programming**  
 Ptolemy II: <http://ptolemy.eecs.berkeley.edu/ptolemyII/index.htm>

NetLogo: <http://ccl.northwestern.edu/netlogo/>

ArgoUML: <http://argouml.tigris.org/>

#### *Modeling infrastructure*

JAMES II: <http://wwwmosi.informatik.uni-rostock.de/mosi/projects/cosa/james-ii>

CC3D: <http://www.compuCell3d.org/>

MASON: <http://cs.gmu.edu/~eclab/projects/mason/>

#### *Actor model<sup>133</sup>*

Erlang: <http://www.erlang.org/>

Scala: <http://www.scala-lang.org/node/242>

#### **Workflow**

Kepler: <http://kepler-project.org/>

Taverna: <http://www.taverna.org.uk/>

Galaxy: <http://galaxyproject.org/>

Triana: <http://www.trianacode.org/>

#### **Graphs**

NetworkX: <http://networkx.lanl.gov/>

JUNG: <http://jung.sourceforge.net/>



## Agent-Based Modeling: A Systematic Assessment of Use Cases and Requirements for Enhancing Productivity in Pharmaceutical Research and Development

Hunt CA, Kennedy RC, Kim SHK, Ropella GEP

Content	pages
I. Use Case	1-5
Use Case References	6-7
II. Supplemental Figure S1	8-9
III. Supplemental Text	10-13
Supplemental Text References	14

### Use Cases

Listed below are specific, current, desired, and envisioned modeling and simulation use cases spanning the pharmaceutical R&D enterprise. The articles referenced are representative of many recent articles that discuss M&S in the context of strategies for reversing the productivity decline.

#### Knowledge Integration

- Integrate knowledge from all aspects of drug development (*in vitro*, *in vivo*, preclinical, clinical and post-marketing), thereby serving as a key decision-making tool, enabling rational, scientifically based choices at critical decision points. [8,9,10,11,12,13,16,18,19,21,22,23,24,28]
- Quantitatively address questions concerning the functional relationship between prognostic factors, dosage, and outcomes. [14,21,24]
  - Integrate/capture causal relationships between disease state, prognostic factors, drug characteristics, individual variability, and drug- and disease-induced changes in biomarkers and clinical end points obtained from all phases of preclinical and clinical development. [1,9,21,12]
  - Integrate sources of variability in exposure-response relationships, which in turn should enable the prospective prediction of the probability of clinical outcomes for specific genotypes and phenotypes and thus eliminate the practice of unguided clinical decisions for pharmacotherapy. [9,20]
- Integrate the basic components to describe and understand the complex interplay between the pharmacology of drug action and the (patho-)physiological systems. [13,21]
  - Allow the integration of the physiological structure, biochemical mechanisms and drug-related properties (e.g., membrane permeability, plasma and tissue protein binding, blood-to-plasma ratio, etc.). [3]
  - Integrate information derived from *in vitro* systems regarding interaction of a drug with enzymes and/or transporters. [16]
- Effectively leverage existing knowledge and guide future research and experiments in order to optimize drug safety and clinical outcomes. [1,2,4,10,11,12,19]
  - Summarize the knowledge in a ready to apply format. [5]
  - Quantitatively summarize all relevant data at the end of a compound's full development. [14]
  - Enable artificial intelligence applications to make inferences that would be difficult or impossible to achieve for human scientists. [30]
- Provide computer-based conceptualizations of the mechanism of disease and of the links between drug treatment and observed effects. [6]

- Provide a comprehensive picture of drug action—both therapeutic and adverse—on cell-, tissue-, and organ-level physiological functions. [9,22]
- Integrate knowledge about growth, development, and maturation of various organs and tissues involved in drug metabolism and elimination across pediatric age groups to predict drug pharmacokinetics in children. [12,11,18]
- Act as a knowledge repository, continuing to be refined and incorporating new data as the candidate molecule progresses into clinical development. [1,2]
  - Provide a framework for integration of data related to drug interaction potential, and their capability to consider the time-varying concentrations of interacting drugs and account for physiological and pharmacokinetic complexities. [2]
  - Incorporate experimental animal data, as well as in-silico-derived and in vitro data, into a coherent framework, from which meaningful and reliable assessments can be made. [8]
  - Provide resources for open-software and open-data, collection and sharing, to facilitate communication and allow for growth of data-driven models [3,24,27,29]
- Integrate knowledge across multiple scales to facilitate the development of drug candidates [9,15,22,23,28]

### **Management of Uncertainty**

- Explore both variability and uncertainty in experiments and clinical trial designs. [1]
  - Account for uncertainty in the underlying assumptions and thus the resulting prediction of drug effects. [4,19]
  - Formalize assumptions, and close gaps in data and knowledge. [24]
- Project the uncertainty and the probability of success in development programs. [21]
  - Illustrate and define the uncertainty in demonstrating efficacy and safety with different development programs as assessed by, for example, the probability of achieving the target clinical efficacy outcomes with various study design. [19]
  - Perform a team-based knowledge-gap analysis, which can provide a forecast of the impact of knowledge gaps on the outcome of future clinical trials, thereby providing management with a method to prioritize the laboratory investigations at the beginning of the development program. [6]

