Spatio-temporal logic, post-translational oscillators, and gene regulatory networks

Chris Banks

September 2014

Introduction

I will give a brief overview of:

- spatio-temporal logic (my Ph.D.),
- how this was applied to systems biology,
- what I'm doing now with Tom Michoel.

Spatio-temporal logic

What is it?

Spatio-temporal logic

What is it?

- First – what is computational logic in general?

Spatio-temporal logic

What is it?

- First – what is computational logic in general?
  - Language of formal, unambiguous, statements.
  - Computable: i.e. a machine can "understand" these statements.

Spatio-temporal logic

What is it?

- First – what is computational logic in general?
  - Language of formal, unambiguous, statements.
  - Computable: i.e. a machine can "understand" these statements.
- **Spatio-temporal logic** is a language such as this which can describe precise properties involving time and space in the model.

Spatio-temporal logic

Very good, but so what?

- The well established field of **model checking** aims answer this:

$$M \models s \ ?$$

  Does a model (M) satisfy ($\models$) a logical statement (s)?

- Of course, for a machine to answer this question our model must also be formal and computational.

- Thankfully, loads of formal modelling tools exist for bio-models.

Great! So now we can ask questions about models and have the computer answer them?

- **Yes**!
- **But**. . . there are lots of common questions in biology that are difficult to express in logic.
- So we design more expressive logics with more specific applications in mind.
- But more expressive = more computationally complex = expensive to compute.

Logic of Behaviour in Context ($\mathcal{LBC}$)

A logical language for expressing:

- Real valued constraints
  - $[A] > c$
- Temporal properties
  - Always, Eventually, Until, ...
  - With time intervals
- Contextual properties
  - inhibitor introduced, combined with another process, ...
- Combinations thereof...

Logic examples

- Within 24h we will see some of species A in the system.
- Between hours 10 and 15 we always have $[A] \leq 0.1$.
- If we introduce species Q, then in 2 to 5 hours $[A] \geq 0.1$.
- If we introduce Q *at any time* up to hour 10 then within 2 to 5 hours we get $[A] \geq 0.1$.

