Multiscale Mathematical Modeling to Support Drug Development

David A. Nordsletten, Beracah Yankama, Renato Umeton, V. A. Shiva Ayyadurai, and C. Forbes Dewey, Jr.*

***Abstract*—It is widely recognized that major improvements are required in the methods currently being used to develop new therapeutic drugs. The time from initial target identification to commercialization can be 10–14 years and incur a cost in the hundreds of millions of dollars. Even after substantial investment, only 30–40% of the candidate compounds entering clinical trials are successful. We propose that multiscale mathematical pathway modeling can be used to decrease time required to bring candidate drugs to clinical trial and increase the probability that they will be successful in humans. The requirements for multiple time scales and spatial scales are discussed, and new computational paradigms are identified to address the increased complexity of modeling.**

***Index Terms*—Biological models, computational biology, drug development, multiscale modeling.**

I. Introduction

IT is now well documented that we need to rethink the drug development process that has been in use for the last 50 years [1], [2]. The traditional research pipeline begins with a single biological target, searches through libraries of potential drugs using *in vitro* methods to find a small number of suitable candidates, uses animal testing to select the most promising of these, proceeds to human trials to first establish dosages and initial toxicity, and finally progresses to broad clinical trials with individual drugs that survive these initial screenings.

Typically, the information created at any one of these steps is not effectively shared with the other portions of the pipeline. Clinical trial data are not incorporated in the modeling such that the next round of animal experiments can be extrapolated with better predictions of the human response. A need exists to define semantic and predictive bridges between these important levels of knowledge.

Manuscript received April 2, 2011; accepted August 24, 2011. Date of publication October 24, 2011; date of current version November 18, 2011. This work was supported in part by the Computational and Systems Biology Program of the Singapore-MIT Alliance and additional funding was provided by the Heart, Lung and Blood Institute of the National Institutes of Health under Grant R01-HL090856-02. *Asterisk indicates corresponding author.*

D. A. Nordsletten is with King's College, London, SE1 8WA, U.K. (e-mail: david.nordsletten@googlemail.com).

B. Yankama and V. A. S. Ayyadurai are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: beracah@mit.edu; vashiva@mit.edu).

R. Umeton is with Sapienza University, 00185 Rome, Italy (e-mail: umeton@mat.unical.it).

*C. F. Dewey, Jr., is with the Department of Mechanical Engineering and Bioengineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: cfdewey@mit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2011.2173245

This paper addresses critical needs in drug development that require multiscale computational modeling. The first is to leverage existing models and to build *predictive models of the biological pathways*. Although the use of modeling in drug development is recognized as a key goal, current approaches are frequently limited to a single drug and a single endpoint. Bottino *et al.* [3] have provided an excellent example of how both *in vitro* data and animal models can be combined using multiscale modeling to predict the proarrythmic risk of a new drug prior to human trials. Modeling is especially required if we are to move from the current single drug discovery paradigm to one in which both therapeutic and toxic response is predicted for treatments that use multiple drugs.

The second key need is to *improve the recycling of information between the different process steps* in drug development. In particular, we see an attractive opportunity to integrate *in silico* predictions of *in vitro* molecular pathway results with current *in vivo* knowledge, to guide and improve the mathematical prediction of biological outcomes. This means predicting how the results in one animal model will be manifest in other test animals and in humans. Heimbach *et al.* [4] provide impressive results to demonstrate the value of combining actual clinical results with predictive models to improve prediction. But allometric [5] and single-point animal–human corrections cannot suffice because many drugs and drug combinations exploit different biological pathways, and the genetics of the animals may vary between the different individual pathways [6]. Use of multiple levels of animal testing incorporating quantitative modeling and appropriate analysis of genetic variation promises to create a robust platform for accurately predicting both efficacy and toxicity before going to Phase 1 clinical trials [2]. A desired end point would be a much higher percentage of successful drugs emerging from Phase 2 and Phase 3 trials. This process can be seen to provide a framework to achieve personalized treatment options for individual genotypes.

Examination of these two needs leads to three specific multiscale problems. First, expanding our fundamental approach to predictive biology even at the level of a single cell entails a *massively complex scale of calculation*. Methods to deal with this complexity are proposed in Section III.

