

# FORMAL REPRESENTATION OF THE HIGH OSMOLARITY GLYCEROL PATHWAY IN YEAST

CLEMENS KÜHN<sup>1</sup>                      K. V. S PRASAD<sup>2</sup>  
clemens.kuehn@biologie.hu-berlin.de    prasad@chalmers.se

EDDA KLIPP<sup>1</sup>                      PETER GENNEMARK<sup>3,4</sup>  
edda.klipp@rz.hu-berlin.de            peterg@chalmers.se

<sup>1</sup>*Humboldt University, DE-10115 Berlin, Germany.*

<sup>2</sup>*Chalmers University of Technology, SE-412 96 Göteborg, Sweden*

<sup>3</sup>*University of Göteborg, SE-412 96 Göteborg, Sweden*

<sup>4</sup>*Uppsala University, SE-751 06 Uppsala, Sweden*

The high osmolarity glycerol (HOG) signalling system in yeast belongs to the class of Mitogen Activated Protein Kinase (MAPK) pathways that are found in all eukaryotic organisms. It includes at least three scaffold proteins that form complexes, and involves reactions that are strictly dependent on the set of species bound to a certain complex. The scaffold proteins lead to a combinatorial increase in the number of possible states. To date, representations of the HOG pathway have used simplifying assumptions to avoid this combinatorial problem. Such assumptions are hard to make and may obscure or remove essential properties of the system.

This paper presents a detailed generic formal representation of the HOG system without such assumptions, showing the molecular interactions known from the literature. The model takes complexes into account, and summarises existing knowledge in an unambiguous and detailed representation. It can thus be used to anchor discussions about the HOG system. In the commonly used Systems Biology Markup Language (SBML), such a model would need to explicitly enumerate all state variables. The *Kappa* modelling language which we use supports representation of complexes without such enumeration.

To conclude, we compare *Kappa* with a few other modelling languages and software tools that could also be used to represent and model the HOG system.

*Keywords:* High osmolarity glycerol pathway; rule-based modelling; Kappa.

## 1. Introduction

The High Osmolarity Glycerol (HOG) signalling pathway in the yeast *Saccharomyces cerevisiae* belongs to the class of Mitogen Activated Protein Kinase (MAPK) pathways found in all eukaryotic organisms, and is of general interest. To unambiguously represent and communicate current hypotheses of the molecular interactions between the signalling proteins in the HOG system is of fundamental interest. This is a challenging problem for several reasons.

First, many details of the molecular interactions are unknown, and alternative hypotheses with various levels of experimental support are present in the literature. The main reason for lack of knowledge is that many phosphorylation states or

protein interactions cannot be experimentally observed. We note in passing that the rapid development of measurement techniques like mass spectrometry has great potential to cure this problem and allow simultaneous observation of multiple species [6].

Second, the system includes complexes and many different phosphorylation states, and an exhaustive formal representation therefore requires a large number of state variables, because of the combinatorial increase of possible combinations. For example, the scaffold protein Pbs2 in yeast has at least three docking sites and can, besides, be single and double phosphorylated (Fig. 1). Scaffold proteins can be thought of as hubs in signalling networks [24, 34, 37, 42]: proteins binding to the docking sites can, in turn, have their own phosphorylation state and additional docking sites. A full representation of Pbs2 and its known binding partners (Sho1, Ssk2/22 and Hog1) and their binding partners (Ste11-Ste50 and Ssk1), requires as many as 1900 state variables.

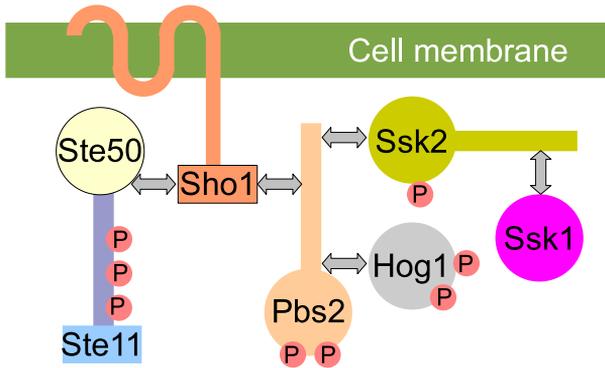


Fig. 1: The yeast scaffold protein Pbs2 has three docking sites and can be phosphorylated on two sites [26, 33, 45]. Double-headed arrows indicate a potential docking of a protein. Small circles denote a potential phosphorylation. To illustrate the combinatorial problem: assuming that all phosphorylation reactions are independent results in 8 states for Ste11-Ste50, 4 states for Pbs2 and for Hog1, and 2 states for Ssk2/22. Besides, some proteins may be absent and Sho1 can be in an inactive or active state. Therefore, in total, 1900 states are possible.

The traditional way to represent the system in journal publications is by an illustration like Fig. 1 and accompanying text in natural language. The text is often extensive and hard to parse and analyse. Representations in this style can therefore be ambiguous.

It is therefore natural to consider more formal approaches. Several mathematical models in the form of ordinary differential equations (ODEs) have been proposed. Using ODEs, we can in principle represent the system unambiguously at the level of molecular interactions. However, the combinatorial problem typically leads to a very

large model as exemplified in Fig. 1. Therefore, in practice, reasonable biological assumptions are used to drastically reduce the number of state variables [4, 5, 7].