### **Decision Support**

- Fulfill traditional roles to support labeling and confirm decision. [20]
  - Document decision making from discovery through to regulatory filing and improve answering postfiling questions and approval, as well as life cycle management of the asset. [24]
  - Support regulatory decision-making on chemical safety and risk of toxicity. [27]
- Provide a decision-making tool for selecting effective and safe doses, optimizing study sample sizes, evaluating alternative trial designs and making rational go/no go decisions based on the probability of achieving predefined study goals. [3,6,7,19,24,25]
  - Learn and confirm the key characteristics embedded within a drug candidate's molecular structure, with the goal of providing explicit, reproducible, and predictive evidence for optimizing drug development plan and resource allocation, enabling critical decision making, and eventually bringing safe and effective medicines to patients. [20,21,30]
- Support decision making on such issues as candidate selection, first-in-human dose finding, assessment of drug-drug interaction potential, and definition of appropriate study designs involving drug-drug interactions or inclusion/exclusion criteria for conducting studies of drugs metabolized by polymorphic enzymes. [7,16]
- Conduct decision analyses of the comparative benefits and risks associated with preventive and treatment strategies. [7,15,24]

## **Preclinical and Clinical Development**

- Guide creation of an efficient clinical development strategy. [19]
  - More effectively define potential drug candidates and optimize clinical development by identifying appropriate patients, dose, and optimal clinical trial designs. [1,29]
  - Facilitate design and/or selection of lead compounds, selection of the first-in-human dose, early clinical trial design, and proof-of-concept studies of experimental drugs and drug combinations. [2,3,10,11,13,29]
  - Use to select the studies that would be most relevant to conduct or to replace the ‘guesswork’ that is common when prospective clinical studies are not feasible. [3,16]
  - Help select compounds for clinical trials by integrating all the information for particular compounds into a mechanistic framework where selection can be made based on the expected human PK profile. [11]
- Evaluate/explore the performance of different designs based on the currently available information about the drug before a specific design is selected for the next study. [3,6,11,16,19]
  - Simulate outcomes of alternative study designs before the experimental investigation commences—incorporating different doses and/or different patients and computing the probability of a successful trial given the characterized patient population and proposed treatment regimens. [3,6,14]
  - Predict undesirable effects for safety assessment and support of lead selection and optimization. [27]
  - Help prioritize compounds prior to in vivo experimentation to reduce the number of in vivo animal studies performed in discovery, and ensure appropriate dose selection is made. [11]
  - Predict first-in-human dosage requirements in conjunction with traditional studies in laboratory animals to establish concentration-effect relationships regarding safety. [16]
  - Predict the toxic potential of chemicals and human adverse effect, and generate hypotheses about the putative molecular mechanisms of chemical-induced injury. [22,27]
  - Project drug PK profiles under various scenarios, and help determine whether there is a need for additional studies as part of risk-benefit assessment of new drugs. [11]
  - Evaluate the full time-course of systemic drug and metabolite exposure under all the conditions and situations likely to be encountered during clinical development and beyond. [17]
  - Relate drug exposure, in the compartment measured, to the pharmacological effect at the site of action. [24,26]
  - Use to understand the relationships between absorption and its associated parameters, and the interactions of drug substance properties and bioavailability, to guide the formulation development process. [10]
  - Perform clinical trial simulations to improve the design of future trials. [1,5,11,14,19]
  - Predict long-term safety early in drug development. [5]
- Translate in vitro findings into in vivo efficacy and allow streamlining of dose finding for Phase I and II studies as well as the assessment of new dosing regimens for their likely clinical efficacy and safety. [19]
- Characterize the relationships between the biomarkers and the preclinical/clinical outcomes, placebo effects, a drug’s pharmacologic effects both on and off-target-wise, and trial execution characteristics for both the desired and undesired response. [5,23,25]
  - Guide the identification and development of biomarkers that can serve to predict efficacy and safety during the late stages of preclinical development or the early stages of clinical development. [24]
- Predict the treatment effect as a function of dose, regimen and study design. [19]
  - Evaluate the impact of protocol deviations from a specific design, including the dropout and compliance behavior. [19]
  - Provide guidance for designing better randomized efficacy trials, by determining the appropriate experimental design (dose selection, treatment duration, population selection, etc.) that should

maximize separation between side effects and efficacy while still maintaining a clinically meaningful benefit. [14]

- Assess population variability of absorption after administration of drugs via other routes of extravascular administration (e.g., dermal). [16]
- Predict human testing results by integration of information of physiology and physicochemical properties of an investigational drug. [18]
  - Predict drug clearance, tissue distribution and rate and extend of absorption, and plasma and tissue concentration time profiles. [11,18]
  - Predict exposure/response to a drug and/or toxics not just in the plasma or blood, but also in remote and/or inaccessible compartments such as the brain and tumor tissues. [12,24]
- Provide the starting point for ‘stratified’ and/or personalized medicine. [11,16]
  - Enable binning of individual patient’s drug response into tranches to develop a limited set of appropriate treatment regimens (precision/individualized medicine). [9]

