Supplementary Methods

We developed a mathematical model of G1 progression using physiological profiles from synchronized cells exposed to constant growth factors. We found that an additional time-dependent mechanism was needed in order to recapitulate observations from our experimental data and targeted gene deletion studies. Below we describe in detail the methodology of formulating the model, fitting the model to experimental data, and using the model to generate in silico predictions. This supplemental section provides details of model construction.

I. Step 1: Formulation of the model

A. State space and reactions

Model diagram in Figure 3a of the article uses the Diagrammatic Cell Language to concisely specify the state space of the system and transformations between binding and conformational states of proteins (Maimon & Browning 2001). In addition, a representation of the model in the graphical notation system proposed by Kitano (Oda et al 2005) is provided in Supplementary Figure 1 (cell designer file with state space and reactions of the model is in Sup_6.xml). Both languages enable you to specify the state space of the system and transformations between binding and conformational states of proteins as well as translation to the Systems Biology Markup Language (SBML) (Hucka et al 2004) (Sup_7.xml provides the SBML representation of the model for simulation).

The Cdks and pRb are represented by boxes with several circles inside denoting their possible states. Open circles are unoccupied binding states or conformational states, black dots are bound states. Proteins with a single internal state are represented with ovals. Binding and unbinding reactions are represented by black dots, synthesis/ degradation – by squiggles (with arrowheads if irreversible). Squiggles inside boxes denote conformational change or degradation of bound components.

An independent set of state transformations is a connected sequence of circles and black dots. A protein can occupy only one state from each independent set of transformations. A set of particular states from each of the independent transformations defines a specific protein species. Protein species are specified using a vector notation whose positions left to right denote independent sets of sequential state transitions and values denote state numbers, in the order top to bottom and left to right in a protein box on the diagram, except the unbound state which always has value 0. For example, given three independent sets of transformation of Cdk2, its 2*2*3=12 possible states are shown to the right of the diagram. Thus Cdk2[0,1,2] denotes a particular combination of states of Cdk2 which is unbound from p27, activated by “Modifier” and bound to Cyclin A. See Supplementary Table 1A for the complete list of protein species in the model.

Hypo- and hyper-phosphorylation of pRb is represented by connecting in a sequence the unphosphorylated state (open circle) with two phosphoryl group-bound states (black dots

connected to circles named P). Double dephosphorylation is denoted by a crème squiggle from the hyper-phosphorylated state back to the unphosphorylated state. Enzymatic control of pRb phosphorylation by specific Cdk species is denoted by connecting the green colored boxes specifying the species to the phosphorylation black dots. Thus hypo-phosphorylation is controlled by active Cdk4 (light green), hyper- – by either active Cdk2 or Cdk1 (dark green). The inability of hyperphosphorylated Rb to bind E2F (conditional binding) is denoted by connecting the hyper- state black dot to the E2F-binding black dot by a blue "control" line.

Activation of Cdk2 and Cdk1 by the "Modifier" is represented by a crème-colored squiggle between two open circles denoting the inactive and active states of a Cdk. A "control" line connecting "Modifier" to the squiggle means that conformational transition occurs only when the "Modifier" is present.

Degradation of Cyclin D bound to Cdk4, Cyclin E and A bound to Cdk2 or Cdk1, of p27 bound to Cdk2 or Cdk1, E2F bound to pRb, and of Emi1 bound to APC/C is represented by gray squiggles connecting the bound state of the degraded component to the unbound state of their binding partners. Degradation of Cyclin A in its bound and unbound states by APC/C is represented by "control" lines connecting APC/C to the gray degradation squiggles.

Control of Cyclin E, A, Emi1 and E2F synthesis by E2F is denoted by the "control" lines connecting E2F to the synthesis squiggles. Constitutive synthesis and degradation of Cyclin D and p27 is denoted by gray reversible squiggles. Total concentrations of pRb, Cdk1, Cdk2, Cdk4 and APC/C were assumed not to change at the time scale of G_1 transition.

A list of chemical reactions corresponding to the set of protein species and state transformations is given in Supplementary Table 1B. Note that enzymatic reactions of pRb phosphorylation and APC/C-controlled degradation of Cyclin A are represented explicitly as reversible binding enzyme and substrate to form an intermediate complex and product release.

B. Model equations

The list of reactions, assuming mass-action kinetics, was used to generate a list of coupled ordinary differential equations (ODEs) for concentrations of protein species and mass conservation constraints (Supplementary Table 2). Concentrations of components were in units of molecules per cell and time was in units of minutes.

Faster degradation of p27 by active Cdk2 is represented by having a different degradation rate constant kd1_p27 for the transformation

$$Cdk2[111] \rightarrow Cdk2[011]$$

The rate is kd_p27 for all other p27 degradation reactions.

While most parameters associated with reactions are single-valued rate constants, some are mathematical functions. Thus, the rate of Cdk2 and Cdk1 activation by "Modifier" is given by the step function

$$\begin{cases} 0, & t < time\_Modifier \\ k\_act, & t \geq time\_Modifier \end{cases},$$

where *time_Modifier* was fitted to data (see below).

