

Everyday Modeling in Biology

Frank D Russo, Ph.D.

Overview

Empowering scientists with the ability to represent, develop and transfer knowledge is the goal of WayCraft's pioneering software platform, WayFarer. This technology, coupled with an intuitive graphical user interface puts the power of biological modeling directly in the hands of bench scientists. Modeling of biological systems need not be an expensive and time consuming task performed by specialists, it can be accomplished by the users who will benefit from it the most: the bench scientists. Allowing researchers to express their ideas precisely will enhance communication and knowledge transfer within the organization and between groups of different knowledge domains. Enabling researchers to explore their ideas in more detail before taking them to the lab will focus their efforts and optimize the experiments they do perform. WayFarer relieves the burden of committing vast numbers of individual facts to the scientist's memory by placing them in a system in which they are easily deposited and retrieved. Once relieved of the tedious and time consuming chores of data gathering and information management, scientists will have more freedom to pursue their ideas and do what they do best: scientific research.

Modeling as a Core Service

Systems biology-based approaches such as modeling and simulation are steadily gaining acceptance within the pharmaceutical industry. But these approaches are still considered the domain of specialists, outside the skill set of most research scientists. Modeling expertise typically resides in a centralized group whose members are allocated to drug development teams on a project-by-project basis. Larger, company-wide initiatives may be undertaken by the systems biology team as a whole or out-sourced to a specialized vendor.

Such a centralized approach makes sense as a way to manage and ration a scarce and expensive form of expertise, but it is not an effective way to disseminate the technology throughout a company. The separation of functions between bench scientists and computational experts introduces substantial inefficiency as scientists with a system to be modeled must cross barriers of both communication and bureaucracy to work with those who can implement that model. Conversely, the specialists in the core group often do not have the depth of biological background to fully understand what the bench scientist needs in a model. These barriers make an expensive resource even more costly as well as less effective. With these added costs, the whole approach can become more trouble than it is worth.

Table of Contents

Overview	1
Modeling as a core service	1
Modeling in everyday workflow	2
Collect knowledge from various sources	2
Synthesize into a model of the system	3
Predict consequences of changes or perturbations	4
Experimentation, analysis and conclusions	4
Iterative refinement and collaboration	5
Possibilities of everyday modeling	6



The root of the problem is that most software is designed not for biologists, but for bioinformaticians, mathematicians, statisticians, or programmers—the burden remains on the user to translate biological questions into the concepts and language of the tool to be applied. Even in the rare cases where one user has expertise that spans the necessary domains, this conceptual mismatch demands a constant shift between the mindset of a biologist and that of a mathematician, statistician, or programmer. At the very least this means a loss of focus on the biological question. At worst, cross-functional individuals find themselves spending all of their time supporting others rather than applying these tools to their own research. The design of the software forces individuals to choose between being a biologist and being a technical specialist.

In other fields, software designed directly for practitioners has represented a turning point in integrating computer-aided methodologies into everyday practice. When professionals no longer have to choose between defining themselves as domain experts or computer specialists, the computer becomes simply a new, more powerful way to express their craft. Architects and engineers use CAD software, photographers use Adobe Photoshop, and desktop publishers use a variety of tools, all relying solely on their own computer skills. With this independence comes the freedom to apply the tool in cases and in a manner that would not even be considered if the practitioner had to justify utilizing a scarce resource to do so. When the individual has direct access to a computational tool, it is used more often and in a wider variety of ways, leading to powerful and often unexpected increases in productivity and creativity. It is with this goal in mind that we have designed WayFarer.

Modeling in Everyday Workflow

Good software is defined not just by broad measures such as ‘power’ and ‘ease of use’, but by how well it enables the tasks that the user actually needs to get done, and how it fits into that user’s overall workflow. No matter how easy the interface may be to use, if the operation being performed is alien to the user’s mindset, it will not succeed in enhancing productivity. The challenge in this case is to design software that thinks like a biologist. Although biologists may be unfamiliar with the technical aspects of modeling a system computationally, the use of models themselves is already central to their science. By focusing less on the mathematical details of a model and more on how this model fits into the overall process of investigation, we can put a powerful computational tool directly in the hands of bench biologists.

