

## METHODOLOGY

# Modelling with ANIMO: between Fuzzy Logic and ODE

Stefano Schivo<sup>1#</sup>, Jetse Scholma<sup>2#</sup>, Paul E. van der Vet<sup>3</sup>, Marcel Karperien<sup>2</sup>, Janine N. Post<sup>2</sup>, Jaco van de Pol<sup>1</sup> and Rom Langerak<sup>1\*</sup>

\*Correspondence:

r.langerak@utwente.nl

<sup>1</sup>Formal Methods and Tools,  
Faculty of EEMCS, University of  
Twente, 7522NH Enschede,  
The Netherlands

Full list of author information is  
available at the end of the article

#Equal contributor

### Abstract

**Background:** The large amount of data which an experimental biologist must deal with makes it nearly impossible to reason on the dynamics of complex systems without relying on computational support. The software tool ANIMO (Analysis of Networks with Interactive MOdelling) provides such computational support to the experimental biologist. Models built with ANIMO allow to reason about the complex networks of signaling events occurring in living cells. The user-friendly interface allows biology experts to exploit formal methods without the need to become experts also in mathematical formalisms.

**Results:** ANIMO allows for a model precision that is intermediate between ordinary differential equations (ODEs) and fuzzy logic. We show this by presenting ANIMO models for two case studies: *Drosophila melanogaster* circadian clock, and signal transduction events downstream of TNF $\alpha$  and EGF in HT-29 human colon carcinoma cells. The models were originally developed with ODEs and fuzzy logic respectively.

**Conclusions:** ANIMO models allow to replicate with good precision the results of both the ODE and fuzzy logic model. Moreover, ANIMO models require less parameters than ODEs and are more precise than fuzzy logic: for this reason, we position the modelling paradigm of ANIMO between ODEs and fuzzy logic.

**Keywords:** modelling; signalling pathway; timed automata; dynamic behaviour

### Background

#### Modelling in cell biology

Executable biology is a young subfield in computational modeling, aimed at constructing models that mimic biological phenomena *in silico*. It provides an interesting paradigm for enhancing network diagrams with an underlying formal description of both network components and their interactions. Executable models typically contain a description of the direct interactions of network components, requiring the modeler to think in terms of “cause and effect”. Biologists reason in similar ways about the molecular mechanisms of network interactions, and this makes construction of an executable model an intuitive process. ANIMO (Analysis of Networks with Interactive MOdelling, [1, 2]) provides an enabling technology to increase the use of computational models by experimental biologists without a thorough mathematical background. Such models are indispensable for formally comparing experimental data to prior knowledge, or for structuring experimental findings into a new theory. When dealing with complex biological networks, executable biology models are

particularly useful to understand the non-linear dynamics and the entailed emergent properties of the networks. In those cases, an ANIMO model can be used as a support to obtain insight based on abstract representations of the interactions occurring inside living cells. Other applications of ANIMO models include performing *in silico* experiments and the storage and transfer of knowledge on biological networks.

## Results

### Modelling oscillation

Results obtained with ANIMO are comparable to results with other modeling approaches. To demonstrate this, Figure 3 represents an ANIMO model of the circadian clock in *Drosophila Melanogaster*, based on [3], where ordinary differential equations (ODEs) were used. The cyclic behavior of the circadian clock is based on the alternating formation and destruction of the CYC/CLK protein complex. Concentration levels of this complex are in turn regulated by a series of proteins which are produced as a consequence of CYC/CLK formation. The CWO protein is central to the functioning of the network, as it degrades the mRNA for most of the involved proteins. As such, CWO acts as an inhibitor that counterbalances the effect of CYC/CLK. The positive influence of the light-regulated cryptochrome CRY on the degradation of TIM is a consequence of the passage between day and night, allowing the circadian clock to synchronize to a time zone (see Suppl. Sect. D.1).

The output of the ANIMO model in Figure 3 closely matches the original ODE model. In particular, the oscillations in both models show the same periods and phases (see Suppl. Fig S5). Details of the comparison between the ANIMO model and the original ODE model is given in Supplementary Section D.2.

### Using ANIMO to generate hypotheses

We constructed a model of the signaling network downstream of TNF $\alpha$  and EGF, formalizing the crosstalk that takes place between the pathways at different levels of cellular regulation. We first modeled the two pathways in isolation (Suppl. Figs. S6, S7), using information on protein interactions from the KEGG [4] and phosphosite [5] databases. These models were fitted to experimental data from previous studies [6, 7]. We then merged the two pathways into a single model and added autocrine crosstalk between the pathways that has been described in [7]. Briefly, stimulation with TNF $\alpha$  (TNFa in the model) leads to a rapid release of TGF $\alpha$  (TGFA), which activates the EGF receptor (EGFR). This activation causes secretion of IL-1 $\alpha$  (IL-1a) at later time points. The effect of IL-1 $\alpha$  is down-regulated by the secretion of IL-1 receptor antagonist (IL-1ra) downstream of TNF $\alpha$ . The resulting model (Fig. ??) was compared to the experimental data for treatments with 100 ng/ml TNF alone and 100 ng/ml EGF alone (data not shown) [6].

At this point, the behavior of the model deviated from the data for some of the nodes. Changing the parameters of the model was not enough to reproduce the behaviour shown by experimental data. This is an interesting situation, as it requires changes in the topology of the model, which can be interpreted as new hypotheses. Below, we give two examples and show how adaptation of the model can be used to generate novel testable hypotheses.

Experimentally, treatment with TGF $\alpha$  alone does not lead to secretion of IL-1 $\alpha$ . Instead, a co-stimulation with TGF $\alpha$  and TNF $\alpha$  is required [7]. However, in the first version of the model, treatment with TGF $\alpha$  was sufficient for IL-1 $\alpha$  expression (Fig. ??). Given the time delay until secretion of IL-1 $\alpha$ , it can be expected that *de novo* synthesis of IL-1 $\alpha$  is required and that both TNF $\alpha$  and TGF $\alpha$  are needed to activate transcription of the IL-1 $\alpha$  gene. JNK1 and ERK signal downstream of TNF $\alpha$  and TGF $\alpha$ , respectively, and are known to affect the activity of multiple transcription factors. We altered the model to make activation of IL-1 $\alpha$  expression dependent on both JNK1 activity and ERK activity (Suppl. Fig. S8, arrows linking JNK1 and ERK to IL-1 $\alpha$  gene). After this modification to the model, IL-1 $\alpha$  was no longer secreted upon stimulation with TGF $\alpha$  alone, which greatly improved the fit between the measured IL-1 $\alpha$  levels and the model (Fig. ??). This hypothesis could now be used to design a new experiment to validate IL-1 $\alpha$  as a target of combined JNK1 activity and ERK activity in HT-29 cells. For example, kinase inhibitors specific to JNK1 and ERK could be used to confirm that activity of both kinases is required for expression and secretion of IL-1 $\alpha$ . Performing the experiment is beyond the scope of this study, but this hypothesis finds support in literature. Transcription factors c-Jun and c-Fos together form a heterodimer known as AP-1 and are activated by JNK1 and ERK, respectively [8, 9]. AP-1 has been reported to bind to the promoter of IL-1 $\alpha$ , providing evidence for a role in the regulation of IL-1 $\alpha$  expression [10]. Based on these findings in literature we included c-Jun and c-Fos in our model as transcriptional activators of IL-1 $\alpha$  (Fig.5).

As a second example, we considered the behaviour of JNK1 and MK2. In the model, both proteins were located downstream of TNF $\alpha$  but not TGF $\alpha$  or EGF. Hence, the model did not show an effect of C225, a pharmacological inhibitor of ligand-EGFR binding, on activation of JNK1 or MK2 after stimulation with TNF $\alpha$ . However, experimental data show that C225 strongly reduces activation of JNK1 and MK2 upon stimulation with TNF $\alpha$  [7]. This fact is indicative of a role for EGFR in activation of JNK1 and MK2. Since both JNK1 and MK2 are located downstream of MEKK1, we hypothesized that activation of MEKK1 is dependent on both TNF $\alpha$ -signalling and TGF $\alpha$ -signalling. In the model we added a new hypothetical node Hyp 2 (hypothesis 2) to link EGFR to MEKK1 (Suppl. Fig. S8). This addition led to an improved fit of the model to the data upon treatment with TNF $\alpha$  + C225: activation of both MK2 and JNK1 was strongly suppressed by C225 (Fig. ??). Stimulation with EGF alone did not lead to activation of JNK1 and MK2. These data support the validity of the modification to the model. Further support for a link between EGFR and MEKK1 was found in literature. Specifically, Ras has been reported as a direct activator of MEKK1 [11]. EGFR is a well-known and potent activator of Ras, which is why it was already in our network [4]. Other studies also report activation of JNK1 and phosphorylation of c-Jun downstream of Ras, which is consistent with an interaction between Ras and MEKK1 [9, 12]. Based on these findings, we adapted our model by removing the Hyp 2 node and creating a direct interaction between Ras and MEKK1 (Fig. 5). Experimentally, the role of Ras could be confirmed by using a pharmacological inhibitor of Ras activity, and measuring the effect of this inhibitor on the activation of JNK1 and MK2. Together, our model suggests that EGFR activity is required but not sufficient for activation of JNK1 and MK2 in HT-29 cells.

There are other nodes for which the experimental data deviates from the model in one or more of the experimental conditions. A comparison between model and experimental data can be found in Figures S9, S10 and S11. A complete deciphering of the signalling events in this biological system is outside the scope of this paper. Instead, we illustrated how interactive modelling of the dynamic behaviour of a signal transduction network can be used to extend previous pathway topologies and can lead to the generation of novel hypotheses.

## Discussion

### Final remarks on the models

In the Results section, we described the construction of an ANIMO model of the circadian clock in *Drosophila Melanogaster*. This model captured the dynamics of the regulatory network and led to similar conclusions as an ODE model that had been published previously [3]. This finding supports the use of the series of modeling abstractions on which ANIMO is based. The biggest difference between the construction of these models is that the model in [3] is constructed by writing a system of mathematical equations, together with an algorithm for simulation. In ANIMO, instead, a number of network nodes is drawn for the molecules involved. These nodes are then linked by directed interactions that represent cause-and-effect relationships, with a single parameter that defines the strength of each interaction. This is a more intuitive approach to construct a model. Further contributing to an interactive modeling process is the compositionality of the model. Each node in the network can be disabled at any time by the user, or extra nodes can be added, without having to change any of the existing interactions.

