A rule-based kinetic model of RNA polymerase II C-terminal domain phosphorylation

Stuart Aitken1,3, Ross D. Alexander2,3,† and Jean D. Beggs2,3

1MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK
2Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK
3SynthSys, University of Edinburgh, Edinburgh EH9 3JU, UK

The complexity of many RNA processing pathways is such that a conventional systems modelling approach is inadequate to represent all the molecular species involved. We demonstrate that rule-based modelling permits a detailed model of a complex RNA signalling pathway to be defined. Phosphorylation of the RNA polymerase II (RNAPII) C-terminal domain (CTD; a flexible tail-like extension of the largest subunit) couples pre-messenger RNA capping, splicing and 3' end maturation to transcriptional elongation and termination, and plays a central role in integrating these processes. The phosphorylation states of the serine residues of many heptapeptide repeats of the CTD alter along the coding region of genes as a function of distance from the promoter. From a mechanistic perspective, both the changes in phosphorylation and the location at which they take place on the genes are a function of the time spent by RNAPII in elongation as this interval provides the opportunity for the kinases and phosphatases to interact with the CTD. On this basis, we synthesize the available data to create a kinetic model of the action of the known kinases and phosphatases to resolve the phosphorylation pathways and their kinetics.

1. Introduction

Ordinary differential equation and stochastic modelling techniques are applicable to RNA processing and signalling pathways. However, a significant obstacle must be overcome: the explosion of entities and reactions that follows when RNA species, their precursors, complexes and states are represented. Rule-based modelling has previously been shown to address combinatorial complexity in protein interactions [1,2]; here, we explore the potential of this new methodology for modelling RNA metabolic pathways. Rule-based models in the Kappa language that we adopt [3] are simulated stochastically, and thus the approach accounts for low molecule numbers and the discrete states that molecules can adopt. We discuss transcription by RNA polymerase II (RNAPII) as an example of a complex system that can be modelled in detail using a rule-based approach.

We present a novel rule-based model of transcription that includes the phosphorylation states of multiple heptapeptide repeats in the representation of the RNAPII enzyme. The model represents the kinases and phosphatases that act on these repeats during initiation, elongation, 3' end maturation and RNAPII recycling. We demonstrate that this model can help explain both genome-wide patterns of RNAPII phosphorylation on yeast genes, and changes in phosphorylation that take place when an intron-containing yeast reporter gene is induced.

All protein-coding genes and most small non-coding RNAs are transcribed by RNAPII [4]. The C-terminal domain (CTD) of the largest subunit of RNAPII contains the heptapeptide repeat, Y1S2P3T4S5P6S7, and functions in coupling transcription to mRNA processing and chromatin modifications [5,6]. In yeast, the CTD consists of 26 repeats of this heptad (52 in humans), each of which can be phosphorylated at the serines (S) in positions 2, 5 and 7, as well as at the tyrosine

© 2013 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/3.0/, which permits unrestricted use, provided the original author and source are credited.
2. The C-terminal domain kinases and phosphatases

The process of messenger RNA transcription, the phases of the CTD cycle and the various states that an individual heptapeptide repeat can adopt in each phase are illustrated in figure 1. In addition to summarizing the rule-based model, figure 1b serves as a reference for the following survey of proteins, their activity as kinases or phosphatases and their substrate specificities. The actions of these entities (referred to as agents in Kappa terminology) are represented as transitions (arrows) between states of the heptapeptide.

Yeast kinase Kin28 phosphorylates RNAPII CTD on serine 5 (Ser5) after the assembly of the preinitiation complex (PIC) at the promoter [6] (denoted by the arrow labelled TFIIH/Kin28 in figure 1b). The phosphorylation of the CTD on Ser5 is important for the recruitment of the mRNA capping enzyme complex, which may contribute to the processivity of the elongation complex. Kin28 and Srb10 are indistinguishable in terms of their ability to phosphorylate the CTD on Ser5, yet Kin28 is a positive regulator, whereas Srb10 is a negative regulator of transcription. The phosphorylation of the CTD by Srb10 prior to the formation of a stable PIC is proposed to inhibit transcription, thus it is the timing of CTD phosphorylation that determines the mode of regulation [13]. In figure 1b, Srb10 competes with Kin28 to prevent RNAPII activity as both kinases act on completely dephosphorylated substrates (Ser2~U, Ser5~U, Ser7~U) at the promoter.