Collect knowledge from various sources

Any project begins by bringing together what is already known about a given system. The sources for such information are many, including scientific literature, reference databases, discussions with collaborators, etc. The goal is to collect individual facts, observations, and connections that can eventually be



fitted into a consensus view of the system. In current practice, this phase is typified by queries against PubMed and other reference databases, followed by downloading abstracts or full papers (where available), as well as sequences, images, presentations, or anything else that may be relevant. Some information may only be available in hard copy, which is stored in the individual scientist's file cabinet; still other information may be thought of as 'loose' facts, captured manually in a lab notebook, on a Post-it note, or simply remembered.

In WayFarer, individual facts are collected for a project in three main ways: by import, by manual creation, or by reuse. Import fits the existing process by bringing in data downloaded from external sources. Manual creation may be necessary for items that are not available electronically, for whatever reason. And any item that has previously been imported or created in the context of one project may be reused for another. The most immediate benefit is to collect explicitly in one place information that would otherwise be scattered across the researcher's computer, desk, notes, and thoughts. More importantly, this information is now available for the rest of the computer-aided workflow.

Synthesize into a model of the system

As individual facts are collected, the scientist assembles them into an overall 'picture', a model that synthesizes multiple lines of observation and reasoning into an understanding of how the target biological system works. Depending on how well the system has been characterized, this model may be highly detailed or extremely rough. In current practice, the main repository for such models is in fact the minds of the researchers themselves. These mental models may be captured at least partially through drawings in notebooks, on whiteboards, or on PowerPoint slides, but even with such figures, a certain amount of verbal explanation is required to describe the dynamics, context, or other nuances of the model as the scientist understands it. In most cases, detailed mental models must be passed directly from scientist to scientist over the course of multiple discussions and presentations.

Synthesizing facts into models is where the WayFarer approach really begins to pay off. The core of the platform is a modeling framework, the Scientific Knowledge Integration Framework (SKIF), designed explicitly to represent knowledge the way a biologist thinks about it. It is a common perception that modeling makes sense only for well-characterized systems, where highly mechanistic computational models may be developed. But biologists' mental models guide their investigations even at the roughest stage. By formalizing even high-level models, we extend the power of the computational approach much earlier into the process of investigation, without sacrificing the mindset of the biologist in the process.



Predict consequences of changes or perturbations

Ultimately, the biologist's goal is to understand how the mechanistic behavior of the system produces macroscopic, functional, or phenotypic effects. This means understanding not only the static structure of the system, but its dynamic behavior under multiple conditions. Even more informative are modifications or other perturbations to components of the system, and their impact on the behavior of the system as a whole. A large part of biological research consists of predicting the effects of such changes, then testing these predictions by either finding or engineering such variants to observe their actual impact. In current practice, such predictions generally take the form of thought experiments, using qualitative or semi-quantitative reasoning based on one's mental model of the system. Figures on paper or a whiteboard are often used as visual aids in this process, whether as part of a discussion with collaborators or for the benefit of the individual scientist.

WayFarer has a sophisticated system for generating and accounting for variant forms of entities, processes, or systems, at any level of abstraction from individual molecules to whole organisms. At a molecular level, molecules or specific activities may be added to or removed from a model system. At the organismal level, perturbation may consist of stimulating or destroying whole cells, organs, or other structures. The impact of such perturbations is observed by monitoring simulated changes in abundance, form, or other properties of some indicator, depending entirely on the specific nature of the system. The use of computational tools to aid in the prediction process lets the scientist explore a wide range of possible experimental manipulations before undertaking any of them in the lab. It also allows the investigation of variants that may be difficult or impossible to generate experimentally, as well as access to indicators in the simulation that may not be easily observable in the real system.

Experimentation, analysis and conclusions

The mechanics of lab work and data collection tend to occupy the majority of a biologist's time and attention. A wide variety of tools and methodologies may be used to collect the data, and to process it to a point where interpretation may begin. But the real payoff comes when the data are analyzed in the context of a model, since predictions based on that model were the impetus for the experiments in the first place. As with the prediction process, this analysis in context has traditionally taken place primarily in the mind of the scientist, again perhaps with the assistance of manually drawn figures as visual aids.