In the Results section we also showed the construction of an executable model of signaling events downstream of TNF $\alpha$  and EGF in human colon carcinoma cells. This data set has been used for previous modeling studies, based on partial least-squares regression and fuzzy logic [13, 14]. The partial least-squares regression model describes an abstract data-driven model that uses statistical correlations to relate signal transduction events to various cellular decisions. This type of modeling is very useful in uncovering new and unexpected relations. It is also successful in making predictions, but gives little direct insight in the dynamic behavior of the network. Fuzzy logic analysis led to a model that gave a better fit to the dynamic network behavior than discrete logic (Boolean) models. Inspection of the inputs to the logical gates that were used to model protein behavior led to the prediction of novel interactions between proteins, showing the usefulness of this approach. For most of the proteins, such as JNK1, time was used as an input parameter. For example, the logical gates “if TNF $\alpha$  is high AND time is low, then JNK1 is high” and “if TNF $\alpha$  is high AND time is high, then JNK1 is low” were used to describe the dynamic behavior of JNK1. Although this leads to a representative description of the dynamic behavior of JNK1, peaks in protein activity at early time points were not reproduced by the fuzzy logic model. Moreover, it gives no insight in the molecular interactions that are involved in activation or inhibition.

In this study, we used a data set based on the experiments described in [6]. We used the resulting experimental data, together with knowledge from curated databases [4, 5] to construct an executable model of the biological system. In contrast with the two approaches described above, ANIMO is aimed at the construction

of more mechanistic models, mimicking biochemical interactions *in silico*. This way of modeling gives a different type of insight. In the process of model construction, we extended a prior-knowledge network with time-dependent extracellular crosstalk that has been reported previously [7]. To come up with possible explanations for a disagreement between the model and the experimental data, two additional layers of crosstalk were introduced, at the signal transduction and transcriptional level. These modifications improved the fit of the model to the data and can be interpreted as novel testable hypotheses. Finally, we proposed new experiments that could be carried out to test these hypotheses, closing the empirical cycle. Together, our model sheds more light on the intricate entanglement between the TNF $\alpha$  and EGF pathways at multiple cellular levels. But above all, the model provides an excellent starting point for further investigation.

#### Comparison between ANIMO and other modeling tools

Different formalisms are in use in the field of computational modeling of biological systems, each with their specific characteristics. Many of these formalisms have been implemented into software tools to support modeling efforts. In order to compare ANIMO with existing tools, we have selected a number of mathematical formalisms, each connected to a supporting tool. With an emphasis on the modeling process rather than the final model, we compared these tools on the basis of the following parameters:

- 1 **Hidden formalism:** a knowledge of the underlying formalism is not required in order to use the tool
- 2 **Visual modeling:** the tool allows the user to model using a visual interface, and is not exclusively founded on formula-, text- or table-based input forms
- 3 **Qualitative parameters:** parameters for reactions can be input as approximated estimations, and not exclusively as numbers
- 4 **Tight coupling with topology:** models are tightly and clearly coupled to the networks they represent, showing the visual representation of the model in a shape similar or comparable to the representation currently used by biologists for signaling pathways
- 5 **User-chosen granularity:** if discretization is applied during the modeling process, the user can change the granularity with which such discretization is made, possibly for each component of the model separately

Table 1 shows the comparison between ANIMO and the selected tools.

Going beyond the user interface, there are a number of “pros and cons” for using ANIMO and Timed Automata in the biological context. First and foremost, being Timed Automata an executable formal language, a state space can be derived from a Timed Automata model. This means that state space-related analyses such as model checking can be performed on Timed Automata: as explained in the Results section, this can be done directly in ANIMO, bypassing the need for a deeper understanding of Timed Automata as ANIMO acts as an intermediary towards the powerful model checking tool UPPAAL.

The smaller amount of parameters needed to build an ANIMO model, if compared e.g. to a differential equation-based approach, makes the modeling process faster, allowing the biologist to obtain interesting results in less time. However, a

less parameter-intensive modeling approach is also less precise: in our case we abstract sequences of biochemical reactions into activation or inactivation interactions, simplifying any relation between two entities in a network as either activating or inhibiting. While this allows for some simplifications in the networks and the underlying formal models, it does not allow to represent in full detail more specific concepts such as reactant concentration or reaction affinity. In case such concepts are of primary importance, we encourage the interested reader to turn to more precise modeling approaches, such as those based on differential equations.

All reactants modeled in ANIMO are assumed to be present at the same (unitary) concentration. While this assumption is not always applicable, it encourages abstract thought: many biological processes can be represented as networks driven by *activity-based* interactions. Even if with a limited scope, ANIMO can be applied also in the analysis of metabolic processes, considering the production of protein *A* as the activation of node *A*, with degradation being represented as inhibition. As stated before, it cannot be expected from such models to be a completely realistic representation of their target biological processes, but they can still be an useful tool. This can be seen for example in the *Drosophila* model in the Results section, where mRNA and protein concentrations were abstracted to activity-driven processes in the ANIMO model.

## Conclusions

ANIMO is not the first modeling tool to provide an interface to a modeling formalism. Such interfaces exist in many other tools (as shown in Tab. 1). With its focus on user-friendliness and intuitive modeling, ANIMO's main contribution lies in making computational modeling more accessible to experts in biology. Making use of the visual interface provided by Cytoscape, network representations subscribe to biological conventions. Model parameters are kept to a minimum and can be directly accessed by mouse-clicking on nodes and edges. Because of the automatic translation of the network topology and user-defined parameters into an underlying formal model, training in the use of formal methods is not needed.

## Future plans

In order to enforce the concept of user interaction as a primary focus of the tool, we plan to extend ANIMO by adding support for parameter sensitivity analysis and automatic parameter sweeps. We already allow for only one numeric parameter per reaction, but an automatic way of adjusting such parameters to fit experimental data will allow the user to concentrate more on defining a sensible topology for a pathway. Such an automatic search could be limited by setting specific bounds on reaction rates based on previous knowledge, or on a trade-off with performances. Moreover, inspired by works on automata learning [15], we plan to add also the possibility to automatically derive a network topology based on experimental data and user-defined restrictions based on previous knowledge.

We aim at widening the available set of model checking queries, in order to allow biologists to perform in silico experiments on an already fitting model and to obtain answers to more relevant questions. This would increase the usefulness of a model as a help to drive wet-lab investigation. In order to allow for meaningful in silico

experiments, we plan to purposefully introduce user-defined non-deterministic parts in our models, which would allow for drug dosage investigations through model checking. This can be done e.g. through the definition of intervals for the values of some reaction kinetic constants, adding considerable uncertainty in the timing of those reactions. A model of this type could then be interrogated with queries like “How much X, Y and Z do we need to provide, and at which time points, in order to obtain targets A, B and C to be active?”. More useful results can be achieved through proper application of the statistical model checking part of UPPAAL [16] to an extended version of our model to include realistically significant stochastic behavior. Finally, in order to further improve performances, the applicability a multi-core model checking approach based on the works by Dalsgaard et al. [17] is under study.

## Methods

### Modeling interactions with Timed Automata

Timed Automata have been shown to be a powerful formalism to model biological processes [18, 19, 20]. A timed automaton consists of locations and transitions between these locations (see Fig. 6), and a system of timed automata can be used to model a system of interacting molecules. At any time, each automaton is in a specific location, and together these locations represent the current state of the biological system. Each timed automaton can have one or more local clocks associated to it, allowing temporal control of transitions between locations. These transitions are used to represent interactions between molecules. Fast interactions take less time than slow interactions to perform an activation or inhibition step. We have previously described in detail how the scenarios presented in Section C.1 can be used to calculate the timing of molecular interactions to give a description of network dynamics [1]. Figure 6 presents a small example that illustrates the basic properties of TA. This model describes the activation of ERK by MEK<sup>[1]</sup>.

### Example: building a model based on data

To illustrate the use of ANIMO in a practical environment, we will demonstrate the generation of a basic version of the model described in Section . The model is based on a literature compendium of signal transduction events in HT-29 human colon carcinoma cells [6]. This data set comprises triplicate measurements of 11 different protein activities or post-translational modification states at 13 time points after treatment with different combinations of tumor necrosis factor- $\alpha$  (TNF $\alpha$ ), epidermal growth factor (EGF) and insulin. The data set contains relative protein levels and activities, and no absolute quantities, which is the typical situation in biochemistry. To start, we normalized measurements for each protein to the maximum value in the complete experiment, resulting in a nondimensionalized data set that is suitable for use with ANIMO.

In Figure 7, we show the stepwise construction of a model of a small part of the network that is able to account for measured variations in activity of inhibitor of nuclear factor kappa-B kinase (IKK), c-Jun N-terminal kinase-1 (JNK1), mitogen-activated protein kinase-activated protein kinase 2 (MK2), Caspase 8 (casp-8) and

---

<sup>[1]</sup>All acronyms used in this paper and their corresponding UniProt IDs are listed in Suppl. Sect. A.

Caspase 3 (casp-3) upon stimulation with 100 ng/ml TNF $\alpha$ . In this example we aimed for inclusion of a minimum number of nodes in the network, while preserving biological relationships. Multistep cascades were aggregated into a single step when possible. Parameters for all reactions were set manually, resulting in a close match between the model and the patterns observed in the dataset.

#### Competing interests

The authors declare that they have no competing interests.

#### Author's contributions

S.S. designed and performed the experiments, developed the Cytoscape integration, wrote the manuscript; J.S. conceived, designed and performed the experiments, wrote the manuscript; P.E.vdV. initiated the study, conceived the Cytoscape implementation, supervised the project; M.K. designed experiments, analyzed data and wrote the manuscript; J.N.P. designed experiments, analyzed data sets, contributed in particular to the application of ANIMO for large biological data, wrote the manuscript, and supervised the project; J.vdP. contributed to the strategy and methodology in the manuscript, in particular the connection with formal methods; R.L. Contributed to the methodology and supervised the project.

#### Acknowledgements

We would like to thank Christof Francke for valuable discussions.

#### Author details

<sup>1</sup>Formal Methods and Tools, Faculty of EEMCS, University of Twente, 7522NH Enschede, The Netherlands.