Both serine 2 (Ser2) kinase Ctk1 and phosphatase Fcp1 can be cross-linked to promoter and coding regions, but not to the non-transcribed regions, and they have opposing effects on elongating polymerases [14]. Fcp1 also has a role in the recycling (dephosphorylation) of RNAPII for subsequent rounds of transcription.

Biochemical investigation of the specificity of Ctk1 shows this kinase to react with synthetic repeats phosphorylated at either Ser5 or Ser2 to phosphorylate Ser2 and Ser5, respectively. Unphosphorylated repeats are phosphorylated at Ser5 by Ctk1, but to a lesser extent [10].

Biochemical analysis of Fcp1 from Schizosaccharomyces pombe reveals this phosphatase to have greater activity on synthetic heptad repeats phosphorylated on Ser2 than on Ser5 [15], and phosphatase is efficiently released by Fcp1 from doubly phosphorylated repeats [15,16]. Fcp1 directly recognizes CTD repeats, in addition to binding to RNAPII at a functionally

Figure 1. Phosphorylation of serines 2, 5 and 7 (Ser2, Ser5 and Ser7) on the CTD of RNAPII. (a) Ser5 and Ser7 are phosphorylated at the promoter by TFIIH/Kin28. Ser2 phosphorylation increases towards the 3′ end of the gene and is a requirement for 3′ end maturation. Serine phosphorylation has been linked to cotranscriptional splicing (the excision of an intron, in red, is illustrated). (b) The actions of the kinases Kin28, Ctk1, Bur1 and Srb10, and the phosphatases Rtr1, Fcp1 and Ssu72 on an individual heptapeptide are indicated by a change of state of one of Ser2, Ser5, Ser7 to P from U, or vice versa. The specific actions of these proteins at the promoter, during elongation, at the 3′ UTR and on the free RNAPII (recycling) must be distinguished. Higher reaction rates are indicated by broader arrows. Two reactions supported by biochemical evidence but not by in vivo data are indicated by dashed arrows.
distinct site that may mediate its effect on elongation [16]. The
dephosphorylation of Ser5 by Fcp1 after the termination of
transcription, in vitro, can be explained by the lower specificity
of Fcp1 for a ternary complex of DNA, RNAPII, TFIIH in com-
parison with that for free RNAPII [17]. Consequently, Fcp1 is
modelled as acting on Ser2→P during elongation and on both
Ser2→P and Ser5→P during recycling.

The Ser2 kinase Bur1 is recruited to CTD repeats of
elongating RNAPII phosphorylated on Ser5. Bur1 augments
the major Ser2 kinase (Ctk1) and is capable of phosphorylating
CTD repeats at Ser2, when Ctk1 is inactivated [18]. Thus,
during elongation, Bur1 acts only on repeats phosphorylated
on Ser5, whereas Ctk1 potentially acts on three of the four
possible CTD states (figure 1b).

Rtr1 is a Ser5 phosphatase that localizes to DNA coding
regions [19]. Rtr1 regulates the transition in the phosphoryl-
ation of the CTD that takes place between the promoter,
where serines 5 and 7 are phosphorylated, and around
1000 nt from the transcription start site (TSS), by which point
point Ser2 phosphorylation has increased and Ser5 phosphorylation
has diminished [19]. Rtr1 mutants show a large increase in Ser5
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5
non-catalytic role in CTD dephosphorylation [20]. Consequently, Fcp1 is
modelled as acting on Ser2→P during elongation and on both
Ser2→P and Ser5→P during recycling.

Ser2 kinase Bur1 is recruited to CTD repeats of
elongating RNAPII phosphorylated on Ser5. Bur1 augments
the major Ser2 kinase (Ctk1) and is capable of phosphorylating
CTD repeats at Ser2, when Ctk1 is inactivated [18]. Thus,
during elongation, Bur1 acts only on repeats phosphorylated
on Ser5, whereas Ctk1 potentially acts on three of the four
possible CTD states (figure 1b).

