Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

Data integration in logic-based models of biological mechanisms

Benjamin A Hall[1] **& Anna Niarakis**[2,3,*]

Affiliations

[1]*Department of Medical Physics and Biomedical Engineering, UCL, London, UK*
[2]*GenHotel, Univ. Évry, University of Paris-Saclay, Genopole, 91025, Évry, France*
[3]*Lifeware Group, Inria Saclay-île de France, Palaiseau 91120, France*
* *corresponding author*

Abstract

Discrete, logic-based models are increasingly used to describe biological mechanisms. Initially introduced to study gene regulation, these models evolved to cover various molecular mechanisms, such as signalling, transcription factor cooperativity, and even metabolic processes. The abstract nature and amenability of discrete models to robust mathematical analyses make them appropriate for addressing a wide range of complex biological problems.

Recent technological breakthroughs have generated a wealth of high throughput data. Novel, literature-based representations of biological processes and emerging machine learning algorithms offer new opportunities for model construction. Here, we review recent efforts to incorporate omic data into logic-based models and discuss critical challenges in constructing and analysing integrative, large-scale, logic-based models of biological mechanisms.

Keywords

Logic-based models, Boolean models, executable models, qualitative dynamical modelling, omic data integration, *in silico* simulations, formal verification

Highlights

- **Logic-based models are powerful tools to decipher complex biological processes**
- **High-throughput data can be used to enrich, validate, contextualise and infer logic-based models**
- **Efficient omic data integration and rigorous formal methods for large-scale dynamic analysis are paramount challenges in systems biology**

Introduction

Logic-based models have made significant contributions to our understanding of a wide range of biological processes in health and disease. Initially introduced in the 60s to describe gene regulatory circuits [1-3], logic-based models have evolved substantially over the past five decades to cover various biological processes, such as signalling cascades, ion channels, coregulation of transcription factors and even metabolism. With the growing body of data available due to technological breakthroughs, new methods are being developed to integrate different biological

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

scales and expand the size and complexity of discrete models. Additionally, efforts to create formalised, large-scale representations of network "maps" open avenues for rapidly repurposing these datasets to serve as scaffolds for qualitative models [4].

Logic-based models use logical operators, such as AND, OR and NOT, to describe the functions that govern the regulation of the biological entities. While detailed mechanistic knowledge is not a prerequisite, the type of regulation (positive or negative) between the biological entities and the directionality of these regulations is necessary to construct the regulatory graph [5]. In the logical formalism, genes, proteins, and other biomolecules are assigned discrete values that correspond to activity thresholds (binary values for Boolean models, multivariate values for logical models), and logical rules define the evolution of the system in the next time step. Time is implicitly modelled using updating schemes that, together with the logical rules, define the emergent behaviour of the system [6, 7].

In silico simulations of the logic-based discrete models give insights into the dynamics of the modelled system and allow in-depth analysis, like the searching of "attractors"- terminal states of the system such as steady states or cycles [8]. Simple attractors represent fixed points that correspond to the system's stable states. These states can be linked to cellular decision-making processes, such as apoptosis, cell proliferation, migration, chemotaxis. Complex attractors represent terminal cycles that can be linked to biological oscillations, like, for example, the p53 MDM2 interactions [9-11]. The absence of parameters makes logic-based models suitable for large-scale biological networks where little or no kinetic information is available. Nevertheless, as their size and complexity scale up, their analysis can prove to be challenging.

Technological advancements including high-throughput methods have led to an overwhelming amount of biological data. Such data has created a pressing need to develop tools and methodologies that could integrate omic data into the modelling pipelines. These new approaches can be broadly divided into two distinct categories. In the first, omic data are used in combination with small-scale experiments and prior knowledge for model enrichment, contextualisation and validation, while in the second category, omic data are used as source input to infer network structure and functions.