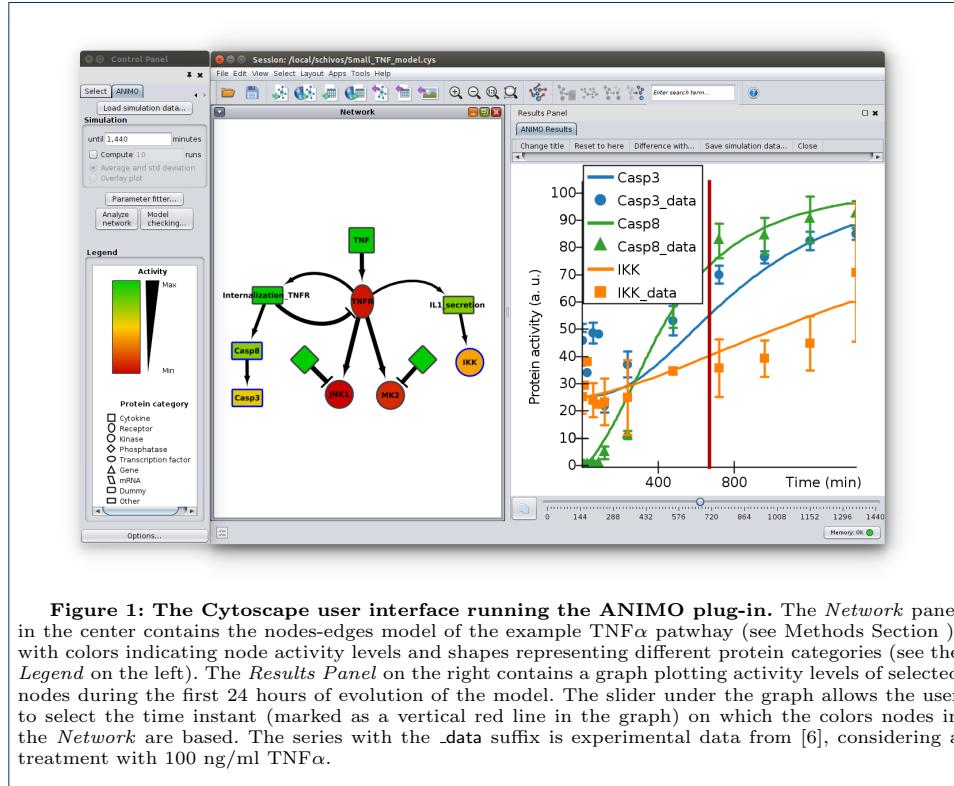
<sup>2</sup>Developmental BioEngineering, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, 7522NH Enschede, The Netherlands. <sup>3</sup>Human Media Interaction, Faculty of EEMCS, University of Twente, 7522NH Enschede, The Netherlands.

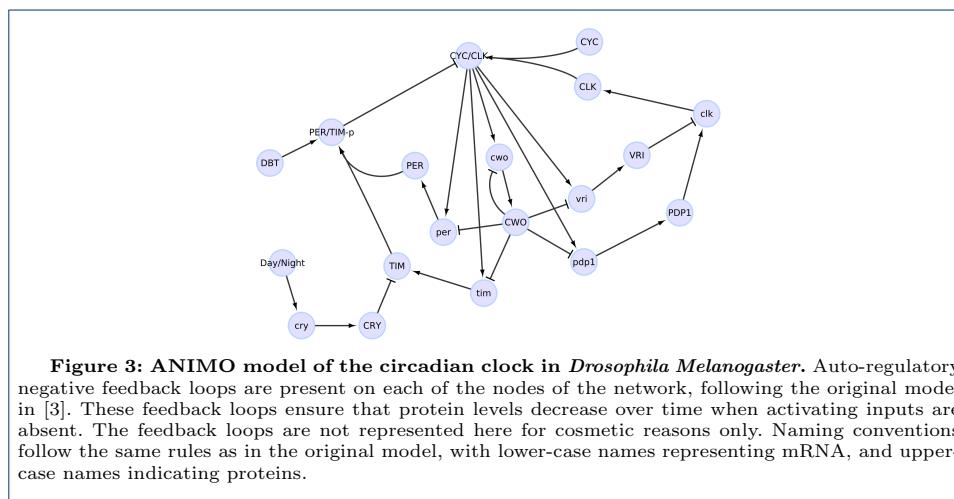
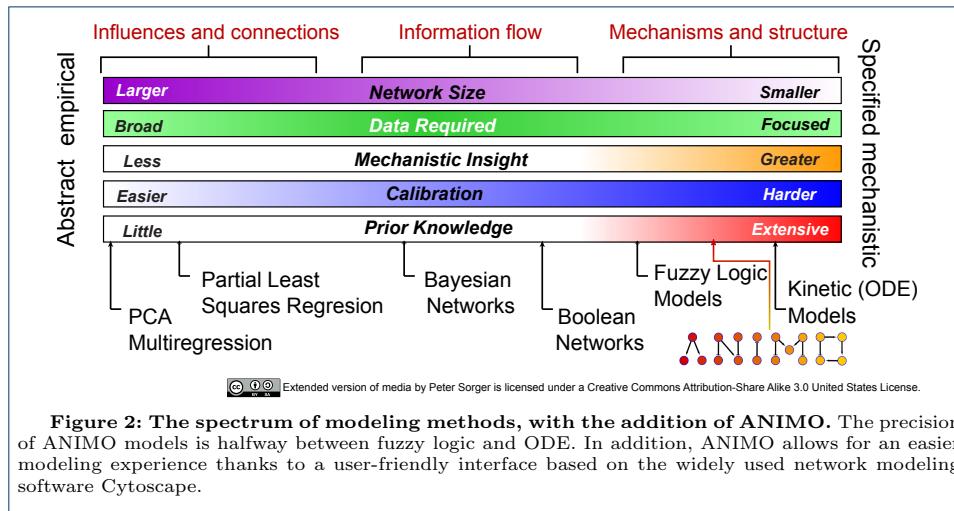
#### References

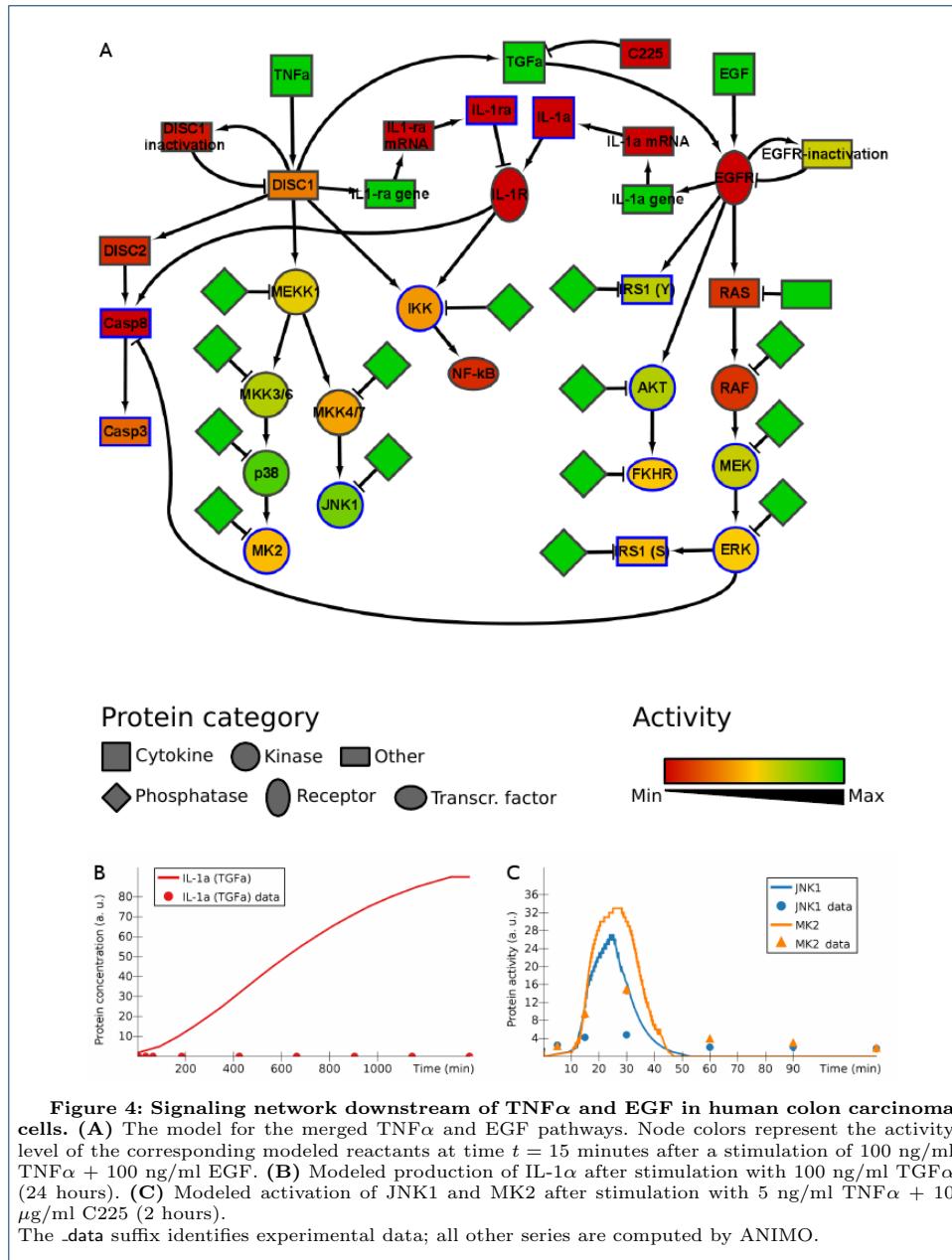
1. Schivo, S., Scholma, J., Wanders, B., Camacho, R.A.U., van der Vet, P.E., Karperien, M., Langerak, R., van de Pol, J., Post, J.N.: Modelling biological pathway dynamics with Timed Automata. *IEEE Journal of Biomedical and Health Informatics* **In press** (2013)
2. Scholma, J., Schivo, S., Camacho, R.A.U., van de Pol, J., Karperien, M., Post, J.N.: Biological networks 101: Computational modeling for molecular biologists. *Gene* **533**(1), 379–384 (2014)
3. Fathallah-Shaykh, H.M., Bona, J.L., Kadener, S.: Mathematical model of the drosophila circadian clock: Loop regulation and transcriptional integration. *Biophysical Journal* **97**(9), 2399–2408 (2009)
4. Kanehisa, M., Goto, S.: KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**(1), 27–30 (2000)
5. Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E., Zhang, B.: PhosphoSite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**(6), 1551–1561 (2004)
6. Gaudet, S., Janes, K.A., Albeck, J.G., Pace, E.A., Lauffenburger, D.A., Sorger, P.K.: A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol Cell Proteomics* **4**(10), 1569–1590 (2005)
7. Janes, K.A., Gaudet, S., Albeck, J.G., Nielsen, U.B., Lauffenburger, D.A., Sorger, P.K.: The response of human epithelial cells to TNF involves an inducible autocrine cascade. *Cell* **124**(6), 1225–1239 (2006)
8. Davis, R.J.: Signal transduction by the JNK group of MAP kinases. *Cell* **103**(2), 239–252 (2000)
9. Bannister, A.J., Brown, H.J., Sutherland, J.A., Kouzarides, T.: Phosphorylation of the c-Fos and c-Jun HOBD motif stimulates its activation capacity. *Nucleic Acids Res* **22**(24), 5173–6 (1994)
10. Bailly, S., Fay, M., Israël, N., Gougerot-Pocidalo, M.A.: The transcription factor AP-1 binds to the human interleukin 1 alpha promoter. *European Cytokine Network* **7**(2), 125–128 (1996)
11. Russell, M., Lange-Carter, C.A., Johnson, G.L.: Direct interaction between Ras and the kinase domain of mitogen-activated protein kinase kinase kinase (MEKK1). *Journal of Biological Chemistry* **270**(20), 11757–11760 (1995)
12. Dérijard, B., Hibi, M., Wu, I.-H., Barrett, T., Su, B., Deng, T., Karin, M., Davis, R.J.: JNK1: A protein kinase stimulated by uv light and ha-ras that binds and phosphorylates the c-Jun activation domain. *Cell* **76**(6), 1025–1037 (1994)
13. Janes, K.A., Albeck, J.G., Gaudet, S., Sorger, P.K., Lauffenburger, D.A., Yaffe, M.B.: A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* **310**(5754), 1646–1653 (2005)
14. Aldridge, B.B., Saez-Rodriguez, J., Muhlich, J.L., Sorger, P.K., Lauffenburger, D.A.: Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/Insulin-induced signaling. *PLoS Comput Biol* **5**(4), 1000340 (2009)
15. Tretmans, J.: Model-based testing and some steps towards test-based modelling. In: Bernardo, M., Issarny, V. (eds.) *Formal Methods for Eternal Networked Software Systems*. Lecture Notes in Computer Science, vol. 6659, pp. 297–326. Springer, Berlin / Heidelberg (2011)
16. David, A., Larsen, K., Legay, A., Mikucionis, M., Wang, Z.: Time for statistical model checking of real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) *Computer Aided Verification*. Lecture Notes in Computer Science, vol. 6806, pp. 349–355. Springer, Berlin / Heidelberg (2011)
17. Dalsgaard, A.E., Laarman, A.W., Larsen, K.G., Olesen, M.C., van de Pol, J.C.: Multi-core reachability for timed automata. In: Jurdzinski, M., Nickovic, D. (eds.) *10th International Conference on Formal Modeling and Analysis of Timed Systems, FORMATS 2012*, London, UK. Lecture Notes in Computer Science, vol. 7595, pp. 91–106. Springer, London (2012)