For example, for the Pbs2 system in Fig. 1, we can assume that Ssk2 is constitutively bound to Pbs2 as indicated in Ref. [45], and that proteins are either fully phosphorylated or not phosphorylated at all. In some cases, depending on the range of experimental scenarios we want to study, assumptions of this kind can be satisfactory, especially since the amount of information also decreases. However, in many cases this solution is unsatisfactory because essential properties of the system may be obscured or removed, and common system modifications (like *pbs2* $\Delta$  and *ssk2* $\Delta$  in the above example) cannot be explicitly represented in the model. It is also the case that assumptions that seem plausible for a certain experimental scenario, may not be realistic for other system modifications and/or input functions that are tried later.

We want therefore to represent the relatively well characterised HOG system by a formal model that takes complexes into account, but without explicitly enumerating all states. Several publications in recent years have suggested clues to the detailed interaction structure of the pathway, but only in the form of figures like Fig. 1 and accompanying text. There is hence a need for a formal unambiguous model of the system, allowing efficient exchange of hypotheses and results. Ideally, such models can subsequently be used for simulation and prediction of experiments, as well as in experimental planning to assist the experimentalist.

There are several different modelling languages that support modelling of complexes without enumerating all states, and some of them even represent the complexes explicitly [2, 3, 9, 28]. There are also corresponding software tools. It is notable that the systems biology mark-up language (SBML) currently lacks support for representing complexes in this respect [20]. However, such support is listed as a possible future enhancement of SBML, highlighting the importance of this issue [21].

We set out the following requirements for our model. It should

- unambiguously summarise current knowledge about the system.
- take complexes into account without enumerating all states.
- be easy for biologists to edit, and to use in discussing the system.
- include dynamic information, allowing simulation when data becomes richer.

Because only a few state variables in the system can be measured, the kinetic rate constants of the system are currently not observable. The problem with unidentifiable parameters is common for these kind of systems, and it will not affect the way we *represent* the structure of the system. However, a future goal is naturally to infer a dynamic model of the HOG system.

In this paper, we propose a generic model of the HOG signalling system based on literature, and an implementation of this model in the *Kappa* modelling language [8, 9] using the software tool Cellucidate (<http://www.cellucidate.com/>). Cellucidate was chosen because it allows all four requirements above to be fulfilled.

Other strong alternatives that we considered were BioNetGen [2] and

PottersWheel [28]. In the Discussion section, we discuss advantages and limitations of these languages and of another relevant language, SPiM [3].

## 2. Modelling the HOG Pathway

The HOG pathway consists of a sensing system, a cascade of protein kinases and output systems such as transcriptional regulators [18]. To give an overview, a static pathway diagram is given in Fig. 2. Here, some causality has been incorporated by re-drawing certain complexes in series and indicating molecule bindings by arrows. As we said, the traditional way to depict models in the biological literature is through figures of this kind, with ambiguities carefully explained in the text.

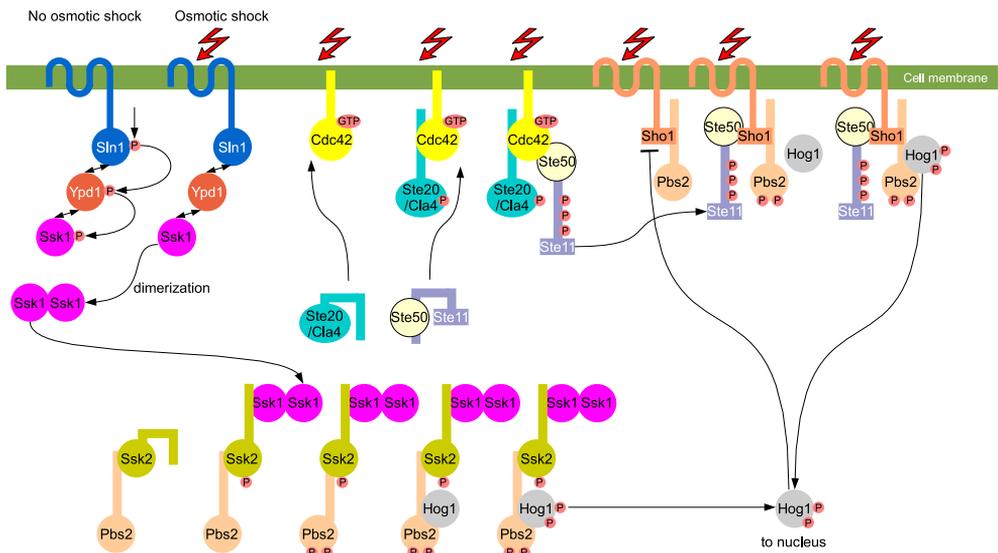


Fig. 2: Traditional model of a signalling pathway reconstructed from the literature. Details and references are given in the text and in Table 1.

The HOG system is activated by osmotic shock, e.g. by the addition of salt to the medium. At least three membrane proteins, Cdc42, Sho1 and Sln1, are activated by the stress, presumably by changes in turgor pressure [40, 44]. Two different branches, the Cdc42-Sho1 branch and the Sln1 branch, merge at the protein Pbs2, which in turn activates Hog1, the last kinase in the pathway. Hog1 activates cytosolic effectors (e.g. Pfk27 [11]) and enters the nucleus and induces transcription. This, in turn, leads to intracellular glycerol accumulation, that counters effects of the osmotic shock by increasing intra-cellular osmotic pressure.

Presently, three scaffold proteins are identified; Cdc42, Sho1 and Pbs2. Some phosphorylation reactions are known to be strictly dependent on the set of species bound to a certain complex. For example,

- the phosphorylation of Ste11 is assumed to require that both Ste50-Ste11 and Ste20 are bound to Cdc42, and besides, that Ste20 is phosphorylated [46].
- the phosphorylation of Pbs2 by Ste11 is assumed to require that both Ste11 and Pbs2 are bound to Sho1, and besides, that Ste11 is phosphorylated [46].
- the phosphorylation of Pbs2 by Ssk2 is assumed to require that unphosphorylated Ssk1 (potentially as a dimer) binds Ssk2 when Ssk2 is in complex with Pbs2 [45].

