Integrative Biology

A Toolbox for Discrete Modelling of Cell Signalling Dynamics

Journal:	*Integrative Biology*
Manuscript ID	IB-TIN-02-2018-000026.R2
Article Type:	Technical Innovation
Date Submitted by the Author:	08-May-2018
Complete List of Authors:	Paterson, Yasmin; University of Cambridge, Department of Biochemistry Shorthouse, David; University of Cambridge, MRC Cancer Unit Pleijzier, Markus; University of Cambridge, MRC Cancer Unit Piterman, Nir; University of Leicester, Department of Informatics Bendtsen, Claus; AstraZeneca UK Ltd, Discovery Science Hall, Ben; University of Cambridge, MRC Cancer Unit Fisher, Jasmin; University of Cambridge, Department of Biochemistry

Integrative Biology
Interdisciplinary approaches for molecular and cellular life sciences

MISSION: To provide a unique venue for research that allows expansion of our knowledge of biology by gaining new insights into important biological and biophysical questions through the application of (novel) enabling quantitative tools and/or technologies (from the nanoscale to the macroscale)

Five year Impact factor: 3.59
Rejection rate: 60%

Technical Innovations - Guidelines to Reviewers

It is essential that all Technical Innovation articles submitted to *Integrative Biology* meet the following assessment criteria:

Technical Innovations should be judged on the *potential* of the method to allow biological insights rather than direct insights. The article should also include extensive biological validation of the method that has not been published elsewhere.

While we expect **full papers** in *Integrative Biology* to present evidence of new biological insights**, Technical Innovations** provide a venue for new methods that may not yet have revealed new insights, but that clearly have the potential to do so, and thus are of strong interest to the biomedical research community. Therefore **Technical Innovations** must provide a sound utility of the technology in the biological world and have clear potential to enable new insights and impact biology research. Thus, these articles are submissions that demonstrate significant potential towards building the bridge between technology and biology — technology or methods that clearly provide new functions that address a significant current challenge or obstacle to elucidating new biological insights!

The "**Insight Box**" should describe how the work presented addresses these criteria (maximum 120 words).

For more information on the scope, please visit http://www.rsc.org/journals-books-databases/about-journals/integrative-biology/

When submitting your report, please:

- provide your report rapidly and within the specified deadline, or inform the Editor immediately if you cannot do so;
- submit your report at **http://mc.manuscriptcentral.com/ib**

The online service for Royal Society of Chemistry authors and reviewers can be found at http://mc.manuscriptcentral.com/rsc

Once again, we appreciate your time in serving as a reviewer. To acknowledge this, the Royal Society of Chemistry offers a **25% discount** on our books: http://www.rsc.org/Shop/books/discounts.asp. Please also consider submitting your next manuscript to *Integrative Biology*.

Best wishes,

Sam Keltie
Executive Editor

Professor Doug Lauffenburger
Editor-in-Chief

The authors have addressed the most of the points that were raised and provided reasonable explanations. However there is one point which needs to be clarified properly before the manuscript can be accepted for publication.

In response to my query about wring diagram of G2/M module in Fig. 3A, the authors provided proper explanation of the causal relationship between Cdk1-CycB and Wee1 and also between Cdk1-CycB and Cdc25. They have also correctly mentioned in many places (Figure caption of Figure 3, line 629, lines 659-661) that Cdk11-CycB and Cdc25 creates a mutual activation module. But influence diagram does not reflect this causal relationship as Cdc25 is connected to the Cdk1-CycB_Pool by an inhibitory arrow and thus contradictory to the text.

We thank the reviewer for pointing this out. We would note that in the QN formalism the type of arrow has no impact, and indeed the target function for Cdc25 shows an activation-type effect (min(10*(var(bound_Cdk1)-var(CKI_active)),3)+(var(phos_Cdk1)-var(Wee1))*floor(var(cdc25)/10)). The decision to render it as an inhibitor was made on the basis of intuition regarding its role in the early development of the model- however we have altered the figure to reflect this.

Now Cdc25 is inhibited (by phosphorylation) upon DNA damage leading to decrease in active Cdk1-CycB complex due to the loss of mutual activation with Cdc25 thus it is consequential but not a direct inhibition.

In order to be biologically consistent, I think, the proper connection would have been like this: phos_Cdk1 -> Cdc25 -> Cdk1-Cycb_Pool and DNA damage -| Cdc25. This will also be fully consistent with the original research paper (Sveiczer et al, PNAS, 97, 7865 (2000)) and the reference # 21 on mathematical model of fission yeast by Tyson group.

We thank the reviewer for this suggestion. However, we note that the model has a variable indicating that DNA is **undamaged** and as such their suggestion regarding these nodes/edges does not apply. With regard to the wiring leading from phos_Cdk1, we do not disagree with the reviewers comments. However we made the explicit choice to encode the network as we did to minimise the complexity of the model. We have added this discussion to the text (line 671-675).

1

A TOOLBOX FOR DISCRETE MODELLING OF CELL SIGNALLING DYNAMICS

Yasmin Z. Paterson[1^], *David Shorthouse*[2^], *Markus W. Pleijzier*[2^], *Nir Piterman*[3], *Claus Bendtsen*[4], *Benjamin A. Hall*[2*] & *Jasmin Fisher* [1,5*]

[1] Department of Biochemistry, University of Cambridge, Cambridge, CB2 1GA, UK

[2] MRC Cancer Unit, University of Cambridge, Cambridge, CB2 0XZ, UK

[3] Department of Informatics, University of Leicester, Leicester, LE1 7RH, UK

[4] Quantitative Biology, Discovery Sciences, IMED Biotech Unit, AstraZeneca, Cambridge, UK

[5] Microsoft Research, Cambridge, CB1 2FB, UK

[^]These authors contributed equally to this work

*Correspondence and request for materials should be addressed to B.A.H (bh418@cam.ac.uk) or J.F. (jf416@cam.ac.uk)

6435 words

ABSTRACT

In an age where the volume of data regarding biological systems exceeds our ability to analyse it, many researchers are looking towards systems biology and computational modelling to help unravel the complexities of gene and protein regulatory networks. In particular, the use of discrete modelling allows generation of signalling networks in the absence of full quantitative descriptions of systems, which are necessary for ordinary differential equation (ODE) models. In order to make such techniques more accessible to mainstream researchers, tools such as the BioModelAnalyzer (BMA) have been developed to provide a user-friendly graphical interface for discrete modelling of biological systems. Here we use the BMA to build a library of discrete target functions of known canonical molecular interactions, translated from ordinary differential equations (ODEs). We then show that these BMA target functions can be used to reconstruct complex networks, which can correctly predict many known genetic perturbations. This new library supports the accessibility ethos behind the creation of BMA, providing a toolbox for the construction of complex cell signalling models without the need for extensive experience in computer programming or mathematical modelling, and allows for construction and simulation of complex biological systems with only small amounts of quantitative data. (199 words)

Insight Box

A limitation of popular ODE models is that they require complete networks and detailed kinetic parameterisation. An alternative is the use of discrete, executable models, in which nodes are assigned discrete value ranges, and the relationship between them defined with logical operations. A fundamental question for executable models however is whether the high level of abstraction substantially reduces expressivity relative to continuous approaches. Here, we present a canonical library of biological signalling motifs, initially defined by Tyson et al (2003), expressed using the BioModelAnalyzer. We show that; 1) these motifs are easily and fully translatable from continuous to discrete models, 2) Combining these motifs generates a fully functional and predictive model of the yeast cell cycle.

(116 words)

INTRODUCTION

We are in an era of ever-increasing biological data. With data available from genomic studies, through to metabolomic studies, the size, scale and heterogeneity of the resources available present many triumphs in terms of advancing high-throughput technologies but also many challenges. Despite the enormous multitude of available data, our understanding of how such information encoded in a cell's genome is used to carry out the complex biological interactions found between genes and gene products is still lacking. It is therefore no surprise that a central goal of modern biology in this post-genomic era is to understand the structural and temporal nature of these control networks. Not only would this allow us to translate 'Big Data' into working models of biological systems, but also equip us with a better understanding of biological mechanisms, allowing the exploration of emergent behaviours and consequences of genomic variants, with an aim to develop real-world hypotheses for experimental validation.

If we are to meet these challenges, new tools, techniques and ways of working need to be adopted. Whilst experimental procedures using a traditional reductionist approach, focusing on the study of individual proteins or genes in isolation from other network interactions have proved useful in uncovering specific elemental functions of various cellular mechanisms, many disease processes continue to elude us. This has fuelled the growth of new lines of scientific inquiry. The wide-ranging, vast improvements in computing power brought about at the beginning of the twenty-first century has led biologists down the path of Systems Biology as a means to organise this

biological data more holistically. This strategy therefore seeks to combine traditional biological thinking with more interdisciplinary, integrated, synthetic approaches allowing for larger-scale simulations of complex systems, which could revolutionise biomedical discovery.

Computational modelling therefore presents a powerful and novel approach to combat these challenges. The application of standard mathematical modelling, such as through stochastic or ordinary differential equations (ODEs), have been faithfully reproducing the interplay between genes and proteins in small regulatory networks with relative success. Prominent examples of ODE models include that of bacterial chemotaxis [1], the lactose operon control system in *Escherichia coli* [2] and the process of X chromosome inactivation [3], the cell cycle in yeast [4], and the generation of amyloid fibrils [5].Such models employ complex kinetic equations to describe relationships between proteins or genes over time, and require highly accurate and intensive experimental data for their development as input. The complexity of such equations and experimental data required can provide a lot of dynamical detail however this complexity also begs the question of whether this approach will scale well when constructing much larger, more intricate networks in the future.

