A Standard Parts List for Biological Circuitry

Background and Overview

One of the hallmarks of biochemical circuits found in nature is analog, asymmetric, asynchronous design. That is, there is little standardization of parts, e.g. all the promoters have different strengths and kinetics, transcription factors are designed to have different effects at different loci, and each enzymatic reaction has its own idiosyncratic mechanism and rates. In addition, all of the heterogeneous circuit elements are executing their functions concurrently and asynchronously. Biological circuits are seemingly designed to deal with the fluctuating delays, different time-scales and energy requirements associated with each component process of the overall network. These factors also make design of novel biochemical circuitry from existent parts difficult to achieve. Without standardization, the qualitative design methods used in other engineering fields are simply inapplicable. The *de facto* design methodology for biological circuitry is natural selection. Rational design of biological systems by humans has remained restricted to rather small or hit-or-miss efforts and has often relied on the ability to "select" for biochemical parts that fulfill some criteria. In practice however biological-designers are rare, and solutions are usually realized through an expensive stepwise trial and error approach or through mutation and selection. Furthermore, these otherwise practical approaches are limited in terms of the problems they can solve. We believe that implementation of designed biological circuitry is limited by issues of practice.

We believe that implementation of designed biological circuitry is, in large part, limited by issues of practice. To address this deficiency, we propose herein a program to produce a set of well-characterized and systematized biological components that can be generically assembled to create custom biological circuitry. Our goal is to lead to a toolkit of biological reagents that can be assembled into biological circuitry in the same way resistors, capacitors, and transistors can be assembled to make electronic circuitry.

This proposal has four specific aims:

1. The selection, refinement, and standardization of existing biological components such that they become useful for implementing biological circuitry,
2. The need-driven creation of novel components,
3. The extensive characterization of each component's "device physics" sufficient to allow the reduction to component input/output in design simulations of complex circuitry.
4. The construction of a small set of test circuits to demonstrate the efficacy of the part design.

The material deliverables here are a set of "pluggable" reagents that may be assembled into circuits. By "pluggable" we mean that are designed in such a way that they may be particularized by the clipping in of specific DNA sequence (either genes or regulatory regions). A good example of a "pluggable" system already in general use is the two-hybrid system that allows the "plugging-in" of two test proteins into specially designed vectors that, in the properly engineered strain of yeast, can test for protein-protein interaction. We envision using constructs of similar spirit to attach well-known degradation or localization signals to proteins, or which reverse the two hybrid experiment to allow fusion of two proteins (say a kinase and response regulator that do not naturally interact) to well-known interaction domains. Similarly, we envision constructs that allow the specific engineering of the translational efficiency and degradation rate of RNA molecules by the "pluggable" swapping of 3' and 5' untranslated regions (UTRs). We describe in a little more detail a variety of such constructions below. The characterization of the reagents and the input/output simulation models of these reagents are also deliverables of this proposal.

It is imperative, that we reach a level of engineering facility with these systems comparable to that we have for mechanical and electronic systems if we are to achieve a number of central industrial and medical goals. There are major industrial efforts in pathway engineering of plants, animals and bacteria for production of biochemicals for pharmaceutical and agricultural needs, for bioremediation and for improvement of the organism itself as an agricultural product. These efforts currently suffer from the lack of well characterized

We will confine ourselves initially to Escherichia coli (E. coli) and Saccharomyces cerevisiae (yeast), as our test beds for our constructs. These two organisms are central model organisms as well as being of industrial interest. In addition, organisms have unique properties that make them more or less amenable to cellular engineering.

Refinement of Existing Components and Benchmarking

As the fields of biochemical and genetic engineering are already well established, a major part of our proposal involves the collection, characterization, and systematization of existing biological components. What makes this useful is the level of component standardization and experimental characterization we will apply throughout the process. Below are described some of the components we will initially pursue. We will focus in the course of this proposal only on those components that aid in building our test circuits. As time and need progresses we will collect and engineer more the components we discuss below. The basic idea in all cases will be to produce set of vectors that allow engineering at three levels of biochemical circuitry: transcriptional control, translational control and post-translational reactions. Diagrammatically, there are a number of regions in a vector that we define. We show the eukaryotic case because the prokaryotic case is a defined subset.

