A diversity-aware computational framework for systems biology

Abstract

Systems biology poses several challenges to computational sciences. As a research method, it implies they need to deal with biological complexity emerging from multi-level, multi-scale, non-linear, quantitative and stochastic systems having dynamic hierarchies of regulative layers. As a research domain, it is very diverse, and many sub-domains compose it, each one with its representational specificities. The scope of this work is to allow for actual collaboration in this context, by a computational framework for knowledge management in systems biology. The framework includes a modeling approach, mostly responding to knowledge inference requirements, and a domain-specific model description language, complementing the modeling approach by dealing with knowledge representation and exchange requirements. While a variety of computational approaches to systems biology exist, they often tackle a limited subset of the existing requirements, the proposed framework intends to respond to them comprehensively.

The proposed modeling approach targets knowledge inference and representation requirements by combining the strengths of state-of-the-art solutions. Relying on the Nets-within-nets (NWNs) formalism, it provides the definition completeness of mathematical models while allowing for direct execution like computational models. Supporting different information specifications, it allows for model composition processes like hybrid models but preserving formalism uniformity. Also, it supports hierarchy and flexible abstractions. These capabilities support the construction of multi-level and multi-context models: they represent not only different organization levels from the system but also different views over them. Thanks to this versatility it is possible to model in an explicit way the role of the spatial and process contexts respectively over the system at each level. Thanks to the fact these features are made explicit, the landscape of regulations and their dynamic hierarchy emerges during execution. The proposed approach initially develops to target complex biological processes such as ontogenesis. At first, an application example for developmental biology, the VPC specification in C. Elegans, is provided. It shows excellent flexibility in representation capabilities. Two more application examples follow: the first one targets a cultured synthetic biological system and the second one focuses on the spread of antibiotic resistance within the microbiota. A limitation is that models following the proposed approach work as knowledge bases only for researchers with a background in computer science.

To make them accessible for non-expert users as well, the Biological System Description Language (BiSDL) has a high-level syntax recapitulating the domain-specific language of experimental biologists. At the same time, it also covers the low-level formalism elements. Also, BiSDL supports modularity: a description can make use of other descriptions, representing interconnected and nested models. The expert user can build up models under the multi-level, multi-context approach using BiSDL, creating re-usable

modules corresponding to biological structures and processes. They can store these modules in libraries. Non-expert users can access libraries and access the knowledge stored in existing modules, as well as re-use, customize and combine them into high-level models by merely connecting them, and tuning their parameters. A custom compiler generates NWNs models from BiSDL descriptions, and a custom simulator directly simulates them. In this way, system dynamics is accessible as well to the non-expert user.

The proposed computational framework devises a modeling approach that collects contributions from the different subdomains involved, and a high-level model description language making models accessible for the non-expert users in the field. The goal is to foster true interdisciplinarity in systems biology by creating a common playground for all the stakeholders. The resulting genuinely shared perspective should allow to ask new questions, and orient the growing technological capabilities both on the computational and high-throughput analysis techniques fronts. Ultimately, the proposed framework wants to contribute, as an enabler, to the cultural shift from multi-disciplinarity to inter-disciplinarity in systems biology.

The framework at the moment provides a prototypical version of the complete flow from BiSDL descriptions to the simulation of NWNs models. In the future, the modeling approach should be tested for scalability, considering both a broader spectrum of intracellular mechanisms and a more significant number of cells in the system. Also, the simulator should adapt to parallel computations, so to handle more computational complexity. The framework should devise complexity reduction strategies to improve computational performances. Bioinformatic pipelines should support partly data-driven models construction processes involving not only parameter identification but also model architecture. The framework should also include model analysis routines to explore models formally. A smart user interface should embed the full flow from BiSDL descriptions to simulations allowing easy model exploration and design. This interface could also rely on a visual version of BiSDL, and simulation outcomes visualization.