An Autonomous Linear DNA Clock

Richard J. Crossland

OVERVIEW OF PROPOSAL

I propose my idea for an autonomous linear DNA clock as a component for synthetic biology. Since the conception of the original idea in February 2010, I have already made significant progress with four designs and a literature review for parts. However the project has recently adopted the Software engineering cycle so I have structured this document according to this approach.

Requirements

A 'construct that sequentially regulates gene expression after a time delay'.

1. Allow gene expression to be ordered.
2. Program duration is consistent.
3. Ability to alter total time in the program and the relative time of each step.
4. Level of gene expression at each step is controllable.
5. Can be destroyed after program is complete.
6. There is a mechanisms by which the program can be initiated.

Objectives

With reference to the requirements list:

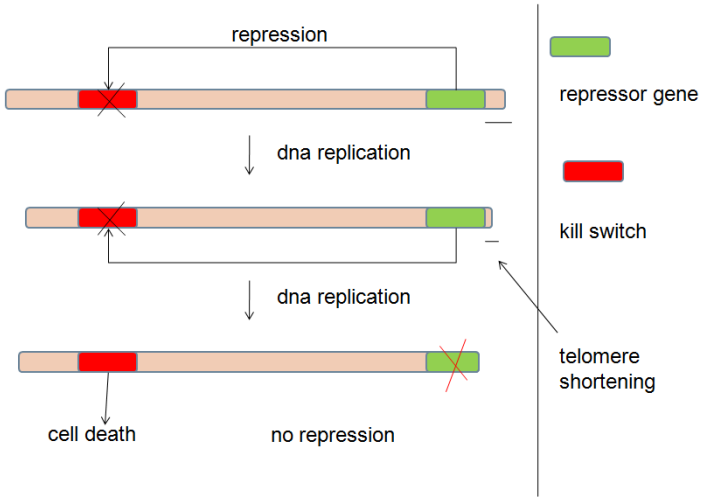
- to design and simulate several alternative models
- to list and specify the parts for my models
- to locate these parts from the scientific literature
- to list parts that need synthesising by synthetic biology
- to evaluate the best model in terms of the requirements and availability of existing parts.

With reference to the experimental work (optional):

- to demonstrate the validity of the concept of a shortening piece of (linear) DNA

Design 1: gap model

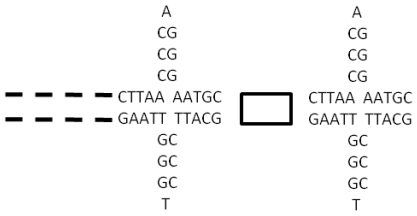
This mechanism is autonomous to the cell, encoded in a piece of linear DNA. However, the telomere repair mechanism is disarmed so that a 5' gap remains following DNA replication. After successive replications, the DNA becomes shorter and shorter from one, or both ends. This property is used to control the behaviour of the cell through the successive destruction of genes lying at the end of the DNA. The timing between these events is determined by the length of the structures on the DNA.



Mechanism for the linear DNA clock showing how telomere shortening eats away into genes that repress cell death pathways. Upon the destruction of the repressor by unrepaired telomere shortening, the cell-death pathway becomes uninhibited to kill the cell.

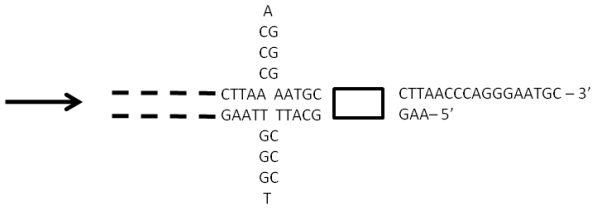
Design 2 : the ER2/gap model

At the end of a linear plasmid is a sequence of genes and secondary structures that are resistant to exonuclease attack (ER2s). An ER2 could be a folding of the DNA (e.g. a hairpin) or a protein complex that recognises a folded or unfolded DNA motif.



Before and after DNA replication/repair, exonuclease enzymes will attack both ends of linear DNA molecules, removing any unprotected nucleotides. Telomeric repair structures and ER2s--but not end-positioned genes--are protected from exonuclease attack.

Because of the end replication problem at the telomeres, DNA replication creates a gap at the 5' end of each newly-synthesised strand which, in the absence of telomere repair mechanisms, disrupts end-positioned ER2s rendering them susceptible to exonuclease (or endonuclease) attack.



Exonucleases remove all nucleotides up to the next resistant ER2, including disrupted ER2s and whole genes.