### **Drug-Disease Modeling**

- Use to understand the relationship between drug effect and the natural progression of the underlying disease. [4,13,19,21]
  - Explore clinical trial designs and methods of data analysis to differentiate symptomatic versus disease modifying effects. [19]
  - Perform longitudinal analysis of disease progression, with all its components, such as placebo effect, lack of adherence and dropout. [14]
  - Simulate the impact of specific diseases (in individuals, clinical trials, and entire populations), and predict pharmacokinetics in specific disease states defined by etiology and severity. [11,12,17,18]
- Describe macrophysiological processes within a particular disease state and use to understand likely modulation of those processes with specific interventions. [1]
- Incorporate dual or multiple drug effects on disease complementary domains. [14]

### **Drug-Drug Interaction, Special Populations**

- Guide investigations in pediatric, the elderly and other special populations. [4,12,16,17,18,19,23]
  - Anticipate pharmacokinetic differences in pediatric patients relative to adult patients and assist in the selection and optimal design of in vivo investigations. [17]
  - Explore “what if” scenarios to determine the most likely cause of altered pharmacokinetics in children. [12]
  - Help design and optimize conductance of clinical trials in special populations such as pediatrics, where optimal planning is needed to minimize the ethical and technical difficulties; for example, in pediatric clinical trials, by suggesting first dose/dose range or optimal sampling times, which is of great importance, as taking blood samples is more difficult than in adults and a main challenge in conducting pediatric clinical trials. [3,12]
  - Use for intra-human scaling from healthy adults to special subpopulations (diseased, the elderly, pediatrics, genetically unique subpopulations) to define differential organ dosimetry as it relates to efficacy and/or safety. [4,11]
- Dynamically evaluate the impact of a DDI with respect to the full concentration-time profiles of the interacting compounds. [17]
  - Assess more complex scenarios involving simultaneous dose-dependent inhibition and induction as well as competition for plasma binding, the inhibitory effects of both parent drug and metabolites. [17]
  - Analyze various elements of complex DDIs involving metabolites and nonlinearity with time. [16]
  - Forecast age-dependency and disease-dependency of pharmacokinetics, explore the variability expected in different patient populations, and provide guidance in the design of DDI studies. [2]

- Capture the complex interplay between the parent compound and its metabolites involved in time-dependent inhibition (TDI). [2]
- Identify and assess complex DDIs early in drug development so that clinical studies could be planned or prioritized to assess the risk. [7]
  - Detect serendipitous connections between approved drugs and harmful or beneficial interactions when taken by the same patient. [30]

## Prediction

- Allow for exploration of the PK properties across species and compounds [13,14,17,23]
  - Extrapolate knowledge from one drug class to another and more quickly address new therapeutic targets. [13]
  - Provide interpretation of inter-animal and inter-species PK differences (for instance, owing to different absorption or disposition), and extrapolation to humans. [14]
  - Identify, characterize, and explain the relationships between genetic and epigenetic changes and alterations in targets of interest within species-specific mechanistic contexts. [9]
  - Extend the above characterizations to patients to enable individualized &/or cohort tailored medicine. [9]
- Map formulation components and in vitro release profiles to in vivo PK. [10]
  - Use animal and potentially human ADME data to define the differential dosimetry that may ultimately aid in definition of tolerable daily intake allowances or risk profiles when linked to human exposure. [4]
  - Guide dosing and optimal sampling in first-in-human trials and in human health risk assessment to predict blood and organ concentrations that may be linked to safety end points. [4]
- Quantitatively evaluate temporal influences of covariate combinations (sources of the variability) (including genetics, BSA, age, ethnicity, gender, pregnancy, obesity, environmental factors, smoking, etc.) on mechanisms of interest, including ADME. [16,21,23]
- Predict outcomes of treatment protocols, not just in an “average” individual but in any cohort of interest that. [2,3,16]
  - Describe and/or predict drug PK and target exposure in cohorts characterized by different physiological, pathological, and/or morbidity attributes. [7,11,12,18]
- Ultimately predict human PK from in silico, in vitro and physicochemical data. [11]
  - Predict the drug PK, efficacy and side effects in a given individual with a known genotype make-up for relevant ADME and pharmacology proteins. [11]