The HOG system has previously been characterised by mathematical models at various levels of detail [13, 15, 16, 22, 23, 29]. Of these, the most detailed model is that of Klipp *et al.* in Ref. [22].

In Table 1, we propose a generic model of the HOG signalling system built from information in the recent literature, and experimentally supported. Our generic model differs from that of Klipp *et al.* in three main respects:

- We include all known signalling branches.
- We model complexes.
- We focus on the signalling pathways. Gene regulation, metabolism and biophysics of osmoregulation are not included.

At this stage, we have chosen not to take activation of phosphatases and localisation of Hog1<sup>PP</sup> into account. We could also not yet account for the implications made in the very recent literature [25]. Table 1 should be read in conjunction with Fig. 2.

In Table 1, states of molecules (phosphorylated:  $X^P$ , unphosphorylated:  $X^U$ ) are only specified if required to unambiguously define a particular reaction. To allow simulation, we define an input variable called *Osmostress* representing the biophysical changes that lead to activation of Sln1, Sho1 and Cdc42. Adaptation is mainly driven by Hog1<sup>PP</sup> dependent accumulation of intracellular glycerol, here modelled by a negative feedback of Hog1<sup>PP</sup> on *Osmostress*.

### 3. Implementing the HOG pathway in Kappa

In this section we describe how we implemented the generic HOG model of Table 1 in Kappa, using the web-based application, Cellucitate, and the corresponding stand-alone program, Kappa Factory [8, 9]. The Kappa code is listed in the Supplementary material and on the model repository of Cellucitate's web site (<http://www.cellucitate.com>).

Kappa is a process calculus in the tradition started by the Calculus of Communicating Systems, referred to as CCS [30]. CCS allows description and analysis of concurrent systems. In Cellucitate a rule-based representation is used, and state variables are not explicitly enumerated [8, 9]. Instead, the model is fully represented by a list of agents (e.g. proteins) and their individual docking sites and activation (e.g. phosphorylation) sites, and a list of rules corresponding to a generalised form of a chemical reaction. Basically, the rules are defined on the same

Table 1: Generic Model of the HOG Pathway.

Reaction	Comment and references
$Cdc42 \longrightarrow Cdc42\text{-act}$	Act. [46] by Osmostress
$Cdc42\text{-act} \longrightarrow Cdc42$	Deact.
$Cdc42 + Ste20 \longrightarrow Cdc42\text{-Ste20}$	Ass. Req. $Cdc42\text{-act}$ [46]
$Cdc42\text{-Ste20} \longrightarrow Cdc42 + Ste20$	Diss.
$Ste20^U \longrightarrow Ste20^P$	Phos. [46] by $Cdc42\text{-act}$
$Ste20^P \longrightarrow Ste20^U$	Dephos.
$Cdc42 + Ste11 \longrightarrow Cdc42\text{-Ste11}$	Ass. Req. active $Cdc42$ [46]
$Cdc42\text{-Ste11} \longrightarrow Cdc42 + Ste11$	Diss. <sup>1</sup>
$Ste11^U \longrightarrow Ste11^P$	Phos. Req. $Ste11\text{-Cdc42}\text{-Ste20}^P$ [46]
$Ste11^P \longrightarrow Ste11^U$	Dephos.
$Sho1 \longrightarrow Sho1\text{-act}$	Act. [46] by Osmostress
$Sho1\text{-act} \longrightarrow Sho1$	Deact.
$Sho1 + Ste11 \longrightarrow Sho1\text{-Ste11}$	Ass. Req. $Sho\text{-act}$ [37, 46]
$Sho1\text{-Ste11} \longrightarrow Sho1 + Ste11$	Diss.
$Sln1 \longrightarrow Sln1^P$	Phos. Inh. by Osmostress [27]
$Sln1 + Ypd1 \longrightarrow Sln1\text{-Ypd1}$	Ass. Req. $Sln1^U, Ypd1^P$ or $Sln1^P, Ypd1^{U^2}$ [38]
$Sln1\text{-Ypd1} \longrightarrow Sln1 + Ypd1$	Diss.
$Sln1^P\text{-Ypd1}^U \longleftrightarrow Sln1^U\text{-Ypd1}^P$	Phosphotransfer <sup>3</sup>
$Ypd1 + Ssk1 \longrightarrow Ypd1\text{-Ssk1}$	Ass. Req. $Ypd1^P, Ssk1^U$ or $Ypd1^U, Ssk1^{P^2}$ [38]
$Ypd1\text{-Ssk1} \longrightarrow Ypd1 + Ssk1$	Diss.
$Ypd1^P\text{-Ssk1}^U \longleftrightarrow Ypd1^U\text{-Ssk1}^P$	Phosphotransfer <sup>3</sup>
$Ssk1^P \longrightarrow Ssk1^U$	Depho. <sup>4</sup>
$Ssk1 + Ssk1 \longleftrightarrow Ssk1\text{-Ssk1}$	Dimerization [19]
$Ssk1 + Ssk2 \longleftrightarrow Ssk1\text{-Ssk2}$	Ass., Diss. <sup>5</sup>
$Ssk1^U\text{-Ssk1}^U\text{-Ssk2}^U \longrightarrow Ssk1^U\text{-Ssk1}^U\text{-Ssk2}^P$	Phos.
$Ssk2^P \longrightarrow Ssk2^U$	Dephos. Req. unbound $Ssk2$
$Ssk2 + Pbs2 \longleftrightarrow Ssk2\text{-Pbs2}$	Ass., Diss. [45]
$Sho1 + Pbs2 \longrightarrow Sho1\text{-Pbs2}$	Ass. Req. $Sho\text{-act}$ [37, 46]
$Sho1\text{-Pbs2} \longrightarrow Sho1 + Pbs2$	Diss.
$Pbs2^{UU} \longrightarrow Pbs2^{PP}$	Phos. Req. $Ste11^P\text{-Sho1}\text{-Pbs2}$ [26, 46] or $Ssk2^P\text{-Pbs2}$ [45]
$Pbs2^{PP} \longrightarrow Pbs2^{UU}$	Dephos.
$Pbs2^{PP} + Hog1 \longrightarrow Pbs2^{PP}\text{-Hog1}$	Ass. Req. $Ssk2^P\text{-Pbs2}^{PP}$ [45] or $Ste11^P\text{-Sho1}\text{-Pbs2}^{PP}$ [37]
$Pbs2\text{-Hog1} \longrightarrow Pbs2 + Hog1$	Diss.
$Pbs2^{PP}\text{-Hog1}^{UU} \longrightarrow Pbs2^{PP}\text{-Hog1}^{PP}$	Pho. Req. $Ssk2^P\text{-Pbs2}^{PP}$ [33, 45]
$\longrightarrow Pbs2^{PP}\text{-Hog1}^{PP}$	Req. $Ste11^P\text{-Sho1}\text{-Pbs2}^{PP}$ [37]
$Hog1^{PP} \longrightarrow Hog1^{UU}$	Dephos. <sup>6</sup> [33]
$Hog1^{PP} + Sho1\text{-act} \longrightarrow Hog1^{PP}\text{-Sho1}\text{-act}$	Ass. [15]
$Hog1\text{-Sho1} \longrightarrow Hog1 + Sho1$	Diss.
$Hog1^{PP}\text{-Sho1}\text{-act} \longrightarrow Hog1^{PP}\text{-Sho1}$	Inactivation
$Hog1^{PP}, Osmostress \longrightarrow Hog1^{PP}$	Simplified Adaptation