Rtr1 is a Ser5 phosphatase that localizes to DNA coding
regions [19]. Rtr1 regulates the transition in the phosphoryl-
ation of the CTD that takes place between the promoter,
where serines 5 and 7 are phosphorylated, and around
1000 nt from the transcription start site (TSS), by which point
point Ser2 phosphorylation has increased and Ser5 phosphorylation
has diminished [19]. Rtr1 mutants show a large increase in Ser5
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.

A recent report has failed to identify an active site on Rtr1
[20]. Experiments to demonstrate CTD phosphatase activity
were unsuccessful which suggests that Rtr1 may have a
non-catalytic role in CTD dephosphorylation [20]. Thus we
consider a complex involving Rtr1 to mediate the Ser5→P
phosphorylation, indicating a loss of phosphatase activity, and
an increase in Ser2 phosphorylation in the coding region rela-
tive to the wild-type. If Ser2 phosphorylation primarily
occurs through repeats that are unphosphorylated at both ser-
ines 2 and 5 (i.e. through the state that results from Rtr1
phosphate action downstream of the promoter) then Ser2
phosphorylation would be expected to reduce considerably,
rather than increase in a Rtr1 mutant. This finding, together
with the lack of biochemical evidence to support the generation
of Ser2 phosphorylation from doubly unphosphorylated
repeats, indicates that Rtr1 may compete kinetically with
the Ctk1 and Bur1 CTD kinases to regulate the Ser5 to Ser2
transition by removing their preferred substrate.
species-based model would require 1280 species and 5056 reactions to encode two repeats each capable of adopting four states at 80 locations.

Rules in the Kappa language also allow new entities to be created. This is particularly useful in modelling RNA pathways as transcription adds nucleotides to the nascent RNA as the polymerase traverses the gene. The synthesis of splice sites, the branchsite, splicing enhancer and splicing suppressor sites all provide the substrates for reactions that may be concurrent with transcription. These reactions determine the final sequence of the RNA and the protein product. Kappa provides a readily understandable rule language, and tools for stochastic model simulation, facilitating model development. Alternative rule-based languages that provide comparable support include BioNetGen, NFSim and ML-Rules [27–29]. The promise of these approaches is described further in Bachman & Sorger [2].

4. Results

A detailed Kappa model of 219 rules describing the entire transcription pathway has been constructed (see figure 1, §6 and electronic supplementary material, file model.tar.gz). Defining the transcription pathway allows rules specifying the interaction of the kinases and phosphatases with sites on the CTD of RNAPII to take the context of the CTD into account as it progresses through the transcription cycle. As discussed earlier, these proteins may act on the CTD in specific ways at the promoter, during elongation, at the 3' UTR and on free RNAPII, and hence the model must distinguish these contexts. The states that an individual heptad may adopt, and the transitions between them, are indicated in figure 1b for each of the four contexts.

We now consider the phosphorylation patterns in two different kinds of yeast data: metagene profiles obtained from genome-wide data [11], and kinetic data obtained from the induction of a reporter gene [12]. As the reporter gene contains a single intron (for the study of splicing), we focus on single-intron genes in the genome-wide analysis.

4.1. Metagene profiles

Genome-wide profiling of highly expressed genes shows Ser5~P to reduce rapidly towards a steady state as RNAPII leaves the promoter, and Ser2~P to increase [11]. We find that this pattern also occurs in yeast genes with a single intron where the intron is long (300–600 nucleotides; figure 2). This subset of the genome is highly expressed on average. The distributions of exon, intron and gene lengths in
single-intron yeast genes are plotted in the electronic supplementary material, figure S1, and the CTD phosphorylations of all highly expressed genes are presented as a scatter plot in the electronic supplementary material, figure S2, for comparison with figure 2b–d (data from [11]).