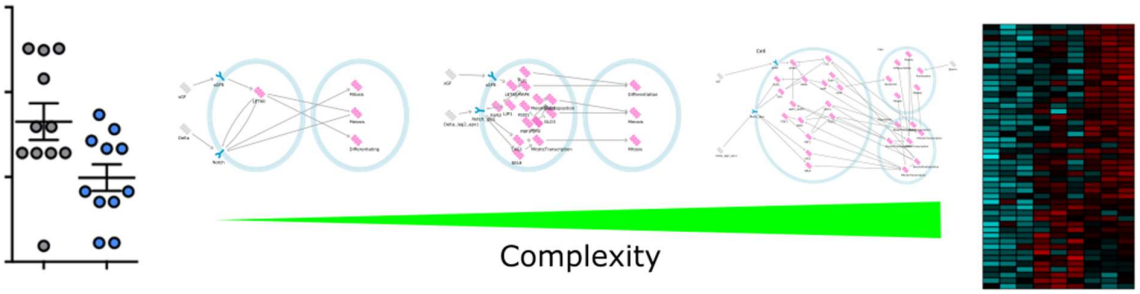

Figure 1. High-throughput data allow for the building of increasingly complex models that can be used to describe a complete picture of the cell.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

High-throughput data integration into logic-based models

Efforts to combine high throughput data with discrete logic-based modelling depend heavily on the model purpose and the data availability and include model enrichment, validation and contextualisation. A typical approach consists of using omic data to expand existing models with entities of interest that can be measurable and comparable in different conditions. Early attempts to combine high throughput data with logic-based models consisted mainly of using the data as a guide to identify key genes and biomolecules to include in the model. An example is the use of high-throughput data to construct a prior-knowledge based, discrete logic model to study mast cell activation in the context of allergy [12]. To build the regulatory graph, the authors used proteomic data, pointing to novel SLP76 interactants identified for the first time in mastocytes [13]. A combination of small-scale experiments, such as quantitative PCR, Western blots, EMSA, together with data from genome-wide assays, such as RNA-sequencing and ChIP-sequencing, was used to assemble a comprehensive regulatory network to study the reprogramming of pre-B cells into macrophages [14]. An iteration of model predictions and *in vitro* validation led to the update of the model with new knowledge and a better understanding of B cell reprogramming mechanisms. In the same line, researchers developed a methodology that integrates several -omics datasets to identify candidate genes, serving as seeds for network modelling. They analysed multi-omics data from the Consensus Molecular Subtypes (CMSs) [15,16] study of colorectal cancer (CRC) to expand a previously built generic cell-fate decision network [17].

In many studies, omic data is used as a source of biomarker signatures compared against stable states to validate phenotypic outcomes. In this case, the regulatory graph of the discrete model is usually built manually through curation of the literature, text mining, and pathway database interrogation. The logical formulae describing specific mechanisms of gene activation are derived from the results of small-scale experiments. Then different types of omic data are analysed and compared against the model behaviour for validation. Recent examples include the enrichment of a logical model of macrophage polarisation to describe cancer cell-macrophage interactions and its training using microarray expression data from *in vitro* co-culture experiments [18-19]. A similar methodology is employed for the building of a logical model for cancer cell invasion and migration. Alongside model building, researchers propose matching transcriptomics data to the attractors and validating the model on cell line experiments [20]. Going one step further and focusing on the role of ion channels in cancer, an executable model of osmotic regulation and membrane transport was proposed predicting behaviour from expression data [21-22].

In a recent commentary, the need for personalised models and the challenges that lie in incorporating high-throughput data into mechanistic dynamic models were highlighted [23]. An example of this is the framework developed to tailor logical models to a particular biological sample. The approach focuses on integrating mutation data, copy number alterations (CNA), and expression data (transcriptomics or proteomics) into logical models [24]. Using this method, the researchers propose a logical model to study the mechanisms of resistance to BRAF inhibition between melanomas and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

colorectal cancers. The model was built using literature mining and pathway integration and was contextualised for 100 melanoma and colorectal cell lines using available omics data, including mutations and RNAseq data [25]. Cell-specific logic-based models have also been employed to recapitulate experimentally tested dynamic proteomic changes and phenotypic responses in diverse Acute Myeloid Leukaemia (AML) cell lines treated with a variety of kinase inhibitors [26]. To improve patient stratification, researchers assembled a network of logical relationships linking genes that are mutated frequently in AML patients and contextualised the model with genomic data inferring relevant patient-specific clinical features [27].