1. Rate can be assumed higher when  $Ste11$  is phosphorylated.
2. Requires reaction partners not bound to members of  $Sln1\text{-Ypd1}\text{-Ssk1}\text{-Ssk2}$  phosphorelay [19].
3. Rate left to right assumed 4 orders of magnitude higher than right to left.
4. Requires  $Ssk1$  to be not bound to  $Ypd1$  and  $Ssk2$ .
5. Dissociation rate assumed increased after phosphorylation of  $Ssk2$ .
6. Requires  $Hog1$  not bound to  $Pbs2$  [33].

level of abstraction as the reactions in our generic model in Table 1, and each rule may represent several chemical reactions. For detailed information about the language we refer to [9]. A very similar modelling framework is BioNetGen [2]

Cellucidate can accept text input directly from the keyboard, or from a file.

There is also an intuitive graphical user interface (GUI) which can easily be used by biologists to input and edit models. No prior knowledge or experience in mathematical modelling is required for setting up a formal unambiguous description of a system.

We implemented the HOG model in Kappa by translating each generic reaction in Table 1 to a corresponding rule. To illustrate the Kappa syntax and to give a flavour of our model, we give some examples here.

To denote that Sho1 has a state called  $x$  that is activated ( $a$ ) and, furthermore, that Sho1 has a docking site for Ste11, we write

$\text{Sho1}(x\sim a, \text{Ste11})$

Similarly, to denote that Ste11 is phosphorylated ( $p$ ) and has docking sites for both Sho1 and Cdc42, we write

$\text{Ste11}(x\sim p, \text{Sho1}, \text{Cdc42})$

To model the association of active Sho1 to Ste11, it is reasonable to require that Ste11 is not bound to Cdc42, and that the phosphorylation state of Ste11 is arbitrary. We write the rule as

$\text{Sho1}(x\sim a, \text{Ste11}), \text{Ste11}(\text{Sho1}, \text{Cdc42}) \rightarrow$   
 $\text{Sho1}(x\sim a, \text{Ste11!1}), \text{Ste11}(\text{Sho1!1}, \text{Cdc42}) @ 2.0$

Docking is indicated by an index after the reserved symbol  $!$ , and the kinetic rate of the rule is defined at the end, after the reserved symbol  $@$ . Note that the phosphorylation state of Ste11 is simply omitted, and that the binding site for Cdc42 on Ste11 is unoccupied on both sides of the arrow.

The formulation is very compact, since only the known interactions and constraints are incorporated. For instance, let us reconsider the example discussed in Section 2. The phosphorylation of Pbs2 by Ste11 is assumed to require that both Ste11 and Pbs2 are bound to Sho1, and furthermore, that Ste11 is phosphorylated. This is encoded as

$\text{Ste11}(x\sim p, \text{Sho1!1}), \text{Sho1}(\text{Ste11!1}, \text{Pbs2!2}), \text{Pbs2}(x\sim u, y\sim u, \text{Sho1!2}) \rightarrow$   
 $\text{Ste11}(x\sim p, \text{Sho1!1}), \text{Sho1}(\text{Ste11!1}, \text{Pbs2!2}), \text{Pbs2}(x\sim p, y\sim p, \text{Sho1!2}) @ 1.0$

where  $x$  and  $y$  refer to a phosphorylation site,  $u$  and  $p$  refer to unphosphorylated and phosphorylated, respectively. The rule says nothing about the activation status of Sho1 and potential bindings to the other two docking sites of Pbs2 (sites for Hog1 and Ssk2/22). It represents only our current understanding of the system; we do not need to write out all possibilities. When new experimental evidence appears, the model can easily be modified.