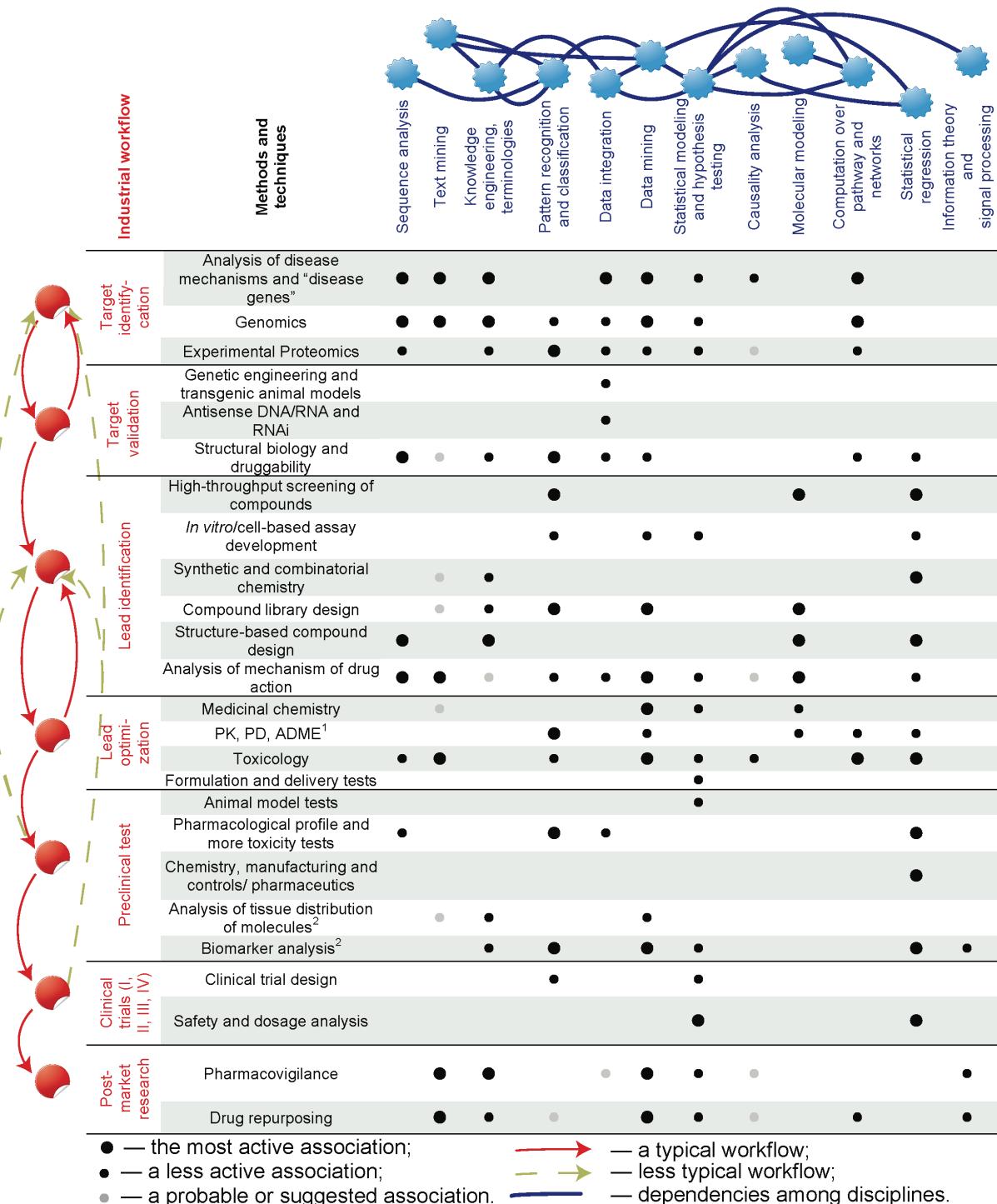
## Use Cases References

Listed in the same order as in the manuscript

1. Allerheiligen SR. Next-generation model-based drug discovery and development: quantitative and systems pharmacology. *Clin Pharmacol Ther* 88(1):135-137 (2010).
2. Baneyx G, Fukushima Y, Parrott N. Use of physiologically based pharmacokinetic modeling for assessment of drug-drug interactions. *Future Med Chem* 4(5):681-693 (2012).
3. Bouzom F, Ball K, Perdaems N, Walther B. Physiologically based pharmacokinetic (PBPK) modelling tools: how to fit with our needs? *Biopharm Drug Dispos* 33(2):55-71 (2012).
4. Edginton AN, Joshi G. Have physiologically-based pharmacokinetic models delivered? *Expert Opin Drug Metab Toxicol* 7(8):929-34 (2011).
5. Gobburu JV, Lesko LJ. Quantitative disease, drug, and trial models. *Annu Rev Pharmacol Toxicol* 49:291-301 (2009).
6. Grasela TH, Slusser R. Improving productivity with model-based drug development: an enterprise perspective. *Clin Pharmacol Ther* 88(2):263-268 (2010).
7. Grillo JA, Zhao P, Bullock J, Booth BP, Lu M, Robie-Suh K, Berglund EG, Pang KS, Rahman A, Zhang L, Lesko LJ, Huang SM. Utility of a physiologically-based pharmacokinetic (PBPK) modeling approach to quantitatively predict a complex drug-drug-disease interaction scenario for rivaroxaban during the drug review process: implications for clinical practice. *Biopharm Drug Dispos* 33(2):99-110 (2012).
8. Huisenga W, Telgmann R, Wulkow M. The virtual laboratory approach to pharmacokinetics: design principles and concepts. *Drug Discov Today* 11(17-18):800-805 (2006).
9. Iyengar R, Zhao S, Chung SW, Mager DE, Gallo JM. Merging systems biology with pharmacodynamics. *Sci Transl Med* 4(126):126ps7 (2012).
10. Jiang W, Kim S, Zhang X, Lionberger RA, Davit BM, Conner DP, Yu LX. The role of predictive biopharmaceutical modeling and simulation in drug development and regulatory evaluation. *Int J Pharm* 418(2):151-160 (2011).
11. Jones HM, Dickins M, Youdim K, Gosset JR, Atkins NJ, Hay TL, Gurrell IK, Logan YR, Bungay PJ, Jones BC, Gardner IB. Application of PBPK modelling in drug discovery and development at Pfizer. *Xenobiotica* 42(1):94-106 (2012).
12. Khalil F, Läer S. Physiologically based pharmacokinetic modeling: methodology, applications, and limitations with a focus on its role in pediatric drug development. *J Biomed Biotechnol* 2011:907461 (2011).
13. Mager DE, Jusko WJ. Development of translational pharmacokinetic-pharmacodynamic models. *Clin Pharmacol Ther* 83(6):909-912 (2008).
14. Nucci G, Gomeni R, Poggesi I. Model-based approaches to increase efficiency of drug development in schizophrenia: a can't miss opportunity. *Expert Opin Drug Discov* 4(8): 837-856 (2009).
15. Prokop A, Michelson S. Development: multiscale CSB—simulation tools. *Systems Biology in Biotech & Pharma, SpringerBriefs in Pharmaceutical Science & Drug Development Volume 2:* pp 87-102 (2012).
16. Rostami-Hodjegan A. Physiologically based pharmacokinetics joined with in vitro-in vivo extrapolation of ADME: a marriage under the arch of systems pharmacology. *Clin Pharmacol Ther* 92(1):50-61 (2012).