The advantage of this design is that it overcomes the immortal strand hypothesis problem (Rando, 2007) because exonucleases eat into both the Watson and Crick strand simultaneously. In addition, the design turns the problem of background exonuclease degradation to an advantage; whole gene sequences can be removed in a single time step (20 minutes) and the exact time delay can be set very accurately by a number of contiguous ER2s.

Design 3: the ER2 only model

This does not require 5' telomere shortening to disrupt ER2s. It requires two kinds of exonucleases, each with their own specific ER2s. An alternating switch coding for each of these exonucleases would control the shortening of linear DNAs that have a complementary pattern of genes and specific ER2s in its structure.

It is a very exciting prospect that this design solves the problem of individual clocks becoming asynchronous, ensuring that all linear DNA molecules are eaten away at a very similar rate. The time between each step could be independent of cell division if controlled by alternative process internal or external to the cell.

Design 4: the exonuclease model

This is a simpler model consisting of a piece of linear DNA, the shortening of which is governed by the activity of exonuclease enzymes present in the cell or encoded by the linear DNA itself.

This mechanism is independent of cell division and has no requirement for disarmed telomere repair mechanisms or ER2s.

Implementation (modelling of the designs)

Due to the specific mechanism of linear DNA shortening, I have chosen to model the designs in Java or R rather than by using modelling software. My progress so far has been to model the underlying shortening of the linear DNA for the gap and ER2/gap models. But in addition to the shortening of the linear DNA, I look forward to including a translational/transcriptional layer simulating how changes in the length of the linear DNA modulates the levels of individual gene expression over time.

Having established alternative designs that can be simulated as a general computer model, it will be necessary to simulate more specific and realistic models. In order to achieve this, I need a specification for the list of parts in the model, which will form a guideline for literature searches.

Verification/testing (literature review and experimental verification)

I aim to select parts from the literature or a synthetic biology repository that can be connected together as biobricks so that the linear DNA clock is itself a biobrick that is easily customised.

The modelling and experimental verification is hoped to be complementary. Mathematical analysis of the experimental data will allow predictions to be made about how the model might need to be improved via repetitions of previous steps in the Software engineering cycle. As an optional achievement, I would like to demonstrate that the concept of a shortening piece of DNA is a valid one. (The maintenance step will not applicable after the duration of the project.)

There are currently three ideas through which a linear DNA clock could be realised:

1. a non-bacteriophage-related extra-chromosomal linear plasmid;
2. a linearised bacterial chromosome;
3. a linear phage that infects the bacterial cell.

Since preliminary work has already been carried out in, it adds support to the practical applications of the idea if I outline briefly my most important parts so far identified. Note that there are a great number of papers which contain relevant information to this project. An extensive literature review is planned for the point in the cycle after the general models have been simulated.

Linear plasmid

Baker *et al.* (2007a, 2007b) have demonstrated that it is possible to transform *E.coli* with a 27 kbp linear plasmid from *Salmonella Typhi*. Along with a kanamycin cassette, the resulting plasmid, *pBSSB2*, was stably inherited for over 500 generations, even in the absence of antibiotic selection. It was not inserted into the chromosome at a detectable level. It has been sequenced and is predicted to contain 33 coding sequences (Accession number AM419040).

The ends of *pBSSB2* contain tandem inverted repeats (tirs) which are probably capped with covalently-bound protein, and not closed hairpin loops. These are implicated in telomere repair. A final interesting feature is that the GC skew and coding bias suggest that *pBBSB2* replicates from an central internal origin. Both these facts suggest that this plasmid is very similar to Streptomyces linear plasmids which are well studied in the literature (Meinhardt,F. *et al.,* 1997). An additional similarity revealed so far is that one tir of both *pBBSB2* and Streptoymces plasmids can be removed without any detrimental effects on DNA replication (Rauland *et al., 1995;* Baker *et al.*, 2007a).

Linearised circular genome
Cui *et al.* (2007) linearised the circular genome of *Escherichia coli* using the lysogenic λ-like phage N15. Cells with a linearised genome were viable, showed no appreciable differences in growth or gene expression, and their genome structure was stable for at least 170 generations. Linearisation was mediated by the TelN protein which binds to a *tos* site, making staggered cuts on palindromic sequences either side of *telRL.* This produces two separate DNA ends which, through self-annealing and sealing by TelN, produces hairpin termini.

Areas/problems already identified for literature research
Nuclease activities:

- Identify nuclease enzymes that attack linear ds-DNA
- Quantify the rate of exonuclease attack on exposed linear DNA ends.
- Find structures (ER2s) which prevent or slow down exonuclease attack

Linear DNA:

- Identify genes that are essential for maintaining and replicating linear DNA. Can this be predicted by homology to linear molecules that have already been characterised?
- How to prevent two exposed linear DNA ends from forming a circular molecule.
- How long is the 5' gap at the telomeres following DNA replication?
- How to induce the linear clock mechanism. Will modifying the palindromes of the *telRL* site prevent TelN recognition?
- How can the telomere repair mechanism be disarmed. What proteins should be targeted? Which endonuclease sites should be introduced into the DNA sequence.

