Proofs of concept: the implementation of a NAND gate in *Escherichia coli* and the use of 'ribolocks' in controlling protein transcription

Kyliah Clarkson
V00172740
Micr 470 (iGEM)

Introduction

The logical operation 'not and' (NAND) is a function that takes two inputs, which may be either 0/false or 1/true, and sets output to 0/false only when both inputs are 1/true.

input A	input B	output
0	0	1
0	1	1
1	0	1
1	1	0

The NAND operation is functionally complete - that is, any other logical operation can be implemented by combining two or more NAND operations. The NAND operation can also be implemented inexpensively in hardware, by using two transistors in series to form a NAND gate. These facts make the NAND operation one of the bases of modern electronics; operations on binary numbers, flash memory, computer processors - all made from sufficiently large collections of NAND gates.

The construction of transistors relies upon the availability of certain metals and semi-conductors. There is a limited supply of these elements on Earth, and recycling programs for electronics are not yet widespread. Although rather science fiction in flavour, biological processes could theoretically be harnessed for use as computers, both simple and complex (i.e. Adleman). The field is in its infancy at the present. What functions can be best handled with DNA programming? With substrate detection and production? With cellular networks and intercellular communication? Crude proofs of concept still need to be laid for the many possibilities available before more elegant processes can be developed.

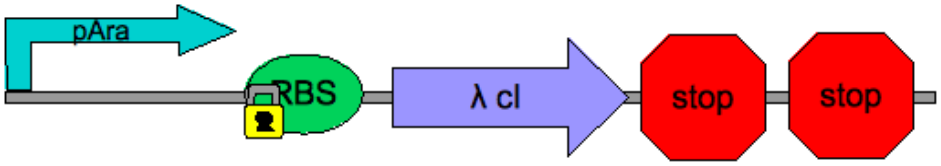
The Registry of Standard Biological Parts is a growing repository of DNA sequences which can be assembled in host cells to generate biological machines. The DNA 'parts' are available in a standardized format, called 'BioBricks', designed to permit assembly-line-style construction of system processes. However, many of the available parts still require further characterization to ensure that they will work as intended, no matter the system into which they have been integrated.

Project Proposal

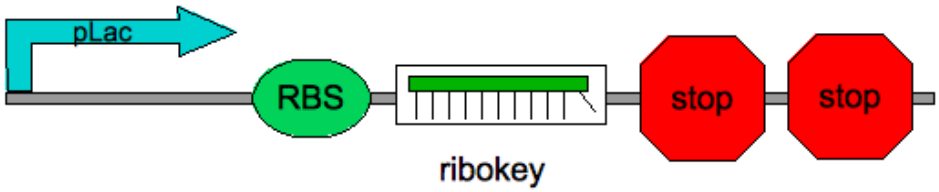
Using BioBrick parts, it should be possible to 'program' a strain of *E. coli* cells to function as simple NAND gates. Such a strain would need to detect two signals (light; temperature; quorum-sensing molecules; metabolites; something else), and respond with the production or non-production of a third signal (a dye, fluorescent protein, or other visible molecule; quorum-sensing molecules; a cell-termination signal; something else).

One possible setup which should function as a NAND gate would contain the following three segments:

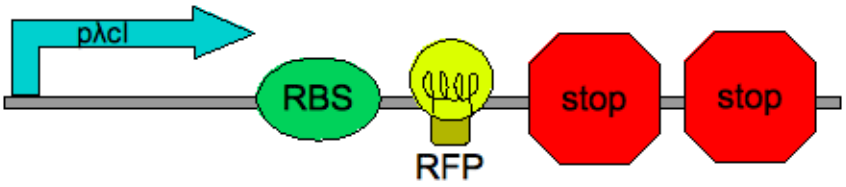
- Detection of input A: the promoter from the ara_{BAD} operon, controlling the transcription of one of the lambda phage repressor proteins (λcI). Instead of a standard ribosome binding site (RBS), there would be a 'ribolock': the RNA transcript will fold in upon itself and form secondary structure at the RBS. This would strongly hinder, or even prevent, translation of the repressor protein unless the structure is forced to unfold.



- Detection of input B: the promoter form the *lac* operon, controlling the transcription of a 'ribokey': an RNA sequence with regions which are complementary to certain upstream regions of the ribolock. When the ribokey associates with the ribolock, it causes the secondary structure to unfold, thus exposing the RBS to ribosomes and allowing the transcription to occur.



- Output: a constitutive promoter which is repressed by λcI, controlling the transcription of a red fluorescent protein (RFP).



The net result with this setup is that RFP will be continuously expressed except for when arabinose is present in the medium (allowing the transcription of mRNA for λcI) at the same time as is lactose (allowing the production of the ribokey which permits translation of λcI), and glucose is absent (forcing the cells to an alternate food source). When both sugars are available in the culture medium, the transcription of the RFP gene will be repressed and there will be no fluorescence.