17. Rowland M, Peck C, Tucker G. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu Rev Pharmacol Toxicol* 51:45-73 (2011).
18. Sinha V, Kimko HHC. Recent Developments in Physiologically Based Pharmacokinetic Modeling. *Clinical Trial Simulations, AAPS Advances in the Pharmaceutical Sciences Series Volume 1, Part 5:* pp 483-499 (2011).
19. Suryawanshi S, Zhang L, Pfister M, Meibohm B. The current role of model-based drug development. *Expert Opin Drug Discov* 5(4):311-321 (2010).
20. Zhang L, Pfister M, Meibohm B. Concepts and challenges in quantitative pharmacology and model-based drug development. *AAPS J* 10(4):552-559 (2008).

21. Zhang L, Sinha V, Forgue ST, Callies S, Ni L, Peck R, Allerheiligen SR. Model-based drug development: the road to quantitative pharmacology. *J Pharmacokinet Pharmacodyn* 33(3):369-393 (2006). added from QuantPharmSet4
22. Jack J, Wambaugh J, Shah I. (2013). Systems toxicology from genes to organs. *Methods in molecular biology* (Clifton, NJ) 930:375-397.
23. Kuepfer L, Lippert J, Eissing T. Multiscale mechanistic modeling in pharmaceutical research and development. *Advances in Systems Biology*, 543-561 (2012).
24. Wetherington J D, Pfister M, Banfield C, Stone JA, Krishna R, Allerheiligen S, Grasela D. M. Model-based drug development strengths, weaknesses, opportunities, and threats for broad application of pharmacometrics in drug development. *J Clin Pharmacol* 50(9 suppl):S31-S46S (2010).
25. van der Graaf PH, Benson N. Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharmaceutical research*, 28(7):1460-1464 (2011).
26. Morgan P, van der Graaf PH, Arrowsmith J, Feltner DE, Drummond KS, Wegner CD, Street SD. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov Today*. 17(9-10):419-24, 2012.
27. Valerio LG. In silico toxicology for the pharmaceutical sciences. *Toxicol Appl Pharm* 241(3):356-370, 2009.
28. Vicini P. Multiscale modeling in drug discovery and development: future opportunities and present challenges. *Clin Pharm Therap* 88(1):126-129, 2010.
29. van der Graaf PH. CPT: Pharmacometrics and Systems Pharmacology. CPT: Pharmacomet Syst Pharmacol 1(9):e8, 2012.
30. Yao L, Evans JA, Rzhetsky A. Novel opportunities for computational biology and sociology in drug discovery. *Trends Biotechnol* 28(4):161-70 (2010).

**Supplemental Figure S1**



The figure above is reproduced with permission from [Yao L, Evans JA, Rzhetsky A.

Novel opportunities for computational biology and sociology in drug discovery. *Trends in Biotechnology* 2010, **28**(4):161-70. doi: 10.1016/j.tibtech.2009.06.003].

The following legend is adapted from the original.

This matrix diagram characterizes various roles of computational technologies within the drug discovery process, and indicates how they impact drug discovery. Traditional linear processes are shifting to become more parallel, simultaneous and cyclical. Red arrows indicate the traditional process and yellow dashed arrows suggest novel workflows that are being used increasingly to increase productivity.