Data-driven discrete model inference

Whilst high-throughput datasets offer new ways to build and analyse models following bottom-up approaches; reverse engineering methods can also be applied to infer models from experimental data. Different algorithms have been developed to reconstruct logic-based models, and specifically Boolean models, from high-throughput data. There exist two broad categories; combinatorial optimisation methods, which include integer or answer set programming (ASP) and allow for full exploration of the search space to identify the model that best explains the experimental data, and methods that implement heuristic approaches. The first category has the drawback of not scaling well due to computational explosion, while the second one tends to focus on specific conditions and stable states to ease the calculation burden.

Recently, the caspo time series (caspo-ts) method [28, 29], which allows learning of BNs from phosphoproteomic time series data given a Prior Knowledge Network (PKN), was applied to data from four breast cancer cell lines (BT20, BT549, MCF7, UACC812). Based on ASP and model-checking, the method could handle a large PKN with 64 nodes and 170 edges [30]. Another popular software for building logic-based models of signalling networks using prior knowledge and phosphoproteomic data is CellNOptR. CellNOptR supports multiple formalisms, from Boolean models to differential equations, in a common framework [31,32]. GABNI (Genetic Algorithm-based Boolean Network Inference) is a method that searches for an optimal Boolean regulatory function by exploiting a mutual information-based Boolean network inference (MIBNI). If this step fails, then a genetic algorithm (GA) is applied to search an optimal set of regulatory genes on a broader solution space [33]. BONITA (Boolean Omics Network Invariant-Time Analysis (BONITA)) is a new algorithm for signal propagation, signal integration, and pathway analysis capable of modelling heterogeneity in transcriptomic data. The logical rules of the model are inferred by the genetic algorithm and are refined by local search. Application of BONITA pathway analysis to previously validated RNA-sequencing studies identifies additional relevant pathways in in-vitro human cell line experiments and in-vivo infant studies [34]. Single-cell expression data has also been used to infer the underlying model of blood development from the mesoderm. The expression of 40 genes, measured using qRT-PCR data in 3934 cells, was discretised and used to infer a Boolean network consisting of 20 transcription factors, giving insight into the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

independent roles of Hox and Sox in Erg activation [35]. Lastly, BTR, an algorithm for training asynchronous Boolean models with single-cell expression data using a novel Boolean state space scoring function, was recently proposed. BTR refines existing Boolean models and infers new by improving the match between model prediction and expression data [36].

Scalability in inference and analysis of logic-based models

Understanding complex biological processes, such as immunometabolism, the tumour microenvironment, chronic or acute inflammation, or autoimmunity, requires models that do not comprise only a handful of nodes but can be adapted accordingly to incorporate hundreds of nodes and reactions. Advancements in the field reflect the tendency to scale up in terms of size and complexity to create models of more realistic performance. Recently, the development of the tool CaSQ bridged the gap between static and dynamic representations of disease mechanisms, with the inference of large-scale Boolean models from molecular interaction maps [37]. The automated inference of large-scale Boolean models creates new challenges in analysing these models, pushing the limits of the existing tools and methodologies. Commonly used software such as GINsim [38] can handle Boolean and multivariate logic-based models; however, the attractor's search can be challenging when scaling up, relying on model reduction techniques to deal with large systems.

Several platforms offer different approaches to dealing with large complex systems, focused on different problem areas. Cell Collective [39] efficiently handles large-scale Boolean models for simulations but does not offer attractors search. In contrast, BoolNet, an R/ Bioconductor package, offers a collection of options for the analysis of Boolean models and a set of heuristics for attractors search when the size and the complexity of the model is considerably large [40]. These heuristics focus on retrieving stable states in lieu of searching the whole state space and significantly reducing the calculation burden, though the results are limited to analysing stable states. BMA [41,42] focuses on analysing stable states and, more particularly, fixed points, offering several highly scalable algorithms for model analysis, including stability proof, cycle searching, and linear temporal logic [43-45]. The specialisation of tools emphasises the importance of commonly agreed standards for model storage.