Not all vectors will have all components. Different vectors will have different purposes (fusion construction, expression, etc.). These vectors will be engineered with compatible restriction sites between regions of interest. We break engineering vectors into regions as follows:

1. Resistance Cartridge	For passage between, *E. coli* and yeast, for maintainance of multiple plasmids in *E. coli*
2. Proteotrophy	To allow selection in yeast, different proteotrophies allow maintenance of multiple plasmids in yeast
3. Promoter Control	To allow selection of polymerase and transcription factor specificities
4. 5' UTR	For engineering of RNA translational efficiency and stability (e.g. stem-loops, antitermination sites)
5. N'-end fusion	For engineering the N-end of particular proteins (e.g. to allow degradation or localization signals)
6. Gene	Insertion point for protein of interest.
7. C-end fusion	C-end fusion protein (e.g. for fusion of domains with specific activity)
8. 3' UTR	To allow engineering of RNA stability
9. Term./IRES	To allow engineering of elongation control elements. (IRES= internal ribosome entry site)
5' UTR	Next RNA or protein gene...

Transcriptional Design Components

Components of this type allow for the regulation of transcription. Engineering of these component parts concerns boxes, 3,4, and 9 on the diagram above. We have identified a number of existent components that may be useful in creating standardized reagents with specific transcriptional activities:

Promoter Control Engineering

The section concerns engineering of box 3 on the above diagram and the collection of proteins that can interact with elements within box 3. The elements within box 3 include RNA polymerase binding sites and operator regions (or upstream activating, repressing sequences, U(A,R)Ss). The goal here is to create a set of well-characterized transcription initiation control "gates" with different molecular specificities and activities such that complex circuits can be built from them.

1. We propose to collect a set of fixed orthogonal RNA polymerase-promoter series. Orthogonal in the sense that the polymerase from one series does not recognize promoters from another. Series in the sense that each polymerase-promoter pairing is available over a range of activation levels created by mutation in the promoter and/or polymerase. Fixed in the sense that variation in activity requires genetic modification of either the polymerase or the promoter,[1, 2]
2. These will be complemented by a set of variable orthogonal RNA polymerase-promoters. As above except that the activity of each RNA polymerase-promoter pairing is subject to the level of an activating or repressing species. [2-5]
3. For yeast constructions, we will collect a set of endogenous and exogenous transcription factor binding sequence for direction of protein to the promoter (e.g. synthetic gal4 UAS, lexA binding motif, repressor binding motif [6]) and a set of promoters that are active under different conditions, such as

P_{Gal1} that must be activated using the gal4 activation domain and P_{CYC1} that can be repressed by SIN3 fusions [7].

For each of these series of transcription control elements, adaptor sequences will be designed that allow the mixing and matching of relevant pieces into vectors designed to accept them.

Translational and RNA Design Components

This section concerns the engineering of boxes 4-9 in the diagram above. Components of this type allow for the regulation of RNA activity, stability, translation efficiency and even localization. Without going into detail, such signals that may be encoded include:
1. A set of 5' and 3' untranslated stem loop structures that affect the translational efficiency and stability of RNA, [8-14]
2. A set of localization signals that affect transport of functional RNA, [15, 16]
3. Possible production of splice junctions for control of swappable protein domains, [17, 18]
4. We propose the creation of Internal Ribosome Entry Segments (IRES') for use in yeast either by producing strain deficient in IRNA [19] which may have unanticipated side effects, or by selection of a functional IRES. One possible mode of selection is construction of a vector bearing resistance for, say, tetracycline and zeomycin separated by a known picornavirus IRES [20-22]. Colonies are then selected on tet/zeo plates with varying amounts of zeomycin.

As discussed briefly below, functional RNA's may be included at these sites (spanning boxes 4-9).

Protein Design Components

This sections concerns engineering of boxes 5-7 to produce proteins with specified properties. Components of this type allow for the regulation of protein stability, connectivity, and localization. Such components include creating or collecting and characterizing fusion vectors
1. with different N-end sequences, in the position of box 5 above, following Varshavsky rules for degradation of eukaryotic proteins, [23-25]
2. with different N- and C-terminal sequences that encode localization and secretion signals.
3. with different members of protein-protein interaction pairs (as determined by two-hybrid or other biochemical/genetic assay) or interaction pairs developed from peptide aptamer libraries,[26]
4. with different DNA binding domains (e.g. gal4, lexA, CI repressor) or transcriptional modulation domains (e.g. gal4, sin3), or particularly designed zinc-finger domains [3, 27-30] or designed PDZ protein interaction domains [31] (and their cognate C-terminal targets)

These components are to serve both as transcriptional regulators and the basis for construction of protein-mediated signaling circuits for faster applications.