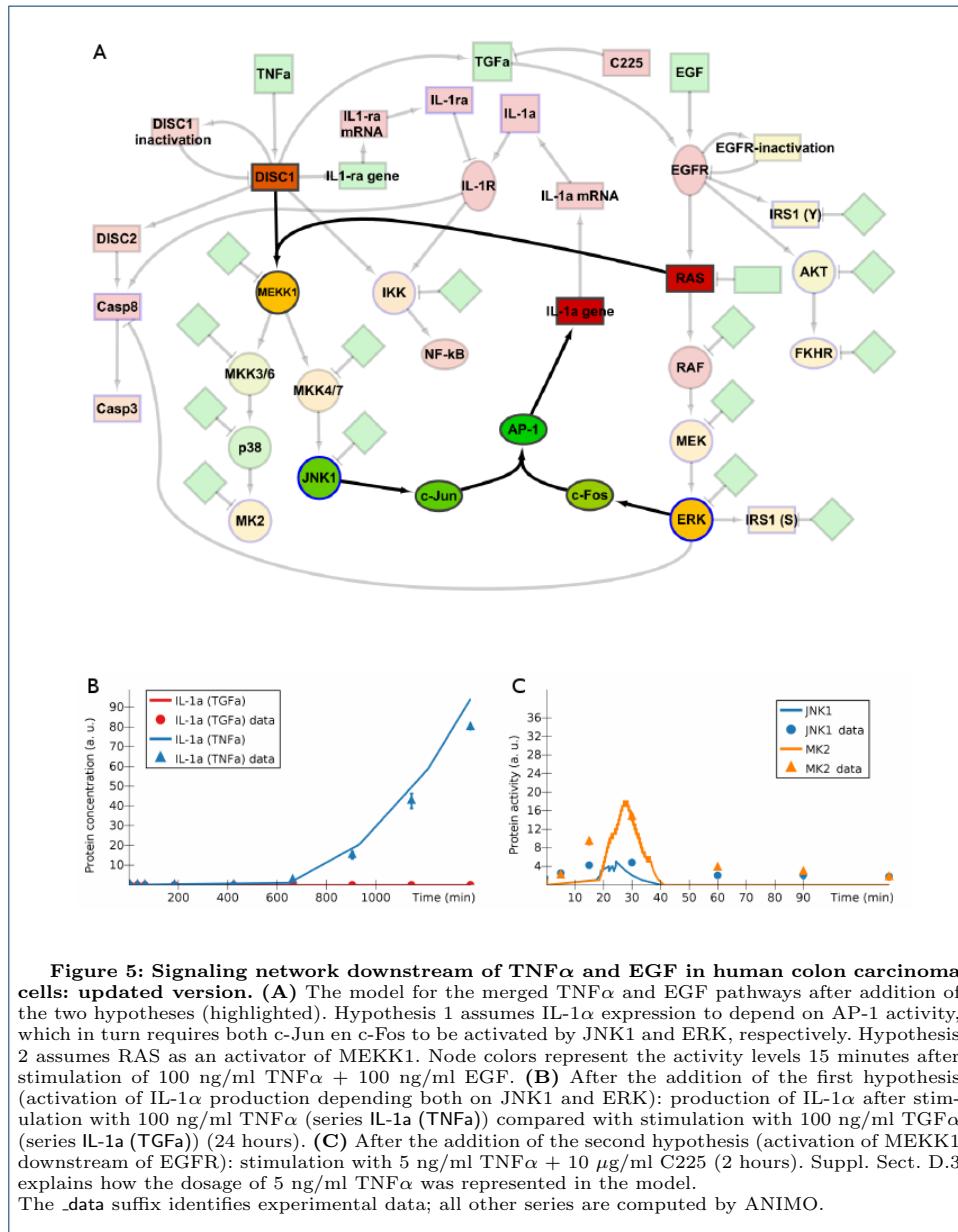
18. Siebert, H., Bockmayr, A.: Temporal constraints in the logical analysis of regulatory networks. *Theor. Comput. Sci.* **391**(3), 258–275 (2008)
19. Bartocci, E., Corradini, F., Merelli, E., Tesei, L.: Model checking biological oscillators. *Electronic Notes in Theoretical Computer Science* **229**(1), 41–58 (2009). Proceedings of the Second Workshop From Biology to Concurrency and Back (FBTC 2008)
20. Batt, G., Salah, R.B., Maler, O.: On timed models of gene networks. In: Proceedings of the 5th International Conference on Formal Modeling and Analysis of Timed Systems. FORMATS'07, pp. 38–52. Springer, Berlin, Heidelberg (2007)
21. Schivo, S., Scholma, J., Wanders, B., Urquidi Camacho, R.A., van der Vet, P.E., Karperien, M., Langerak, R., van de Pol, J., Post, J.N.: Modelling biological pathway dynamics with Timed Automata. In: 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), pp. 447–453. IEEE Computer Society, Los Alamitos, CA, USA (2012)
22. Ciocchetta, F., Duguid, A., Gilmore, S., Guerrero, M.L., Hillston, J.: The Bio-PEPA Tool Suite. International Conference on Quantitative Evaluation of Systems, 309–310 (2009)
23. Nagasaki, M., Saito, A., Jeong, E., Li, C., Kojima, K., Ikeda, E., Miyano, S.: Cell illustrator 4.0: a computational platform for systems biology. *Stud Health Technol Inform* **162**, 160–81 (2011)
24. Mendes, P., Hoops, S., Sahle, S., Gauges, R., Dada, J., Kummer, U.: Computational modeling of biochemical networks using COPASI systems biology. *Methods in Molecular Biology*, vol. 500, pp. 17–59. Humana Press, Totowa, NJ (2009). Chap. 2
25. Chaouiya, C., Remy, E., Mossé, B., Thieffry, D.: Qualitative analysis of regulatory graphs: A computational tool based on a discrete formal framework. In: Benvenuti, L., De Santis, A., Farina, L. (eds.) *Positive Systems. Lecture Notes in Control and Information Sciences*, vol. 294, pp. 830–832. Springer, Berlin / Heidelberg (2003)
26. de Jong, H., Geiselmann, J., Hernandez, C., Page, M.: Genetic Network Analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics* **19**(3), 336–344 (2003)
27. Killcoyne, S., Carter, G.W., Smith, J., Boyle, J.: Cytoscape: a community-based framework for network modeling. *Methods in molecular biology (Clifton, N.J.)* **563**, 219–239 (2009)
28. Larsen, K.G., Pettersson, P., Yi, W.: UPPAAL in a nutshell. *International Journal on Software Tools for Technology Transfer (STTT)* **1**, 134–152 (1997)
29. Elowitz, S. Michael B.Leibler: A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(6767), 335 (2000)
30. Kadener, S., Stoleru, D., McDonald, M., Nawathean, P., Rosbash, M.: Clockwork orange is a transcriptional repressor and a new drosophila circadian pacemaker component. *Genes & Development* **21**(13), 1675–1686 (2007)