In parallel, progress has been made in developing hybrid and multi-scale integrative modelling frameworks, connecting different formalisms, and generating new insights from the emergent, combined properties. FlexFlux, an open-source java software, combines metabolic and regulatory networks based on the identification of steady states. These steady states are further used as constraints for metabolic flux analyses using Flux Balance Analysis (FBA) [46]. A multi-scale framework that couples cell cycle and metabolic networks in yeast was proposed, integrating Boolean models of a minimal yeast cell cycle with a constraint-based model of metabolism. Models are implemented in Python using the BooleanNet and COBRApy packages and are connected using Boolean logic. The methodology allows for the incorporation of interaction data and validation through -omics data [47].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

Community efforts for the reproducibility of discrete models in biology

Recent studies have raised concerns about reproducibility in various scientific fields. In computational systems biology, efforts have been made to identify the problem and propose strategies to tackle it [48]. The Curation and Annotation of Logical Models (CALM) initiative emerged to promote reproducibility, interoperability, accessibility and reusability of the discrete biological models [49]. The initiative promotes reproducibility by linking model components to the underlying experimental papers using proper identifiers like BioModels.net Qualifiers[1]. Furthermore, the CoLoMoTo Interactive Notebook developed by the community relies on Docker and Jupyter technologies to provide a unified and user-friendly environment to edit, execute, share, and reproduce analyses of qualitative models of biological networks [50].

Table 1: Brief overview of relevant modelling software and their main features

Tool	Features	Environment	Annotation Support
CaSQ	Boolean model inference from molecular interaction maps	Python	Yes
GINsim	Logical network analysis, reduction functionality, attractors search	Java	Yes
Cell Collective	Boolean network analysis, real-time simulations, topological analysis	Javascript, web-based	Yes
BoolNet	Boolean network analysis, attractors search, heuristics	R/ Bioconductor	No
BMA	Stability analysis, linear temporal logic	Web-based, optional CLI	Yes
Caspots	Inference of Boolean models from time-series omic data	Python	No
CellNOpt	Inference of Boolean models from time-series omic data	R/Bioconductor	No
BONITA	Inference of Boolean models from transcriptomic data	R/Bioconductor	No
FlexFlux	Boolean and FBA analysis	R/Bioconductor	No

[1] https://co.mbine.org/standards/qualifiers

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

New methods for formal analysis of large-scale logic-based models

The Most Permissive Boolean Networks (MPBN) approach is a promising formal method that addresses the fact that both synchronous and asynchronous dynamical interpretations of Boolean models can miss some predictions of behaviours observed in similar quantitative systems. The MPBNs approach formally guarantees not to miss any behaviour achievable by a quantitative model following the same logic. Moreover, MPBNs significantly reduce the complexity of dynamical analysis, allowing for modelling genome-scale networks. One limitation of the approach can be the generation of over approximated dynamical representations, with only small subjects of the corresponding trajectories effectively observed [51].

The control of Boolean networks offers the possibility to delineate interconnected pathways and specify conditions to determine a functional outcome. Researchers compute a minimal subset of nodes (Cmin) in recent work that allows a BN to be driven from any initial state in an attractor to an attractor of interest by a single step perturbation of Cmin. In their method, they decompose the network into modules, compute the minimal control on the projection of the attractors to these modules, and then compose the results to obtain the global Cmin [52].

Model verification, derived from the broader field of verification in software and hardware, offers a new way to tackle complexity. This approach involves using mathematical proofs to offer guarantees of model behaviour, allowing for more reliable and robust conclusions. Examples include the computation of attractors [53] and proofs of stability [43], where proofs of properties of the whole model are composed of proofs computed on individual components.

Conclusion

The growing availability of high quality, whole-cell biological data has underlined the need to develop rigorous integrative methods that connect observations to fundamental mechanisms of action. Data driven-model synthesis combined with high-quality biocuration could lead to the construction of more accurate and robust models. At the same time, the rapid adoption of increasingly large logic-based models challenges the existing methods and tools used for dynamic analysis. Efficient formalisms and tool implementations will be required to analyse and understand these models and, more widely, the complex biological mechanisms in health and disease.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