Component Standardization, Compatibility and Benchmarking

In order to use all these different types of engineered biocomponents two major tasks must be completed: 1) The design of the various vectors and subsequences must be such that ultimately they may be used to create circuits of relatively high complexity, and 2) Each of the components must be sufficiently biochemically characterized such that their behavior in a larger circuit may be predicted. The first task involved careful planning of adaptor sequences and vector "blanks" into which specific sequences corresponding to elements in boxes 1-9 may be inserted and recombined. This, in itself is a major undertaking, however, we will limit ourselves initially only to those reagents necessary for creation of our test circuitry and for testing some of the components we collect or reproduce. A good existing example of what we mean by vector blanks are the vectors included in two-hybrid system kits which bear particular promoter/UAS, proteotrophy and resistance cartridges, and a fusion sequence with an engineered restriction site into which a particular protein gene may be clipped.

The second task involves direct measurement of the device "physics" for each component. The high water mark for such measurements can be found in papers such as [5, 32-37]. In essence, these papers measure relatively precise binding constants, transcriptional activities, elongation rates, termination effectiveness, RNA stability and translational efficiency, and protein kinetics. Our goal is to select for components that obviate much of the need for detailed physical models of each component. Because of good component design, detailed physical models for transistors are not required for the somewhat accurate simulation of logic circuits. Our goal is to produce a set of biocomponents whose properties are such that model abstractions at a higher level than the molecular kinetics are valid. For example, definitive promoters that become fully active at very low concentration of the cognate polymerase look very much like a logical buffer. Designing and selecting

components that meet certain criterion like these allows better application of existing circuit design theories to biological circuit design.

Development of New Components and Functions

As part of this project we propose to develop and test a number of novel parts and methodologies. These are in addition to the collection, systematization and selection of the parts discussed above. We wish to develop new types of parts that might not be found naturally in biological systems. Here we briefly discuss two examples.

We propose to try to use specially design RNA's as scaffolding to enforce enzymatic specificity and perhaps to cause localization of specific proteins. One example is shown in the figure below. Here two RNA molecules, each of which binds to a particular protein (e.g. one to a kinase and the other to a response regulator), also contain regions complementary to one another. The RNA-protein binding specificity is engineered through SELEX evolution of the RNA aptamer. When all four molecules are expressed, then the RNAs bind to their protein targets and RNA-RNA binding via the complementary regions brings the two proteins together. The hypothesis here is that the high local concentration of kinase and response regulator will cause an increased specificity of the kinase for the chosen target. The two RNAs can be made into one RNA by placing the proper stem-loop structure between the protein-binding regions. In addition, it is possible that by adding another protein binding region to the opposite end of one of the RNAs that the entire scaffolded complex can be targeted to a third localized protein.
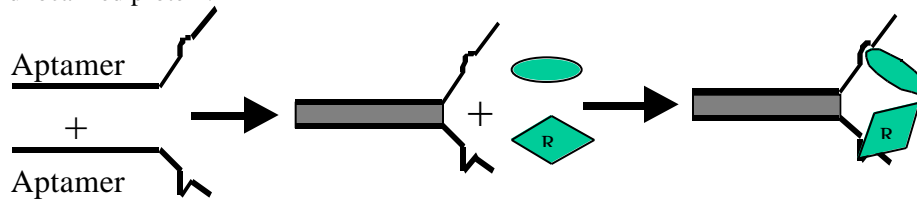
Another approach to creating specific protein-protein interaction pairs is to create peptide aptamer partners. One possible method of pair creation is the use of the two hybrid system. Here, a library of peptide aptamers is fused to DNA binding domain in a two-hybrid vector. A similar library is made with fusion to the transcriptional activation domain. Theoretically, then, cotransformation of the two libraries into yeast bearing a reporter vector in which transcriptional activation leads to production of a resistance gene allows for the screening for pairs that recognize one another. In practice, a small number of "orthogonal"(non-cross reacting) aptamers will provide be used as the bait against which the full library is screened.

Computational Support

In order to facilitate this process and keep track of collected and engineered parts and their status, a database of biomolecular engineering rules and reagents will be created. During the reagent characterization process, dynamical models, at possibly multiple levels of abstraction, will be made of the biochemistry. These will then be used in the creation of a computer-aided biomolecular design package to aid in creation of more complex circuitry.