Executable modelling on the other hand, which describes biological systems as discrete systems, can provide a much simpler class of models [6]. Such models are immediately executable, meaning that any update of the model is formally defined and expressible with formal logic. Executable modelling also

provides the ability to undergo model checking (the ability to prove the existence or non-existence of specific user-defined states and transitions) and other formal verifications with ease [10–13]. Analyses performed using the BMA are highly scalable in that efficient construction of formal proofs for user-specified mathematical properties of large (>50 node) networks is possible [7] through the use of bespoke algorithms [8,9]. One of the oldest and simplest forms of executable network models is based on Boolean states (logical models), where each node of the network represents a single gene or protein which is in one of two states: active/on (1) or inactive/off (0) [14,15]. Abstract models based on this paradigm have proved capable of forecasting dynamic processes. Clear examples include that of the control of segment polarity genes in Drosophila [16] or modelling of the neurotransmitter-signalling pathway between dopamine and glutamate receptors [17]. Yet the activities of cellular networks and signalling pathways are often subtler than this, which has resulted in various extensions being made to this model. One such refinement is *Qualitative Networks* (QNs), which uses discrete variables as opposed to Boolean states, and is able to model a much broader range of interactions by using algebraic target functions [10]. These target functions are composed of simple mathematical operations (e.g. addition, subtraction, division, multiplication) to allow for the generation of models with complex relationships between variables.

The BioModelAnalyzer (BMA) tool is a freely accessible online platform that creates QNs from user's instructions. These instructions are formed using a graphical interface, where different genes or proteins are represented by

simple symbols that can be connected by inhibitory or activatory edges negating the need for extensive experience in computer programming, logical formalisms or mathematical proofs [18]. As a result, the BMA is a highly accessible, unimposing interface that is suitable for experimental biologists, whilst still providing powerful stability checking, simulation and Linear Temporal Logic analysis abilities. Although on the surface BMA may appear to be highly abstract, elaborate biological functions can be robustly modelled such as that of *C. elegans* germline development [19], mammalian epidermis differentiation [10], gene and protein regulatory networks in chronic myeloid leukaemia (CML) [7] and acute myeloid leukaemia (AML) [20]. In the case of CML, a novel therapeutic strategy using an Imatinib and pan-Bcl2 family gene inhibitor combination has been identified, highlighting BMAs ability to work on either a hypothesis-creation or hypothesis-testing basis. Cell line specific differences in the PIM pathway were identified in the case of AML, leading to clinically relevant predictions about resistance and how to overcome it.

Although BMA provides the ability to encode complex dependencies between different genes or proteins via the use of algebraic target functions, this task can still seem quite onerous to many biologists. In 2003, Tyson, Chen and Novak [21] published a review outlining a concise mathematical vocabulary of common cellular interactions and pathways using ODEs. In their article, they identify a number of simple functional motifs, akin to electrical circuits which are found at the base of a variety of key biological processes and can be easily combined in order to model complex regulatory interactions. Here we outline a target function library which translates the ODEs outlined by Tyson

158 et al. [21] into discrete equations encoded within nodes of a BMA model. In
159 order to investigate whether these target functions are capable of modelling
160 cellular behaviours of greater complexity when combined, we then created a
161 BMA model of eukaryotic cell cycle regulation similar to Tyson et al [21]. *In silico*
162 over-expression and knockouts of combinations of genes and genetic
163 interactions, which were not used to generate the model, were then carried
164 out to highlight the sensitivity of our model. A key benefit of using discrete,
165 executable modelling is that complex systems can be simulated and analysed,
166 and experimentally testable hypotheses can be generated in the absence of
167 large amounts of quantitative data required for ODE models.
168
169 This library of ODE translations to discrete target functions also complements
170 the accessibility ethos behind the creation of the BMA. By providing simple
171 building blocks that can be “plugged” into a set of specific nodes, much time
172 and effort will be saved allowing biologists to construct elaborated valid
173 models of biological phenomena, which can guide and direct hypotheses and
174 ultimately drug treatments.
175

METHODS

Qualitative Networks

Qualitative networks (QNs) are an extension of Boolean models. In Boolean Networks, nodes are able to be in either and active (1), or inactive state (0), and are connected via functions that describe the mathematical relationship between them in an abstract way. Boolean Networks can be synchronous or asynchronous, that is – they may update every node simultaneously when a change is introduced in the system, or they can update in sequence from a propagation point. Qualitative networks are analogous to a synchronous Boolean Network, except that nodes are able to vary over a wide range of discrete values (called a granularity). Simple networks may be represented as Boolean, but Qualitative Networks may involve nodes with a greater range of values. For example, a node may have a range of 0-2 (granularity 3), where a value of 1 represents “normal activity” of an enzyme or gene product, and 0 and 2 represent low and high values respectively. This can be extended for much larger granularities, for example 0-10, where 10 represents maximal activity, and 0 represents minimal activity, with each discrete value in between representing a different concentration.

Nodes within a Qualitative Network are associated with either activatory or inhibitory relationships. Activatory relationships generally result in a response being high when a stimulus is high, and inhibitory relationships result in a response being low when a stimulus is high. Relationships between nodes are controlled by simple mathematical functions that describe the value that a

node should represent, given its current inputs, and this function is called a target function. The values of nodes within a Qualitative Network are updated simultaneously when the network is simulated, and nodes will change their values by a single integer (increase or decrease) each calculation step, in order to reach their target function gradually. Due to the synchronous and defined nature of Qualitative Networks, they are deterministic, and susceptible to formal verification techniques. QN's can stabilize and reach a single self-perpetuating state (called a stable point), but can also give rise to cycles, oscillations, and bifurcations.

Models and motifs described in this document are available in supplementary information and at https://github.com/shorthouse-mrc/biomodelanalyzer_targetfunctionlibrary.

The BioModelAnalyzer (BMA) Platform

The BMA is an accessible, publicly available (www.biomodelanalyzer.org) graphical tool for discrete modelling and analysis of Qualitative Networks. The platform, with its user-friendly graphical interface, uses visual notations familiar to specialists in biology. BMA models are constructed on a gridded canvas upon which one or more cells, and cell elements (i.e. membrane receptor, cellular proteins etc.) can be placed and connected together with activatory or inhibitory links. To create a model, the user starts by dragging and dropping a cell onto the gridded canvas. These cells have no functional role in the analysis, being purely a visual aid to assist model design clarity. Cell elements are then placed in or outside of these cells, which can represent internal proteins, external proteins or membrane bound receptors.

Connections between these cell elements can then be made using activatory arrows or inhibitory bar-arrows. Each cell element can then be labelled accordingly, using the simple drop down menus and a finite value range assigned, with the BMA default being [0,1], or Boolean. This range may be altered to add different levels of concentration, for example a range of [0,2] may represent “low”, “normal” and “high” concentrations of a protein or gene. If the user does not specify a target function for a node, then the BMA assigns a default target function. The default target function assigned within the BMA is described as:

$$average(activating\ inputs) - average(inhibiting\ inputs)$$

More complex target functions can be inserted for each node manually using an autocomplete function simplifying the use of correct syntax when referencing variables or using operators.

This underlying QN can then be analysed using simulation, stability analysis or Linear Temporal Logic tools each of which is accessible using the graphical interface. Simulation analysis shows the step-by-step execution of the model starting from a set point, based on either initial values specified by the user or a randomised start point. A graphical representation of all node values as they update over a user-defined number of time steps is produced, as well as a table of the simulation progression values, which can be exported as a CSV file for further analysis. Stability analysis can be used to test general properties of the model. If a model, given all possible starting conformations,

will always result in a same self-perpetuating state, it is considered stable, and the graphical interface presents the user with the “stable values”. If stability is not achieved, however, the interface presents whether the system results in bifurcations (can potentially end in multiple states depending on the starting conformation) or oscillations (results in an infinite cycle). More advanced queries can be asked using the Linear Temporal Logic (LTL) interface, which allows the user to define simple or complex temporal logic queries with a drag and drop interface. LTL queries will return True, True sometimes, False, and False sometimes responses to queries, and the interface allows the user to see examples of systems where the behaviour occurs.

Cell Cycle Model Generation

The model was composed of 3 main modules; G1/S, G2/M and M/G1 linked to a central node representing the level of Cdk1-cycB activity throughout the cycle. Each module was represented by a different cell in the BMA and labelled accordingly. The modules themselves were comprised of 6 key components namely; Cyclin, CKI, Wee1, Cdc25, APC and Cdc20, which regulate this Cdk1-cycB activity and thus the different cell cycle transitions (**Table 2**). These 6 components, modelled in their different chemical states (phosphorylated, active, inactive etc.) thus comprise a 20 node network, including 4 cell behaviours and 3 descriptive nodes linked by 28 interactions **(Supplementary Table 2 & 3).** Three members of the BMA target function library were combined to create the cell cycle model, with the granularity set to 11 (A range of 0-10). This granularity, which differs from the default of 5 in our

target function library, was chosen to accommodate the varying levels of Cdk1-cycB activity required, and to allow for clearer analysis of mutant phenotypes. Modules were initially generated based on the wiring and target functions from the BMA target function library examples, which were linked together through appropriate nodes **(Supplementary Figure 3).** This method resulted in the creation of individual pools of Cdk1-cycB activity at the different cell cycle phases that fed into one central pool of Cdk1-cycB activity. To better represent the biological system, the model was then refined, by simply combining the Cdk1-cycB individual pool target functions into a single node via compound addition of each target function within the target function interface. To allow for multiple rounds of cell division, rather than the simulation of a single cell cycle, modification to the mutual activation target function was required. The mutual activation target function defines a one-way switch, and as such is not reversible. Here only nodes S and A (see figure 1, e, ii) and their associated target functions were used, thus allowing the cell to return from the high state achieved following the critical switch point activation.

Table 2: Cross-species nomenclature of key nodes within each module

Module	Target Function	Node	Mammalian Cells	*Xenopus* embryo	Fission Yeast	Budding Yeast	Function
G1/S	Mutual Inhibition	CKI	$p27^{Kip1}$	Xic1	Rum1	Sic1	Stoichiometric cyclin-dependent kinase inhibitor
G2/M	Mutual Inhibition	Wee1	hWee1	Xwee1	Wee1	Swe1	Inhibitory kinase that inactivate Cdk-cyclin dimer

	Mutual Activation	Cdc25	Cdc25C	Xcdc25	Cdc25	Mih1	Activatory phosphatase that activate Cdk-cyclin dimer
M/G1	Negative Feedback Oscillator	APC	APC	APC	APC	APC	Anaphase-promoting complex
		Cdc20	$p55^{Cdc}$	Fizzy	Slp1	Cdc20	Degrades cyclin in complex with APC

Knock-Out (KO) & Overexpression (OP) Analysis

In order to show if a model can faithfully reproduce known biological perturbations, loss of function and gain of function mutations can be analysed in BMA. A list of genetic perturbations curated from the literature, and not used to generate the model, was created and used to test the model. In the case of KO mutations, the corresponding node within the model range was set to a range of 0-0, corresponding to a permanently inactive state. OP mutations were simulated by setting the corresponding node range to max-max, (i.e. max based on the chosen granularity) simulating a permanently active state. Simulation analysis is then carried out, and the results compared to the wild-type simulation. Differences were then compared to known biological behaviours.