An interesting feature of the Kappa system is the possibility to *refine* a certain rule, that is to give an alternative interpretation given some additional requirement [10]. For example, the rate of dissociation of Ste11 from Cdc42 can be assumed higher when Ste11 is phosphorylated. Naturally, each refinement leads to

an additional rule with a corresponding kinetic rate, and hence increased complexity of the model.

Implementing models of this kind is error prone, and to debug the model we used the following means:

- The static properties of a model were verified by automatically generated dependency graphs where all agents are nodes and their potential binding partners are edges, a so-called contact map.
- The probability and timing of reaching certain events (the 'firing' of a specific set of rules in the model) given certain initial conditions was verified by simulation of so-called stories.

In addition, we used simulation to judge the qualitative behaviour of the model. Cellucitate uses stochastic simulation and does not rely on internal enumeration of all states. In this way, simulation scales well with the number of rules and possible agents. Especially for very large systems, this becomes prominent [8]. As already mentioned, the parameters of our model are not observable and simulation is not our main objective at this stage. However, in the presented model, we have manually tuned the kinetic rates in order to qualitatively resemble real data. Relative protein abundance data was collected from [14]. The model can be simulated resulting in time-series of number of molecules of various species. A sample plot is given in Fig. 3.

One difficulty arising in simulation models with many states is setting up the initial concentrations. Although not every possible state has to be encoded manually for the rules in Kappa, they have to be for initial conditions. The approach we use here is to run simulations in the unstressed state, starting with unbound and inactive molecules. Assuming the system evolves to a steady state (defined by characteristic distributions of the molecule numbers), we collect the results of a number of simulations. The initial conditions for the simulation under stress are then sampled from these collected results. This approach has the advantage that stochasticity is already included in the initial conditions and that this approach is highly automatable. Drawbacks are that considerable computation is needed to compute the initial steady state and that not all systems necessarily evolve to one steady state (e.g. bistable systems).

#### 4. Discussion

This paper has presented in the language Kappa a formal executable model of the HOG pathway in *S. cerevisiae*, the first such that takes complexes into account. The model, like all Kappa programs, represents the HOG system in terms of agents and rules. It has the desired properties listed earlier:

- It represents current knowledge about the HOG system (summarised as a detailed generic model in Table 1 and as a Kappa model in the Supplementary material) compactly and without the ambiguity of traditional pathway diagrams

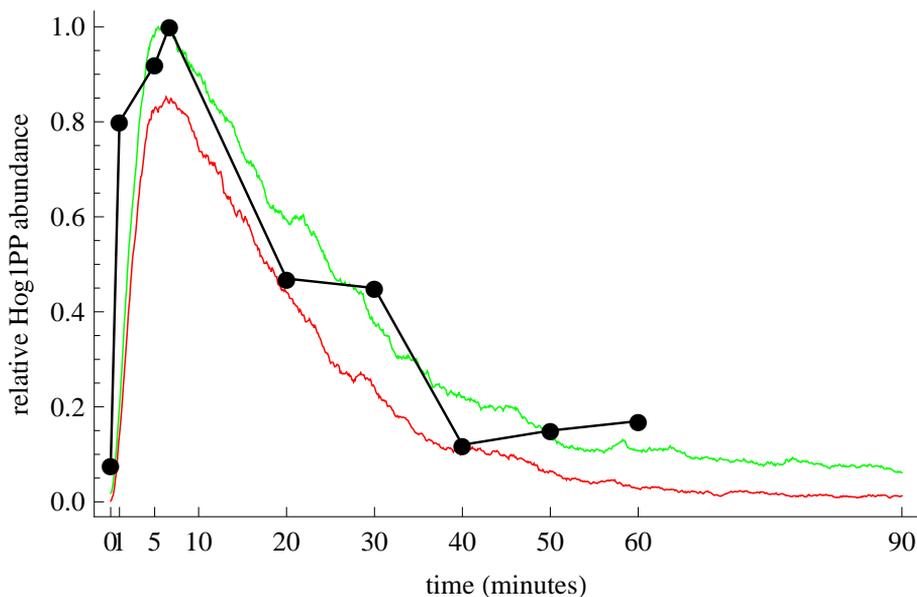


Fig. 3: Simulated and experimental time series for Hog1<sup>PP</sup> after an osmotic shock. Simulated data was obtained from ten stochastic simulations, and here plotted as  $\pm$  one standard deviation. Experimental data (rings, connected by lines) is obtained by the standard Western blot technique.

like in Fig. 2. In particular, our model takes complexes into account without enumerating all states, and captures phosphorylation reactions known to be strictly dependent on the set of species bound to a certain complex.

- The model can easily be edited by biologists, who can use it to anchor discussions about the system. The Kappa formalism allows biological knowledge to be encoded directly into a mathematical model [17]. With the GUI, this encoding requires no prior knowledge of mathematical modelling, and supporting routines for error detection are available. Our model is publicly available for collaborative work on the Cellucitate web site.
- The model includes dynamic information, allowing simulation. Kinetic rates are manually tuned to make simulations resemble experimental data. The model is hence prepared for the natural next step, to be fitted to data. Given current data, this is the best we can do since parameters are not observed when modelling at this level of detail.

#### 4.1. Constructs Needed to Model Pathways

Protein interactions alone do not suffice to properly model the dynamics of signalling pathways. Additional model constructs and analysis features may be needed.