List of Example Applications

Detailed descriptions of several sample applications that would result from the successful application of the research we outline above are available. These include:
1. A biological NAND gate created in of *S. cerevisiae.* (Arkin, Data available).
2. A biological Flip-Flop, again in *S. cerevisiae,*
3. A new model-bacteriophage infecting *E. coli.* (Endy)

Implementation and Timeline

Tasks to develop and deploy standardized biomolecular engineering toolkit:

1. Collate all experimentally validated rules for biomolecular engineering of DNA, RNA, and proteins. Choose a subset of these for creation of standardized reusable "engineering" reagents.
2. Implement of reagent design at all levels of the problem from vector to organism. Characterize the biochemical/kinetic properties of these components to the level that they may be accurately modeled in a simulation package.
3. Test the RNA scaffolding component design. Develop streamlined aptamer-pair development apparatus. Develop protocols for insert synthesis.
4. Demonstrate use of components by creation of test circuits.

All four tasks are even now being carried out in parallel as need arises. For example, we will be finalizing construction of the yeast flip-flop in the next two months, and then optimizing the NAND components for pluggability of different DNA binding sites, proteotrophies, and fusion proteins.

Though this project is very ambitious in scope, we will begin by strongly confining ourselves to the subset of components necessary for test circuit creation. During construction, reagents will be produced with an eye to reusability as well as reliability. In creation of the small biological test circuits, we will be beginning the process of learning how to program in the biological language and how to deal with the specialized problems of engineering with asynchronous, analog and stochastic designs. Just as writing such programs as "Hello World" is useful when learning a new programming language, creation of these small and not so small biological circuits will be invaluable to learning the key to standardized cellular programming.

Authors

Adam Arkin (aparkin@lbl.gov)
Department of Bioengineering and Chemistry, University of California *and*
Physical Biosciences Division, Lawrence Berkeley National Laboratory
Berkeley, CA 94720

Drew Endy (dendy@molsci.org)
The Molecular Sciences Institute
2168 Shattuck Avenue, Berkeley, CA 94704

References

1. Lyakhov, D.L., *et al.*, *Mutant bacteriophage T7 RNA polymerases with altered termination properties.* J Mol Biol, 1997. **269**(1): p. 28-40.

2. Rong, M., *et al.*, *Promoter specificity determinants of T7 RNA polymerase.* Proc Natl Acad Sci U S A, 1998. **95**(2): p. 515-9.

3. Segal, D.J., *et al.*, *Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2758-63.

4. Beerli, R.R., *et al.*, *Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks.* Proc Natl Acad Sci U S A, 1998. **95**(25): p. 14628-33.

5. Reichardt, L.F., *Control of bacteriophage lambda repressor synthesis: regulation of the maintenance pathway of the cro and cI products.* J Mol Biol, 1975. **93**(2): p. 289-309.

6. Serebriiskii, I., V. Khazak, and E.A. Golemis, *A two-hybrid dual bait system to discriminate specificity of protein interactions.* J Biol Chem, 1999. **274**(24): p. 17080-7.

7. Wang, H. and D.J. Stillman, *Transcriptional repression in Saccharomyces cerevisiae by a SIN3-LexA fusion protein.* Mol Cell Biol, 1993. **13**(3): p. 1805-14.

8. Stefanovic, B., C. Hellerbrand, and D.A. Brenner, *Regulatory role of the conserved stem-loop structure at the 5' end of collagen alpha1(I) mRNA.* Mol Cell Biol, 1999. **19**(6): p. 4334-42.

9.   Rott, R., *et al.*, *The sequence and structure of the 3'-untranslated regions of chloroplast transcripts are important determinants of mRNA accumulation and stability.* Plant Mol Biol, 1998. **36**(2): p. 307-14.

10.   Mertens, N., E. Remaut, and W. Fiers, *Increased stability of phage T7g10 mRNA is mediated by either a 5'- or a 3'-terminal stem-loop structure.* Biol Chem, 1996. **377**(12): p. 811-7.

11.   Klaff, P., D. Riesner, and G. Steger, *RNA structure and the regulation of gene expression.* Plant Mol Biol, 1996. **32**(1-2): p. 89-106.

12.   Honda, M., E.A. Brown, and S.M. Lemon, *Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA.* Rna, 1996. **2**(10): p. 955-68.