The Kappa model of the CTD cycle is capable of reproducing the observed pattern of Ser2, Ser5 and Ser7 phosphorylation under different assumptions about the ability of Ctk1 to act as a Ser5 kinase during elongation (a controversial issue in RNAPII biology). These reactions have been observed in vitro, but in vivo evidence is lacking. To explore the inclusion, or otherwise, of the possible Ctk1 Ser5 kinase activity in the model, we examine the fit of four versions of the model to the genome-wide data. We consider models that lack any Ctk1 Ser5 kinase activity (see the electronic supplementary material, figure S3), include Ctk1 Ser5 kinase activity from the Ser2~U, Ser5~U, Ser7~P substrate (see the electronic supplementary material, figure S4), include Ctk1 Ser5 kinase activity from the Ser2~U, Ser5~U, Ser7~P substrate (see the electronic supplementary material, figure S5), and finally include Ctk1 Ser5 kinase activity from both Ser2~U, Ser5~U, Ser7~P and Ser2~P, Ser5~U, Ser7~P substrates (see the electronic supplementary material, figure S6). To widen the basis of model comparison, we also add three assumptions: that Rtr1 binds faster than, at the same rate or slower than the rate at which Ctk1 and Bur1 bind when acting as Ser2 kinases. For Rtr1 to mediate the Ser5~P to Ser2~P transition as proposed by Mosley [19], it is to be expected that the Rtr1 binding rate would be faster than that of the competing Ser2 kinases, and hence that the fit of the model to the data should be better for faster Rtr1 binding.

Qualitatively, the loss of Ser5~P and gain of Ser2~P along the gene is maintained under the alternative modelling assumptions. However, the kinetics of these changes differ considerably when the model predictions are mapped to the genome-wide data. The possible Ser5~P kinase reaction from substrate Ser2~P, Ser5~U, Ser7~P to Ser2~P, Ser5~P, Ser7~P does not improve the fit of the model to the data, either by itself or in conjunction with that from Ser2~U, Ser5~U, Ser7~P, and is not included in the model. In contrast, the addition of the reaction from Ser2~U, Ser5~U, Ser7~P to Ser2~U, Ser5~P, Ser7~P quantitatively improves the fit of the model to the data and maximizes the dynamic range of predicted values that correspond to observed values. Consequently, this Ctk1 Ser5 kinase reaction is included in the model henceforth.

The predictions of the model (figure 2a) are indicated by the solid black lines in the scatterplots of the genome-wide data in figure 2b–d. Ser7~P is predicted to be constant as no Ser7 kinase or phosphatase has been proposed to act during elongation, and hence there are no rules for the modification of this residue. Figure 2d shows that a number of intron-containing genes have elevated Ser7~P at the TSS and on exon2, which may indicate a regulatory role [30]. A general tendency for Ser7~P to increase along the gene can be seen in the electronic supplementary material, figure S2 (potentially attributable to Bur1 [31]).

In the Kappa model, changes in phosphorylation are not dependent on the length of the gene, but on the time RNAPII spends in elongation which provides the opportunity for the kinases and phosphatases to act. Hence the model provides mechanisms to explain genome-wide observations, including the proposed regulatory role of Rtr1 [19].

4.2. Sensitivity analysis

The model specifies that the rate at which Bur1 and Ctk1 bind to sites on RNAPII is half that of Rtr1 binding, and consequently Rtr1 tends to bind first and reduce the extent of Ser5~P. Nonetheless, some RNAPII heptads are directly phosphorylated on Ser2 by Bur1 and Ctk1, and Ser5~P can be restored by Ctk1 allowing these kinases to act. When these binding rates are reduced, few heptads are phosphorylated on Ser2 and, as this is a requirement for 3' end maturation, RNAPII accumulates at the 3' end. In contrast, when the binding rates are increased, Ser2~P and Ser5~P reach their steady-state values within 150 nt of the promoter (that is, more rapidly than observed).

The effects of varying the two parameters that govern the binding and release rates of the kinases and phosphatases that act during elongation were explored systematically by sensitivity analysis. Figure 3 presents heatmaps showing the difference of simulated RNAPII, Ser2~P and Ser5~P with respect to simulations based on the parameters used in the baseline model (figure 2). The binding rate p0 (see §6) was varied over two orders of magnitude (0.01–1) and the unbinding rate p3 over four orders of magnitude (0.1–1000). The binding rates for all kinases were kept at a constant fraction of p0 as defined in the model; for example, the rate for Bur1 was kept at p0/2, thus p0 acts as a ‘meta-parameter’. 