- Pathways are naturally integrated with other cellular processes like gene

regulation and metabolism, and with the environment. For example, biophysical interactions, gene regulation and glycerol production are key features in adaptation to osmotic shock [22], and cell size is key to cell cycle regulation [1].

- The internal structure of the cell is very complex. Compartments like the cytosol, nucleus and vacuole have individual properties and species concentrations.
- Biological processes at the molecular level are stochastic, and deterministic simulation is not always an adequate approximation. Clearly, hybrid simulation methods are of interest when combining signalling pathways with metabolic systems in one model. When experimental data is gathered from a population of cells (as opposed to single cell experiments), the composition of the population and its temporal development are both stochastic. This population effect is usually not accounted for in mathematical models.

Given this diverse list of requirements, what languages and tools are available and what properties do they have?

## 4.2. Comparison of Languages and Softwares

### 4.2.1. Kappa

Kappa/Cellucitate is very good as far as it goes, but has several shortcomings. It lacks global variables and support for compartments, which can only be implicitly defined, e.g. creating two species  $A_{\text{cytosolic}}$  and  $A_{\text{nuclear}}$  each with its own interaction possibilities and transitions from one to the other. Furthermore, the stochastic simulation may be tedious and unnecessary when modelling metabolites with high concentrations.

Finally, Kappa/Cellucitate lacks programmability, so that to program a system with symmetric parts  $Q(a)$  and  $R(b)$ , differing only in the parameter, each part has to be coded separately. In a system like SPiM (below), this can be done in the usual programming way: write a single program  $P(x)$ , and the two parts are then  $P(a)$  and  $P(b)$ .

When it comes to repeatedly running stochastic simulations, the user has to go through the Cellucitate web-service. Although it is free to use at the moment, Cellucitate and KappaFactory are closed source and extensive use of the web-service might become associated with a fee.

### 4.2.2. BioNetGen

BioNetGen [2, 12] shares many properties with Cellucitate, in particular almost identical input languages. For example, we have relatively easily transcribed the HOG model to the BioNetGen syntax.

The stochastic simulation engine of BioNetGen currently uses a less powerful algorithm than Cellucitate. The new solver under development, nfsim (Sneddon,

Faeder and Emonet, private comm.), implements a generalized version of the algorithm presented in Ref. [47].

In contrast to Cellucitate, BioNetGen is open source, and (fully enumerated) models can be exported to SBML. Models can then be simulated using standard deterministic integration methods. This is an important feature when it comes to parameter estimation, where simulation speed often is crucial.

BioNetGen and its web interface GetBonNie (<http://getbonnie.cs.unm.edu>) feature less sophisticated GUIs compared to KappaFactory/Cellucitate. However, all options of the software are directly available from the command line, making repeated simulations of stochastic models easy to set up.

#### 4.2.3. SPiM

The Stochastic Pi Machine (SPiM) [3, 35, 36, 43] runs the stochastic version of the pi-calculus [31, 32]. The aim of the pi-calculus is to describe concurrent computations whose configuration may change during the computation. That the stochastic pi-calculus could be used for biochemistry was first discovered by Aviv Regev [39, 41].

The SPiM programming language allows parameterisation of symmetric code, as mentioned above. It has global variables, though not fully dynamic ones. Compartments (even changing ones) can be represented, but coded in a way that might not be transparent to a biologist. Similarly, a link between two molecules A and B can be encoded in various ways, by moving both into new states where a parameter records the linkage. We can choose whether to say just that A is linked on this contact point x, or add details such as to what kind of molecule it is linked to, or even that it is specifically linked to B. Thus the programmability is a gain, but at the cost of direct representation of chemistry.

#### 4.2.4. Rule-based Modelling in PottersWheel

PottersWheel [28] (PW) is based on ODEs and deterministic simulation, and includes a range of analysis tools like parameter estimation, sensitivity analysis and model discrimination. Besides, PW offers a syntax for describing complexes by rule based reactions. The full set of ODEs is then automatically generated and used in simulations.

A fundamental difference between Kappa and rule-based modelling in PW is that Kappa by default considers all possible reactions, while in PW the alternatives must be explicitly defined. In general, the rule-based modelling in PW offers the same flexibility as Kappa, but the code is less compact, and requires additional manual work.

As a modelling software tool, PW offers more general functionality compared to Kappa/Cellucitate, e.g. algebraic equations, arbitrary input functions, and compartments. Besides, PW is very strong on model analysis and parameter

estimation, while it is currently lacking support for stochastic simulation and a GUI for rule-based modelling.

### **4.3. Concluding Remarks**

Mathematical models enable a formal representation of biological systems and thorough analysis of hypotheses generated from experiments. The success of such models demands further refinement of mathematical methods and computational tools, as well as of experimental methods.

There are now several modelling frameworks, and it is important to choose the one appropriate to the purpose at hand, to the detail at which the system is described, and to the nature of the system. Our Kappa model of the HOG system is valuable as an unambiguous description of the protein interactions involved, and therefore as a basis for discussion. It cannot currently be used to study the complete adaptation of a yeast cell to hyperosmotic shock, from sensing the biophysical changes in turgor changes to accumulation of intracellular glycerol. But this is exactly the strength of such descriptions. Each allows us to 'zoom in' on a specific part of the problem. How to integrate the results in a higher level description of the system is a topic for new research.

Mathematical modelling of signalling pathways requires a large amount of experimental data, computation power, and the coordinated work of many scientists. This coordination will eventually have to go beyond formulation of common standards and storage of models in databases. One way is to work directly on a common, executable model that unambiguously represents current knowledge. Our HOG model in Kappa is such a model, and the web-based approach of Cellucitate allows it to be directly annotated and discussed by multiple users, for example to jointly incorporate new implications as those mentioned in [25].