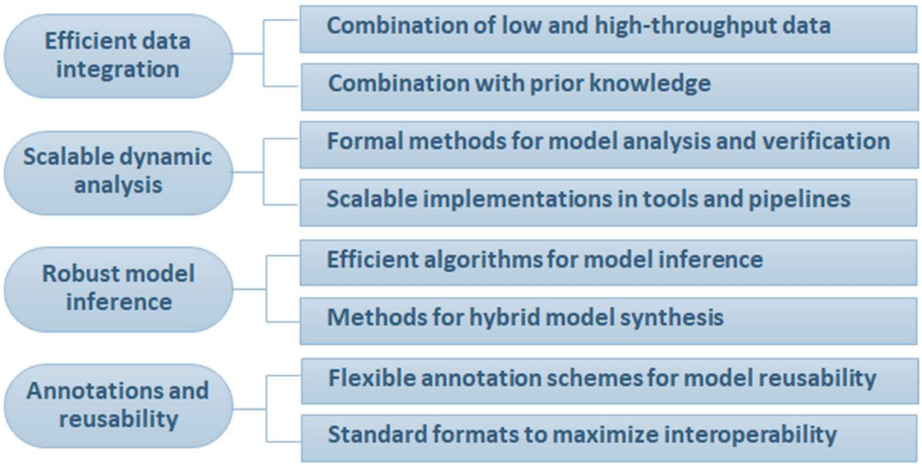

Figure 2. Key challenges in integrating high-throughput data in logic-based models

Acknowledgements

The authors would like to thank Professor Denis Thieffry, IBENS, Paris, for the valuable comments and exchanges. BAH is supported by a Royal Society University Research Fellowship (grant number UF130039).

Conflict of interest statement

The authors declare no conflict of interest.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

* * of outstanding interest

1. Sugita M, Fukuda N: **Functional analysis of chemical systems in vivo using a logical circuit equivalent. 3. Analysis using a digital circuit combined with an analogue computer.** *J. Theor. Biol.* 1963, **5**:412–425.

2. Kauffman SA: **Metabolic stability and epigenesis in randomly constructed genetic nets.** *J. Theor. Biol.* 1969, **22**:437–467.

3. Thomas R: **Boolean formalisation of genetic control circuits.** *J. Theor. Biol.* 1973, **42**:563–585.

4. Ostaszewski M, Gebel S, Kuperstein I, Mazein A, Zinovyev A, Dogrusoz U, Hasenauer J, Fleming RMT, Le Novère N, Gawron P, et al.: **Community-driven roadmap for integrated disease maps.** *Brief. Bioinformatics* 2019, **20**:659–670.

5. Wynn ML, Consul N, Merajver SD, Schnell S: **Logic-based models in systems biology: a predictive and parameter-free network analysis**

method. *Integr Biol (Camb)* 2012, **4**:1323–1337.

6. Abou-Jaoudé W, Traynard P, Monteiro PT, Saez-Rodriguez J, Helikar T, Thieffry D, Chaouiya C: **Logical modeling and dynamical analysis of cellular networks.** *Front. Genet.* 2016, **7**:94.

7. Niarakis A, Helikar T: **A practical guide to mechanistic systems modeling in biology using a logic-based approach.** *Brief. Bioinformatics* 2020, doi:10.1093/bib/bbaa236.

8. * Schwab JD, Kühlwein SD, Ikonomi N, Kühl M, Kestler HA: **Concepts in Boolean network modeling: What do they all mean?** *Comput. Struct. Biotechnol. J.* 2020, **18**:571–582.

Provides a comprehensive overview of all concepts used in Boolean modelling.

9. Suarez OJ, Vega CJ, Sanchez EN, González-Santiago AE, Rodríguez-Jorge O, Alanis AY, Chen G, Hernandez-Vargas EA: **Pinning Control for the p53-Mdm2 Network Dynamics Regulated by p14ARF.** *Front. Physiol.* 2020, **11**:976.

10. Choi M, Shi J, Jung SH, Chen X, Cho K-H: **Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage.** *Sci. Signal.* 2012, **5**:ra83.

11. Proctor CJ, Gray DA: **Explaining oscillations and variability in the p53-Mdm2 system.** *BMC Syst. Biol.* 2008, **2**:75.

