Large Scale Biological Models in Rodin

Usman Sanwal[1], Thai Son Hoang[2], Luigia Petre[1], and Ion Petre[3,4]

[1] Faculty of Science and Engineering, Åbo Akademi University, Finland
[2] School of Electronics and Computer Science, University of Southampton, UK
[3] Department of Mathematics and Statistics, University of Turku, Finland
[4] National Institute for Research and Development in Biological Sciences, Romania

1 Introduction

Biological systems are typically very large and complex, so much that it is remarkably difficult to capture all the necessary details in one modeling step. The concept of refinement – gradually adding details to a model while preserving its consistency – is thus instrumental. We provide an Event-B based modeling framework for the modeling of large and complex biological systems. Other formal modeling methods such as process algebra, Petri nets and many others has also been used for biomodeling see [1], [3]. The advantage that Event-B brings is that it has refinement as the key concept of the development method. System details can be introduced in several steps and the tool manages all the links between all the intermediary models. Consistency of refinement ensures that all the properties of a model M_i are still valid in its direct refinement successor M_{i+1}. At each refinement step, one can focus on the new elements that are introduced and on their consistency with the previous model.

In this work, we model two biological systems using refinement in Event-B, i.e., we first model a simple, more abstract model of the system and then we add more details in a correct-by-construction manner. The two systems we address are the heat shock response and the ErbB signaling pathway. Modeling the heat shock response in Event-B succeeded before [4]: we started with the abstract model having 10 variables and 17 events and ended up with the concrete model having 22 variables and 57 events. Modeling the ErbB signaling pathway only succeeded earlier [2] for the abstract model, with 110 variables and 242 events. The concrete model would have 1320 events, which was not supported by Rodin. With our current approach we are able to handle such a big model.

2 Modelling reaction networks in Event-B

We model reaction networks as sets of biochemical reactions, where each reaction specifies its reactants, products, and possibly inhibitors and catalyzers. These reactions can be either reversible or irreversible and each reaction could also have an associated flux, describing the rate at which its products are produced and its reactants consumed. With these assumptions, a reaction r can be written as a rewriting rule of the form:

$$r: \quad m_1X_1 + m_2X_2 + ... + m_nX_n \rightarrow m'_1X_1 + m'_2X_2 + ... + m'_nX_n, \qquad \text{(R)}$$

where $\mathcal{S} = \{X_1, \ldots, X_n\}$ is the set of *reactants* and $m_1, ..., m_n, m'_1, ..., m'_n \in \mathbb{N}$ are non-negative integers.

To model reaction networks in Event-B: every reactant is modeled by a variable and every reaction is modeled by an event. Invariants ensure the correctness of each reactant and biological properties of interest, for instance the mass conservation rule that ensure that the number of certain reactants is constant.

Thus, $X_1, X_2, ..., X_n$ are the variables of the model, their type being specified by corresponding invariants. Initial values for all of these variables are set in the initialisation event. For each reaction r of the reaction network, we specify in its guard that it must have enough of each reactant in order for the reaction to be enabled, while the action of the event specifies the changes to happen to each variable. This general scheme can be applied to model any reaction network in Event-B. We have demonstrated its applicability for the case studies of heat shock response model and the ErbB signalling pathway model in [4] and [2] respectively.

In this work, we revisited the Event-B model of heat shock response model and the ErbB signalling pathway model to make them more scalable. We demonstrate here that how a particular modeling feature of Event-B – the common mathematical function – enables us to significantly reduce the concrete models sizes. The relation between the abstract and the concrete forms of a reactant is captured with a function. This enables us to model the concrete reactions more elegantly and concisely, and as a result, the total number of events in the refined model is reduced significantly. In the case of the heat shock response, the complete model is described through 21 events, instead of the 57 events of the model in [4]. The difference in the case of the ErbB model is drastic, as we now need only 242 events for the full model of the ErbB signaling pathway in Rodin, instead of 1320 events. We have successfully implemented the model in Rodin and all of the proof obligations were dischrged automatically. To the best of our knowledge, this is the largest-ever model built in Rodin.

References

1. Federica Ciocchetta and Jane Hillston. Process algebras in systems biology. In Marco Bernardo, Pierpaolo Degano, and Gianluigi Zavattaro, editors, *Formal Methods for Computational Systems Biology*, volume LNCS 5016, pages 265–312, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
2. Bogdan Iancu, Usman Sanwal, Cristian Gratie, and Ion Petre. Refinement-based modeling of the erbb signaling pathway. *Computers in Biology and Medicine*, 106:91 – 96, 2019.
3. Ina Koch, Wolfgang Reisig, and Falk Schreiber. *Modeling in systems biology: the Petri net approach*, volume 16. Springer Science & Business Media, 2010.
4. Usman Sanwal, Luigia Petre, and Ion Petre. Stepwise construction of a metabolic network in event-b. *Computers in Biology and Medicine*, 91(C):1–12, 2017.