A secondary aspect to this project could be to further characterise the ribolock system. The Berkeley team from the 2006 iGEM competition developed a number of ribokeys and ribolocks, some of which had very good preliminary results. However, they only tested the locks and keys when under the control of the tetracycline-responsive (tetR-repressed) promoter; did not release their best lock to the Registry, only one of the precursors; and seem to have tested the efficiency of the locks only under standard culture conditions (i.e., only at 37°C).

This project as organised above can be extended to any or all of these three goals. First, this project will place the ribolock and the ribokey under the control of two different promoters, neither of which is the one used by Berkeley. Any aspects of ribolock efficiency or expression pattern which are affected by the promoter sequence should be revealed.

Second, the difference between Berkeley's best ribolock and the best available is a matter of nine base pairs on the 5' non-complementary tail. There is no simple way to add that small a number of bases using only BioBrick parts. However, a lock variant could be constructed by prepending the shortest non-RBS sequence in the registry to the available lock; the additional length would be approximately twenty-four bases. It would be interesting to attempt this, to see what sort of lock can be constructed only from 'off-the-shelf' BioBricks. Alternatively or additionally, the ribolock sequence could

be synthesised by a company, at which point it could be put into BioBrick format and submitted to the Registry if if proved to be worthwhile.

Third, the ribolock depends upon complementary binding of sequence in the mRNA transcript. Different culture temperatures might contribute to different efficiencies of repression in the absence of the key, or to different efficiencies of unlocking in its presence. It may be worthwhile to culture *E. coli* containing the ribolock at different points along its viability range, to determine the effects of temperature on efficiency.

Materials and Methods

BioBrick parts all have a standardised prefix, with restriction sites for EcoRI, NotI, and XbaI; and a standardised suffix, with restriction sites for SpeI, NotI, and PstI. XbaI and SpeI both form the same 3' overhangs, which can anneal together to form a 'scar' site. This scar is subsequently vulnerable to neither enzyme. The XbaI-SpeI method is the standard method used to join together two BioBricks.

BioBricks are shipped as components inside plasmids. These plasmids are guaranteed to carry at least one antibiotic resistance gene, to allow screening of *E. coli* cells which have been transformed with a plasmid.

A wide variety of BioBrick parts could come to be used in this project; a list of them follows on a later page. These parts are all available in the 2009 iGEM distribution, except for the part representing the best ribolock from Berkeley. However, some of the core parts are in plasmids which can be induced to high copy number when exposed to IPTG. Since IPTG will also activate the *lac* promoter which is a part of the system, the parts will need to be assembled in a specific order.

For example, the assembly of the λcI-repressed promoter-RBS-RFP composite component could be implemented as follows:

1. In separate reactions, transform *E. coli* cells with the plasmids bearing BioBrick parts R0051 (promoter), B0034 (RBS), E1010 (RFP), and B0014 (transcription termination). Unsuccessful transformations can be screened by culturing the cells on medium containing ampicillin.

2. After culturing the transformed cells, purify the four plasmid types by using spin columns. After use, the columns can be sterilised and re-used.

3. In separate reactions, cut R0051 and E1010 with EcoRI and SpeI, and B0034 and B0014 with EcoRI and XbaI.

4. Perform gel electrophoresis on the products of these four reactions. For R0051 and E1010, recover and purify the shorter sequence; for B0034 and B0014, recover and purify the longer sequence.

5. Set up ligation reactions to connect R0051 to B0034, and E1010 to B0014.

6. Transform *E. coli* cells separately with the R0051-B0034 and E1010-B0014 plasmids. Again, ampicillin can be used to screen unsuccessful transformations.

7. Purify the two plasmid types with spin columns.

8. Cut R0051-B0034 with EcoRI and SpeI, and E1010-B0014 with EcoRI and XbaI.

9. Perform gel electrophoresis. Recover and purify the shorter sequence from R0051-B0034, and the longer from E1010-B0014.

10. Set up a ligation reaction to connect R0051-B0034 to E1010-B0014.

11. This should be the finished component. With the correct primers, a round of PCR could be used to amplify the insert for sequencing, to ensure that no errors have crept in. *E. coli* must be transformed with the plasmid as well. Ampicillin can be used for screening; cells containing plasmids which have assembled correctly should also express RFP.

Similar steps can be followed for the construction of every other composite component required by the project. In fact, the various assemblies can be performed in parallel. This can save time and reagents by preventing the need to mix solutions on multiple occasions.

The final product (i.e., a strain with all three composite components diagrammed above) should be able to be assembled onto a single plasmid. There will, naturally, be a large number of intermediate steps. The estimated length of the simple and composite parts required for the initial proposal are listed later as a reference.

Discussion

The project as laid out above is not without its potential pitfalls. Some of these issues are considered here.

- The ribokey, the ara_{BAD} promoter, and the histidine tag are all reported to have 'inconsistent' sequence reads in this year's distribution. This may reflect legitimate sequence error. However, secondary structure is the more likely culprit. The ribokey is designed to fold in upon itself a great deal. The promoter is designed to be held in a hairpin by a protein dimer (Schleif); it is plausible that the same sequence is used for each recognition site, and could lead to secondary structure during PCR. The histidine tag is very short and repetitive, and silppage may have prevented a good sequence read. However, since it is the length of the sequence that matters in this case, it doesn't make much difference if the identity is accurate or not.