Figure 3. Sensitivity of (a) RNAPII, (b) Ser2~P and (c) Ser5~P to the binding rate p0 and unbinding rate p3. The colour scale white to red indicates the effect on the model prediction (lowest to highest sum of square differences) for each combination of p0 and p3 surveyed, taking p0 = 0.1 and p3 = 100 as the reference.
Similarly, all unbinding reactions occur at $p3$. All other rates were kept constant, most importantly, those governing the progression of RNAPII along the gene as this defines the window of opportunity for CTD modifications. This computational experiment determines the robustness of the predictions to parameter values.

The heatmap in figure 3a shows that the predicted profile of RNAPII is insensitive to the phosphorylation parameters except with values for $p0$ of 0.01 or less: should there be no Ser2 phosphorylation at the 3′ end of the gene as the result of the low binding rate, the model requires RNAPII to pause as 3′ end maturation cannot proceed without Ser2~P and this results in an accumulation of RNAPII at the 3′ end. The heatmaps for Ser2~P and Ser5~P show a greater sensitivity for these predictions on values of $p0$ greater and less than those of the baseline model. The value of $p3$ does not have an impact on the predictions while $p3$ is 10 times the value of $p0$ or greater. Lower values of $p3$ combined with greater values of $p0$ give rise to predictions comparable with those of the baseline model as the duration of the binding/unbinding sequence is comparable. Thus, the potential trade-off between binding rates illustrates the importance of direct measurement of binding constants (e.g. as in [32]) as their values cannot be inferred with confidence from model fitting.

4.3. Kinetic profiles

Having established that the typical profiles of intronless and intron-containing genes are similar at steady state, and that the Kappa model reproduces this profile, we apply the model to data from the induction of a single gene. Data from individual genes often provide insights into regulation and opportunity for more detailed investigations than genome-wide datasets, which typically reflect steady-state conditions.

The ChIP data on a splicing-related pause in the Ribo1 yeast gene provide an opportunity to explore the relationship between RNAPII density and serine phosphorylation [12]. Briefly, Ribo1 is a chimeric yeast gene that contains the single intron from the ACT1 gene and the 3′ end processing signal from PGK1. The reporter gene is integrated in the genome, transcribed under the control of a doxycycline-responsive promoter in a doxycycline-inducible strain of Saccharomyces cerevisiae. Cells in culture respond to doxycycline in synchrony, allowing the evidence in the literature, and this leads to the novel opportunity for more detailed investigations than genome-wide investigations that influence their recruitment.

RNAPII density, Ser2~P and Ser5~P signals along Ribo1 are shown in figure 4a, and the results of simulating Ribo1 are shown in figure 4b. The model now includes a rule that briefly pauses RNAPII elongation 600 nucleotides from the TSS. The distance of the pause site from the TSS is important as it determines the extent to which Ser2 and Ser5 are phosphorylated, and hence the degree to which phosphorylation will apparently increase on paused RNAPII. This factor is accurately accounted for in the simulation. An increase in the absolute level of phosphorylation is expected where RNAPII pauses; however, additional changes in phosphorylation observed at the 5′ end of exon 2 (see point 5′-E2 in figure 4a) that occur concurrently with a decrease in RNAPII are highly significant as they may indicate signalling that occurs subsequent to the pause, rather than changes that can be ascribed to RNAPII density [12].

![Figure 4. Paused RNAPII, Ser2 and Ser5 phosphorylation in the coding regions of single-intron genes. (a) Phosphorylation and RNAPII density along the Ribo1 gene at 240 s after induction (data from [12]). Phosphorylations at the promoter (Prom), exon1 (E1), 3′ splice site (3SS), 5′ end of exon2 (5′-E2) and 3′ end of exon2 (3′-E2) of Ribo1 are indicated. Note that Ser2~P, Ser5~P and the density of RNAPII are each normalized to their respective values on induction, and hence are not comparable to each other on an absolute scale. (b) Model simulation of Ribo1 with a pause introduced 600 nucleotides downstream of the TSS. (a) Real and (b) simulated pauses in RNAPII lead to an accumulation of RNAPII that reverses the typical reduction in the absolute level of Ser5~P and raises the level of Ser2~P.](http://rsif.royalsocietypublishing.org/Downloaded from http://rsif.royalsocietypublishing.org/)