### **Acknowledgments**

The project was done as part of a Marie Curie EST school in Göteborg 2008. CK was funded by the International Research Training Group *Genomics and Systems Biology of Molecular Networks*, supported by the German Research Foundation (DFG). PG was supported by the Göteborg University quantitative biology platform and the Swedish Strategic Research Foundation through Göteborg Mathematical Modeling Center. Thanks to Vincent Danos for introducing us to Kappa, to James R. Faeder for support on BioNetGen and to Jörg Schaber for comments on the model. *Author contributions:* CK and PG conceived the modelling concept, and constructed the basic model. CK significantly improved the model and tuned it to data. KVSP contributed technical expertise and essential modelling language background information. EK contributed experience with the Hog-Pathway and modeling thereof. PG wrote the paper with significant contributions from CK and KVSP.

## References

- [1] Barberis, M., Klipp, E., Vanoni, M., Alberghina, L., Cell Size at S Phase Initiation: an Emergent Property of the G1/S Network, *PLoS Comput. Biol.*, 3(4): e64, 2007.
- [2] Blinov, M., Faeder, J., Goldstein, B., Hlavacek, W., Bionetgen: software for rule-based modeling of signal transduction based on the interactions of molecular domains, *Bioinformatics*, 20(17):3289–3291, 2004.
- [3] Blosssey, R., Cardelli, L., Phillips, A., A Compositional Approach to the Stochastic Dynamics of Gene Networks, *Transactions in Computational Systems Biology (TCSB)*, 3939:99–122, 2006.
- [4] Borisov, N., Markevich, N., Hoek, J., Kholodenko, B., Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity, *Biophys. J.*, 89(2):951–966, 2005.
- [5] Borisov, N., Markevich, N., Hoek, J., Kholodenko, B., Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains, *Biosystems*, 83(2-3):152–166, 2006.
- [6] Choi, M.Y., Kang, G.Y., Hur, J.Y., Jung, J.W., Kim K.P., Park, S.H., Analysis of Dual Phosphorylation of Hog1 MAP Kinase in *Saccharomyces cerevisiae* Using Quantitative Mass Spectrometry, *Mol. Cells*, 26:200–205, 2008.
- [7] Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Kholodenko, B., Gilles, E., A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks, *BMC Bioinformatics*, 7:3, 2006.
- [8] Danos, V., Feret, J., Fontana, W., Krivine, J., Scalable Simulation of Cellular Signaling Networks, *APLAS 2007, the Fifth ASIAN Symposium on Programming Languages and Systems*, 2007.
- [9] Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J., Rule-based modelling of cellular signalling, *CONCUR, the 18th International Conference on Concurrency Theory*, 2007.
- [10] Danos, V., Feret, J., Fontana, W., Harmer, R., Krivine, J., Rule-based modelling, symmetries, refinements, *Lecture Notes in Computer Science*, 5054:103–122, 2008.
- [11] Dihazi, H., Kessler, R., Eschrich, K., High osmolarity glycerol (HOG) pathway-induced phosphorylation and activation of 6-phosphofructo-2-kinase are essential for glycerol accumulation and yeast cell proliferation under hyperosmotic stress, *J. Biol. Chem.*, 279:23961–23968, 2004.
- [12] Faeder, J.R., Blinov, M.L., Hlavacek, W.S., Rule-based modeling of biochemical systems with BioNetGen, *Methods Mol. Biol.*, 500:113–167, 2009.
- [13] Gennemark, P., Nordlander, B., Hohmann, S., Wedelin, D., A simple mathematical model of adaptation to high osmolarity in yeast, *In Silico Biol.*, 6(3):193–214, 2006.
- [14] Ghaemmaghami S., Huh W.K., Bower K., Howson R.W., Belle A., Dephoure N., O’Shea E.K., Weissman J.S., Global analysis of protein expression in yeast, *Nature*, 425(6959):737–741, 2003.
- [15] Hao, N., Behar, M., Parnell, S.C., Torres, M.P., Borchers, C.H., Elston, T.C., Dohlman, H.G., A systems-biology analysis of feedback inhibition in the Sho1 osmotic-stress-response pathway, *Curr. Biol.*, 17(8):659–667, 2007.
- [16] Hersen, P., McClean, M.N., Mahadevan, L., Ramanathan, S. Signal processing by the HOG MAP kinase pathway, *Proc Natl Acad Sci U S A.*, 105(20):7165–7170, 2008.
- [17] Hlavacek W.S., How to deal with large models?, *Molecular Systems Biology*, 5(240), 2008.
- [18] Hohmann, S., Osmotic stress signaling and osmoadaptation in yeasts, *Microbiol Mol Biol Rev.*, 66(2):300–372, 2002.
- [19] Horie, T., Tatebayashi, K., Yamada, R., Saito, H., Phosphorylated Ssk1 prevents