13.   Carrier, T.A. and J.D. Keasling, *Mechanistic modeling of prokaryotic mRNA decay.* J Theor Biol, 1997. **189**(2): p. 195-209.

14.   Carrier, T.A. and J.D. Keasling, *Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli.* Biotechnol Prog, 1999. **15**(1): p. 58-64.

15.   Narayanan, A., *et al.*, *Nucleolar localization signals of Box H/ACA small nucleolar RNAs.* Embo J, 1999. **18**(18): p. 5120-5130.

16.   Narayanan, A., *et al.*, *Role of the box C/D motif in localization of small nucleolar RNAs to coiled bodies and nucleoli.* Mol Biol Cell, 1999. **10**(7): p. 2131-47.

17.   Long, M., S.J. de Souza, and W. Gilbert, *The yeast splice site revisited: new exon consensus from genomic analysis [letter].* Cell, 1997. **91**(6): p. 739-40.

18.   Spingola, M., *et al.*, *Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae.* Rna, 1999. **5**(2): p. 221-34.

19.   Das, S., *et al.*, *Inhibition of internal entry site (IRES)-mediated translation by a small yeast RNA: a novel strategy to block hepatitis C virus protein synthesis.* Front Biosci, 1998. **3**: p. D1241-52.

20.   Paz, I., L. Abramovitz, and M. Choder, *Starved Saccharomyces cerevisiae cells have the capacity to support internal initiation of translation.* J Biol Chem, 1999. **274**(31): p. 21741-5.

21.   Clarke, N.J., *et al.*, *pPV: a novel IRES-containing vector to facilitate plasmid immunization and antibody response characterization.* Immunotechnology, 1997. **3**(2): p. 145-53.

22.   Schmid, M. and E. Wimmer, *IRES-controlled protein synthesis and genome replication of poliovirus.* Arch Virol Suppl, 1994. **9**: p. 279-89.

23.   Varshavsky, A., *The N-end rule pathway of protein degradation.* Genes Cells, 1997. **2**(1): p. 13-28.

24.   Byrd, C., G.C. Turner, and A. Varshavsky, *The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor.* Embo J, 1998. **17**(1): p. 269-77.

25.   Varshavsky, A., *The N-end rule: functions, mysteries, uses.* Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12142-9.

26.   Colas, P., *et al.*, *Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2.* Nature, 1996. **380**(6574): p. 548-50.

27.   Kim, J.S., *et al.*, *Design of TATA box-binding protein/zinc finger fusions for targeted regulation of gene expression.* Proc Natl Acad Sci U S A, 1997. **94**(8): p. 3616-20.

28.   Kim, J.S. and C.O. Pabo, *Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2812-7.

29.   McColl, D.J., C.D. Honchell, and A.D. Frankel, *Structure-based design of an RNA-binding zinc finger.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9521-6.

30.   Wang, B.S. and C.O. Pabo, *Dimerization of zinc fingers mediated by peptides evolved in vitro from random sequences.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9568-73.

31.   Schneider, S., *et al.*, *Mutagenesis and selection of PDZ domains that bind new protein targets [see comments].* Nat Biotechnol, 1999. **17**(2): p. 170-5.

32.   Shea, M.A. and G.K. Ackers, *The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation.* J Mol Biol, 1985. **181**(2): p. 211-30.

33.   Brenowitz, M., *et al.*, *"Footprint" titrations yield valid thermodynamic isotherms.* Proc Natl Acad Sci U S A, 1986. **83**(22): p. 8462-6.

34.   Burz, D.S. and G.K. Ackers, *Cooperativity mutants of bacteriophage lambda cI repressor: temperature dependence of self-assembly.* Biochemistry, 1996. **35**(10): p. 3341-50.

35.   Merabet, E. and G.K. Ackers, *Calorimetric analysis of lambda cI repressor binding to DNA operator sites.* Biochemistry, 1995. **34**(27): p. 8554-63.

36.   Maurer, R., B. Meyer, and M. Ptashne, *Gene regulation at the right operator (OR) bacteriophage lambda. I. OR3 and autogenous negative control by repressor.* J Mol Biol, 1980. **139**(2): p. 147-61.

37.   Meyer, B.J., R. Maurer, and M. Ptashne, *Gene regulation at the right operator (OR) of bacteriophage lambda. II. OR1, OR2, and OR3: their roles in mediating the effects of repressor and cro.* J Mol Biol, 1980. **139**(2): p. 163-94.