Biomarkers and analysis of the tissue distribution of target molecules are the most recently introduced checkpoints; the FDA does not require them. Computational biology methods are listed along the top. Blue lines illustrate how methods are related.

The impact of each computational technique on each drug discovery stage is classified into three categories: actively or heavily used (large black dot), less actively used (small black dot) and our suggestion (small gray dot).

Chemical informatics is not emphasized because it relates to issues from chemistry and not biology. Chemical informatics comprises a wide range of approaches from computational and combinatorial chemistry that model lead properties and their interaction with targets. Examples include chemical structure and property prediction; structure–activity relationships; molecular similarity and diversity analysis; compound classification and selection; chemical data collection, analysis and management; virtual drug screening; and prediction of *in vivo* compound characteristics.

## Supplemental Text

The following text is adapted from Hunt CA, Ropella GEP, Lam TN, Gewitz AD. Relational grounding facilitates development of scientifically useful multiscale models. *Theoretical biology & medical modelling* 2011B, 8:35.

<http://www.tbiomed.com/content/8/1/35/>

### Essential roles of relational grounding in satisfying requirements

The units, dimensions, and/or objects to which a variable or model constituent refers establish groundings. Absolute grounding: variables, parameters, and I/O are in real-world units. Each term is foundational and maps to a tacit entity having a conceptual or biological meaning. Relational grounding: variables, parameters, and I/O are in units defined by other system components. Variables and components may map to counterparts outside the system or have meanings that are unrelated to the real world. Biology uses relational grounding. Grounding issues do not pose problems for conventional models because they are most often focused narrowly on a single aspect of a system (e.g., a PBPK or gene network model). However, when the model is an analog that aims to describe multiple aspects (i.e., different phenotypic attributes), grounding problems emerge. To satisfy the above requirements, a spectrum of model classes, methods, and groundings will be needed: absolute grounding occupies one extreme; relational grounding occupies the other. The conventional model classes cited in the reviews, mostly pharmacometric and inductive, data-focused ODEs, which use absolute grounding, will remain important and will see critical use cases, but how those methods are used within the framework will differ from those described in the reviews.

Inductive, data-focused, mechanistic, ODE models are accretions of at least three different model types: 1: conceptual models of biological mechanisms, features, and aspects; 2: equations that describe temporal phenomena of type 1; 3: measurement units that provide a quantitative 1:1 mapping from type 2 to measures of biological phenomena. The conceptual models are grounded to the biology via the literature, proprietary information, and expert opinion. Both types 1 and 2 may contain bias and/or questionable assumptions. Good science requires all three be made specific and concrete. Such model accretion reduces flexibility and makes reuse problematic. Progress in satisfying the Box 2 use cases requires a different three-step approach. 1) Map specific biological features to actual, concrete, relationally grounded software objects and spaces that have execution protocols. They will become analogs. 2) Measure actual phenomena generated during executions. 3) Specify quantitative mapping models to relate measures of in silico phenomena to real (not derived: see below) measures of referent phenomena. Such model separation provides the required flexibility and encourages reuse.

Requiring relational grounding is essential because so doing enables synthesizing flexible, easily adapted, extensible, hierarchical analogs of the systems they mimic. Grounding to metric spaces and real world units—absolute grounding—provides simple, interpretive mappings between output, parameter values, and referent data. However, complex issues must be addressed each time one of the following occurs: expand the model to include additional phenomena; combining models to form a larger system; and/or model context changes. Expansions are challenging, even infeasible, when center-left in Fig. 1. A PBPK model, for example will have limited reusability when experimental conditions change or when an assumption made in the original model formulation has been falsified. Reusability is hindered in part because the conflated semi-mechanistic ODE model and the model-to-referent mapping model have

different uses. The components and processes in discrete event, object and agent oriented, biomimetic analogs need not have assigned units [1-6].

Grounding decisions impact managing mechanistic and related uncertainties within analogs. In ODE models grounded absolutely, variables and parameters are often expressed as precise mathematical values, even though their networked uncertainties are usually significant. Examples include the physiological parameters in mechanistic PBPK models. Representing uncertainty within such models is mathematically complex and can be problematic for large models. Integrating models from different contexts can require that the whole model be refitted, which should involve re-examination of the cumulative consequences of the networked assumptions. In contrast, probabilistic functions within analogs can represent inherent uncertainties conveniently.

An expectation running through the use cases listed in Supplement—Use Cases is that models will evolve (become more complicated) as R&D advances and new mechanistic insight accumulates, which will require changing, adding, and removing component linkages within analogs. We have recommended [3] striving to keep component groundings relational. So doing facilitates component replacement, limiting any one component formulation solely to its coupling with the others. Any component can be replaced at will as long as the minimal I/O matching requirements are met. Starting with relationally grounded analog components allows the modeler to iterate progressively from qualitative to quantitative validation.