12. Niarakis A, Bounab Y, Grieco L, Roncagalli R, Hesse A-M, Garin J, Malissen B, Daëron M, Thieffry D: **Computational modeling of the main signaling pathways involved in mast cell activation.** *Curr. Top. Microbiol. Immunol.* 2014, **382**:69–93.

13. Bounab Y, Hesse A-M-, Iannascoli B, Grieco L, Couté Y, Niarakis A, Roncagalli R, Lie E, Lam K-P, Demangel C, et al.: **Proteomic analysis of the SH2 domain-containing leukocyte protein of 76 kDa (SLP76) interactome in resting and activated primary mast cells [corrected].** *Mol. Cell. Proteomics* 2013, **12**:2874–2889.

14. Collombet S, van Oevelen C, Sardina Ortega JL, Abou-Jaoudé W, Di Stefano B, Thomas-Chollier M, Graf T, Thieffry D: **Logical modeling of lymphoid and myeloid cell specification and transdifferentiation.** *Proc Natl Acad Sci USA* 2017, **114**:5792–5799.

15. Tsirvouli E, Touré V, Niederdorfer B, Vázquez M, Flobak Å, Kuiper M: **A Middle-Out Modeling Strategy to Extend a Colon Cancer Logical Model Improves Drug Synergy Predictions in Epithelial-Derived Cancer Cell Lines.** *Front. Mol. Biosci.* 2020, **7**:502573.

16. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, et al.: **The consensus**

molecular subtypes of colorectal cancer. *Nat. Med.* 2015, **21**:1350–1356.

17. * Niederdorfer B, Touré V, Vazquez M, Thommesen L, Kuiper M, Lægreid A, Flobak Å: **Strategies to Enhance Logic Modeling-Based Cell Line-Specific Drug Synergy Prediction.** *Front. Physiol.* 2020, **11**:862.

Describes the use of logic-based models to predict cell line-specific drug combination effects based on omic-inferred baseline calibration data.

18. Palma A, Jarrah AS, Tieri P, Cesareni G, Castiglione F: **Gene regulatory network modeling of macrophage differentiation corroborates the continuum hypothesis of polarisation states.** *Front. Physiol.* 2018, **9**:1659.

19. Marku M, Verstraete N, Raynal F, Madrid-Mencía M, Domagala M, Fournié J-J, Ysebaert L, Poupot M, Pancaldi V: **Insights on TAM Formation from a Boolean Model of Macrophage Polarization Based on In Vitro Studies.** *Cancers (Basel)* 2020, **12**.

20. Cohen DPA, Martignetti L, Robine S, Barillot E, Zinovyev A, Calzone L: **Mathematical modelling of molecular pathways enabling tumour cell invasion and migration.** *PLoS Comput. Biol.* 2015, **11**:e1004571.

21. Riedel A, Shorthouse D, Haas L, Hall BA, Shields J: **Tumor-induced stromal reprogramming drives lymph node transformation.** *Nat. Immunol.* 2016, **17**:1118–1127.

22. Shorthouse D, Riedel A, Kerr E, Pedro L, Bihary D, Samarajiwa S, Martins CP, Shields J, Hall BA: **Exploring the role of stromal osmoregulation in cancer and disease using executable modelling.** *Nat. Commun.* 2018, **9**:3011.

23. Saez-Rodriguez J, Blüthgen N: **Personalised signaling models for personalised treatments.** *Mol. Syst. Biol.* 2020, **16**:e9042.

24. Béal J, Montagud A, Traynard P, Barillot E, Calzone L: **Personalisation of Logical Models With Multi-Omics Data Allows Clinical Stratification of Patients.** *Front. Physiol.* 2018, **9**:1965.

25. * Béal J, Pantolini L, Noël V, Barillot E, Calzone L: **Personalised logical models to investigate cancer response to BRAF treatments in melanomas and colorectal cancers.** *PLoS Comput. Biol.* 2021, **17**:e1007900.

Describes a comprehensive pipeline from clinical question to a validated mechanistic model that uses different omics data types and adapts to dozens of different cell lines.

26. Silverbush D, Grosskurth S, Wang D, Powell F, Gottgens B, Dry J, Fisher J: **Cell-Specific Computational Modeling of the PIM Pathway in Acute Myeloid Leukemia.** *Cancer Res.* 2017, **77**:827–838.