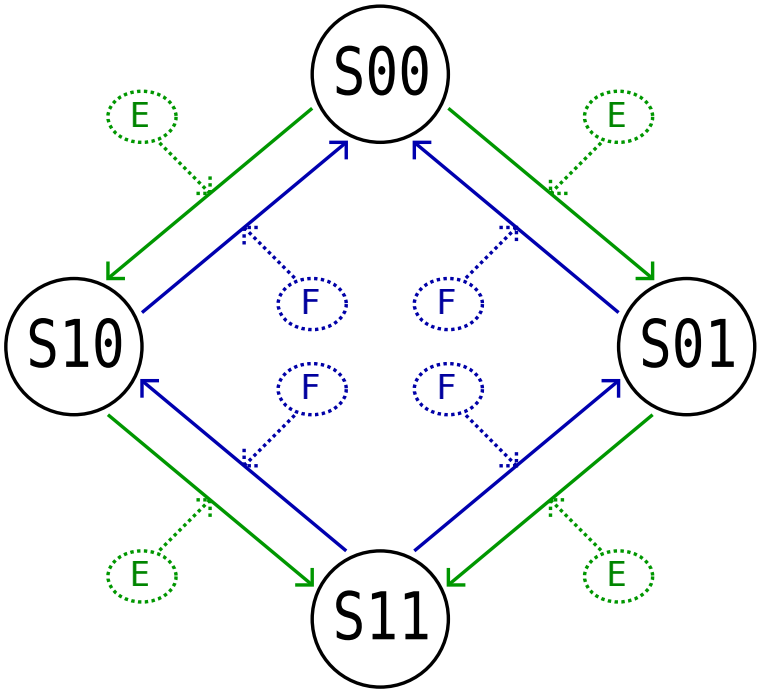
Logic examples

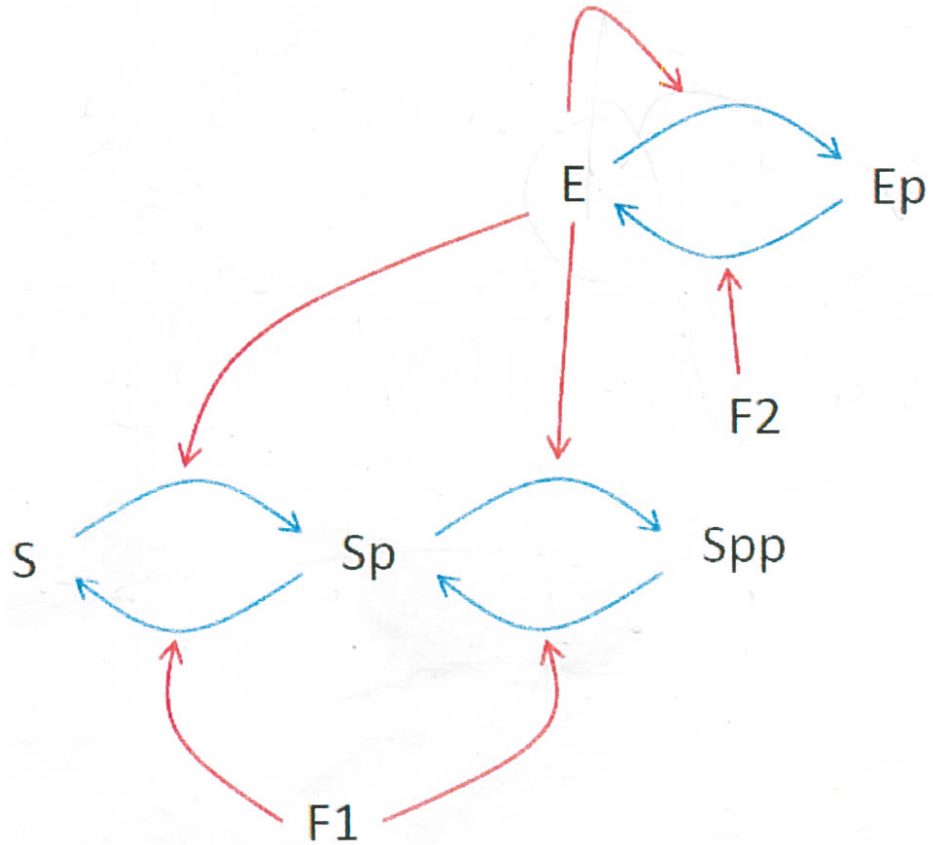
- Within 24h we will see some of species A in the system.
  - $\mathbf{F}_{24}([A] > 0)$
- Between hours 10 and 15 we always have $[A] \leq 0.1$.
  - $\mathbf{G}_{[10,15]}([A] \leq 0.1)$
- If we introduce species Q, then in 2 to 5 hours $[A] \geq 0.1$.
  - $Q \triangleright (\mathbf{F}_{[2,5]}([A] \geq 0.1))$
- If we introduce Q *at any time* up to hour 10 then within 2 to 5 hours we get $[A] \geq 0.1$.
  - $\mathbf{G}_{10}(Q \triangleright (\mathbf{F}_{[2,5]}([A] \geq 0.1)))$
- From this we can build even more complicated properties
  - Oscillation, inhibitor response, phase response, ...