Validation will be based on similarity of phenomena generated during analog simulations to corresponding referent system phenomena. The similarity spectrum can range from qualitative to quantitative. With qualitative similarity, objects will either possess some quality, or they will not. Simulation and referent attributes are considered similar if they have (almost) the same qualities. With quantitative similarity, attributes are categorically the same but vary by magnitude and can be compared by some ordering relation (e.g., less than or greater than). It is always the case that a valid qualitative description is a prerequisite for (and provides constraints for) quantitative descriptions, in the sense that any quantities defined must relate to one or more qualities. Validation (or lack thereof) should act as an important determinant of what type of analog or model (and therefore, what type of grounding) to employ at different stages during project R&D.

Relational grounding facilitates scaling between wet-lab and animal models used during R&D (and between different cohorts of individuals). Consider scaling metabolic clearance of a particular compound in mice (ml/ min/g) to enable human prediction when clearance is grounded absolutely on concentration and time. Scaling to human clearance values requires applying mass, volume, and time scaling factors to all parameters simultaneously, knowing that each scaling factor is imprecise and uncertain. When scaled predictions deviate significantly from observed values (common), there is no way to ascertain which scaling factor(s) and/or which scaled parameter(s) is problematic. Within a relationally grounded analog, mass and volume scaling can be done separately and validated independently. Setting the scale for one variable in accordance with trusted validation data can help set the scale for other, related variables. Time scaling is more complicated [3]. It may require a separate scaling for each probability parameter. However, satisfying the requirements opens the door to using available methods and tools (Box 3) to automate analog-to-analog scaling for analogs grounded relationally.

Relational grounding will also facilitate development of theories of translation. One relationally grounded analog exhibiting the above requirements can be morphed into another. That morphing models

conceptual translation and thus stands as a tentative theory of cause-effect translation. It can be directly challenged and falsified (or not). Importantly, gaining exploitable insight into normal-to-disease transitions can be best achieved with analog components that use relational grounding. Park et al. [4] provide an example. Such transitions will require change in how components at multiple levels are parameterized and/or interact. Alternative mechanistic scenarios will need to be explored and challenged. Having components grounded absolutely makes such exploration problematic. Reliance on relational grounding simplifies mechanism exploration and makes the process more intuitive.

Satisfying the requirements requires that all analogs be articulated: consist of distinct parts, modules, and/or components, some of which are autonomous. However, articulation issues are orthogonal to grounding issues. A component that initiates and maintains its own run-time is autonomous. Mammalian cells can be autonomous *in vitro*. When analog components are quasi-autonomous, they can be effectively replaced by other components for which component I/O requirements are specified in ways consistent with biology. The extent to which a component is autonomous is handled by the clear specification and maintenance of component use cases (aspects, phenotypic attributes, experiment protocols, etc.) or, collectively, by the component's phenotype. A model will lack any autonomy as long as there is only one use case for given component and a single use case for all connected components. Autonomy can be established regardless of how a module is grounded, but only when targeted phenotypic attributes are clearly defined, which can be challenging when left-of-center in Fig. 1.

Relational grounding facilitates referent knowledge embodiment within analog mechanisms. Computational biology markup languages standardize relationships between the terms. To fulfill the requirements, we must have explicit ontologies enabling specialization into technical programming (Box 3) and biomedical domain expertise. An expert must be able to examine an analog, its components, and simulation events without needing computational expertise. Achieving the knowledge repository will have parallels to the evolution of modern biomedical research. Biomedical scientists design and perform complex experiments without becoming experts in laboratory equipment, reagent design, and production. Progression from custom-built experimental to standardized lab products allowed scientists to effectively compress complicated methods. They were subsumed by engineering, production, and validation processes. That subsumption reduced experimental variability, facilitated experiment replication, and lowered costs, which freed scientists to build increasingly more sophisticated experiments atop complicated equipment (e.g., a cell sorter, a confocal microscope, monoclonal antibodies, transfection reagents, automated DNA sequencer). An analogous progression, progressed in time, must occur for the envisioned productivity gains to be achieved.

### **Neither knowledge nor semantics is embedded (encoded) in equation models (e.g., PBPK)**

**Ready to use software tools.** Bayer's PK-Sim [7] and its relationship with MoBi [8] provide a good example for understanding where knowledge is encoded in “ready to use” software tools and packages. MoBi is a generic PBPK simulation tool that can express a large variety of specifically designed systems of ODEs based on low-level structures (e.g., molecules and reactions). PK-Sim restricts the set of possible models by providing higher-level structures (e.g., individuals, populations, compounds, etc.). PK-Sim and MoBi are designed so that the more restrictive PK-Sim models can be imported into MoBi and further refined to provide high-level models with use case specific, low-level, modifications.

In order to draw a complete picture of where the knowledge is encoded in such models, we can contrast a PK-Sim model, imported and modified by MoBi against a system of ODEs implemented directly in, say, Matlab. Matlab is an even more generic modeling tool than MoBi. It can not only express PBPK models, but virtually any type of computation. Given these three degrees of expressibility (Matlab, MoBi, and PK-Sim), we see that the higher level constructs act as constraints. The more a set of possible models is constrained to a particular organization of the system of equations and boundary values, the more knowledge is encoded in the system.

Hence, it is the constraining of the set of possible models that encodes the knowledge, not the particular model that obtains at the end of the development process. In other words, it is the PK-Sim and MoBi tools that encode the knowledge, not the equations they numerically integrate during simulation.