Other parts: fluorescent reporter genes, strong and weak repressors/activator genes, intracellular and extracellular signals.

Advantages of the linear DNA clock over alternative clock mechanisms

1. The timing of events is specified by the distance between structures on a single molecule rather than by changes in the concentration of complex intra- or extracellular signalling mechanisms.
2. Unlike other clocks it is autonomous to the DNA and is independent from external signals. For example, this may be an advantage when using bacteria in inaccessible environments such as for the repair of micro-cracks in concrete.
3. Bacteria are transformed once with a custom-made linear DNA biobrick which can be induced at a convenient time.
4. As an ethical consideration, it can prevent of the evolution or escape of GM organisms by a kill switch that is activated after a time delay. If all GM genes were contained on a linear plasmid which ensured its own destruction, the organism would no longer be GM.

RESEARCH PLAN

The iGEM wiki page will be updated with all progress on the project. All .pdf files will be edited by Foxit reader and stored in the iGEM dropbox.

Requirements

Finalised on 13 April 2010. Unlikely to change during the project.

Design

Final cut-off date for implementation of new ideas must occur before designs are shortlisted.

Implementation

All implementations are possible at this early stage of the cycle, but designs 2 and 4 are the most promising. I expected the transcriptional/translational layer of all the models to be almost identical.

All designs are affected by the activities of exonucleases. Therefore design 4 is given priority. However, I do expect that unprotected DNAs will disappear very quickly. If it were possible to find structures from the literature that limit this rate, this could become a good model. It is certainly the first model I wish to test in the lab, using fluorescent reporter genes.

The modelling of all designs will show whether they are viable or contain flaws. If any flawed designs can't be improved, only the parts for shortlisted viable designs will be pursued.

Verification/testing

A list of parts and their specifications will be drafted to aid the literature search. Once as many parts as possible have been found, they will be used to create a specific model for each of the designs.

There is limited potential in this project to conduct an experiment using the specific models of my designs, although I expect to complete the planned literature as early as possible to improve the chances of this happening. Since the plan for lab work is highly dependent upon the availability of parts and the success of the modelling, its planning has been kept minimal. I would however like to achieve my objective of proving the underling concept of telomere shortening.

Literature research	Example search words:
Exonuclease enzymes (1 week)	resistant, structures, (deoxyribo-)(exo-)(endo-)nuclease, DNA modification, stability, digestion, degradation, resistance, hairpins, protection, shortening, in *B.subtilis*/*E.coli*, exonuclease-deficient strains, poly-adenylation
Linear DNA (2 weeks)	linear DNA, linearisation, telomeres (loss), telomerase, tirs, terminal proteins, (N15) phage, shortening, gene deletion, phase changes, eating away, ends, termini, 5' gap/3' overhang, replication, palindrome, homologous recomb.

References

Baker,S. *et al.* (2007a) A novel linear plasmid mediates flaggellar variation in *Salmonella Typhi*. *PLoS Pathog*, **3**(5), e59.