RESULTS

A Target Function Library Accurately Reproduces Expected Biological Behaviour in Simple Networks

We constructed QN models representing the ten major archetypal regulatory and signalling pathways. Networks were generated within the BMA, and signal/response curves compared to previous publications [21–25] for accuracy. Networks are represented by a series of nodes interconnected via activatory (i.e. generally increasing target node value), and inhibitory (generally decreasing target node value) relationships. Nodes in the system can contain values with a granularity of 5 (a range of 0-4), but are generally easily extrapolated to different system ranges. Full details are included in **Supplementary Table 1**, and all models are available in supplementary data.

1. Linear Response

A system where the signal-response is linear (i.e. an increasing signal gives a proportionally increasing response) can be accurately modelled using the default target function. A node with no specified target function will have its value calculated by:

$$average(activating\ inputs) - average(inhibiting\ inputs)$$

A schematic linear signal-response network, from Tyson et al. [21] and built within the BMA is shown in **Fig 1, A, i &ii**, with signal-response curves from both systems shown in **Fig 1, A**, **iii & iv**.

Fig 1. Comparison of Signal-Response Elements. In this illustration, the rows correspond to **(A)** linear response **(B)** hyperbolic response, **(C)** sigmoidal response, **(D)** perfect adaption, **(E)** mutual inhibition, **(F)** mutual inhibition and **(G)** homeostasis as in Tyson et al. [21] The columns correspond to **(i)** Tyson et al. [21] wiring diagrams, **(ii)** BMA wiring diagram translation, **(iii)** Tyson et al. [21] signal-response curves and **(iv)** BMA equivalent signal-response curves; crosses represent stable steady states. Dark lines are interpreted outputs generated through linking stable steady-states, and represent an output that would be seen by sequentially altering the signal. Parts E and F are unique, in that they contain orange and blue crosses, representing increasing and decreasing alterations in the signal respectively. They also contain a shaded region indicating areas of instability. Each BMA wiring diagram contains a unique set of target functions located within particular nodes of the network which can be found in **Supplementary table 1**. For most cases clear comparison between Tyson et al. [21] wiring diagrams (i) and the corresponding BMA wiring diagrams (ii) can be made. Here like in Tyson et al. [21] S indicates the input Signal and R indicates the output Response with, in our case, letters A-C representing intermediate nodes. The graphs in (iv) are derived from simulation analysis carried out in the BMA. For all cases bar (d- iv) and (g-iv) the signal is altered from 0 through to 4 directly within the S node and the output in node R recorded and subsequently plotted. For cases (d- iv) and (g-iv) a simulation is run with a set signal input of 4 as an example, and the response output from the BMA simulation plotted based on the response per calculation time step. Graphs plotted from the BMA model (iv) can then be

compared to ODE counterpart (iii). In (e-iv) and (f-iv) the grey lines represent a series of updates linking fixpoints. In (e-iv) S_{crit}, which is denoted x in our target function (**Supplementary table 1)** represents the signal input where a switch in steady states will occur. The motif reproduces the bifurcation as expected (**Supplementary Figure 1**). Similarly, in (f-iv) S_{crit1} which is denoted y in our target function and S_{crit2} which is denoted z in our target function also correspond to the switch points in stable states.

2. Hyperbolic Response

We generated four ways to discretely model different hyperbolic functions within the BMA. This function describes a "phosphorylation and dephosphorylation" reaction and is modelled using a three-node wiring diagram shown **in Fig 1, B, ii**. A simple function included in node A results in a hyperbolic response as a result of a linearly increasing input. Node A contains the target function:

$$ceil\left(\left(\frac{3}{2}\right)(var(signal))\right)$$

Where $var(signal)$ represents the signal received by the network. This linear approximation captures the rapid initial growth of the response, whilst the plateau is enforced by the maximum value of the node. Additional modifiers (for details see **Supplementary Table 1**) can be included to change the shape and thresholds of the response. Hyperbolic signal-response curves from Tyson et al. [21] and from within the BMA are shown in **Fig 1, B, iii & iv**.

3. Sigmoidal Response

Sigmoidal response curves represent systems that act in a switch like manner, which are reversible and increase continuously with an increasing input. Schematics for sigmoidal signal-response networks are shown in **Fig 1, C, I & ii**. The target function for Node A contains the function describing the sigmoidal response. The wiring diagram is the same as that for the hyperbolic response (as it is in Tyson et al[21])), as alteration of the target function is enough to see the shift in behaviour – similar to ODEs. Multiple functions produce differing sigmoidal curves (for details see **Supplementary Table 1**). The simplest, producing a sigmoidal response from a linear signal is:

$$\left(floor\left(\frac{var(signal)}{x}\right)\right)y$$

Where x is the value at which the system switches between high and low values, and y is the upper value for the sigmoidal response. Signal-response curves are shown in **Fig 1, C, iii & iv**.

4. Perfect Adaptation Response

Adaptation is defined as “a process where a system initially responds to a stimulus, but then returns to basal or near-basal levels of activity after some period of time” [26]. Perfect adaptation is further characterised by the final response of the network returning to the exact pre-stimulus level. Perfect adaptation is used in numerous biological systems, for example, the Friedlander and Brenner [27] model of ion channel activation and inactivation.

The network is characterised by a node activating two further nodes, one of which inhibits the other, after a slight delay. This delay allows a signal to be transmitted, before the network inhibits the response for a short period, despite a sustained signal. Network schematics for perfect adaptation systems can be seen in **Fig 1, D, i & ii**. Perfect adaptation is modelled with the addition of the following target function to Node C:

$$x(var(A) - floor(var(B)))$$

Where x represents the maximal height of the response before the system adapts. In this function, the value of Node A determines the signal strength until it is inhibited by the accumulation of Node B, in the same manner as occurs in Tyson el al. Additionally, in order to ensure that the signal can reach its full strength before it is inhibited, an increase in the granularity of node B is required. In our system with a granularity of 5 (0-4), we increased the granularity of node B to 17 (0-16), to allow sufficient time for a signal to be realised before it is inhibited. Altering the granularity of this node will change the shape of the response in a manner analogous to the biological rate at which B is produced. An additional property which may be of importance to modelling perfect adaptation is desensitisation (that is, each successive peak after an increasing input becomes smaller than the previous). To include this functionality in the model we simply take into account the value of Node B in determining the activity of Node C, represented by altering the target function for Node C to (Signal-response curves can be seen in **Fig 1, D, iii & iv**):

20

$$\Big(x - floor\big(var(B)\big)\Big) * \big(var(A) - floor\big(var(B)\big)\big)$$

Where x represents the maximal height of the response.

5. Mutual Activation Response

Mutual activation behaviour represents irreversible cell switches, i.e. a "point-of-no-return". These discontinuous, one-way switches are typical of cell fate determination. Once a critical signal value (S_{crit}) is reached, the response immediately increases to a high level. A critical feature of mutual activation networks is that the switch is irreversible i.e. if the signal increases beyond S_{crit} and subsequently decreases, the response will not decrease. Network schematics for mutual activation systems are shown in **Fig 1, E, i & ii**. In the BMA schematic, the inclusion of a "self-loop" allows the node to be aware of the value of itself, and thus once a critical threshold is reached, increases to a maximally defined value irreversibly. Inclusion of the following target function in Node A is required:

$$floor\left(\frac{var(signal)}{x}\right) + var(C)$$

Where x is S_{crit} - the value at which the irreversible switch occurs. Also, the addition of the following target function in Node B allows the node to pass a point of no return at which point it cannot be decreased:

$$var(B) + var(A)$$

Analysing the stability of this system at a signal below S_{crit} results in a bifurcation, where the two stable steady states represent cases where a

simulation trace starts below and above the S_{crit} (**Supplementary Figure 1**). This is represented in **Figure 1, E, iv** by the inclusion of blue crosses, indicating state transitions under decreasing signal where the low level response can no longer be reached, and grey arrows representing large state transitions.