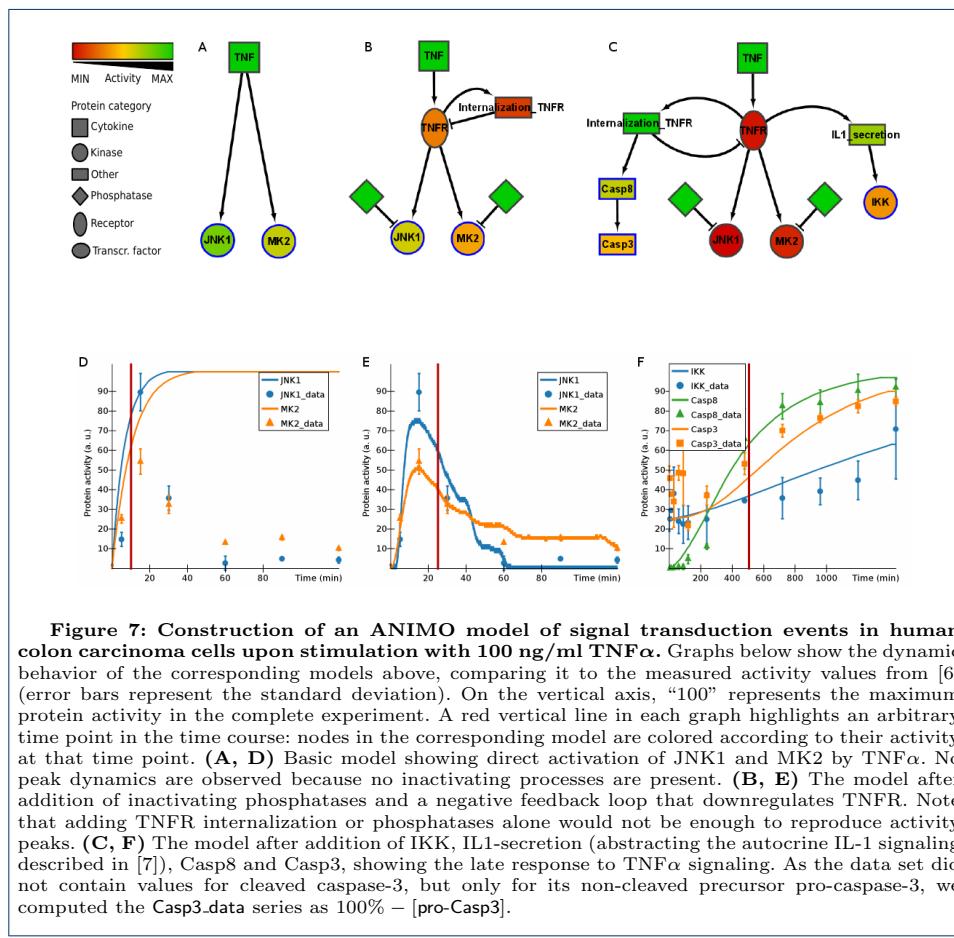
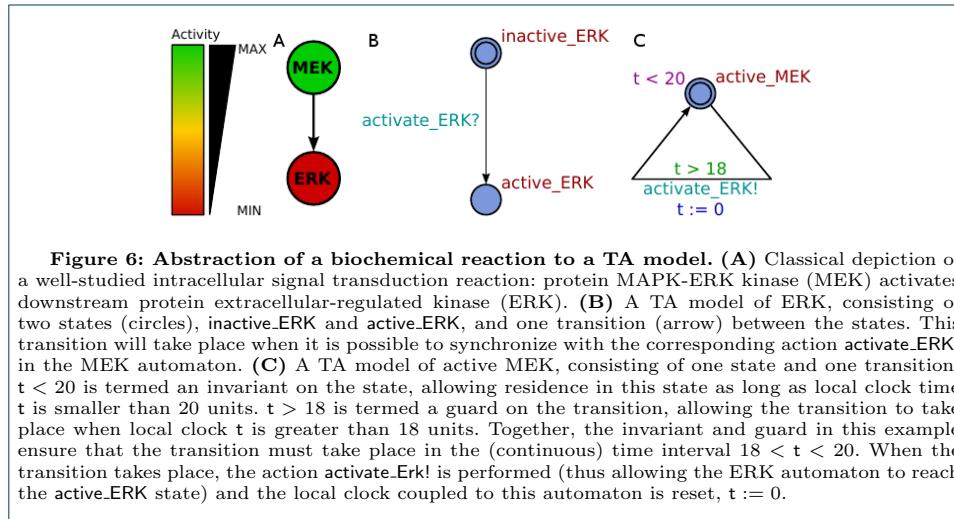
## Figures











## Tables

Tool	Formalism	Hidden formalism	Visual modeling	Qualitative parameters	Tight coupling with topology	User-chosen granularity
ANIMO [21]	Timed Automata	Yes	Yes	Yes	Yes	Yes
Bio-PEPA Workbench [22]	Bio-PEPA	No	No	No	No	Yes
Cell Illustrator [23]	Petri Nets	Yes	Yes	No	Yes	No
COPASI [24]	ODE, stochastic models	No	No	No	No	No
COSBI LAB <sup>1</sup>	BlenX	Yes	Yes	No	Yes	No
GINsim [25]	Boolean Networks	No	Yes	Yes	Yes	Yes <sup>2</sup>
GNA [26] <sup>3</sup> Rhapsody <sup>3</sup>	PLDE Statecharts	No No	Yes Yes	Yes Yes	Yes No <sup>4</sup>	Yes <sup>2</sup> No

**Table 1:** Comparison between ANIMO and some existing approaches to modeling biological systems. A “Yes” under a column indicates that the modeling tool (mostly) fulfills the parameter, “No” indicates very limited or no fulfillment.

<sup>1</sup> COSBILab web page <http://www.cosbi.eu/index.php/research/cosbi-lab>

<sup>2</sup> The user can choose the number of levels for each reactant, allowing to define multi-level models based on Boolean reaction dynamics.

<sup>3</sup> IBM Rational Rhapsody web page <http://www-01.ibm.com/software/rational/products/rhapsody/designer>

<sup>4</sup> Statecharts represent more closely the so-called *transition system* of the model as opposed to the components and interactions occurring among them.

**Additional Files**

Additional file 1 — ANIMO models and data for the case studies

The provided .zip file contains the Cytoscape .cys network files which can be analyzed with ANIMO. In addition, the reference values in .csv format are provided, to compare ANIMO's results with the previously existing models.

## Additional Materials

## **Contents**

<b>Abstract</b>	<b>1</b>
<b>Contents</b>	<b>2</b>
<b>A Naming conventions</b>	<b>1</b>
<b>B Normalizing experimental data for use with ANIMO</b>	<b>3</b>
<b>C ANIMO and Timed Automata</b>	<b>4</b>
C.1 Modeling abstractions . . . . .	4
C.2 ANIMO. . . . .	5
C.3 Timed Automata model . . . . .	5
C.4 Granularity of an ANIMO network node . . . . .	6
<b>D Additional notes</b>	<b>9</b>
D.1 Simulating the day-night cycle . . . . .	9
D.2 ANIMO model of the Drosophila circadian clock . . . . .	9
D.3 Note on the parameters in the TNF $\alpha$ -EGF model . . . . .	11
<b>E Supplementary figures</b>	<b>12</b>
<b>Supplementary figures</b>	<b>12</b>

## Appendix A: Naming conventions

Table S1 explains the abbreviations used in the paper.

**Table S1:** Explanation of the abbreviated names referring to molecular species in the main text.

Abbreviation	Full name	
Akt	protein kinase B	P31749
AP-1	activator protein 1	heterodimer of c-Jun and c-Fos
Casp3	caspase 3	P42574
Casp8	caspase 8	Q14790
c-Fos	proto-oncogene-protein c-fos	P01100
c-Jun	Jun activation domain binding protein	P05412
CLK	clock	O61735
CRY	cryptochrome	O77059
CWO	clockwork orange	Q9VGZ5
CYC	cycle	O61734
CYC/CLK	cycle-clock complex	
DBT	double-time kinase	O76324
DISC1	death-inducing signaling complex 1	
DISC2	death-inducing signaling complex 2	
EGF	epidermal growth factor	P01133
EGFR	EGF receptor	P00533
ERK	extracellular regulated kinase	P27361
FKHR	forkhead box protein O1	Q12778
IKK	inhibitor of nuclear factor kappa-B kinase	O14920
IL-1a	interleukin 1 $\alpha$	P01583
IL-1R	interleukin 1 receptor	P14778
IL-1ra	interleukin 1 receptor antagonist	P18510
IRS1 (S)	insulin receptor substrate 1 (Serine 636)	P35568
IRS1 (Y)	insulin receptor substrate 1 (Tyrosine 896)	P35568
JNK1	c-Jun N-terminal kinase 1	P45983
MEK	MAPK ERK kinase	Q02750
MEKK1	MAPK/ERK kinase kinase 1	Q13233

**Table S1** - continued from previous page

Abbreviation	Full name	
MK2	mitogen-activated protein kinase-activated protein kinase 2	P49137
MKK3/6	dual specificity mitogen-activated protein kinase kinase 3/6	P46734 / P52564
MKK4/7	dual specificity mitogen-activated protein kinase kinase 4/7	P45985 / O14733
NF-kB	nuclear factor kappa-B	P19838
p38	mitogen-activated protein kinase p38	Q16539
PDP1	par-domain protein 1	Q9TVY0
PER	period	P07663
PER/TIM-p	phosphorylated period-timeless complex	
RAF	Raf	P04049
RAS	Ras GTPase-activating protein	P01112
TGF $\alpha$	transforming growth factor $\alpha$	P01135
TNF $\alpha$	tumor necrosis factor- $\alpha$	P01375
TNFR	TNF receptor	P19438
TIM	timeless	P49021
VRI	vrille	O18660

## **Appendix B: Normalizing experimental data for use with ANIMO**

Document S1 in the Supplemental Data of [7] contains three tables, named Replicates, Averages and DPLSR dataset. The data we use to compare the results computed with ANIMO are based on the values from the Averages table. In particular, we compute activity data by performing a normalization on a 0 ... 100 scale using this formula

$$v_{\text{norm}} = \frac{v}{v_{\text{max}}} \times 100$$

where  $v$  is the datum to be normalized taken from the column  $v$  in the Averages table,  $v_{\text{norm}}$  is the normalized value and  $v_{\text{max}}$  is the maximum value over the whole column  $v$  in the Replicates table. For each series, we compute also the standard deviation using the triplicate measurements present in table Replicates. The standard deviation is also normalized using the formula presented for the average. Each table produced with this process contains a subset of the columns from the Averages table, and refers to one treatment condition only. A column with time references is added to the table in first position. Finally, a column named *Number\_of\_levels* containing only the value 100 (see the instructions on page ??) is added at the rightmost position. The tables are all exported in .csv format to be used with ANIMO, and are included in the *Model\_and\_data.zip* file in the additional materials of the present work.

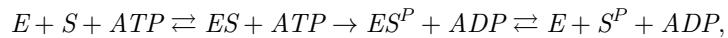
## Appendix C: ANIMO and Timed Automata

### C.1 Modeling abstractions

In living cells, cascades of chemical and physical interactions enable propagation of signals through molecular networks. In this process, the activity of upstream molecules induces a change in the concentration or activity of downstream molecules. For many reactions, the values of the kinetic parameters are unknown or difficult to collect. This lack of knowledge hampers the feasibility of computational models that describe molecular networks in fine mechanistic detail, especially for larger networks. As a solution to this problem, we propose the construction of models at a higher level of abstraction, thereby reducing the number of parameters involved. In choosing a suitable abstraction level, it is important to retain enough descriptive power to give a meaningful formal description of the topology and the associated dynamic behavior of biological networks.

As a first abstraction in ANIMO models, the active and inactive forms of each network component are represented together by a single node in the network. Each of these nodes is characterized by its *activity level*, which represents the fraction of active molecules of that molecular species. When a molecule is known to be constitutively active, changes in concentrations of that molecule are treated as changes in its activity level. Activity levels are discretized into integer variables with a user-defined granularity, ranging from Boolean (2 levels) to near-continuous (100 levels).