- unphosphorylated Ssk1 from activating the Ssk2 mitogen-activated protein kinase kinase kinase in the yeast high-osmolarity glycerol osmoregulatory pathway, *Mol. Cell Biol.*, 28(17):5172–5183, 2008.
- [20] Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., et al., The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network models, *Bioinformatics*, 19(4):524–531, 2003.
- [21] Hucka, M., Finney, A., Hoops, S., Keating, S., Le Novère, N., Systems Biology Markup Language (SBML) Level 2: Structures and Facilities for Model Definitions. SBML Level 2 Version 3, Release 2, 26 September 2007. Section 8.1. <http://sbml.org>, 2007.
- [22] Klipp, E, Nordlander, B, Krüger, R, Gennemark, P, Hohmann, S., Integrative model of the response of yeast to osmotic shock, *Nat Biotechnol.*, 23(8):975–982, 2005.
- [23] Kühn, C., Petelenz, E., Nordlander B., Jörg Schaber, Hohmann, S., Klipp, E., Exploring the Impact of Osmoadaptation on Glycolysis Using Time-Varying Reponse-Coefficients, *Genome Informatics*, 20:77–90, 2008.
- [24] Levchenko A., Bruck J., Sternberg P.W., Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties, *PNAS*, 97(11):5818–5823, 2000.
- [25] Macia, J., Regot, S., Peeters, T., Conde, N., Solé, R., Posas, F., Dynamic Signaling in the Hog1 MAPK Pathway Relies on High Basal Signaling Transduction, *Science Signaling*, 2(63), ra13, 2009.
- [26] Maeda, T., Takekawa, M., Saito, H., Activation of yeast PBS2 MAPKK by MAPKKs or by binding of an SH3-containing osmosensor, *Science*, 269(5223):554–558, 1995.
- [27] Maeda, T., Wurgler-Murphy, S.M., Saito, H., A two-component system that regulates an osmosensing MAP kinase cascade in yeast, *Nature*, 369:242–245, 1994.
- [28] Maiwald, T., Timmer, J., Dynamical modeling and multi-experiment fitting with PottersWheel, *Bioinformatics*, 24(18):2037–2043, 2008.
- [29] Mettetal, J.T., Muzzey, D., Gomez-Urbe, C., van Oudenaarden, A., The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*, *Science*, 319(5862):482–484, 2008.
- [30] Milner, R., *A Calculus of Communicating Systems*, Springer Verlag, ISBN 0-387-10235-3, 1980.
- [31] Milner, R., Parrow, J., Walker, D., A calculus of mobile processes, *Information and Computation*, 100(100), 1–40, 1992.
- [32] Milner, R., *Communicating and Mobile Systems: the Pi-Calculus*, Cambridge Univ. Press, ISBN 0-521-65869-1, 1999.
- [33] Murakami, Y., Tatebayashi, K., Saito, H., Two adjacent docking sites in the yeast Hog1 mitogen-activated protein (MAP) kinase differentially interact with the Pbs2 MAP kinase kinase and the Ptp2 protein tyrosine phosphatase, *Mol. Cell Biol.*, 28(7):2481–2494, 2008.
- [34] Netz, D.J., Pierik, A.J., Stümpfig, M., Mühlhoff, U., Lill, R., The Cfd1-Nbp35 complex acts as a scaffold for iron-sulfur protein assembly in the yeast cytosol, *Nat. Chem. Biol.*, 3(5):278–286, 2007.
- [35] Phillips, A., Cardelli, L., Castagna, G., A Graphical Representation for Biological Processes in the Stochastic Pi-calculus, *Transactions in Computational Systems Biology (TCSB)*, 4230:123–152, 2006.
- [36] Phillips, A., Cardelli, L., Efficient, Correct Simulation of Biological Processes in the Stochastic Pi-calculus, In Proceedings of Computational Methods in Systems Biology (CMSB'07), 4695, 184–199, 2007.
- [37] Posas, F., Saito, H., Osmotic activation of the HOG MAPK pathway via Ste11p

- MAPKKK: scaffold role of Pbs2p MAPKK, *Science*, 276(5319):1702–1705, 1997.
- [38] Posas, F., Wurgler-Murphy, S.M., Maeda, T., Witten, E.A., Thai, T.C., Saito, H., Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 “two-component” osmosensor, *Cell*, 86(6):865–875, 1996.
- [39] Priami, C., Regev, A., Shapiro, E.Y., Silverman, W., Application of a stochastic name-passing calculus to representation and simulation of molecular processes, *Inf. Process. Lett.*, 80(1):25–31, 2001.
- [40] Reiser, V., Raitt, D. C. and Saito, H., Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure, *J. Cell Biol.*, 161:1035–1040, 2003.
- [41] Regev, A., Silverman, W., Shapiro, E.Y., Representation and Simulation of Biochemical Processes Using the pi-Calculus Process Algebra, *Pacific Symposium on Biocomputing*, 6:459–470, 2001.
- [42] Rischitor, P.E., May, K.M., Hardwick, K.G., Bub1 is a fission yeast kinetochore scaffold protein, and is sufficient to recruit other spindle checkpoint proteins to ectopic sites on chromosomes, *PLoS ONE*, 2(12): e1342, 2007.
- [43] SPiM: <http://research.microsoft.com/~aphillip/spim>
- [44] Tamas, M.J., Rep, M., Thevelein, J.M., Hohmann, S., Stimulation of the yeast high osmolarity glycerol (HOG) pathway: evidence for a signal generated by a change in turgor rather than by water stress, *FEBS Lett.*, 472(1):159–65, 2000.
- [45] Tatebayashi, K., Takekawa, M., Saito H., A docking site determining specificity of Pbs2 MAPKK for Ssk2/Ssk22 MAPKKKs in the yeast HOG pathway, *The EMBO Journal*, 22:3624–3634, 2003.
- [46] Tatebayashi, K., Yamamoto, K., Tanaka, K., Tomida, T., Maruoka, T., Kasukawa, E., Saito, H., Adaptor functions of Cdc42, Ste50, and Sho1 in the yeast osmoregulatory HOG MAPK pathway, *EMBO J.*, 25(13):3033–44, 2006.
- [47] Yang, J., Monine, M.I., Faeder, J.R., Hlavacek, W.S., Kinetic Monte Carlo method for rule-based modeling of biochemical networks, *Phys. Rev. E.*, 78:031910, 2008.