6. Mutual Inhibition Signal Response Curve

Mutual inhibition differs from mutual activation in that these systems exhibit hysteresis; if the input decreases below a defined critical value, then the output will return to zero. Tyson et al. [21] describe this type of feedback as a "toggle switch" , where there are two defined critical values; S_{crit1} and S_{crit2}, at which point the response will shift from either upper or lower values to the opposite. This is simplified below:

$$S \geq S_{crit1} \rightarrow R = R_{max}$$

$$S \leq S_{crit2} \rightarrow R = R_{min}$$

$$S_{crit1} < S_{crit2}$$

Essentially, this works similarly to mutual activation, except if S is decreased below Scrit2 then the switch will return to the inactive state. Our model is composed of 6 nodes and is compared to the "traditional" toggle switch schematic in **Fig 1, F, ii**. B is split into two separate nodes representing active and inactive states (B_{active}, and $B_{inactive}$ respectively), and it is the interactions between these 2 states of B that give rise to hysteresis. Whichever state reaches its critical value first overrides the other, resulting in it "winning" directing the network to stabilize at a particular state. This results in single

476 stable states being possible below and above the two critical states, but
477 bifurcations occurring in-between, where the starting state of the system
478 $S \geq S_{crit1}$ or $S \leq S_{crit2}$ determines which end state the system reaches. The
479 target function for the node representing active B (Node B_{active}) in a system
480 with a granularity of 0-4 is:
481

$$\left(x - var(B_{inactive})\right) + (1 - y)$$

482
483 Where x represents the maximal response of the network, and y is S_{crit1}. The
484 target function for the node representing inactive B (Node $B_{inactive}$) in a system
485 with a granularity of 0-4 is:
486

$$x\left(\left(var(A) + (4 - z)\right) - var(B_{active})\right)$$

487
488
489 Where x represents the maximal response for the network, and z is S_{crit2}.
490 Additionally, Node C contains the following target function:
491

$$x(var(A) - var(B_{active})$$

492
493 Where x represents the maximal response for the network. Signal-response
494 curves for this network are shown in **Fig 1, F, iii & iv**. Orange and blue
495 crosses represent increasing and decreasing changes in signal respectively,
496 and grey arrows represent large state transitions.
497

23

7. Homeostasis

Homeostatic regulation involves a network where the network counteracts the activity of the stimulus such that the response is constrained to a very narrow window (in the case of our network, a single value). A schematic homeostasis network is shown in **Fig 1, G, I & ii**. In this network, the granularity of Node B is adjusted such that it is double the range of the other nodes within the system. We find that the increased range leads to a "slower" rate of change, and this stabilizes the system, preventing oscilations between Node A and B, which occur without the granularity difference. In this network, two target functions are required to exhibit homestasis, for Node A:

$$(2 * var(signal)) - floor\left(\frac{2}{3}var(B)\right) - 1$$

And for the node representing the system response (Node Response):

$$ceil\left(\frac{var(A)}{3}\right)$$

In this case, extreme perturbations of the signal (lowest or highest possible in the network – 0 or 4) will result in a change in the response. This is a characteristic of many biological homeostatic networks, such as osmotic regulation, in which perturbation at extreme signals provides the stimulus to enact control mechanisms[28]. In this model, changes in the signal within the "homeostatic" range (i.e not at extremes), results in a transient change in the response, before the difference is rectified (**Figure 1, G, iv**). Enacting a model

24

in which the signal does not alter the response at all (as demonstrated in Tyson et al[21]) is also possible, with the addition of a separate node to slow the passage of signal from Node A to Node R (effectively changing the rate at which changes in Node A are felt by Node R). In this case, changes to the signal within the homeostatic bound (in our model signals of strength 1-3) does not alter the response at all (see model Homeostasis-slow in the supplementary).

8. Negative Feedback Oscillations

Negative feedback oscillations result from similar network wiring as homeostasis, with the result being a system where the response oscillates between 0 and the signal value.

A negative feedback oscillation loop is seen in **Fig 2, A**. No change in the default target functions are required to generate an oscillatory output. For these networks, an input of S will result in an oscillation that tends to between S and 0. The temporal constraints of the network however (in that each node can only update by a single integer value each step) results in cases where the oscillation will not reach the maximal value before the inhibitory portion of the network kicks in. The ultimate range of the oscillations can be tailored however, with either the addition of values to the output node (in order to adjust the oscillation range up or down), or by inserting the following formula into node A:

$$(var(signal) + x) - (var(C)) + y)$$

Where the difference between x and y changes the range of the oscillations. Additionally, the temporal properties of the system, specifically how long it takes to perform each loop, can be adjusted by the addition of more nodes to the loop, with a large number of nodes increasing the number of steps required to complete one oscillation. For an example see **Supplementary Figure 2**.

Fig 2. Comparison of Oscillatory Networks. In this illustration, the rows correspond to **(A)** negative feedback, **(B)** activator-inhibitor and **(C)** substrate-depletion oscillators as in Tyson et al. [21] The columns correspond to **(i)** Tyson et al. [21]wiring diagrams, **(ii)** BMA wiring diagram translation, **(iii)** Tyson et al. [21] signal-response curves and **(iv)** BMA equivalent signal-response curves. Each BMA wiring diagram contains a unique set of target functions located within particular nodes of the network which can be found in **Supplementary table 1**. For most cases clear comparison between Tyson et al. [21] wiring diagrams (i) and the corresponding BMA wiring diagrams (ii) can be made. Here like in Tyson et al. [21] S indicates the input Signal and R indicates the output Response with, in our case, letters A-E representing intermediate nodes. The graphs in (iv) are derived from simulation analysis carried out in the BMA. For all cases bar a simulation is run with a set signal input of 2 as an example, and the response output from the BMA simulation plotted based on the response per calculation time step and are thus not directly comparable, however clear oscillatory behaviour can still be observed.

9. Activator-Inhibitor Oscillations

The activator-inhibitor oscillation relationship is characterised by a positive and negative feedback loop within a system (shown in **Fig 2, B, i and ii**). The interactions of the two loops result in a system that oscillates between a maximal and minimal value, called a hysteresis oscillator. Including the following formula in node A results in an oscillation between the maximal and minimal values of the nodes when I = 2:

$$\Big(x\big(var(signal) - (var(E) + y)\big)\Big) + var(A)$$

Where x is 3 value of the nodes, and y is 0. Adjusting these values will change the range of the oscillations (i.e a range of 3 or 2 is obtained by reducing the value of x), and altering the value of y adjusts the start and end points of the oscillation. The signal-response curves for activator-inhibitor networks are shown in **Fig 2, B, iii and iv**).

10. Substrate-Depletion Oscillations

The substrate depletion oscillation (SDOs) is quite similar to that of negative feedback. However, the number of nodes are reduced to reflect the greater intimacy between enzyme-substrate reactions compared to negative feedback loops. The network schematics for substrate-depletion oscillations are shown in **Fig 2, C, i**. In substrate-depletion oscillations, a small signal produces a

small response and a large signal produces a large response. To model this, the following target function is applied to Node A:

$$1 + \left(floor\left(\frac{var(signal)}{x} \right) \right) * \left(\left(\frac{y}{2} \right) * var(signal) \right) - \left(\frac{z}{2} * var(B) \right)$$

Where x is the starting point of the oscillations and y and z are the range of the oscillations. Signal-response curves are presented in **Fig 2, C, ii**.

Using the BMA Target Function Library to Construct Complex Networks: Eukaryotic Cell Cycle Control

After translating these motifs that were originally defined with ODEs to BMA models and their target functions, we then sought to determine the robustness and reusability of these motifs and their target functions by modelling a complex cellular behaviour: Eukaryotic cell cycle regulation. Based on the wiring diagram presented by Tyson et al. [21], a QN model was constructed using our own BMA target function library (**Fig 3, A**). Clear descriptions of the dynamics of cell cycle regulation can be found in the following review articles: Tyson, Csikasz-Nagy & Novak (2002) [24], Tyson & Novak (2015) [22] and Hochegger, Takeda & Hunt (2008) [29].

Fig 3. Qualitative Network of Eukaryotic Cell Cycle Regulation. (A) BMA Wiring diagram. The network is constructed around a central pool of the major

cell cycle regulator cyclin dependent kinase (Cdk1) and its cyclin partner (cycB). This cell cycle transitions are triggered by changes in the Cdk1-CycB activity, which is regulated by a number of different components. CKI a cyclin kinase inhibitor and Wee1 kinase subunit inactive the Cdk1-CycB complex whereas the Cdc25 phosphatase activates the complex. Cdk1-CycB activity can also be destroyed via the Anaphase-promoting complex (APC) in combination with Cdc20, which target cyclin for degradation. The activities of the Cdk1-CycB activity can then be monitored by 3 extracellular markers; G1S, G2M and MG1. **(B)** BMA simulation of Cdk1-CycB activity. The solid black line indicates the progression of Cdk1-CycB levels through the cycle. Dotted lines and block colours represent distinct phases as determined by the key. The cycle repeats itself if growth conditions remain favourable, as is represented in this simulation.

Pool Module

This module contains the “master molecules” of the cell cycle, that being cyclin-dependent kinases (Cdks) and their cyclin partner, which as the name suggests are required in order to activate the Cdks. Our model is limited to only a single Cdk-Cyclin partnership, Cdk1-CycB for simplicity. This module is fuelled by the growth of cyclin levels which we assume can have unlimited binding capacity to Cdk1. Unlike cycB, intracellular Cdk1 concentration does not fluctuate throughout the cell cycle [30], we therefore model Cdk1 as being at a constant level which can accommodate the variations in cycB levels [22].

G1/S Module

The G1/S module features mutual inhibition between Cdk1-CycB and CKI. This feedback loop is described as a “toggle switch” and is modelled using our BMA mutual inhibition target function. Here we model CKI as being present at high levels in G1 by assigning it an initial value of 10 (max based on our granularity choice). The input in this case is labelled as cyclin, which, as it increases causes an increase in bound_Cdk1 (i.e. heterodimer of CKI, Cdk1 & CycB) due to the initial high levels of CKI. As the CKI doesn’t stop cyclin accumulation and binding to Cdk molecules, the rising Cdk1-cycB levels which are not opposed by CKIs soon tips the balance, phosphorylating the CKI and labelling them for degradation. The values chosen for the switch points can be found in **Supplementary table 2**.