Detailed biochemical reaction mechanisms are abstracted to *interactions*, which can represent either activations ( $\rightarrow$ ) or inhibitions ( $\dashv$ ). This aggregation of elementary reactions into single interaction steps reduces the number of kinetic parameters involved, while preserving cause-and-effect relationships. For example, consider a reaction in which enzyme  $E$  phosphorylates and activates substrate  $S$ , transferring a phosphate group from a molecule of ATP to a molecule of  $S$ . Biochemically, this reaction can be represented as



with conservation condition  $S + S^P = \text{constant}$  and  $ATP + ADP = \text{constant}$ .

Under the assumption of ATP constantly being replenished by the cell, this reaction is abstracted in ANIMO to the corresponding interaction



Each occurrence of the interaction  $E \rightarrow S$  will increase the activity level of  $S$  by one discrete step. Since the activity level is defined as the active fraction of a molecular species, an increase in the active fraction implies a decrease in the inactive fraction. Hence, the original conservation condition is automatically satisfied. The interaction rate,  $R$ , depends on the activity levels of the reactants involved and on a single kinetic parameter  $k$  that is set by the user. The three available interaction scenarios can be interpreted as abstracted kinetic rate laws:

- 1  $R = k \times [E]$ : the interaction rate depends only on the activity level of the upstream node.

- 2  $R = k \times [E] \times [1 - S]$  (activations) or  $R = k \times [E] \times [S]$  (inhibitions): the rate depends on the activity levels of both the upstream and downstream participants. Activations depend on the presence of inactive substrate,  $[1 - S]$ , whereas inhibitions depend on the level of active substrate,  $[S]$ .
- 3  $R = k \times [E_1] \times [E_2]$ : this scenario can be used when the activation or inhibition of a downstream node depends on the simultaneous activity of two upstream nodes. This scenario is comparable to an *AND-gate* in Boolean logic.

We will show in Section that the abstraction proposed here preserves ample expressivity to capture the dynamic behavior of a biological network.

## C.2 ANIMO

The modeling approach described in Section C.1 and Section is implemented in the software tool ANIMO (Analysis of Networks with Interactive MOdeling, [1]) as a plug-in to the network visualization tool Cytoscape [27]. The visual interface of Cytoscape makes the construction, expansion and rewiring of a network topology a fast and user-friendly process.

When a new node is added to the network, it has to be initialized with the number of activity levels and its initial activity. For each interaction, a scenario needs to be selected, together with the corresponding kinetic parameter and the interaction type: activation or inhibition. All settings can be readily adapted by double clicking, or via a table of nodes or interactions.

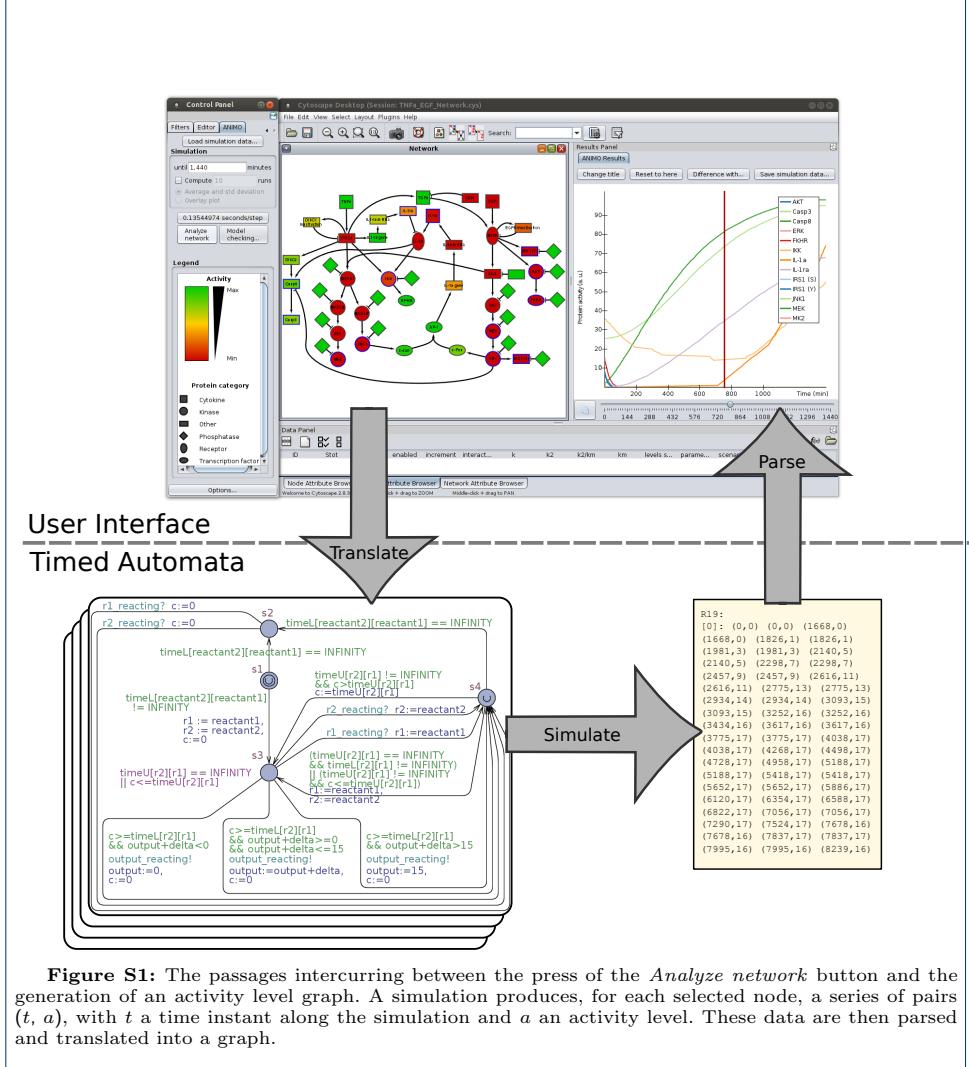
ANIMO automatically translates the user input to a TA model, which is then simulated with the model checking tool UPPAAL [28]. The results are subsequently parsed and translated to a graph that shows the dynamic behavior of nodes in the network. A schematic overview of this process is given in Supplementary Section C. No training or prior knowledge on the use of TA or UPPAAL is needed in order to benefit from ANIMO. Nevertheless, the TA model and the model checking process in UPPAAL can be accessed when desired by the user.

The dynamic behavior of a model can be interactively explored by moving a time slider underneath the graph to highlight time points in a simulation. In the network view, each node will be colored according to its activity level at the selected time point. Experimental data can be compared to the model by importing and superposing these data upon an output graph from the model (Figure 7 B,D,F). The ANIMO user workflow and the features described above are illustrated in Suppl. Video 1.

## C.3 Timed Automata model

The Timed Automata (TA) model underlying an ANIMO network is generated whenever an analysis is requested by the user. Starting from the network represented in the Cytoscape-based user interface, ANIMO automatically generates a TA model to be used with UPPAAL. The analysis result is then parsed and properly presented to the user, for example as a graph of reactant activity levels. This workflow is described in Figure S1.

Each TA model generated by ANIMO contains one automaton for each interaction (activation or inhibition) in the network. A TA representing an interaction performs a cyclic series of steps, continuously updating the target of the interaction



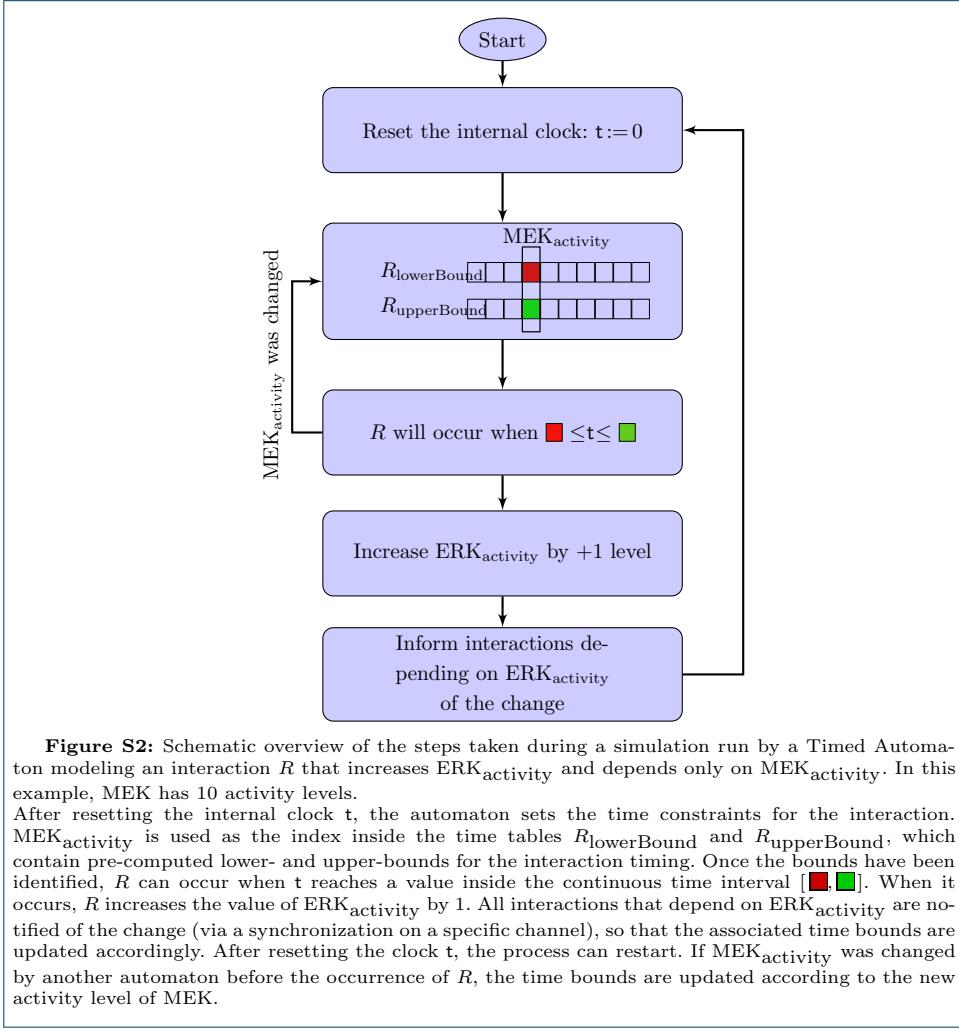
**Figure S1:** The passages intercurring between the press of the *Analyze network* button and the generation of an activity level graph. A simulation produces, for each selected node, a series of pairs  $(t, a)$ , with  $t$  a time instant along the simulation and  $a$  an activity level. These data are then parsed and translated into a graph.