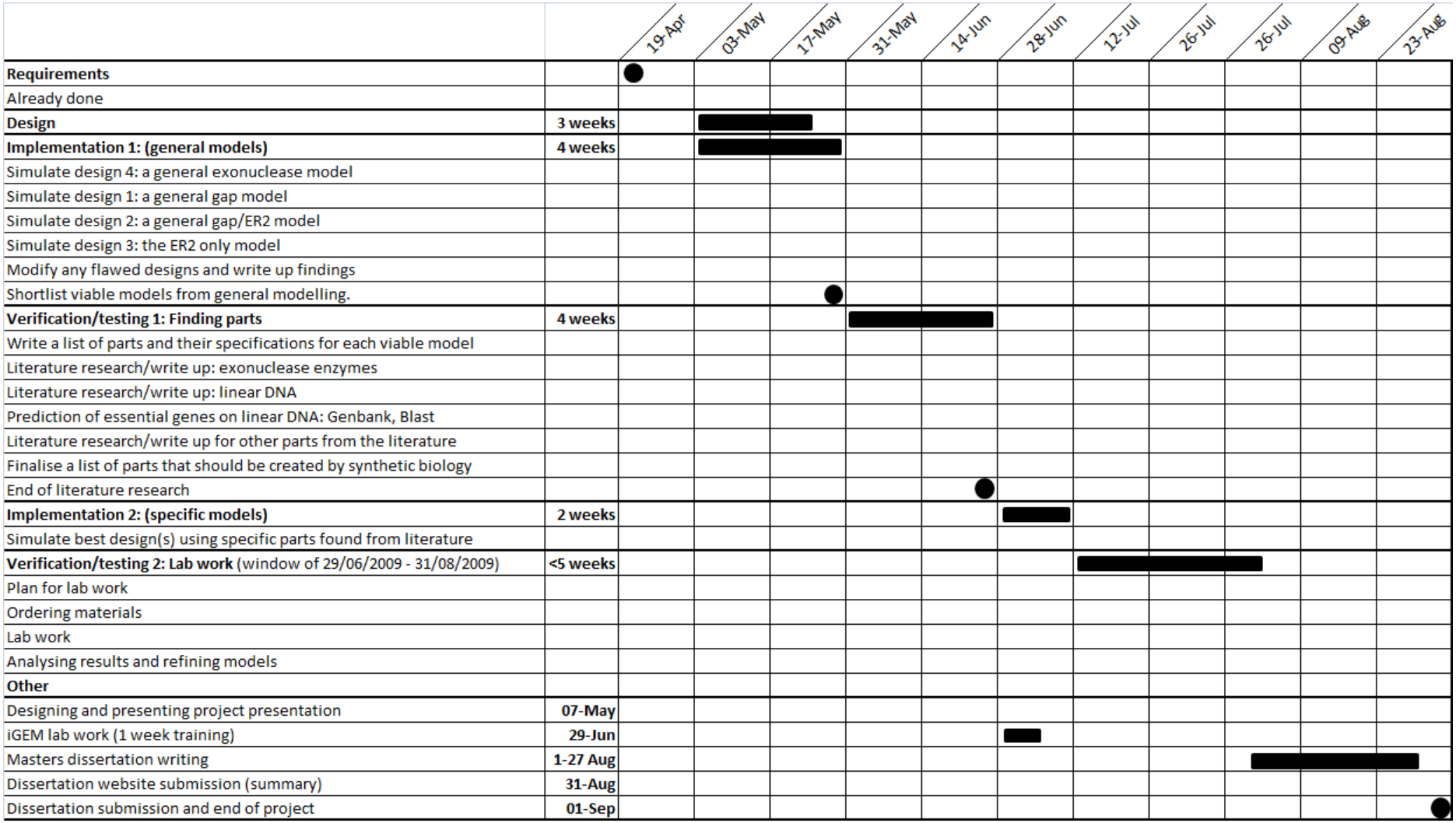
Baker,S. *et al.* (2007b) A linear plasmid truncation induces unidirectional flagellar phase change in H:z66 positive *Salmonella Typhi*. *Molecular Microbiology*, **66**(5), 1207–1218.

Cui,T. *et al.*(2007) *Escherichia coli* with a linear genome. *EMBO reports,* **8**(2), 181–187.

Rando,A.T. (2009) The Immortal Strand Hypothesis: Segregation and Reconstruction. *Cell*, **129***,* 1239–1243.

Rauland,U. *et al.* (1995) DNA amplifications and deletions in *Streptomyces lividans* 66 and the loss of one end of the linear chromosome. *Mol Gen Genet*, **246**, 37–44.

Gannt Chart



Requirements	
Already done	
Design	**3 weeks**
Implementation 1: (general models)	**4 weeks**
Simulate design 4: a general exonuclease model	
Simulate design 1: a general gap model	
Simulate design 2: a general gap/ER2 model	
Simulate design 3: the ER2 only model	
Modify any flawed designs and write up findings	
Shortlist viable models from general modelling.	
Verification/testing 1: Finding parts	**4 weeks**
Write a list of parts and their specifications for each viable model	
Literature research/write up: exonuclease enzymes	
Literature research/write up: linear DNA	
Prediction of essential genes on linear DNA: Genbank, Blast	
Literature research/write up for other parts from the literature	
Finalise a list of parts that should be created by synthetic biology	
End of literature research	
Implementation 2: (specific models)	**2 weeks**
Simulate best design(s) using specific parts found from literature	
Verification/testing 2: Lab work (window of 29/06/2009 - 31/08/2009)	**<5 weeks**
Plan for lab work	
Ordering materials	
Lab work	
Analysing results and refining models	
Other	
Designing and presenting project presentation	**07-May**
iGEM lab work (1 week training)	**29-Jun**
Masters dissertation writing	**1-27 Aug**
Dissertation website submission (summary)	**31-Aug**
Dissertation submission and end of project	**01-Sep**

Key

▬ time taken to complete task ● milestone