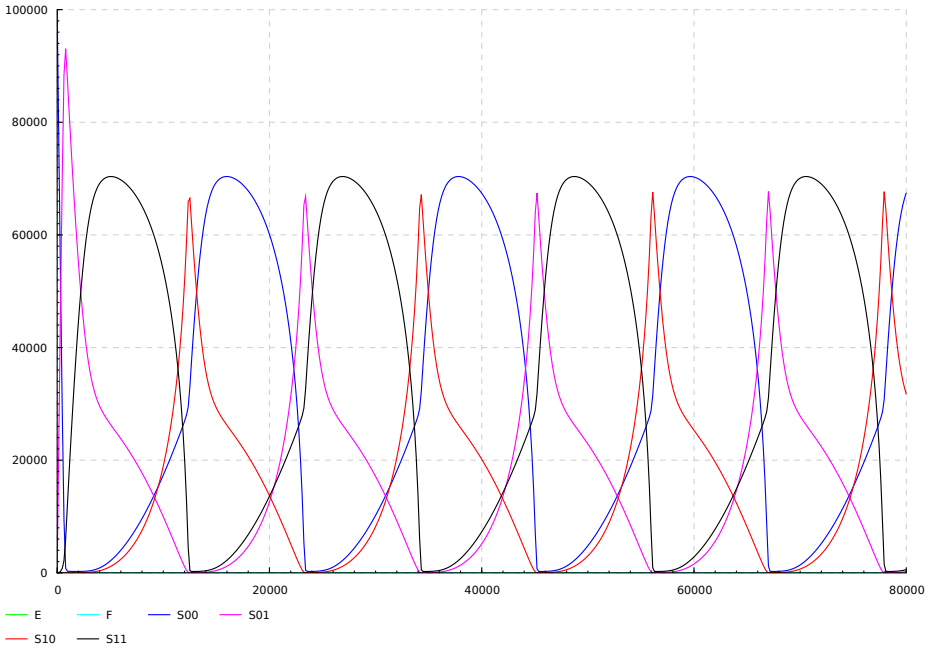
Case study (with Daniel Seaton, Millar Lab)

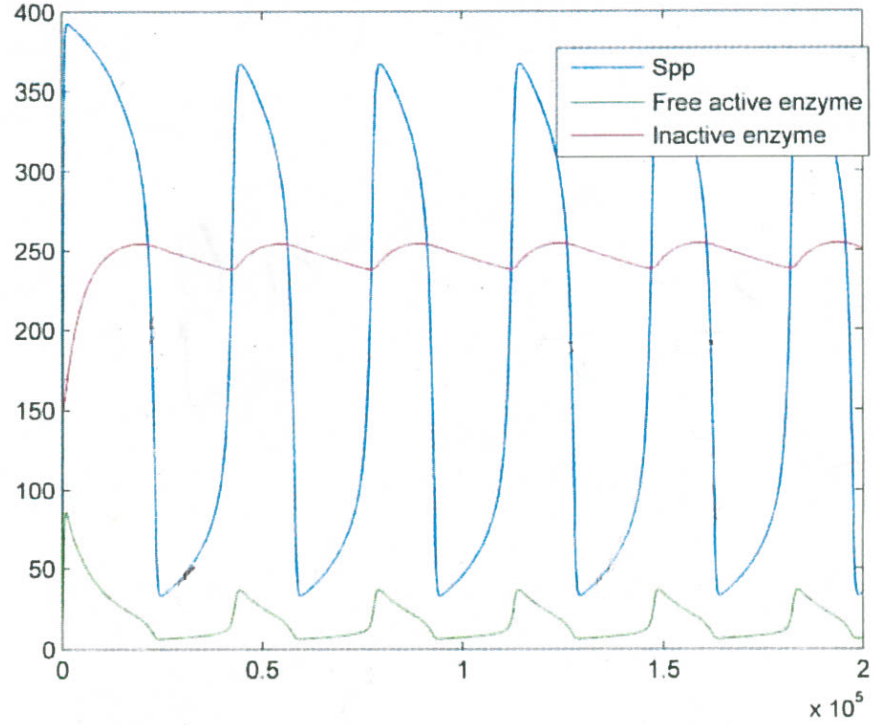
Posttranslational oscillators (PTOs):