5. Conclusions

Messenger RNA transcription is a complex process with many regulatory interconnections acting through RNAPII subunits. The proposed model of the CTD cycle accounts for both the substrate specificity of the CTD kinases and phosphatases, and the contexts that influence their recruitment. The model was created by surveying and reconciling the evidence in the literature, and this leads to the novel observation that Rtr1 regulates the Ser5~P to Ser2~P transition by removing the substrate for the Ser2 kinases on many (but not all) heptads. We demonstrate that the model reproduces the prototypical pattern of phosphorylation observed from the promoter, through to the 3′ UTR, based solely on the binding kinetics of the kinases and phosphatases with the CTD substrate. Modelling suggests that Ctk1 may have a Ser5 kinase activity (as has been demonstrated in vitro) as this improves the fit of the model to the data, and provides a mechanism whereby Ser5~P can be reduced considerably while retaining the possibility of restoring this state as a substrate for the known Ser2 kinases.

Undoubtedly there are other, yet to be discovered, interactions between these proteins and RNAPII, and with each other. The addition of new rules, or the refinement of existing rules, will allow such interactions to be added to the rule-based model in a straightforward fashion. For example,
were it to be shown that Bur1 recognizes the bivalent mark Ser5\textasciitilde P + Ser7\textasciitilde P as hypothesized in [24], the rules for Bur1 binding could be strengthened to represent this finding.

Rule-based modelling addresses combinatorial expansion in systems modelling, and enables the simulation and analysis of reactions whose formal analysis would otherwise be intractable [2,27,28]. The rule-based approach is being extended to multi-level modelling [29,33], where model entities can be nested, and downward and upward causation between different levels is possible, and to the representation of spatial relations in models [34]. Thus, the approach holds great potential at theoretical and practical levels.

6. Material and methods

A detailed Kappa model of 219 rules describing the entire transcription pathway has been constructed (see the electronic supplementary material, file model.tar.gz). This model represents PIC assembly, followed by pre-mRNA capping, elongation, 3' end maturation and polyadenylation, and release of the mature mRNA [4,6], and is based on a validated stochastic model of a yeast reporter [35]. The complex processing represented in the model yields realistic behaviour in terms of time required for RNAPII to transcribe the stimulated gene, and the net synthesis rate of one mature RNA every 4 s previously observed in a yeast reporter [36].

The transcription initiation rules (see the electronic supplementary material, file transcription.ka) define how Activator and Mediator bind to DNA, followed by RNAPII and TFIIH. The phosphorylation of serines 5 and 7 of RNAPII, and capping of the nascent transcript T1 occur during initiation. Rules for elongation specify how RNAPII progresses from one site of DNA to the next, while simultaneously extending the RNA transcript (see the electronic supplementary material, file elongation.ka). Following [35], RNAPII occupies a 30 nt footprint (DNA sites e1–e80 represent spans of 30 nt that RNAPII occupies in sequence) and hence the gene has a maximum capacity for RNAPII. Pausing of RNAPII can cause a queue of polymerase as those behind are prevented from progressing. Rules for the maturation of the RNA through the binding of Pcr11, synthesis of the polyA tail, binding of Rtt103 and release of the mature transcript and RNAPII are defined (see the electronic supplementary material, file maturation.ka). Pcf11, synthesis of the polyA tail, binding of Rtt103 and release is defined in rules maturation before the rules for polyadenylation and poly(A) tail formation.

The CTD kinetics are determined by the structure of the CTD and by the phosphorylation and dephosphorylation of Ser2 and Ser5, and the CTD can be phosphorylated on both Ser2 and Ser5 at the same time. The CTD phosphorylation states are defined in a Kappa model of 219 rules describing the entire transcription pathway.

Endnotes

1. While it is to be expected that the entire dynamic range of phosphorylation can be predicted (0–100%) may not map to distinct observable values, especially for Ser2\textasciitilde P which is around 0 at the transcription start site, a better model will maximize this range.

References