The second is the *multiple time scales* that occur at the molecular level. Fig. 1 is an example of the large variation in time scales observed with living cells. One needs to be able to represent each of the phenomena listed with sufficient time resolution to provide an accurate account of the changes that are taking place. This leads to serious computational problems that are discussed in Section IV.

The third is that there are many different types of interacting tissues and organs, each with *discrete volumes and length scales*,

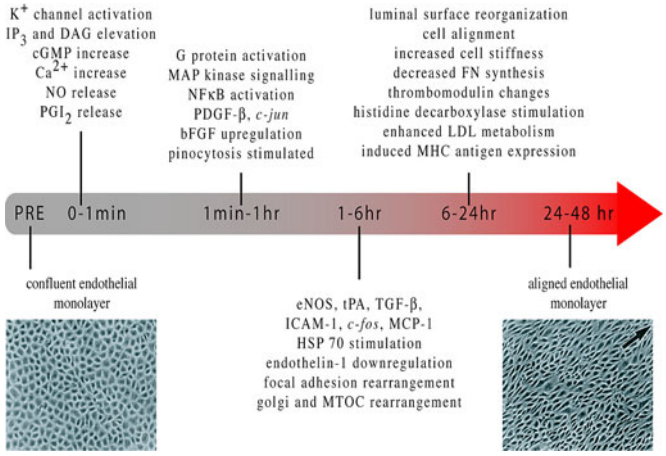

Fig. 1. Time scales describing the response of vascular endothelium to fluid flow (adapted from [7]).

within which the presence or absence of a particular molecule or cell may have different consequences. Section V examines the current successes and limitations of the use of separated but communicating biological compartments in meeting these needs.

II. Requirements for Drug Development

A. Need for Predictive Behavior

Imagine a single monolithic computational model predicting the complete behavior of a single cell with hundreds of proteins interacting inside the cell, tens of compartments such as the nucleus and the golgi, over 30 important integrins for signaling proteins, and dozens of membrane protein complexes that effect ion transport and create transmembrane signaling pathways. One would need probably 10^3 molecular pathways, each with 10–30 specific proteins and molecules, and a total of something on the order of 10^5 ordinary differential equations (ODEs) to describe their interactions.

The complexity of such a model, or any approximation to it, would be enormous. The complexity is compounded by the fact that each molecular pathway is literally a dynamic knowledge domain in itself, with rate constants and reactions that are unique to the particular process being described by the pathway and subject to future revision.

A predictive model does not require all levels of complexity to be visible simultaneously. Thus, details of the protein docking dynamics at the subnanometer level with time scales in microseconds can appear as deterministic rate constants in higher level models [8]. Similarly, complex pathways that govern the production of key molecules may be represented as lumped parametric sources [9].

Quantitative biological pathway models express known relationships and are not useful in identifying new biological connections. One pathway discovery strategy is to use systems biology to design *in vivo* qualitative experiments that can identify logical networks, as exemplified in the response of intestinal epithelium to tumor necrosis factor-α [10]. Quantitative rate constants can be generated from subsequent experiments based on these qualitative networks.

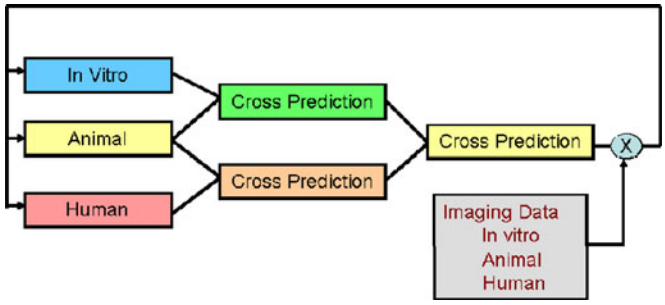

Fig. 2. Design for a flexible and configurable cross-platform *in silico* model to accelerate drug development.

Considerable progress has been made in understanding the docking dynamics of potential drug molecules with known receptor targets [11]. Additionally, the use of *in silico* methods to optimize the drug molecule binding properties has proven to be highly successful, as in the case of alternative cancer drugs to erbitux [12].

B. Need for Communication Between Process Steps

Next-generation drug development will require quantitative models that can predict the major events known about drugs and their targets along with automated means of maintaining and updating the information about the model reactions. When a new toxicity pathway is discovered in clinical trials, the information must be fed back to the predictive models in a consistent and transparent manner. As discussed in Section III, this argues for a system that maintains the integrity of the individual biological pathways and combines them "on the fly" to solve specific problems.