27. Palma A, Iannuccelli M, Rozzo I, Licata L, Perfetto L, Massacci G, Castagnoli L, Cesareni G, Sacco F: **Integrating Patient-Specific Information into Logic**

Models of Complex Diseases: Application to Acute Myeloid Leukemia. *J. Pers. Med.* 2021, **11**.

28. Razzaq M, Paulevé L, Siegel A, Saez-Rodriguez J, Bourdon J, Guziolowski C: **Computational discovery of dynamic cell line specific Boolean networks from multiplex time-course data.** *PLoS Comput. Biol.* 2018, **14**:e1006538.

29. Ostrowski M, Paulevé L, Schaub T, Siegel A, Guziolowski C: **Boolean network identification from perturbation time series data combining dynamics abstraction and logic programming.** *BioSystems* 2016, **149**:139–153.

30. Dorier J, Crespo I, Niknejad A, Liechti R, Ebeling M, Xenarios I: **Boolean regulatory network reconstruction using literature based knowledge with a genetic algorithm optimisation method.** *BMC Bioinformatics* 2016, **17**:410.

31. ** Gjerga E, Trairatphisan P, Gabor A, Koch H, Chevalier C, Ceccarelli F, Dugourd A, Mitsos A, Saez-Rodriguez J: **Converting networks to predictive logic models from perturbation signalling data with CellNOpt.** *Bioinformatics* 2020, **36**:4523–4524.
An updated collection of Bioconductor R packages for building logic-based models of signalling networks from perturbation data and prior knowledge to handle more efficiently large datasets.

32. Terfve C, Cokelaer T, Henriques D, MacNamara A, Goncalves E, Morris MK, van Iersel M, Lauffenburger DA, Saez-Rodriguez J: **CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms.** *BMC Syst. Biol.* 2012, **6**:133.

33. Barman S, Kwon Y-K: **A Boolean network inference from time-series gene expression data using a genetic algorithm.** *Bioinformatics* 2018, **34**:i927–i933.

34. Palli R, Palshikar MG, Thakar J: **Executable pathway analysis using ensemble discrete-state modeling for large-scale data.** *PLoS Comput. Biol.* 2019, **15**:e1007317.

35. Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, Buettner F, Macaulay IC, Jawaid W, Diamanti E, Nishikawa SI, Piterman N, Kouskoff V, Theis FJ, Fisher J, Göttgens B: **Decoding the regulatory network of early blood development from single-cell gene expression measurements.** *Nat Biotechnol.* 2015, **33(3)**:269-276.

36. Lim CY, Wang H, Woodhouse S, Piterman N, Wernisch L, Fisher J, Göttgens B: **BTR: training asynchronous Boolean models using single-cell expression data.** *BMC Bioinformatics* 2016, **17**:355.

37. ** Aghamiri SS, Singh V, Naldi A, Helikar T, Soliman S, Niarakis A: **Automated inference of Boolean models from molecular interaction maps using CaSQ.** *Bioinformatics* 2020, **36**:4473–4482.
Describes a "map-to-model framework" using CaSQ, a software tool that infers Boolean rules based on the topology and semantics of molecular interaction maps, creating annotated, fully executable large-scale Boolean networks.

38. Chaouiya C, Naldi A, Thieffry D: **Logical modelling of gene regulatory networks with GINsim.** *Methods Mol. Biol.* 2012, **804**:463–479.

39. Helikar T, Kowal B, McClenathan S, Bruckner M, Rowley T, Madrahimov A, Wicks B, Shrestha M, Limbu K, Rogers JA: **The Cell Collective: toward an open and collaborative approach to systems biology.** *BMC Syst. Biol.* 2012, **6**:96.

40. Müssel C, Hopfensitz M, Kestler HA: **BoolNet--an R package for generation, reconstruction and analysis of Boolean networks.** *Bioinformatics* 2010, **26**:1378–1380.

41. ** Hall BA, Fisher J: **Constructing and Analysing Computational Models of Cell Signaling with BioModelAnalyzer.** *Curr. Protoc. Bioinformatics* 2020, **69**:e95.
Describes a comprehensive protocol to construct and analyse large-scale Boolean models with BMA.