**Where is the knowledge?** The biomedical domain knowledge that one associates with a PBPK model is found in the interpretations of equations, provided by domain experts, found in the domain literature. Those interpretations provide a semantic grounding of otherwise generic equation models. The semantic grounding of such equations is always separate from their software implementations. However, given “ready to use” software tools of the type described above, it is easy for a “user” of those tools to conflate the knowledge with the equations. It is a good practice to keep them separate. The framework described in the main text will facilitate doing so.

To illustrate the above point, consider Garg and Balthasar’s PBPK model of immunoglobulin-gamma (IgG) kinetics [9]. The authors aimed to predict the distribution of an antibody. The model was discussed in example one in [10]. The authors clearly describe their model in prose supported by sketches. Straightforward ODEs with parameterization details are provided. Variables within their model take meaning regardless of the background of the reader. For example, Equation 2 from [9], shown below, describes the rate of change of IgG plasma concentration. A software engineer (SE) with no biological knowledge would be able to draw on her knowledge of mathematics to build an executable implementation of the model. It is probable that she would first rewrite the equation shown below in pseudocode to facilitate translation to a programming language. The variables in the equation have no immediate meaning to the SE. Her code can be implemented absent any biomedical knowledge. A domain expert typically supplies parameterizations and draws on domain knowledge to interpret the input-output (I/O) using prose and sketches thus providing the semantic grounding. Within “ready to use” tools, separate software features can mimic the domain expert’s activity and provide standardized interpretations. It thus becomes clear that neither knowledge nor semantics is embedded in those equations.

```
Let Lu = Lung, P = Plasma, Lv = Liver, S = Spleen, H = Heart, K = Kidney, Sk = Skin, M = Muscle,
O = Others
```

$$V_p * C_{Pdt} = (Q_{Lu} - L_{Lu}) * CV_{Lu} - ((L_{Gi} + L_S + Q_{LV} + Q_H + Q_K + Q_{Sk} + Q_M + Q_O) * C_P) + ((1 - \sigma_{Lu}) * L_{Lu} * CI_{Lu}) + ((1 - \sigma_{Gi}) * L_{Gi} * CI_{Gi}) + ((1 - \sigma_{S}) * L_S * CI_S) + ((1 - \sigma_{Lv}) * L_{Lv} * CI_{Lv}) + ((1 - \sigma_{H}) * L_H * CI_H) + ((1 - \sigma_{K}) * L_K * CI_K) + ((1 - \sigma_{Sk}) * L_{Sk} * CI_{Sk}) + ((1 - \sigma_{M}) * L_M * CI_M) + ((1 - \sigma_{O}) * L_O * CI_O)$$

## References

[1-7] and [10] are cited in the main text.

1. Engelberg JA, Datta A, Mostov KE, Hunt CA. MDCK cystogenesis driven by cell stabilization within computational analogues. *PLoS Computational Biology* 2011, 7(4):e1002030. doi: 10.1371/journal.pcbi.1002030.
2. Hunt CA, Ropella GE, Lam TN, Tang J, Kim SH, Engelberg JA, Sheikh-Bahaei S. At the biological modeling and simulation frontier. *Pharmaceutical Research* 2009, 26:2369-2400. doi: 10.1007/s11095-009-9958-3.
3. Hunt CA, Ropella GEP, Lam TN, Gewitz AD. Relational grounding facilitates development of scientifically useful multiscale models. *Theoretical Biology and Medical Modelling* 2011, 8:35. doi: 10.1186/1742-4682-8-35.
4. Park S, Kim SH, Ropella GEP, Roberts MS, Hunt CA. Tracing multiscale mechanisms of drug disposition in normal and diseased livers. *Journal of Pharmacology and Experimental Therapeutics* 2010, 334(1):124-136. doi: 10.1124/jpet.110.168526.
5. Sheikh-Bahaei S, Maher JJ, Hunt CA. Computational experiments reveal plausible mechanisms for changing patterns of hepatic zonation of xenobiotic clearance and hepatotoxicity. *Journal of theoretical biology* 2010, 265:718-733.
6. Tang J, Hunt CA. Identifying the rules of engagement enabling leukocyte rolling, activation, and adhesion. *PLoS computational biology* 2010, 6(2):e1000681. doi: 10.1371/journal.pcbi.1000681.
7. <http://www.systems-biology.com/products/pk-sim.html>
8. <http://www.systems-biology.com/products/mobi.html>
9. Garg A, Balthasar JP. Physiologically-based pharmacokinetic (PBPK) model to predict IgG tissue kinetics in wild-type and FcRn-knockout mice. *Journal of Pharmacokinetics and Pharmacodynamics* 2007, 34:687-709. doi: 10.1007/s10928-007-9065-1.
10. Hunt CA, Ropella GEP, Lam TN, Gewitz AD. Relational grounding facilitates development of scientifically useful multiscale models. *Theoretical Biology and Medical Modelling* 2011, 8:35. doi: 10.1186/1742-4682-8-35.