Our experience [13] has convinced us that the key to success in integrating information is to use ontologies as the base descriptions for all data sources. These ontological descriptions can then be used to facilitate machine-based queries and data sharing. This approach can also be used to combine different data repositories without the need to physically move or rearrange the existing primary sources. Further, these processes can be aided and enhanced by machine reasoning and automation[14].

A diagram of our cross-prediction ontological platform is shown in Fig. 2. We are constructing three robust ontological knowledge platforms at the *in vitro*, animal, and human levels. These platforms use existing standards (MIRIAM [15], GO [16], GOA [17], OBO, UniProt [18], and others) and make use of extensive curated ontology collections at the National Center for Biomedical Ontology [19].

III. Multiscale Complexity

In Section II, it was pointed out that comprehensive biological pathway models can require very large numbers of equations, molecular species, and separate biological compartments. A critical task is assembling the known individual pathway

models to create a large ensemble that is computable and normalizes all the information contained within all of the individual models. Our experience in several example problems quickly demonstrated that the unaided human curation of two biological pathways from different sources is marginally possible, but assembling much larger pathways becomes a daunting and error-prone task.

In addition, any complex model formed by merging individual models must be curated as new information is generated about the reactions and their properties. This task can be readily accomplished using the original individual models but becomes nearly intractable when dealing with monolithic merged models.

We have made a significant investment in developing a new computer-aided environment called *Cytosolve* [20] that materially changes the ease and accuracy with which individual biological models can be merged, computed, and curated. It is based on the strategy of keeping information in the individual pathways intact and combining the required pathways at the time of computation. The philosophy is akin to the use of editable and compile-able subroutines in a large software program, where the subroutines can be individually edited and maintained. Details of the merging process, such as the proper normalization and identification of common input and output variables and elimination of duplicate ordinary differential equation (ODE) between collaborating pathways, are embedded in the individual pathways themselves. The merging is materially assisted by a software program called OREMP [21], and the results from OREMP are supervised by a human curator in the same way that the projections of air traffic control radar calculations are supervised by human air traffic controllers. Open use of the Cytosolve package is available at http://cytosolve.mit.edu/.

IV. Complexity of Multiple Time Scales

The problem of managing adaptive time step size in integration of a single monolithic model is well documented and used in current solvers of SBML models—SOSlib and CVODE—which internally vary the time step of the solution to guarantee convergence to the true solution. In that case, the problem of modeling and analyzing response time scales is simply a function of running an ODE simulation for as large a time desired and identifying the characteristic times *a posteriori*.

If we reject the approach that one creates very large monolithic models and we choose to combine smaller models according to their input and output parameters as with Cytosolve, then the entire scheme of adaptive time stepping must be reexamined. One can show that, for a simple set of reactions, separation and simultaneous fixed stepping (of convergent solvers) using a fixed time step no longer guarantees convergence both in transient response and in steady-state conditions.

A. Combining Reaction Concentrations

In the first order, methods that support the joint simulation of separate pathways can be executed explicitly via mass-balance at intermittent time intervals t. That is, between two separate model simulations M_A and M_B, the change in concentration over some fixed Δt in M_A along a shared alignment point (molecular species) must be propagated and combined with M_B concentration change on the same point [20]. While this guarantees conservation of mass, the individual changes in concentration in either M_A or M_B over that Δt may not completely reflect how a fully monolithic model of $M = (M_A \ \& \ M_B)$ would have behaved. A solution must also propagate the rates of change from each model along that Δt back to the centralized controller which determines the stiffness of the transient response over that interval and decides specifically whether to reduce the time step size to lower error growth.