G2/M Module

Following the degradation of CKI and subsequent spike in Cdk1-CycB activity the cycle enters the G2/M module. This module features both mutual activation, between Cdk1-CycB and Cdc25, and mutual inhibition between Cdk1-CycB and Wee1 [21]. The later works in a similar way to that of Ck1-CycB and CKI, with a race occurring being between Cdk1-cycB and Wee1. The Cdk-CycB and Cdc25 mutual activation interaction on the other hand is a type of positive feedback loop, where Cdc25 and Cdk1-CycB activate each other rather than inhibit each other. This is modelled using our BMA mutual inhibition target function combined with the mutual activation target function. Here we model Wee1 as being present at high levels in G2/M by assigning it an initial value of 10. The input in this case comes from the G1_Cdk1 levels,

665 which as it increases causes an increase in phos_Cdk1 (i.e. phosphorylated
666 form of Cdk1-CycB) due to the initial high levels of Wee1. As Wee1 does not
667 stop cyclin accumulation and binding to Cdk molecules, the rising Cdk1-CycB
668 levels (which are not opposed by Wee1) soon tips the balance,
669 phosphorylating the Wee1 and marking them as inactive. Inactive Wee1
670 maintains active Cdc25, thus decreasing Wee1 results in an increase in
671 Cdc25 and thus the switch like activation of Cdk1-CycB. Again, the values
672 chosen for the switch points can be found in **Supplementary table 2**. We
673 note that an alternative network model could connect phos_Cdk1 to cdc25.
674 However, we made the choice to reduce the number of connections by
675 encoding this behaviour through the interactions with Wee1.
676
677 **M/G1 Module**
678 Once the Cdk1-CycB reaches a high level due to Cdc25 activation the cell
679 enters mitosis. In order to exit this phase, the Cdk1-CycB activity must be
680 destroyed and CKI levels stockpiled. This transition is aided by the
681 Cdc20:APC complex, which itself is indirectly activated by Cdk1-CycB activity,
682 causing degradation of CycB. This results in a substantial drop in Cdk1-CycB
683 activity, which then allows CKI to rise again. This relationship is described as
684 an oscillator based on a negative feedback loop, where Cdk1-CycB activates
685 APC, which activates Cdc20, which then degrades CycB [21]. In the BMA, the
686 negative feedback oscillators target function uses the default function.
687 Therefore this was modelled simply by considering the whole cycle as a
688 feedback loop by adding an inhibitory edge back to the cyclin B in order to
689 create the desired response **(Supplementary table 2 & 3).**

31

To measure which stage of the cell cycle the model is in additional nodes were added representing G1S, MG1, and G2M states. G1S responds to the levels of Cdk1_cycB_Pool, MG1 responds to the levels of APC, and G2M is reliant on the model having completed MG1, and G1S before it can be observed.

Comparison to ODE Eukaryotic Cell Cycle Model Predictions

The initial conditions were set so that all nodes remained with an initial value of 0, except for CKI and Wee1 which are given an initial value of 10 (max based on our granularity choice). As Growth, Replicated_DNA, Undamaged_DNA and Aligned_Chromosomes are conditions that can be represented by a binary value, a value of 0 represents the absence of the cell phenotype, whereas a value of 1 corresponds to the presence of the phenotype. The initial values for all four of these phenotypes were therefore set to 1 to represent normal growth conditions. Simulation analysis, starting from this initial state leads to the initiation of a series of network states (ranging between 0 and 10 based on our granularity). These steps correspond to the biological time series of protein activation and inactivation that occur during the wild-type cell cycle (**Fig 3, B**).

Similar to the Tyson et al. [21] signal-response curve, **Fig 3, B** shows the cell progresses through the cycle via a number of steady states. Firstly, at low levels of Cdk1-CycB activity the cell will remain in G1. With increased growth

it will eventually pass a critical point, resulting in the irreversible disappearance of G1. As the cell moves into the S phase the level of Cdk1-CycB continues to grow until it reaches an intermediate level (3 as determined by our target function). Here in the G2 phase the cell will continue to grow until it reaches the next critical threshold, where the G2 state will disappear. This gives rise to a large spike in Cdk1-CycB activity (driving the cell into mitosis) which then decreases as cycB is degraded by APC:Cdc20, signalling cell division and resetting the system for the next round of division. One added benefit of this model is its ability to continuously cycle, as highlighted in **Fig 3, B**.

Simulation of Mutant Phenotypes Replicate Experimental Results Found in the Literature

In order to evaluate the accuracy of our model loss of function (KO) and over-expression (OP) mutations were carried out based on a sample of previous experiments found in the literature (**Table 1**), and not used in the generation of either our model, or the model presented in Tyson et al [21]. In our limited subset of mutant experiments 8 out of 9 cases were able to accurately replicate the experimental results found in the literature without making any modifications to the underlying model described above beyond modelling the mutations (**Fig 4**). For instance in the case of *cdc25 OP*, studies in both yeast and mice have shown that over-production of Cdc25 result in premature entry into mitosis due to early activation of Cdk1-CycB [31,32]. In the in silico experiment, the same result can be discerned. Rather than needing 8 steps to

pass through G2 the *cdc25 OP* model only takes one step. Similarly less time is spent in M phase with only 1 step occurring versus 2 steps for wild types (WT). This results in the mutant model undergoing each cycle in fewer calculation steps, needing only 33 steps compared to the 41 needed in the WT model. Descriptions of the other seven successfully reproduced experiments can be found in **Table 1**. For the case concerning the *cki OP* mutant, experimental results were not as clearly reproduced. Experimental evidence by Moreno & Nurse [33] showed that overexpression of Rum1, a fission yeast CKI, leads to delays in G1, with repeated S-phase and no M-phase. This is partially replicated in our mutant model, with there being a long delay in G1 phase (23 calculation steps compared to 14 in the WT model), as well as no M phase being reached (where Cdk1-CycB hits max value of 10). The model however still runs through the M/G1 phase rather than just repeating the S phase.

Table 1: Mutant simulations reproduce described behaviour from the literature. Summary of experimental results are given in "Expected Outcome", and in silico results are given in "Model Outcome".

Genetic Perturbation	Source	Expected Outcome	Model Output
wee1Δ	Yeast - Nurse 1975[34] Mammal - Tominaga 2006[35]	Premature entry into mitosis, with long G1, short G2, but still viable	G1 same length, Short G2, cell cycles quicker
wee1 OP	Arabidopsis – De Schutter 2007[36]	Cell cycle blocked in G2	Arrest in G2
ckiΔ	Yeast – Lengronne 2002[37]	Short G1, extended G2, increased activation of Cdks	Short G1, extended G2
cki OP	Yeast – Moreno 1994[33]	In yeast delays G1 followed by	Long delay in G1, cycles but

34

		multiple S & no M	no M phase
wee1Δ *ckiΔ*	Yeast – Sveiczer 2000[38]	Cell divides very quickly, cell gets smaller with each division	Divisions occur over less time steps
cdc25Δ	Yeast – Russell 1986[31] Mammal – Lee 2009[39]	Cell cycle blocked in G2	Arrest in G2
wee1Δ *cdc25Δ*	Yeast – Davidich 2013[40]	Cell not viable, cannot enter mitosis	Arrest in G2
cdc25 OP	Yeast – Russell 1986[31] Mammal – Timofeev 2010[32]	Premature entry into mitosis, early activation of Cdk-cyc	Short G2, cell cycles quicker
cdc20Δ	Yeast – Kim 1998[41] Mammal – Li 2007[42]	Lethal, cannot complete mitosis	Arrest in M

758

759 **Figure 4. Mutant Phenotype Simulation Analysis.** Depicts the temporal
760 evolution of the network following perturbation of particular nodes. Each
761 mutant perturbation can be compared to the wild type, which is listed first.
762 Each distinct cell cycle phase is coloured coded according to the key
763 provided. Each time step corresponds to each calculation step recoded in the
764 BMA simulation which is exported as a CSV file.

35

DISCUSSION

We present a library of novel Qualitative Network modules that can accurately replicate the biological behaviour of core, ubiquitous network motifs. We generate and compare our library based on biological behaviours defined previously [21], and confirm the modular nature of the library with the generation of a model for the eukaryotic cell cycle produced using motifs from the library. By simulating known genetic perturbations we further test this novel qualitative eukaryotic cell cycle model, highlighting its capacity to accurately replicate many well-known mutant phenotypes without the need for explicit parameterisation, as would generally be needed for ODE models. This study constitutes both a toolbox for biologists to construct elaborate networks with ease, but also an example of its application to a relevant biological system. The QN presented has much wider applications, with our working model having the potential to be adapted in order to provide much more dynamic details on the regulation of these core cell cycle components. Such a model could then be utilised to provide new insights into cell cycle regulation allowing the prediction of novel mutant phenotypes that have not been previously investigated. Not only could this provide a more thorough understanding of the underlying cell cycle regulatory principles, but also assist in the identification of a host of mutants that contribute to cancers or other pathologies, potentially allowing for the identification of novel therapeutics, as has been demonstrated previously [43]. Similarly, in compiling this simple, easy to use BMA target function library we hope to encourage experimentalists to adopt this type of QN modelling as part of mainstream biological research. This would offer a wealth of advantages in terms of consolidating what is known about large

networks into concise descriptions, as well as by allowing the generation of novel predictions about systems in the absence of large amounts of data and thus help focus experimental design. In particular, the accessibly, and fact that the generated networks have a finite set of states, is an attractive concept, particularly for studying diseases of "rare events", such as cancer. Additionally, the ability of the BMA to construct complex models in the absence of precise kinetic data means that resultant models can be tested through proof-based analyses, and thus may represent a more appropriate abstraction in some cases than the equivalent network in ODEs, where fitting is used to generate functions to produce desired results. Another potential use of qualitative modelling is in multiscale systems, as an intermediary between Boolean and mathematical systems, or linking spatial dynamics to signalling, as has been published previously [19,44].

Through the construction of our BMA motif/target function library we have been able to capture the dynamic behaviours of simple cell signalling pathways. Although networks can be modelled using ODEs, with behaviours being predicted using numerical simulations, this requires more complex and harder-to-obtain biological data and may equally appear mathematically complex to many biologists. As shown through the analysis of a model of eukaryotic cell cycle regulation, a relatively simple QN model can capture many of the advanced dynamic features of ODE models, including multistability and bifurcations. Simulation analysis of the described model shows strong similarities to that of the quantitative biological signal response curve, first proposed by Stern & Nurse [45], which was based on the results of

multiple Cdk and cyclin knockout experimental studies. Like our model, they described the cycle as having three distinctive phases of Cdk activity, with the Cdk1-CycB levels transitioning through the cell cycle via different levels or bifurcations [24,45.] These levels or bifurcations are representative of firstly a stage of inactivity (G1) where Cdk activity remains low, secondly a stage of moderate Cdk activity sufficient to trigger S phase, and lastly a stage of high Cdk activity sufficient to initiate mitosis, all of which can be easily recognised in our model simulation [45]. This ability to model varying levels of Cdk activity sets our model apart from its Boolean counterparts, where only two levels of detail ("on" or "off") can be captured. Through simple manipulation of our target functions we were also able to capture an extra layer of detail, by allowing our model to continue cycling over unceasing divisions when conditions remain favourable, a behaviour which has not always been replicated in previous studies [21,40,46]. The addition of this extra layer of complexity, showing sustained cell cycle oscillations, results in a model that is more representative of the true clock-like oscillatory nature of the cell cycle [47]. It is worth noting, however, that our model does not contain continuous, biologically measurable values for components, and as such is limited in its ability to interpret continuous experimental data.