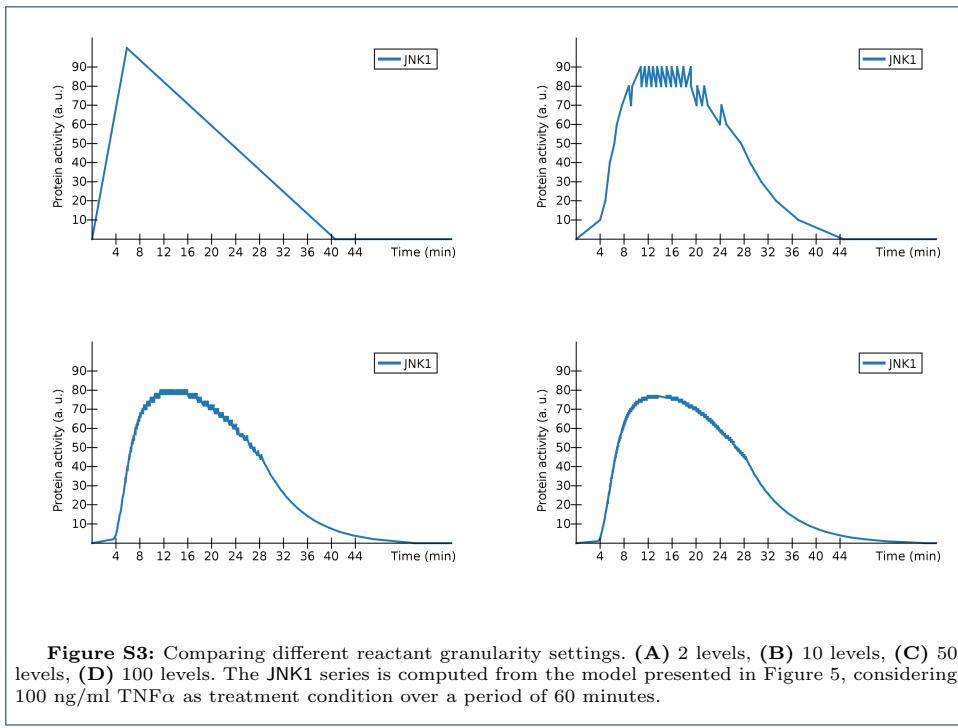
it represents, and adapting the timing of the next update according to the user-defined dynamics. Synchronizations between different automata occur when the activity level of a network component (e.g. ERK) changes: this allows the automata depending on that component to update their time settings.

The abstract behavior of the interaction MEK → ERK in the TA model used in ANIMO is described in Figure S2. There, the activity levels of MEK and ERK are represented by variables called, respectively,  $\text{MEK}_{\text{activity}}$  and  $\text{ERK}_{\text{activity}}$ . A more detailed description of the TA model underlying ANIMO was presented in the IEEE Journal of Biomedical and Health Informatics [1].

#### C.4 Granularity of an ANIMO network node

Figure S3 shows the differences between different choices for the number of levels of a node. This allows to adapt a model to the quality of experimental data.



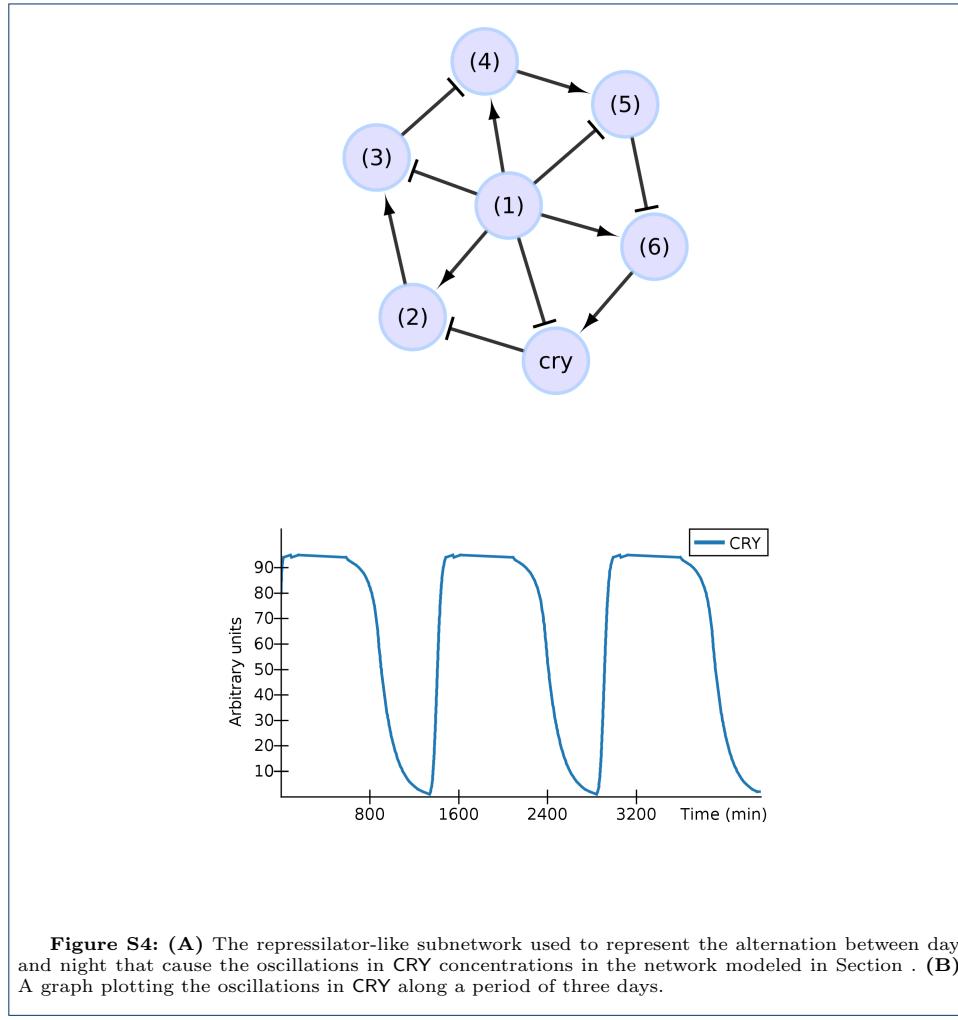


## Appendix D: Additional notes

### D.1 Simulating the day-night cycle

The model presented in Figure 3 contains a node labeled Day/Night. That node abstracts our representation of the cyclic alternation of day and night, which causes the variations in cryptochrome (*cry*): these oscillations allow the network to synchronize to a time zone. Note that the network oscillates also when the node *cry* is not included in the model.

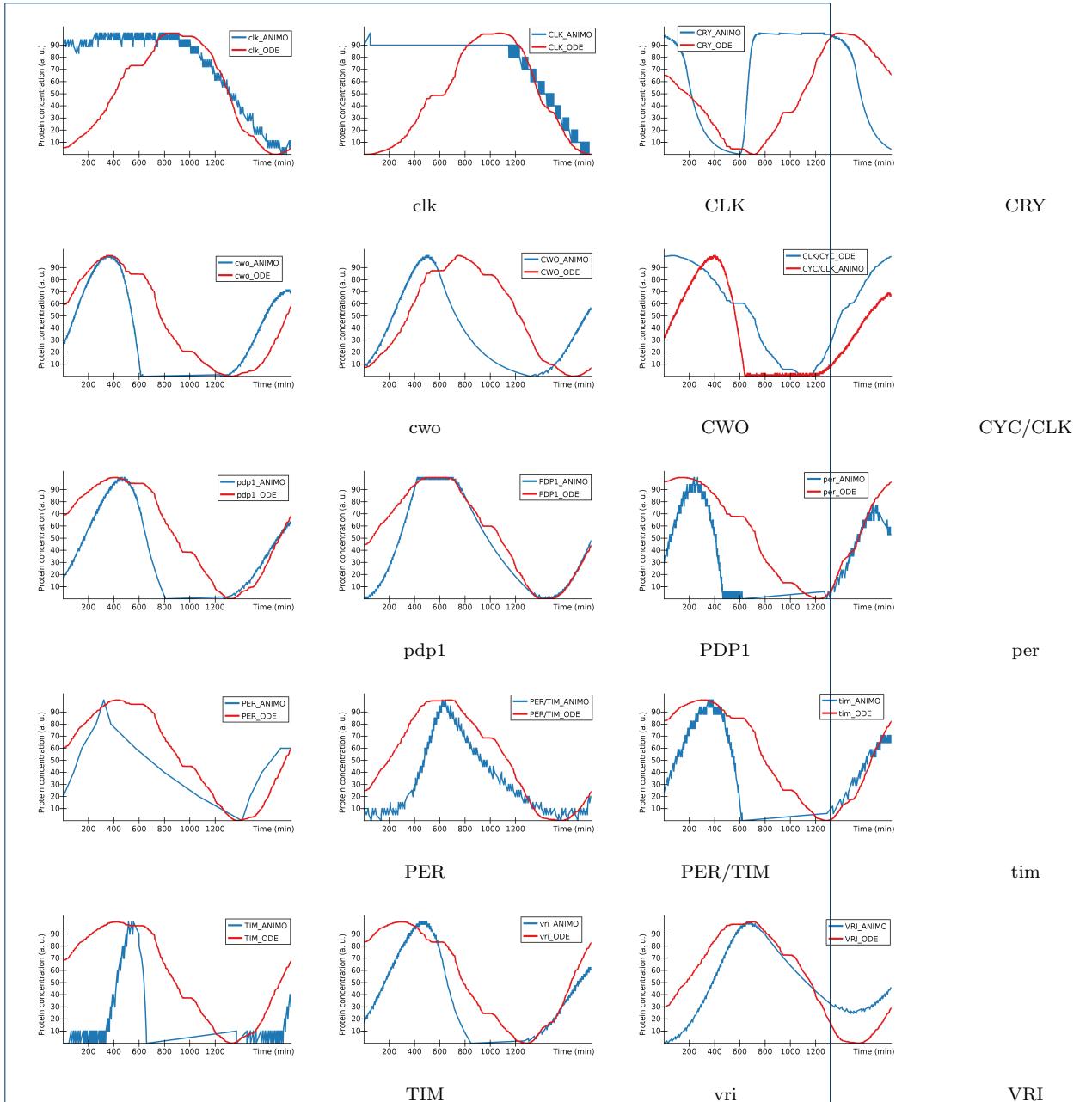
The alternation between day and night is represented in our model with a repressilator-like [29] subnetwork, as can be seen in Figure S4. In the model in [3] a specific function was introduced in the equations to approximate the experimental data from [30].



**Figure S4:** (A) The repressilator-like subnetwork used to represent the alternation between day and night that cause the oscillations in CRY concentrations in the network modeled in Section . (B) A graph plotting the oscillations in CRY along a period of three days.