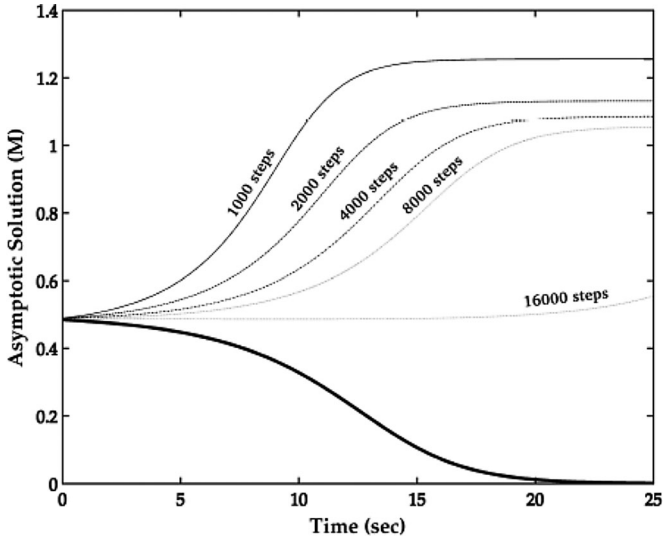

Fig. 3. Behavior of intermittent time synchrony of mass balance for two asymptote problem. The solid line represents the true solution for M(0) slightly < 0.5. Dotted lines are the solutions for x fixed steps.

B. Problem of Two Asymptotes

To illustrate the difficulties arising with multiple solvers operating in parallel, we pose a test case in which what is a sufficient time step for one solution is not necessarily a sufficient time step for the other. This means that small errors in one model calculation may lead to large errors in another linked model calculation unless the time step is reduced to a very small value or a special algorithm is used.

We consider a model consisting of two asymptotes and an unstable nullcline at 0.5. In the true (monolithic) solution, if the initial condition of this model starts at the nullcline $M(0) = 0.5$, then it will remain there. If perturbed, or started above or below the 0.5 value, the solution will increase or decrease to 1.0 or 0.0, respectively. When the models are separately simulated, the time step size may produce an error which drives the solution off of the unstable asymptote onto the incorrect steady-state value. In Fig. 3, the initial condition is started slightly *below* the nullcline, and the error is sufficient to drive the solution from the lower asymptote to the upper asymptote. This model is a simplistic demonstration of the fact that new time-stepping methods must be used. A solution has been obtained in restricted cases and a more general solution and is currently in the final research phase prior to publication [22].

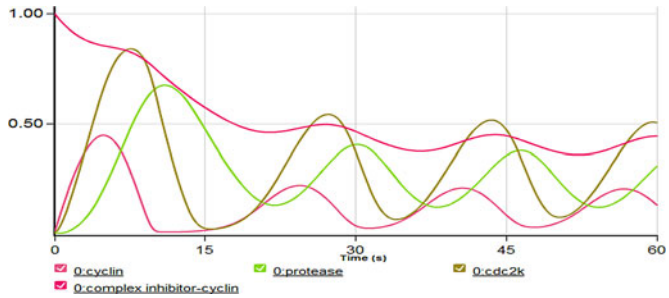

Fig. 4. Individual simulation of BIOMD..8. The response rate of *cyclin* and *complex inhibitor-cyclin* yields ∼4 cycles over 60 s, with a decreasing cyclin concentration.

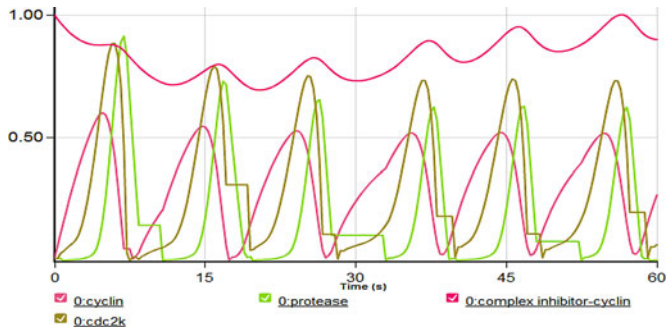

Fig. 5. Joint simulation of BIOMD..8+BIOMD..4. The frequency of *complex inhibitor-cyclin* is 50% higher and more triangular, and *cyclin* no longer purely decreases.

C. Complexity in Combinations—Public Models

Tied to the problem of time scales is the problem of system stiffness and complex coupled behavior. Consider the example combination of two models from Biomodels.net [23] involving cell mitosis, BIOMD..4 & BIOMD..8, where the reactions are aligned on the species *cyclin*, *protease*, and *cdc2k*. Though both models are designed *a priori* to be tuned mitotic oscillators, Figs. 4 and 5 demonstrate that the interaction of the different individual reactions significantly alters both the magnitude as well as the tuned frequency of the oscillator.