As a means to further validate the model, loss of function and overexpression mutants were simulated, with the simplicity and generality of the model limiting the number of mutant phenotypes studies. Regardless of the simplification of using discrete modelling to represent continuous protein concentrations and interactions, the BMA model was capable of correctly modelling 8 out of 9

mutant phenotypes studied. All knockout mutations were correctly reproduced, with the model capturing dynamic properties such as phase length changes. For over-expression models, where the corresponding node range is set to max-max, 90% of the OP mutants studied corresponded accurately with experimental data, with *cki OP* mutants being in partial agreement. This partial agreement is likely due to the minimalistic nature of our model, and could likely be overcome by using additional nodes to model the CKI interaction in more detail. Overall, the model produced using the BMA target function library accurately represents not only the WT regulation patterns of the general cell cycle control engine, but also the dynamic changes resulting from a number of mutants. This showcases our BMA target function library's ability to be easily manipulated in order to model complex networks. Of particular note is the ability of the method to accurately generate protein behaviour through the simple addition of target functions from different modules that act on the same proteins, as is the case with the Cdk1-CycB node in our QN. This ability to draw together simple motifs to create realistic and useful biological networks demonstrates the validity of the approach and the opportunities that executable modelling makes available.

39

ACKNOWLEDGEMENTS

We thank the Fisher and Hall groups for useful discussions. Work was supported by Medical Research Council core funding (D.S.) and Royal Society grant UF130039 (B.A.H).

AUTHOR CONTRIBUTIONS

BH, CB, NP and JF conceived the study. DS and MP generated the target function library, YP developed the model of the cell cycle. DS and YP co-wrote the manuscript. All authors were responsible for editing of the manuscript.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interest.

REFERENCES

1 T. M. Yi, Y. Huang, M. I. Simon and J. Doyle, *Proc. Natl. Acad. Sci.*, 2000, **97**, 4649–53.

2 A. Esmaeili, T. Davison, A. Wu, J. Alcantara and C. Jacob, *BMC Bioinformatics*, 2015, **16**, 311.

3 C. Webb and J. Wang, *Nat. Publ. Gr.*, 2016, **6**, 1–12.

4 K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak and J. J. Tyson, *Mol. Biol. Cell*, 2004, **15**, 3841–62.

5 F. Ortega, J. Stott, S. A. G. Visser and C. Bendtsen, *J. Biol. Chem.*, 2013, **288**, 785–92.

6 J. Fisher and T. A. Henzinger, *Nat. Biotechnol.*, 2007, **25**, 1239–49.

40

7 R. Chuang, B. A. Hall, D. Benque, B. Cook, S. Ishtiaq, N. Piterman, A. Taylor, M. Vardi, S. Koschmieder, B. Gottgens and J. Fisher, *Sci. Rep.*, 2015, **5**, 8190.

8 K. Claessen, J. Fisher, S. Ishtiaq, N. Piterman and Q. Wang, .

9 B. Cook, J. Fisher, E. Krepska and N. Piterman, 2011, pp. 134–149.

10 M. A. Schaub, T. A. Henzinger and J. Fisher, *BMC Syst. Biol.*, 2007, **1**, 4.

11 Clarke Edmund M., O. Grumberg and D. Peled, 1999.

12 J. Fisher, N. Piterman, A. Hajnal and T. A. Henzinger, *PLoS Comput. Biol.*, 2007, **3**, e92.

13 S. Nusser-Stein, A. Beyer, I. Rimann, M. Adamczyk, N. Piterman, A. Hajnal and J. Fisher, *Mol. Syst. Biol.*, 2012, **8**, 618.

14 S. A. Kauffman, *J. Theor. Biol.*, 1969, **22**, 437–467.

15 L. Glass and S. A. Kauffman, *J. Theor. Biol.*, 1973, **39**, 103–129.

16 M. Chaves, R. Albert and E. D. Sontag, *J. Theor. Biol.*, 2005, **235**, 431–449.

17 S. Gupta, S. S. Bisht, R. Kukreti, S. Jain and S. K. Brahmachari, *J. Theor. Biol.*, 2007, **244**, 463–469.

18 D. Benque, S. Bourton, C. Cockerton, B. Cook, J. Fisher, S. Ishtiaq, N. Piterman, A. Taylor and M. Y. Vardi, Springer, Berlin, Heidelberg, 2012, pp. 686–692.

19 B. A. Hall, N. Piterman, A. Hajnal and J. Fisher, *Biophys. J.*, 2015, **109**, 428–438.

20 D. Silverbush, S. Grosskurth, D. Wang, F. Powell, B. Gottgens, J. Dry and J. Fisher, *Cancer Res.*, 2017, **77**, 827–838.

41

21 J. J. Tyson, K. C. Chen and B. Novak, *Curr. Opin. Cell Biol.*, 2003, 15, 221–231.

22 J. J. Tyson and B. Novák, *BMC Biol.*, 2015, **13**, 46.

23 A. Csikász-Nagy, D. Battogtokh, K. C. Chen, B. Novák and J. J. Tyson, *Biophys. J.*, 2006, **90**, 4361–79.

24 J. J. Tyson, A. Csikasz-Nagy and B. Novak, *BioEssays*, 2002, 24, 1095–1109.

25 J. J. Tyson, K. Chen and B. Novak, *Nat. Rev. Mol. Cell Biol.*, 2001, **2**, 908–916.

26 J. E. Ferrell, *Cell Syst.*, 2016, 2, 62–67.

27 T. Friedlander and N. Brenner, *Proc. Natl. Acad. Sci.*, 2009, **106**, 22558–22563.

28 L. K. Putney, S. P. Denker and D. L. Barber, *Annu. Rev. Pharmacol. Toxicol.*, 2002, **42**, 527–552.

29 H. Hochegger, S. Takeda and T. Hunt, *Nat. Rev. Mol. Cell Biol.*, 2008, **9**, 910–916.

30 D. L. Fisher and P. Nurse, *EMBO J.*, 1996, **15**, 850–60.

31 P. Russell and P. Nurse, *Cell*, 1986, **45**, 145–153.

32 O. Timofeev, O. Cizmecioglu, F. Settele, T. Kempf and I. Hoffmann, *J. Biol. Chem.*, 2010, **285**, 16978–16990.

33 S. Moreno and P. Nurse, *Nature*, 1994, **367**, 236–242.

34 P. Nurse, *Nature*, 1975, **256**, 547–551.

35 Y. Tominaga, C. Li, R.-H. Wang and C.-X. Deng, *Int. J. Biol. Sci.*, 2006, **2**, 161–170.

36 K. De Schutter, J. Joubès, T. Cools, A. Verkest, F. Corellou, E.

42

933 Babiychuk, E. Van Der Schueren, T. Beeckman, S. Kushnir, D. Inzé and
934 L. De Veylder, *Plant Cell*, 2007, **19**, 211–225.
935 37 A. Lengronne and E. Schwob, *Mol. Cell*, 2002, **9**, 1067–1078.
936 38 A. Sveiczer, A. Csikasz-Nagy, B. Gyorffy, J. J. Tyson and B. Novak,
937 *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7865–70.
938 39 G. Lee, L. S. White, K. E. Hurov, T. S. Stappenbeck and H. Piwnica-
939 Worms, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4701–4706.
940 40 M. I. Davidich and S. Bornholdt, *PLoS One*, 2013, **8**, e71786.
941 41 S. H. Kim, D. P. Lin, S. Matsumoto, A. Kitazono and T. Matsumoto,
942 *Science*, 1998, **279**, 1045–1047.
943 42 M. Li, J. P. York and P. Zhang, *Mol. Cell. Biol.*, 2007, **27**, 3481–3488.
944 43 J. C. Sible and J. J. Tyson, *Methods*, 2007, **41**, 238–47.
945 44 C. C. (2014) P.L. Varela, P.T. Monteiro, N.D. Mendes, A. Fauré, in
946 *Poster at 13th European Conference on Computational Biology*
947 *(ECCB'14), Strasbourg, France*, 2014.
948 45 B. Stern and P. Nurse, *Trends Genet.*, 1996, 12, 345–350.
949 46 C. Hong, M. Lee, D. D. Kim, D. D. Kim, K.-H. Cho and I. Shin, *BMC*
950 *Syst. Biol.*, 2012, **6**, 129.
951 47 J. E. Ferrell, T. Y. C. Tsai and Q. Yang, *Cell*, 2011, 144, 874–885.

952

953

954 ## SUPPLEMENTARY INFORMATION

955

956 **Supplementary Figure 1. Bifurcation proof for Mutual Activation Motif.** In
957 this figure, shown is the network for mutual activation in the BMA. Stability

analysis has been run (right), and the further testing interface used to explore why the model does not have a single stable state. Shown is the result of the further testing, proving that the model has two stable states (a bifurcation), showing that the model can stabilize in two conditions, dependent on whether the signal previously passed the critical threshold.

Supplementary Figure 2. Adjustment of Negative Feedback Oscillatory Module. In this illustration, the rows correspond to **(A)** a one node system **(B)** a two node system and **(C)** three node system of negative feedback. The columns correspond to **(i)** the BMA wiring diagram translation and **(ii)** the BMA response curves. Each BMA wiring diagram contains a unique set of target functions located within particular nodes of the network which can be found in **Supplementary table 1**. S indicates the input Signal and R indicates the output Response with, in our case, letters A-D representing intermediate nodes. The graphs in (ii) are derived from simulation analysis carried out in the BMA. For all cases a simulation is run with a set signal input of 2 as an example, and the response output from the BMA simulation plotted based on the response per calculation time step. Comparison of the three different node length systems highlights that with increased number of nodes there is an increased length of oscillation response, as shown in the 20 calculation steps graphed, with the three node system **(C)** carrying out only 2 full oscillations compared to the one node system **(A)** which carries out 4 oscillations in the 20 calculation steps.