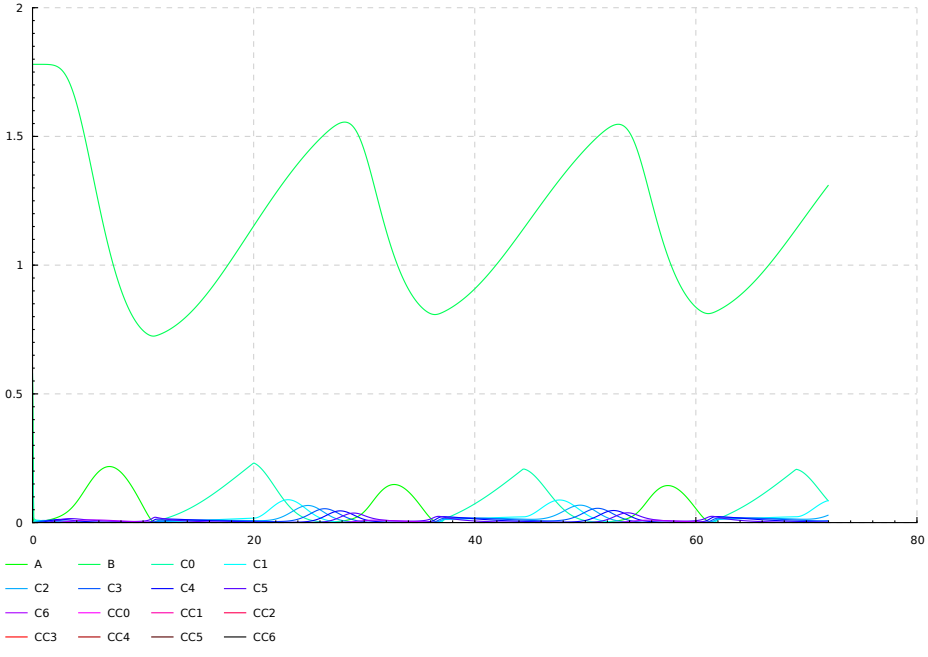
- The **Kai Circadian Clock** is a large, well studied model of the circadian clock mechanism in a cyanobacteria.
- **Jolley's PTO** is candidate mechanism for a circadian PTO
- **Seaton's PTO** is novel oscillator mechanism based on auto-inhibition






100000
80000
60000
40000
20000
0
0
20000
40000
60000
80000
E
F
S00
S01
S10
S11


400
350
300
250
200
150
100
50
0
0
0.5
1
1.5
2
x 10^5
Spp
Free active enzyme
Inactive enzyme


2
1.5
1
0.5
0
0
20
40
60
80
A
B
C0
C1
C2
C3
C4
C5
C6
CC0
CC1
CC2
CC3
CC4
CC5
CC6

Case study

- Studied the composition of models (using context)
  - properties of coupled oscillators
  - one oscillator in the context of another

Case study

- Studied the composition of models (using context)
  - properties of coupled oscillators
  - one oscillator in the context of another
- Formalised comparison of behaviour between models
  - circadian clock behaviour vs. conjectured mechanisms
  - oscillation, inhibitor response, and phase response

Case study

- Studied the composition of models (using context)
  - properties of coupled oscillators
  - one oscillator in the context of another
- Formalised comparison of behaviour between models
  - circadian clock behaviour vs. conjectured mechanisms
  - oscillation, inhibitor response, and phase response
- $\mathcal{LBC}$ efficiently analyses more than just one model run
  - e.g. behaviour when coupled *at any point in time*
  - concise, precise definition of "computational experiments"
  - "experiments beyond simulation"—higher level than standard simulation experiments