V. Simulating Spatial Variation

Nearly all of the existing quantitative biological pathways are written as ordinary differential equations in time. The reaction A + B → C assumes the absence of spatial gradients of the reactants, i.e., they occupy a well-mixed reactor. Even the languages adopted to describe these reactions (SBML, CellML, and MML) and the solvers used for the ODE solutions do not support the concept of spatial extent. FieldML has been developed to support spatial gradients in biological models, but it is not widely used.

Support for spatial variation has traditionally been incorporated within the ODE environment by defining separate compartments and labeling the molecular components separately for each compartment. A particularly successful example is the quantitative comparison of the clearance rates of erythroprotein and an analog molecule called ESPN [24], where the model differentiates between concentrations both inside and outside of the red cell membrane. In another example, Ca++ in the fluid and tissue surrounding a cell, in the cytosol inside that cell, and in the golgi of the cell are considered independent species, with ODE components describing the transition of one of one pseudospecies to another [25], [26]. In this example, external Ca++ could not interact with cytosolic components until transformed into cytosolic Ca++ by an ODE-modeled calcium channel.

Just as the molecular dynamics calculations to determine protein–protein interaction strengths require levels of approximation to the atomic force fields to solve the many-body calculations, the multiple-compartment approach to spatial variation is a first step toward a more complex 3-D description. It has proven very useful with many models, as evidenced in the collection in Biomodels.net [23]. However, more complex problems involving true spatial variation, such as the spreading of the heart depolarization wave in time across the myocardium, require a full 4-D calculation to be meaningful [27]. Even at that level, the smaller cellular scales are represented by lumped models and heuristics that provide species transport between lumped elements. A comprehensive vocabulary for representing multiple compartments in ODEs is not yet in place.

Acknowledgment

The authors would like to thank R. I. Kitney of Imperial College, S. S. Bhowmick of Nanyang Technological University in Singapore, P. Hunter of the University of Auckland, A. Popel of Johns Hopkins University, G. Nicosia of the University of Catania, N. L. Novere of the European Bioinformatics Laboratory, Hingston, U.K., and T. Clark of Harvard Medical School. Many of the ideas expressed in this paper have been refined by active use with their professional colleagues.

References

[1] M. C. Fishman and J. A. Porter, "Pharmaceuticals: A new grammar for drug discovery," *Nature*, vol. 437, pp. 491–493, Sep. 2005.
[2] E. C. Butcher, E. L. Berg, and E. J. Kunkel, "Systems biology in drug discovery," *Nat. Biotechnol.*, vol. 22, pp. 1253–1259, Oct. 2004.
[3] D. Bottino, R. C. Penland, A. stamps, M. Traebert, B. Dumotier, A. Georgiva, G. Helmlinger, and G. S. Lett, "Preclinical cardiac safety assessment of pharmaceutical compounds using an integrated systems-based computer model of the heart," *Prog. Biophys. Molecular Biol.*, vol. 90, pp. 414–443, Jan./Apr. 2006.
[4] T. Heimbach, S. B. Lakshminarayana, W. Hu, and H. He, "Practical anticipation of human efficacious doses and pharmacokinetics using in vitro and preclinical in vivo data," *AAPS J.*, vol. 11, pp. 602–614, Sep 2009.
[5] H. Tang, A. Hussain, M. Leal, M. Mayersohn, and E. Fluhler, "Interspecies prediction of human drug clearance based on scaling data from one or two animal species," *Drug Metabol. Disposit.: Biol. Fate Chem.*, vol. 35, pp. 1886–1893, Oct 2007.
[6] B. M. Agoram, "Use of pharmacokinetic/pharmacodynamic modelling for starting dose selection in first-in-human trials of high-risk biologics," *Br. J. Clin. Pharmacol.*, vol. 67, pp. 153–160, Feb. 2009.
[7] C. F. Dewey, Jr., "Haemodynamic flow: Symmetry and synthesis," *Biorheology*, vol. 39, pp. 541–549, 2002.
[8] S. C. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, "Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems," *Annu. Rev. Phys. Chem.*, vol. 62, pp. 42–64, Apr. 2011.
[9] K. Chen, R. N. Pittman, and A. S. Popel, "Nitric oxide in the vasculature: Where does it come from and where does it go? A quantitative perspective," *Antioxid Redox Signal*, vol. 10, pp. 1185–1198, Jul. 2008.