42. Paterson YZ, Shorthouse D, Pleijzier MW, Piterman N, Bendtsen C, Hall BA, Fisher J: **A toolbox for discrete modelling of cell signalling dynamics.** *Integr Biol (Camb)* 2018, **10**:370–382.

43. Cook B., Fisher J, Krepska E, Piterman N: **Proving Stabilization of Biological Systems**. In: Jhala R., Schmidt D. (eds) Verification, Model Checking, and Abstract Interpretation. VMCAI. *Lecture Notes in Computer Science* 2011*,* **6538:** 134-149.

44. Cook B., Fisher J., Hall B.A., Ishtiaq S., Juniwal G., Piterman N: **Finding Instability in Biological Models.** In: Biere A., Bloem R. (eds) Computer Aided Verification. CAV 2014. *Lecture Notes in Computer Scienc*e 2014, **8559:** 358-372.

45. Claessen K, Fisher J, Ishtiaq S, Piterman N, Wang Q: (2013) **Model-Checking Signal Transduction Networks through Decreasing Reachability Sets**. In: Sharygina N., Veith H. (eds) Computer Aided Verification. CAV 2013. *Lecture Notes in Computer Science* 2013, **8044:** 85-100

46. Marmiesse L, Peyraud R, Cottret L: **FlexFlux: combining metabolic flux and regulatory network analyses.** *BMC Syst. Biol.* 2015, **9**:93.

47. * van der Zee L, Barberis M: **Advanced Modeling of Cellular Proliferation:**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 May 2021 doi:10.20944/preprints202105.0219.v1

Toward a Multi-scale Framework Coupling Cell Cycle to Metabolism by Integrating Logical and Constraint-Based Models. *Methods Mol. Biol.* 2019, **2049**:365–385.
Describes a computational, multi-scale framework that couples off-the-shelf logical (Boolean) models of the yeast cell cycle with a constraint-based model of metabolism.

48. Tiwari K, Kananathan S, Roberts MG, Meyer JP, Sharif Shohan MU, Xavier A, Maire M, Zyoud A, Men J, Ng S, et al.: **Reproducibility in systems biology modelling.** *Mol. Syst. Biol.* 2021, **17**:e9982.

49. * Niarakis A, Kuiper M, Ostaszewski M, Malik Sheriff RS, Casals-Casas C, Thieffry D, Freeman TC, Thomas P, Touré V, Noël V, et al.: **Setting the basis of best practices and standards for curation and annotation of logical models in biology-highlights of the [BC]2 2019 CoLoMoTo/SysMod Workshop.** *Brief. Bioinformatics* 2021, **22**:1848–1859.
Roadmap for the building of interoperable, reusable and reproducible logic-based models

50. Naldi A, Hernandez C, Levy N, Stoll G, Monteiro PT, Chaouiya C, Helikar T, Zinovyev A, Calzone L, Cohen-Boulakia S, et al.: **The colomoto interactive notebook: accessible and reproducible computational analyses for qualitative biological networks.** *Front. Physiol.* 2018, **9**:680.

51. ** Paulevé L, Kolčák J, Chatain T, Haar S: **Reconciling qualitative, abstract, and scalable modeling of biological networks.** *Nat. Commun.* 2020, **11**:4256.
Describes the Most Permissive Boolean Networks (MPBNs) that provides a formal guarantee not to miss dynamic behaviours observable in quantitative models following the same logic. MPBNs can reduce the complexity of dynamical analysis, enabling to model of large-scale networks.

52. Baudin A, Paul S, Su C, Pang J: **Controlling large Boolean networks with single-step perturbations.** *Bioinformatics* 2019, **35**:i558–i567.

53. * Hernandez C, Thomas-Chollier M, Naldi A, Thieffry D: **Computational Verification of Large Logical Models-Application to the Prediction of T Cell Response to Checkpoint Inhibitors.** *Front. Physiol.* 2020, **11**:558606.
Describes two approaches to cope with analysing complex, large-scale logic-based models. Local model verification is inspired by unit testing, and input propagation helps to assess the impact of constraints on the dynamical behaviour.