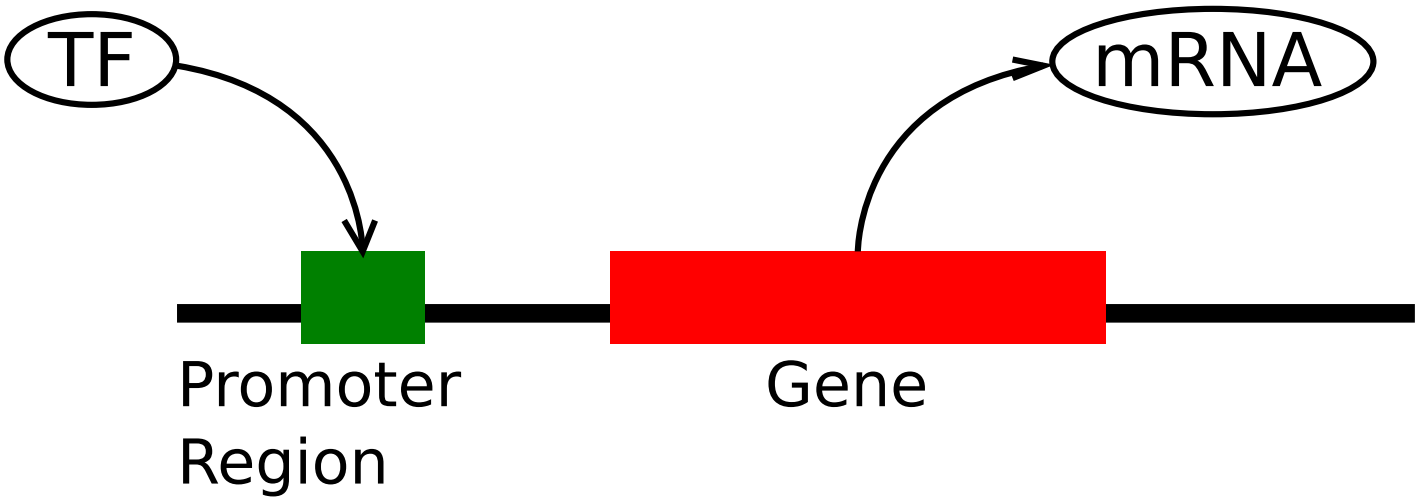
References

- Banks, C.J., & Stark, I. (2014).
  A Logic of Behaviour in Context.
  Information and Computation, 236, pp.3–18.

- Banks, C.J., Clark, A., Georgoulas, A., Gilmore, S., Hilston, J., Milios, D., & Stark, I. (2013).
  Stochastic modelling of the Kai-based circadian clock.
  Electronic Notes in Theoretical Computer Science, 296, pp.43–60.

- Forthcoming journal paper (in prep.) with D. Seaton on the PTO study.

Reconstructing gene regulatory networks (Current work with Tom Michoel)



- One mechanism for gene regulation involves the binding of a Transcription Factor into the promoter region for a gene,
- but which genes are regulated by a given TF?
- and some binding sites may not be functional, which ones are?

Knockouts and ChIP-seq

- **Knockout** experiments give us the genes which are expressed with a given TF,
  - but not the direct mapping between binding sites and expression.
- **ChIP-seq** experiments give us the locations to which a TF binds.
  - Unfortunately ChIP-seq alone does not give very good results:
    - Some binding sites may be non-functional.
    - Some genes may be regulated *indirectly* by a TF binding.