Supplementary Figure 3. Wiring Diagram. Wiring diagram composed using the network topology specified from the BMA target function library modules. Cdk-cycB activity is subdivided into individual pools which link to a central Cdk1-cycB pool. Subsequent models combine the target functions of the G1 Cdk-cycB activity and the G2-cdk-cycB activity together into one node which better represents the true biology.

Supplemental Table 1. List of networks assessed in the manuscript. Paper reference refers to the figure in which the network occurs, included are the target functions for each node (if not the default), and comments. Included are the filenames and model names where the specific network can be found.

Supplemental Table 2. List of nodes in the network. Network ID refers to the internal label of the node. Full name is the common name found in the literature while Network Name is the name given in construction of model.

Supplemental Table 3. List of nodes in the network. Network ID refers to the internal label of the node. Full name is the common name found in the literature while Network Name is the name given in construction of model. The target function located in each node is found under the "Target Function" heading.

Model Files. We have additionally included all model files in .json format within an enclosed .zip file, they are also available at https://github.com/shorthouse-mrc/biomodelanalyzer_targetfunctionlibrary. Importing any file into the BMA will load the model and allow manipulation and simulation/stability analysis. Each file is named explicitly in Supplementary

45

Table 1, with some files containing multiple models, which are referenced independently.

Insight Box

A limitation of popular ODE models is that they require complete networks and detailed kinetic parameterisation. An alternative is the use of discrete, executable models, in which nodes are assigned discrete value ranges, and the relationship between them defined with logical operations. A fundamental question for executable models however is whether the high level of abstraction substantially reduces expressivity relative to continuous approaches. Here, we present a canonical library of biological signalling motifs, initially defined by Tyson et al (2003), expressed using the BioModelAnalyzer. We show that; 1) these motifs are easily and fully translatable from continuous to discrete models, 2) Combining these motifs generates a fully functional and predictive model of the yeast cell cycle.

(116 words)

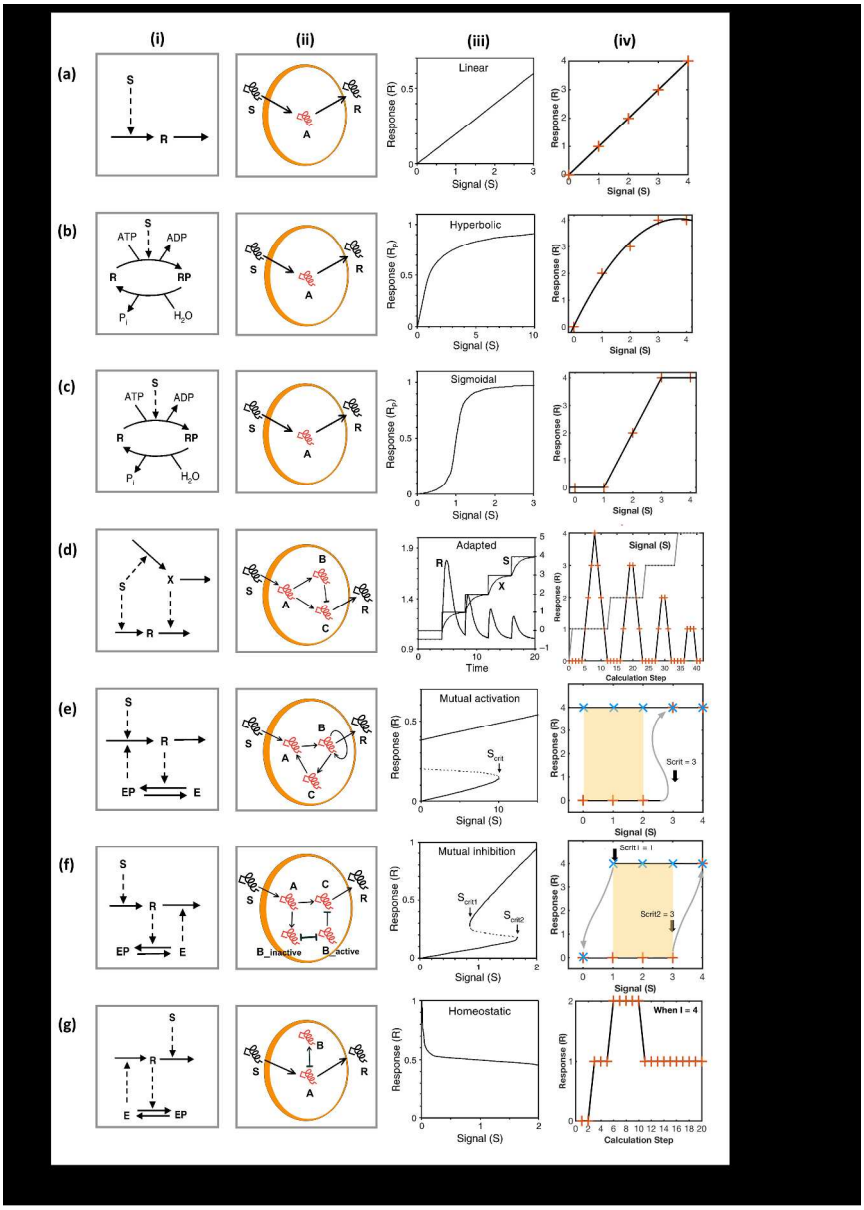


Fig 1. Comparison of Signal-Response Elements. In this illustration, the rows correspond to (A) linear response (B) hyperbolic response, (C) sigmoidal response, (D) perfect adaption, (E) mutual inhibition, (F) mutual inhibition and (G) homeostasis as in Tyson et al. 21 The columns correspond to (i) Tyson et al. 21 wiring diagrams, (ii) BMA wiring diagram translation, (iii) Tyson et al. 21 signal-response curves and (iv) BMA equivalent signal-response curves; crosses represent stable steady states. Dark lines are interpreted outputs generated through linking stable steady-states, and represent an output that would be seen by sequentially altering the signal. Parts E and F are unique, in that they contain orange and blue crosses, representing increasing and decreasing alterations in the signal respectively. They also contain a shaded region indicating areas of instability. Each BMA wiring diagram contains a unique set of target functions located within particular nodes of the network which can be found in Supplementary table 1. For most cases clear comparison between Tyson et al. 21 wiring diagrams (i) and the corresponding BMA wiring diagrams (ii) can be made. Here like in Tyson et al. 21 S indicates the input Signal and R indicates the output Response with, in our case, letters A-C representing intermediate nodes. The graphs in (iv) are derived from

simulation analysis carried out in the BMA. For all cases bar (d- iv) and (g-iv) the signal is altered from 0 through to 4 directly within the S node and the output in node R recorded and subsequently plotted. For cases (d- iv) and (g-iv) a simulation is run with a set signal input of 4 as an example, and the response output from the BMA simulation plotted based on the response per calculation time step. Graphs plotted from the BMA model (iv) can then be compared to ODE counterpart (iii). In (e-iv) and (f-iv) the grey lines represent a series of updates linking fixpoints. In (e-iv) Scrit, which is denoted x in our target function (Supplementary table 1) represents the signal input where a switch in steady states will occur. The motif reproduces the bifurcation as expected (Supplementary Figure 1). Similarly, in (f-iv) Scrit1 which is denoted y in our target function and Scrit2 which is denoted z in our target function also correspond to the switch points in stable states.