- The *lac* operon is somewhat 'leaky' in nature, since its repressor protein is under its control. This may result in some ribokey being produced when lactose or IPTG is not present. If this has too large an effect on RFP levels, an alternate composite can be constructed using one of the less-efficient RBSes. If RNA polymerase recognises the binding site less efficiently, it should balance out the increased opportunity for recognition.

- The λcI protein in the Registry has an LVA tag appended. Such a tag induces the shuffling of the protein into the degradation and recycle process. If RFP continues to be made, regardless of the amount of arabinose and lactose which are present, it may be that the λcI protein is being degraded too fast to allow for repression to occur. In that case, the lambdaphage-derived promoter and repressor could be substituted out for analagous, non-LVA-tagged, parts from phage 434.

- Culturing the cells on glucose for rapid growth will allow production of RFP. However, this excess RFP could be retained for a long period of time after being transferred to a medium with arabinose and lactose as carbon sources. Should this occur, the tests would yield a false failure of the construction. If it does seem that the lifetime of RFP molecules is too long, then, an alternate construction using a TVA-tagged cyan fluorescent protein (CFP) may suffice.

- If it turns out that one of the sugar promoters is simply not suited to the task at hand, or that both promoters simply refuse to activate simultaneously, there are alternatives. One is to construct a constitutively-expressed *lux* regulator protein, replace the problem promoter with the *lux* right promoter, and use AHL as an input signal in the medium.

- It may turn out that having all three of the components directly adjacent to one another will cause the regulator molecules to hinder each other. Expression levels would in this case be strangely affected. Should this happen, then one or more of the components would need to be shuffled to a different plasmid. The new plasmid(s) would need to bear a different origin of replication, to ensure that every plasmid would be retained. Such a variety of plasmids are available in the 2009 distribution of parts.

Parts Needed

Table 1: Part Identities

Serial number	Part	Notes on load-bearing plasmid
I0500	ara_{BAD} promoter with araC repressor	kanamycin resistant, IPTG-inducible copy number
R0010	*lac* promoter	ampicillin resistant
E1010	RFP	kanamycin resistant, IPTG-inducible copy number
B0034	standard RBS	ampicillin resistant
B0014	double transcription terminator	ampicillin+kanamycin resistant
J23066	ribokey	ampicillin resistant
J23032	ribolock	ampicillin resistant
C0051	λcI (rapid degrade)	ampicillin resistant
R0051	λcI-repressed promoter	ampicillin resistant
B0030	RBS (60% standard rate)	ampicillin resistant
B0032	RBS (30% standard rate)	ampicillin resistant
C0056	434cI	ampicillin resistant
R0052	434cI-repressed promoter	kanamycin resistant, IPTG-inducible copy number
K128005	histidine tag	ampicillin resistant
E0022	promoter plus CFP (rapid degrade)	ampicillin resistant
C0062	*lux* regulator protein	ampicillin resistant
R0062	*lux* right promoter	ampicillin resistant
J23078	best ribolock	N/A

Table 2: Estimated Part Lengths of Intermediates and Final Constructs

Serial number	Part	Length (bp)
I0500	ara_{BAD} promoter	1210
R0010	*lac* promoter	200
E1010	RFP	681
B0034	standard RBS	12
B0014	double transcription terminator	95
J23066	ribokey (includes terminator)	336
J23032	ribolock	43
C0051	λcI (rapid degrade)	750
R0051	λcI-repressed promoter	49
I0500-J23032	p_{BAD}-ribolock	1261
C0051-B0014	λcI-term	853
I0500-J23032-C0051-B0014	p_{BAD}-ribolock-λcI-term	2120
R0010-B0034	p_{lac}-RBS	220
R0010-B0034-J23066	p_{lac}-RBS-ribokey	564
E1010-B0014	RFP-term	786
B0034-E1010-B0014	RBS-RFP-term	804
R0051-B0034-E1010-B0014	$p_{\lambda cI}$-RBS-RFP-term	861
I0500-B0034	p_{BAD}-RBS	1230
I0500-B0034-C0051-B0014	p_{BAD}-RBS-λcI-term	2089

Conclusion

A broad outline of the methods needed for the assembly of a simple NAND gate in a strain of *Escherichia coli* has been made. The two input signals have been proposed to be lactose or IPTG, and arabinose; the output signal has been proposed to be red fluorescent protein, which is visible to the unaided eye at higher concentrations. Further, the initial planning for a rigorous testing phase has also been laid out.

References

Adleman, L. (1994). Molecular computation of solutions to combinatorial problems. *Science* **266**:5187, pp. 1021-1024.

Schleif, R. (2000). Regulation of the L-arabinose operon of *Escherichia coli*. *Trends in Genetics* **16**:12, pp. 559-565.