

## Short Communication

## Modeling and Simulating Constrained Protein Interaction Networks

Bianca K. Stöcker<sup>1,\*</sup>, Johannes Köster<sup>1,2</sup>, Eli Zamir<sup>3</sup>, Sven Rahmann<sup>1,\*</sup><sup>1</sup>Genome Informatics, Institute of Human Genetics, University of Duisburg-Essen, Germany<sup>2</sup>Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, USA<sup>3</sup>Department of Cellular Biophysics, Max Planck Institute for Medical Research, Heidelberg, Germany

\*Correspondence: Bianca.Stoecker@uni-due.de, Sven.Rahmann@uni-due.de

Received 2017-10-15; Accepted 2017-10-27

## SUMMARY

Cellular functions of biochemical reactions are enabled by protein interactions. In addition to the protein interactions themselves, dependencies between these interactions such as allosteric activation or mutual exclusion contribute to the complexity and functionality of these systems. We introduce a model of constrained protein interaction networks that uses propositional logic to combine protein networks with interaction dependencies. Further, we present an efficient model, enabling a fast simulation and analysis of many proteins in large networks. This allows to simulate perturbation effects (over-expression/knockout of single or multiple proteins, protein concentrations changes). A comparison of simulation results with known dependencies against simulated complex formation without dependencies shows that interaction dependencies limit the resulting complex sizes. Further, we show how propagation of perturbation effects is influenced by the interplay of network topology and interaction dependencies and how to analyze this with our model.

## KEYWORDS

Protein-protein interactions; Interaction dependencies; Propositional logic; Stochastic simulation; Integrin adhesome

## BODY

Cellular functions of biochemical reaction systems are enabled by protein interactions. In addition to the protein interactions themselves, dependencies between these interactions such as allosteric activation or mutual exclusion contribute to the complexity and functionality of these systems. Hence, the integration of such dependencies in formal models is important.

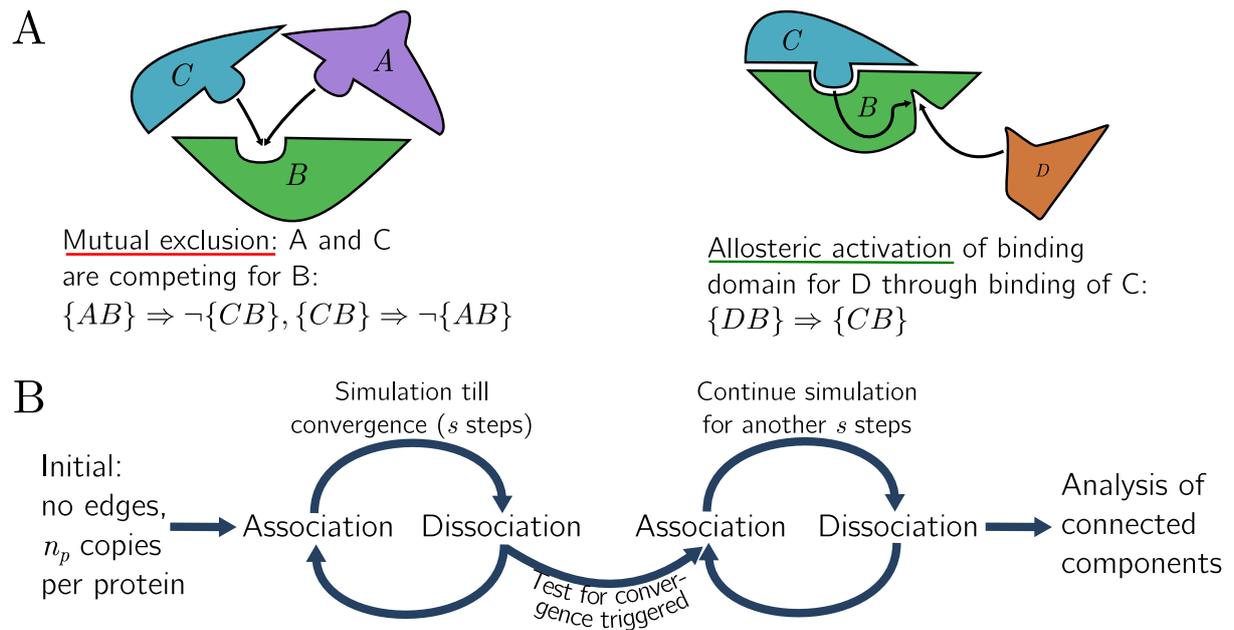
We introduce a model of *constrained protein interaction networks* ("constrained networks") that uses propositional logic to combine protein networks with interaction dependencies. We define a *constraint* as logic implication  $i \Rightarrow \psi$ , where  $i$  is an interaction and  $\psi$  a propositional logic formula over the set of all interactions. The top row of Figure 1 visualizes

two types of interaction dependencies and shows the corresponding constraints. Arbitrary combinations of mutual exclusion and allosteric effects can be achieved through the combination of different constraints.