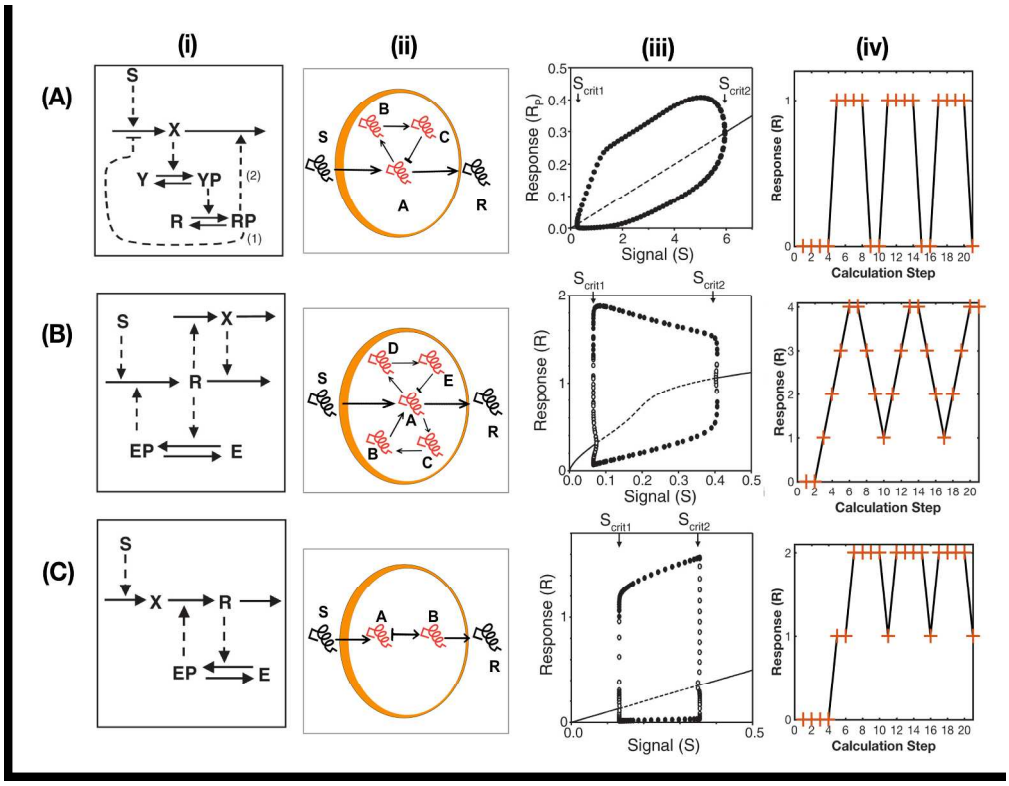
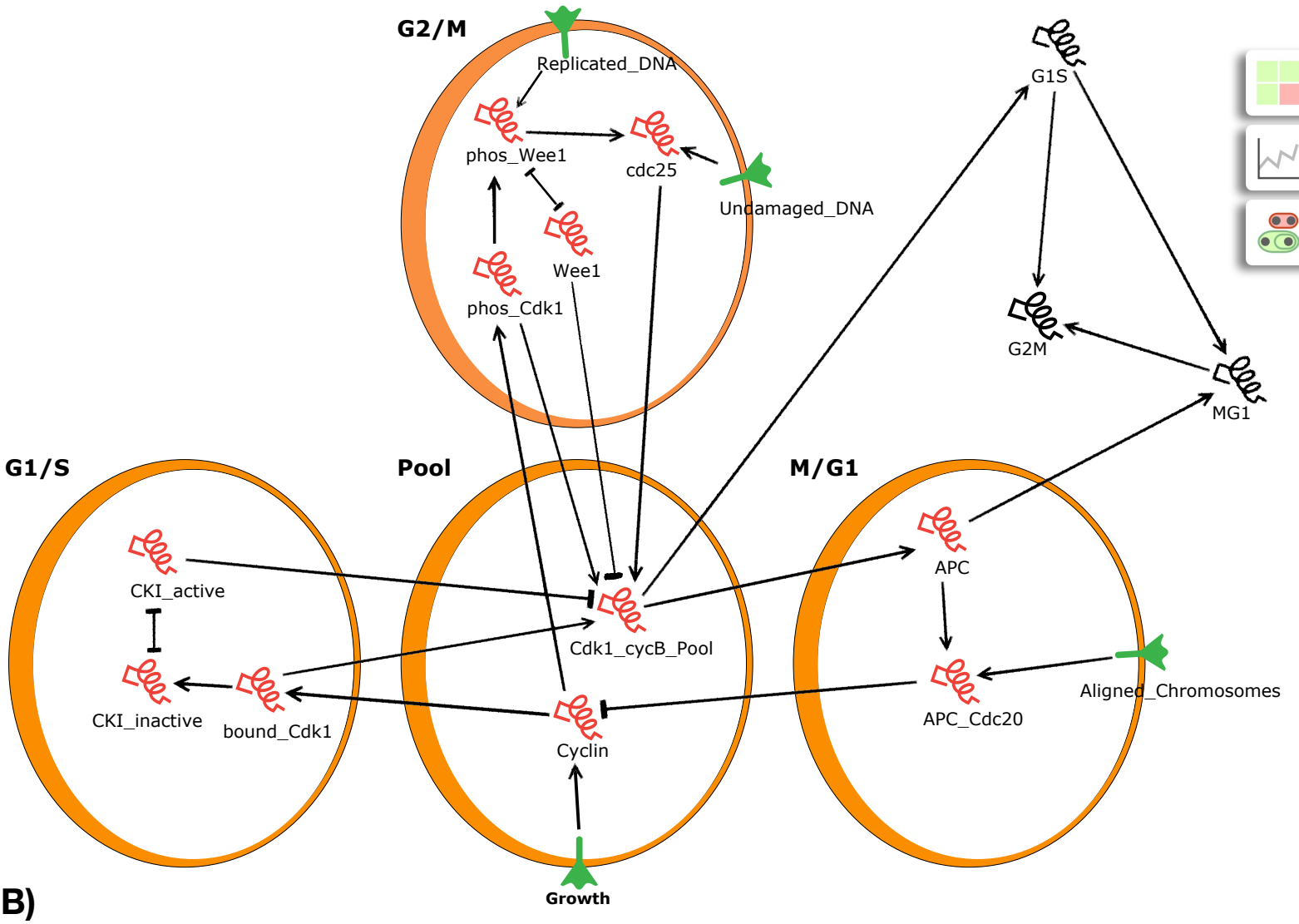
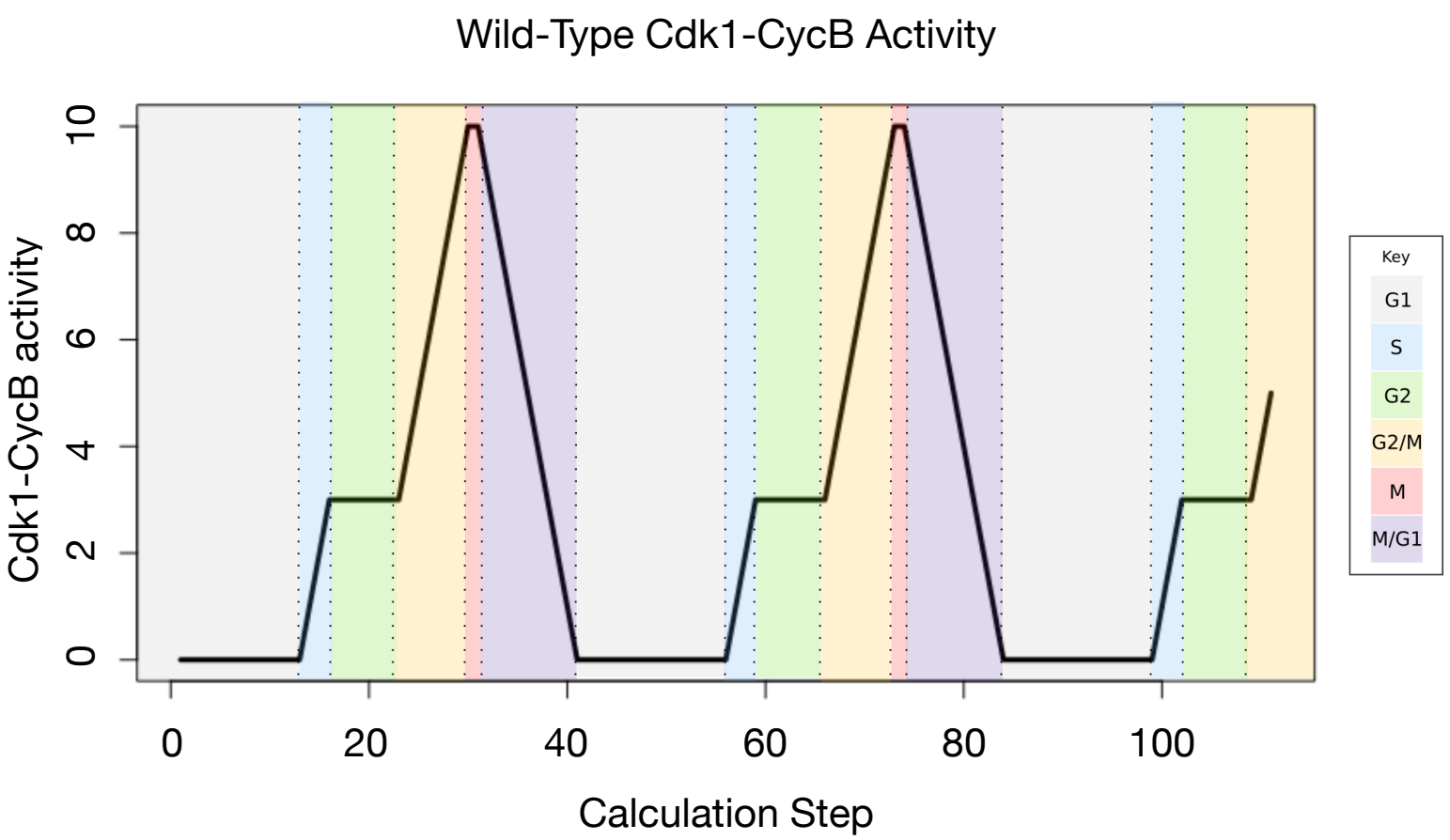


Fig 2. Comparison of Oscillatory Networks. In this illustration, the rows correspond to (A) negative feedback, (B) activator-inhibitor and (C) substrate-depletion oscillators as in Tyson et al. 21 The columns correspond to (i) Tyson et al. 21wiring diagrams, (ii) BMA wiring diagram translation, (iii) Tyson et al. 21 signal-response curves and (iv) BMA equivalent signal-response curves. Each BMA wiring diagram contains a unique set of target functions located within particular nodes of the network which can be found in Supplementary table 1. For most cases clear comparison between Tyson et al. 21 wiring diagrams (i) and the corresponding BMA wiring diagrams (ii) can be made. Here like in Tyson et al. 21 S indicates the input Signal and R indicates the output Response with, in our case, letters A-E representing intermediate nodes. The graphs in (iv) are derived from simulation analysis carried out in the BMA. For all cases bar a simulation is run with a set signal input of 2 as an example, and the response output from the BMA simulation plotted based on the response per calculation time step and are thus not directly comparable, however clear oscillatory behaviour can still be observed.

A)


G2/M
Replicated_DNA
phos_Wee1
cdc25
Undamaged_DNA
Wee1
phos_Cdk1
G1S
G2M
MG1
G1/S
Pool
M/G1
CKI_active
CKI_inactive
bound_Cdk1
Cdk1_cycB_Pool
Cyclin
APC
APC_Cdc20
Aligned_Chromosomes
Growth

B)


Wild-Type Cdk1-CycB Activity
Cdk1-CycB activity
0
2
4
6
8
10
0
20
40
60
80
100
Calculation Step
Key
G1
S
G2
G2/M
M
M/G1

Mutation	Time Step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61
	Wild Type	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3
	Wee1Δ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	10	10	9
	Wee1 OP	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	CKI Δ	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3
	CKI OP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	7	6	5	4	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	CKI & Wee1Δ	0	0	0	0	1	2	3	3	3	3	3	3	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	4	5	6	7	8	9	10	10	9	8	7	6	5	4	3
	Cdc25 Δ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Cdc25 & Wee1Δ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Cdc25 OP	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	10	9	8	7	6	5	4	3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	4	5	6	7	8	9	10	9	8	7	6
	Cdc20 Δ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	4	5	6	7	8	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	No Growth	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Unreplicated DNA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Damaged DNA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Misaligned Chromosomes	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	3	3	3	3	3	3	3	3	4	5	6	7	8	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10

Key:

G1	S	G2	G2/M	M	M/G1