This is one area where some progress has been made in using software to assist biologists directly. Tools such as GeneSpring, Ingenuity Pathway Analysis, and several others can overlay microarray expression data on a drawing of a biological pathway, letting the biologist put the results in the context of a system. The WayFarer approach is to generalize this functionality for multiple



data types, such as proteomics or metabolomics data. More importantly, WayFarer's underlying representation of the biological system is formal and computable, enabling comparison between hypothetical and actual results. If the model needs to be modified or extended based on these results, changes can be incorporated directly. As the scientist's mental model evolves, so does their WayFarer model.

Iterative refinement and collaboration

The discussion above is presented as if this workflow were purely linear, but in practice it is cyclical. Only when embarking on an entirely unfamiliar project does a scientist start from scratch in collecting external knowledge. The starting point is more often an existing model, which is extended over time as additional information becomes available, through internal experiments or external sources. Information and models contained in a WayFarer project represent a local snapshot of the researcher's understanding. This snapshot may be updated by revisiting designated external sources, either to re-import updated versions of existing items, or to retrieve new items associated with those that are already part of the project. The overall process may be thought of as similar to maintaining a reference library through a tool such as EndNote, although in the case of WayFarer the information being managed goes beyond literature references to the knowledge contained within them.

As a scientist's mental model of a system evolves, these changes must be conveyed to collaborators and eventually to the broader scientific community through publication. Much of this collaboration is interactive, involving discussions and presentations of varying degrees of formality, from hallway conversations to group meetings to seminars. And invariably, these presentations involve figures, with PowerPoint being the currently predominant vehicle. In other words, PowerPoint drawings are the *de facto* standard for knowledge representation today.

WayFarer fits into this current practice directly, as a means to articulate the scientist's evolving mental model, for presentation or publication. In fact, the drawing functionality has a look and feel that should be familiar to users of PowerPoint. The key difference is that the figures so created are not simply drawings; they represent underlying knowledge and relationships in a structure that more closely reflects the mental model of the biologist. As a consequence, WayFarer models can not only be presented as static figures, they can be shared with collaborators as a richer and more efficient medium to communicate knowledge from one researcher to another.

Furthermore, collaborators can work from the same model, both drawing from and adding back knowledge that can be immediately presented and understood by other scientists. In this way, incomplete and misunderstood information is minimized and the transfer of knowledge from researcher to researcher or group to group is more precise and effortless.



Possibilities of Everyday Modeling

A modeling system put directly in the hands of individual researchers can bring the power of the computational approach out of the realm of restricted application through specialists and into everyday practice in the lab. There is no intrinsic barrier to biologists using computers themselves: they routinely use the Web to explore scientific literature and databases, and regularly use desktop applications such as Excel and PowerPoint to organize thoughts, plan experiments, analyze results, and present conclusions. Despite the perception that modeling and prediction are relevant only in narrowly defined cases, mental models and thought experiments are in fact routine and essential to the biologist's craft. A software system that manages this activity will simply augment the process the scientist is already using.

A formal modeling system will also improve communication with peers and collaborators. Current practice relies almost exclusively on inert drawings using loose conventions, coupled with sometimes lengthy verbal descriptions to explain a scientist's mental model. Extending this practice to include formal, active models allows greater richness of expression, and enables more precise and complete knowledge transfer from person to person or group to group.

Perhaps more importantly, the ability to 'play' with a system is crucial to building an intuitive understanding of how it works, generating ideas that would never come up otherwise. In fields such as architecture, mechanical engineering, and graphic arts, the advent of computer-aided methodologies has led to explosions in creativity, as practitioners are freed up to think beyond earlier bounds. Model-based biology will have many immediate and predictable benefits, but its potential extends well beyond what can be contemplated from the confines of existing practice.

WayCraft Biosoftware, Inc.

www.WayCraftBio.com

Frank.Russo@WayCraftBio.com

(408) 733-2468