We simulate the behavior of real proteins in a cell using a constrained network. In the simulation we use  $n_p$  copies of each protein  $p$  as nodes, while the edges represent the current interactions between them. Thus, connected components of the graph denote protein complexes. Initially, there are no interactions and each protein is a singleton. We omit the spatial location of the proteins, i.e., protein position and movement in the cell are not simulated and the formation of an interaction is independent of it. The stepwise simulation repeats the execution of two phases: (1) Association phase: Each protein copy can randomly form new associations corresponding to possible interactions and interaction constraints. (2): Dissociation phase: Existing interactions probabilistically dissociate, potentially splitting large complexes into smaller ones. These phases are repeated until steady state is reached, i.e., until neither the distribution of complex sizes nor the total number of interactions (edges) in the simulation network changes. This process is visualized in the bottom row of Figure 1. In the association phase, only interactions that do not violate constraints are allowed. To ensure this, we transform the logic formula resulting from all constraints of the interaction  $i$  to the equivalent *disjunctive normal form* (DNF) and evaluate its satisfiability using a fast bitvector approach.

With our efficient model we are able to conduct a fast simulation and analyse many proteins in large networks. Using an extension of the human adhesome network with 718 different proteins and 1000 copies per protein, one simulation run takes  $\approx 4$  minutes and requires  $\approx 1.1$  GB RAM on a single Intel Core i7-4790K with 4.00GHz. This allows to simulate perturbation effects (overexpression/knockout of single or multiple proteins, protein concentrations changes).

The construction of the constrained networks is based on public interaction databases [1] (binary interactions only) and known [2] as well as text-mined interaction dependencies [3]. Using the human adhesome network as input network, we adjusted our simulation parameters to match properties of known human



**Figure 1:** **A:** Simultaneously possible protein interactions are limited through interaction dependencies. Left: Competition of two or more proteins on the same binding domain. Right: One interaction depends on another, allosteric, interaction that induces a conformational change. **B:** Visualisation of the simulation procedure. We keep alternating association and dissociation phase until the convergence criterion is first met after  $s$  steps and then continue the simulation for another  $s$  steps. At the end of simulation each connected component denotes one protein complex.

protein complexes [4] to demonstrate the benefits of our model. A comparison of simulation results with known dependencies against simulated complex formation without dependencies shows that interaction dependencies limit the resulting complex sizes. Further, we show how propagation of perturbation effects is influenced by the interplay of network topology and interaction dependencies and how to analyse this with our model.

At the moment, information on interaction dependencies is quite sparse, but given more data, it may become possible to analyse diseases that are based on failing constraints in addition to failing interactions, e.g., an allosteric effect not taking place due to a small change in conformation on one protein. For this analysis, one may add the removal of certain constraints as an additional type of perturbation and evaluate the effects on the resulting complexes.

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**REFERENCES**

1. Das J, Yu H. **HINT: High-quality protein interactomes and their applications in understanding human disease.** BMC Systems Biology. 2012;6:92. doi:10.1186/1752-0509-6-92.
2. Ceol A, Chatr-aryamontri A, Santonico E, Sacco R, Castagnoli L, Cesareni G. **DOMINO: a database of domain-peptide interactions.** Nucleic Acids Research. 2007;35(suppl 1):D557–D560. doi:10.1093/nar/gkl961.
3. Koster J, Zamir E, Rahmann S. **Efficiently mining protein interaction dependencies from large text corpora.** Integrative Biology. 2012;4:805–812. doi:10.1039/C2IB00126H.

4. Ruepp A, Waegle B, Lechner M, Brauner B, Dunger-Kaltenbach I, Fobo G, et al. **CORUM: the comprehensive resource of mammalian protein complexes – 2009.** Nucleic acids research. 2010;38(suppl 1):D497–D501. doi:10.1093/nar/gkp914.