### D.2 ANIMO model of the Drosophila circadian clock

We compared the simulation results from the ANIMO model presented in Figure 3 with the ODE model described in [3]. The raw data coming from the two models were aligned to have a roughly close initial point, and all amplitudes were normalized following the procedure described in Suppl. Sect. B. The results of this comparison can be seen in Figure S5.



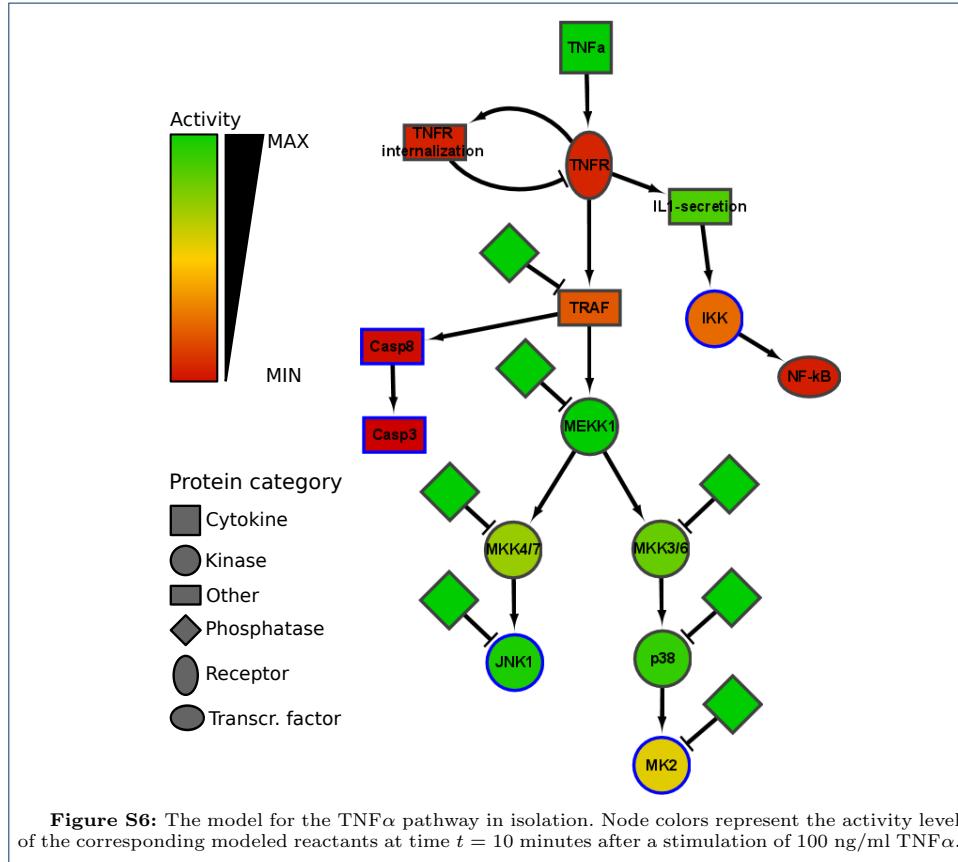
**Figure S5:** Comparing the simulation results from the ANIMO and ODE models of *Drosophila Melanogaster* circadian clock. Blue lines (.ANIMO series) represent data from the ANIMO model in Figure 3, while red lines (.ODE series) represent data from the ODE model in [3].

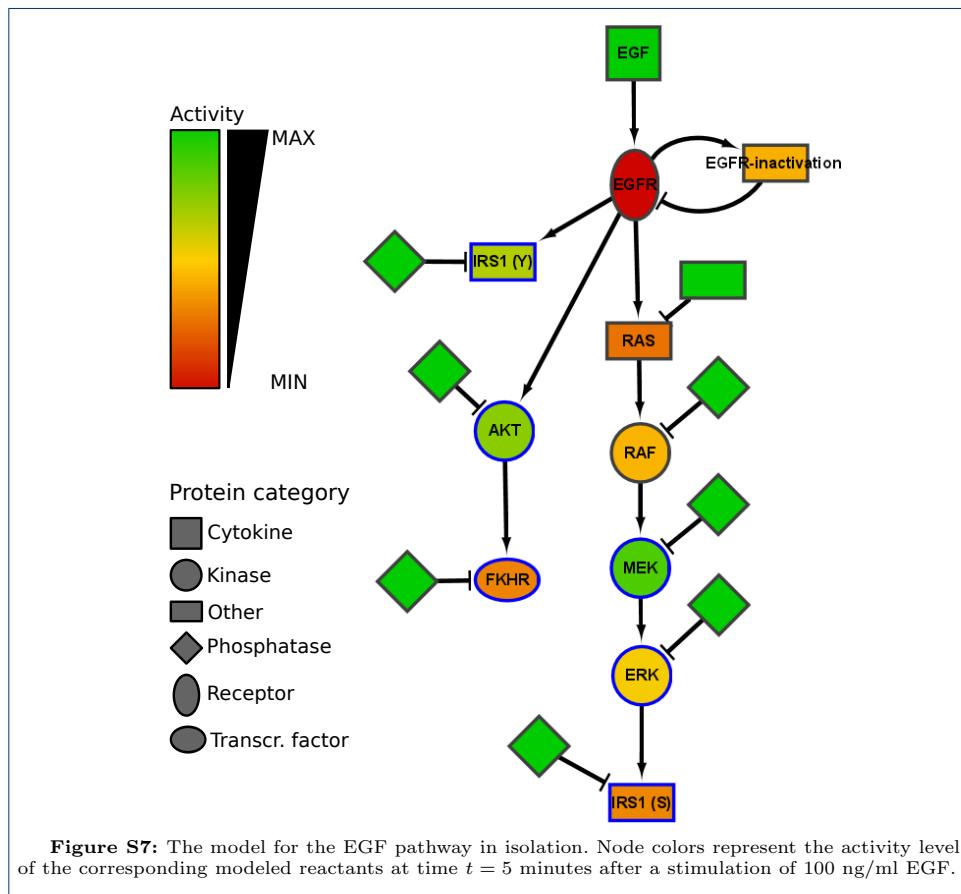
Most of the molecules represented in the two models evolve with the same period and phase. CLK and clk in the ANIMO model have a small oscillation range (their values change by around 10% during a simulation), so their behavior match the continuous model less precisely.

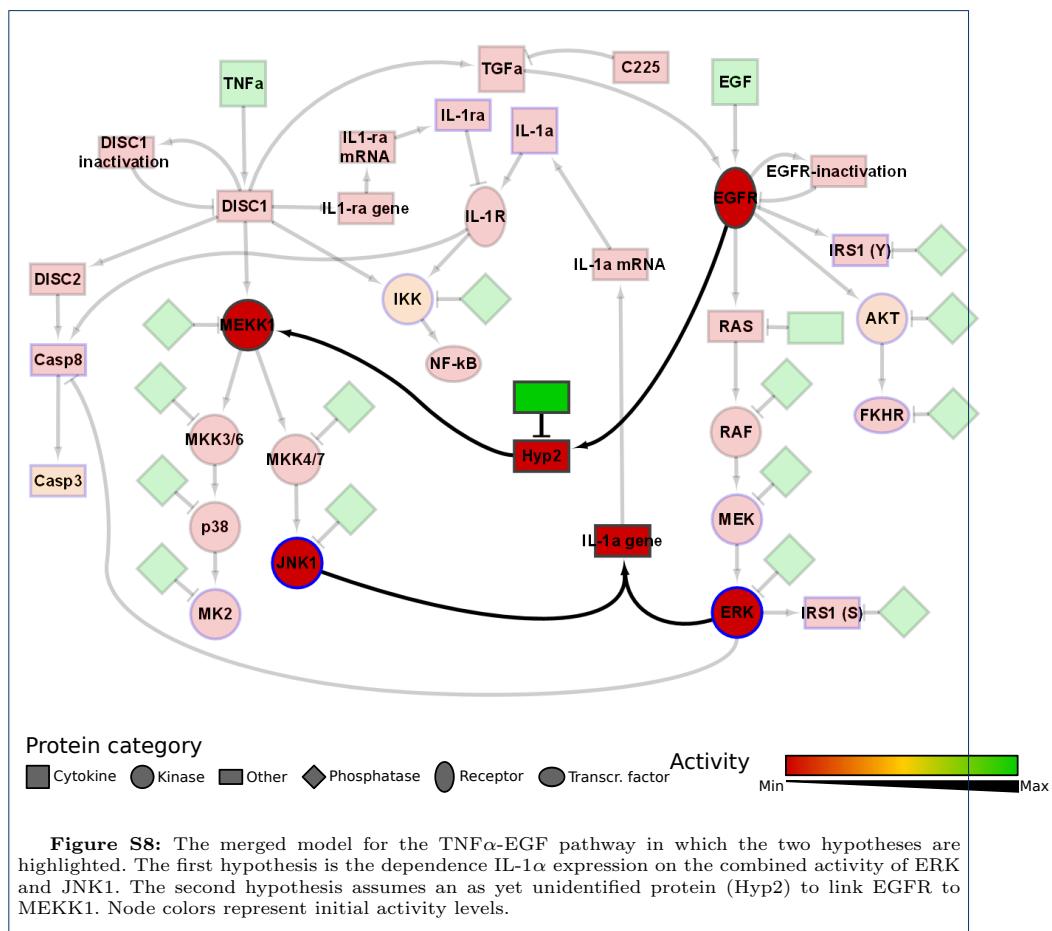
### D.3 Note on the parameters in the TNF $\alpha$ -EGF model

The parameters in the model in Figure 5 have been set by fitting the model to the experimental data for conditions with 100 ng/ml TNF $\alpha$ . In the model we have set the starting level of TNF $\alpha$  at 100 out of 100 for these conditions. This level is a dimensionless quantity that indicates the maximum activity level in the data set. We found that setting the initial level of TNF $\alpha$  at level 8 out of 100 gave slightly better results for the condition with 5 ng/ml TNF $\alpha$  than level 5 out of 100. We believe that this has to do with the fact that 100 ng/ml is a highly supra-physiological concentration of TNF $\alpha$ , that will rapidly cause activation of all receptors present. Fitting the model to this experimental condition may have resulted in slight deviations in the parameter values. Nevertheless, the modeling results illustrate that building a model with basic kinetic rate laws can give useful predictions over a range of concentrations. Figures ?? and ?? show the modeling results with TNF set at 8 out of 100.

## Appendix E: Supplementary figures







**Figure S8:** The merged model for the TNF $\alpha$ -EGF pathway in which the two hypotheses are highlighted. The first hypothesis is the dependence IL-1 $\alpha$  expression on the combined activity of ERK and JNK1. The second hypothesis assumes an as yet unidentified protein (Hyp2) to link EGFR to MEKK1. Node colors represent initial activity levels.