We assumed that the rate of synthesis is proportional to the occupancy of the E2F-controlled promoter regions. Thus the rate of E2F-controlled protein synthesis ks_CyclinE, ks_CyclinA, ks_E2F, ks_Emi is given by a saturable sigmoid function with a non-zero baseline rate of synthesis ks0:

$$ks = ks0 + ks1 \ \frac{E2F}{ksM + E2F} \ ,$$

where E2F is the concentration of free E2F, $ks1$ is the maximum rate and ksM is E2F concentration that gives half-maximum rate.

The ODEs were first solved for the steady-state with nonzero concentrations of "ground" states of proteins without synthesis and degradation terms as initial conditions, which served as the initial condition for simulations with the "Modifier".

II. Steps 2 and 3: Quantitative measurements and fitting of parameters to data

Model parameters were fitted to data by minimizing the cost function of the form

$$C = \sum_{ij} \ \frac{c_{ij}}{N_i} \ , \qquad c_{ij} = \tfrac{1}{2} \ \frac{d_{ij}^{\ 2}}{s_{ij}^{\ 2}} \ ,$$

where d_{ij} is the difference between the measurement of the i-th quantity at the j-th time-point and the corresponding simulation value, s_{ij} is the corresponding standard error, and N_i is the number of time-points available for the i–th quantity (Supplementary Table 1 C).

Measurements of change over time of total concentrations of Cyclin E, Cyclin A, p27 and Emi1, and hypo- and hyperphosphorylated pRb, Cyclin E and A mRNA, and Cdk2 activity were incorporated into the cost function as follows. The data normalized to actin are available in arbitrary units so the means of the replicates (and corresponding standard deviations) of these measurements were first normalized to the maximum over the course of the experiment. To match the relative change of protein levels to simulation, which is done in dimensional units, the normalized data was multiplied by "reference" parameters which have the dimension of concentration. A separate "reference" parameter was used to convert to dimensional units ratios

of total Cyclins approximately D:E:A = 0.4:0.2:1 at 16 hr time point (data not published). The ratios are expressed in terms of total Cyclin A and gave rise to three separate terms d_{ij} in which data values were ratios 1, 0.2, 0.4, and simulation values were total Cyclin A, Cyclin E, and Cyclin D (standard deviations were chosen to give reasonable weights to these terms). See Supplementary Table 1 D for normalized data, definitions of model quantities that were matched to each piece of data, and ranges of "reference" parameter values.

"Reference" parameters were found at each iteration of the main optimization algorithm as a solution of an optimization subproblem in which kinetic parameters were held fixed. The effect of such optimization is to effectively slide the observed relative change in protein levels to best match the simulation with a given kinetic parameter set. The locally found values of "reference" parameters were then passed to the global optimizer, which evaluates the value of the cost function and manipulates the sets rather than individual kinetic parameters.

Two terms d_{ij} of the cost function were allocated to constrain the simulation to two additional pieces of data. First, ability of the model to explain data in the realistic region of parameter space (for optimization to converge when fitting to our time course data), is determined by total Cyclin A being less than total Cdk2 concentration as is also suggested by our experimental observations. Since Cdk2 concentration is fixed, we chose total Cyclin A to be approximately 80% of Cdk2 level, at the 16 hr time point. To account for uncertainty of this assignment, one term in the cost function was the difference between the simulation output total Cyclin A and 80,000 with a standard deviation chosen to give a reasonable weight to this term. Note that precisely for this reason, we did not fix the total Cyclin A "reference" parameter when fitting the time-course data, or ratios of Cyclins.

Second, a sharp rise of Cyclin A concentration at the time of G_1-S transition is physiologically important for constraining the model. In our data the Cyclin A increase from 10 to 14 hr was 2.8, 3.7, and 8.6 fold in the three replicates. However, because of the large scatter at 14 hr, the corresponding d_{ij} term is minimized when total Cyclin A is flat. To capture this information we added a separate term to the cost function for the fold change of Cyclin A given by

$$C_{10-14} = 8 \times (3.5 - r)^2 \text{ where } \mathrm{r} = \mathrm{MIN}(3.5, \frac{A_{14}}{A_{10}}),$$

where A_{14} and A_{10} are total Cyclin A at 14 and 10 hr respectively. This term effectively constrains the solutions to have at least 3.5 fold change. Factor 8 was chosen to give a reasonable weight to this term.

Supplementary Figure 2 shows the solutions to the optimization runs for total Cyclin A concentrations under no deletion, a Cyclin E deletion, and a CDK2 deletion. Blue lines indicate Cyclin A levels for the six models that meet the filtering requirement; the six sets of curves in red are representatives of population members that do not meet filtering requirements.

References

1. Maimon, R. & Browning. S. Diagrammatic Notation and Computational Grammar for Gene Networks. In Proc. Second Intl. Conf. on Systems Biology, (Omnipress) 311-317 (2001).

2. Oda K, Matsuoka Y, Funahashi A, Kitano H 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. Mol Syst Biol 1 E1-E17 (Nature Publishing Group) (2005).

3. Hucka M, Finney A, Bornstein BJ, Keating SM, Shapiro BE, Matthews J, Kovitz BL, Schilstra MJ, Funahashi A, Doyle JC, Kitano H. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. Syst Biol (Stevenage) 1(1):41-53 (2004).