[10] K. S. Lau, A. M. Juchheim, K. R. Cavaliere, S. R. Philips, D. A. Lauffenburger, and K. M. Haigis, "*In vivo* systems analysis identifies spatial and temporal aspects of the modulation of TNF-α-induced apoptosis and proliferation by MAPKs," *Sci. Signal*, vol. 4, p. ra16, 2011.

[11] C. M. Song, S. J. Lim, and J. C. Tong, "Recent advances in computer-aided drug design," *Brief Bioinform*, vol. 10, pp. 579–591, Sep. 2009.

[12] S. M. Lippow, K. D. Wittrup, and B. Tidor, "Computational design of antibody-affinity improvement beyond in vivo maturation," *Nat. Biotechnol.*, vol. 25, pp. 1171–1176, Oct. 2007.

[13] C. F. Dewey, Jr., S. Ayyadurai, V. Rouilly, C. Poh, S. Bhowmick, J. Evans, and R. I. Kitney, "Footprints in the sand: Supporting external analysis of medical and biological databases," presented at the World Congr. Med. Phys. Biomed. Eng., Seoul, Korea, 2006.

[14] J. Quackenbush, C. Stoeckert, C. Ball, A. Brazma, R. Gentleman, W. Huber, R. Irizarry, M. Salit, G. Sherlock, P. Spellman, and N. Winegarden, "Top-down standards will not serve systems biology," *Nature*, vol. 440, p. 24, Mar. 2006.

[15] N. Le Novere, A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence, and B. L. Wanner, "Minimum information requested in the annotation of biochemical models (MIRIAM)," *Nat. Biotechnol.*, vol. 23, pp. 1509–1515, Dec. 2005.

[16] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, pp. 25–29, May 2000.

[17] D. Barrell, E. Dimmer, R. P. Huntley, D. Binns, C. O'Donovan, and R. Apweiler, "The GOA database in 2009—An integrated gene ontology annotation resource," *Nucleic. Acids Res.*, vol. 37, pp. D396–D403, Jan. 2009.

[18] Uniprot_Consortium, "The universal protein resource (UniProt) 2009," *Nucleic. Acids Res.,* vol. 37, pp. D169–D174, Jan. 2008.

[19] National Center for Biomedical Informatics Research, Stanford School of Medical, Stanford, CA 94305-5479. Available: http://bmir.stanford.edu/.

[20] V. A. Ayyadurai and C. F. Dewey, "CytoSolve: A scalable computational method for dynamic integration of multiple molecular pathway models," *Cell Mol. Bioeng.*, vol. 4, pp. 28–45, Mar. 2011.

[21] R. Umeton, B. Yankama, G. Nicosia, C. F. Dewey, and Jr., "OREMP: Ontology reasoning engine for molecular pathways," in *Proc. 1st Int. Workshop Ontology Repositories and Editors for the Semantic Web*, Heraklion, Crete, Greece, 2010, pp. 26–30.

[22] D. A. Nordsletten, Personal communication, 2011.

[23] N. LeNovère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schlistra, B. Shapiro, J. L. Snoep, and M. Hucka, "BioModels database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems," *Nucleic. Acids Res.*, vol. 34, pp. D689–D691, Jan. 2006.

[24] A. W. Gross and H. F. Lodish, "Cellular trafficking and degradation of erythropoietin and novel erythropoiesis stimulating protein (NESP)," *J. Biol. Chem.*, vol. 281, pp. 2024–2032, Jan. 2006.

[25] T. F. Wiesner, B. C. Berk, and R. M. Nerem, "A mathematical model of cytosolic calcium dynamics in human umbilical vein endothelial cells," *Amer. J. Physiol.*, vol. 270, pp. C1556–C1569, May 1996.

[26] R. J. WinslowRL, S. Jafri, E. Marban, and B. O'Rourke, "Mechanisms of altered excitation-contraction coupling in canine tachycardia-induced heart failure—Part II: Model studies," *Circ. Res.*, vol. 84, pp. 571–586, 1999.

[27] D. A. Nordsletten, S. A. Niederer, M. P. Nash, P. J. Hunter, and N. P. Smith, "Coupling multi-physics models to cardiac mechanics," *Prog. Biophys. Mol. Biol.*, vol. 104, pp. 77–88, Jan. 2011.