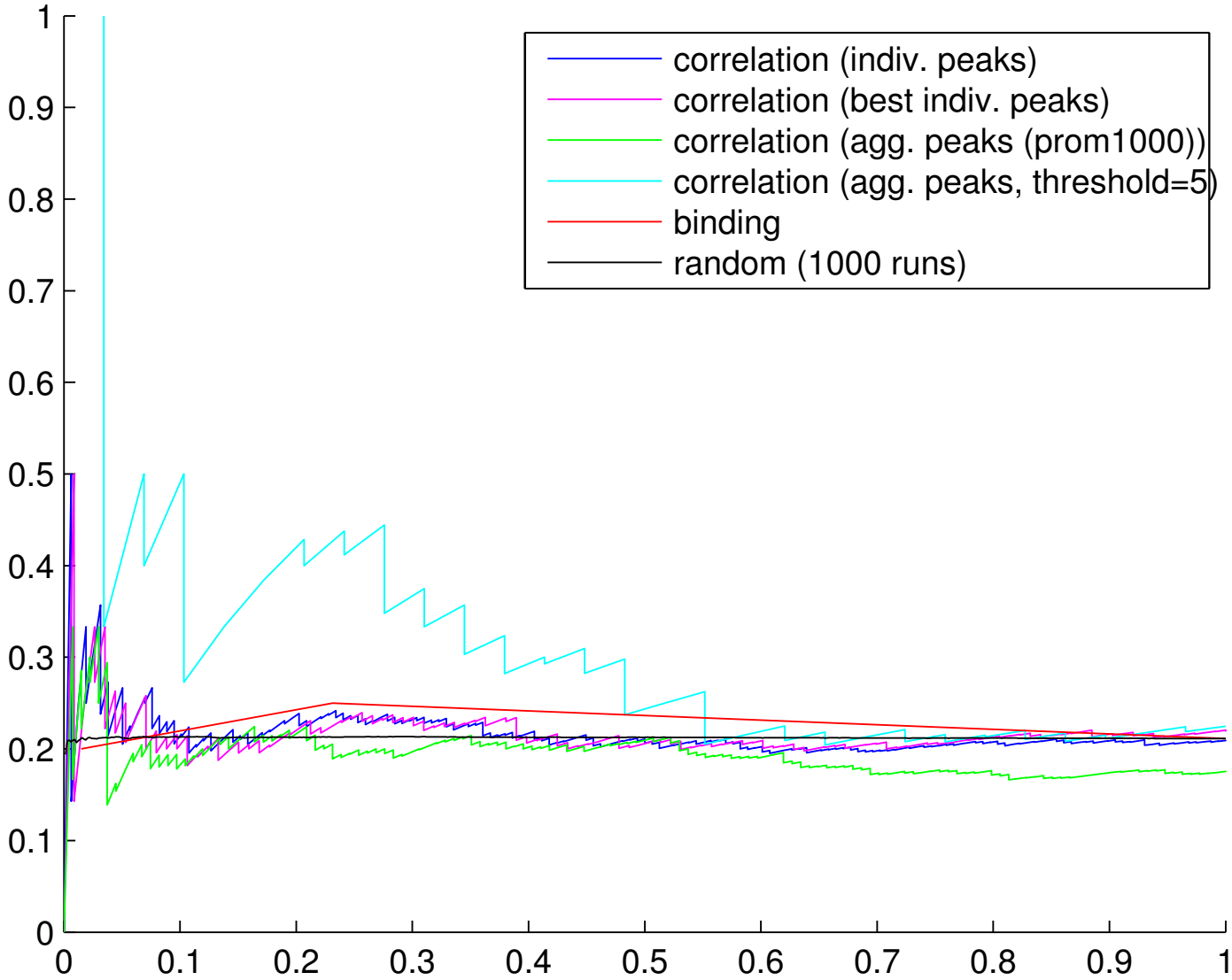
Tom's idea

- **RNA-seq** gives us genes expressed.
- **Correlation** between **ChIP-seq** and **RNA-seq** for a number of cell lines may give a better prediction.

Results so far

- Tried this on a number of datasets.
- ChIP/RNA correlation does seems to give better results for a number of data sets.
  - We're trying to identify why the best cases work.
- Placing a threshold on the ChIP-seq peak levels improves results.
  - But how to find the right threshold?
- Using individual ChIP-seq peaks doesn't work as well as using the aggregation of all binding sites for a gene.
  - This suggests that binding to combinations of sites within a promoter region is important.

Precision vs. Recall plots for ENCODE/EZH2



To do:

- What are the properties of the datasets where this works well, vs. where it doesn't?
- Optimise thresholds for data where we have KO using a precision/recall measure.
  - Can we learn something general about how to set the threshold?
  - If not, can we train a machine learning algorithm on the known datasets?
- . . . ?