Pacific Symposium on Biocomputing 15:385-396(2010)

CO-DESIGN IN SYNTHETIC BIOLOGY: A SYSTEM-LEVEL ANALYSIS OF THE DEVELOPMENT OF AN ENVIRONMENTAL SENSING DEVICE

DAVID A. BALL[1*], MATTHEW W. LUX[1*], RUSSELL R. GRAEF[2], MATTHEW W. PETERSON[2], JANE D. VALENTI[1], JOHN DILEO[2], JEAN PECCOUD[1†]

[1]*Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA (USA)*
[2]*Emerging Technologies Office, The MITRE Corporation, McLean, VA (USA)*

The concept of co-design is common in engineering, where it is necessary, for example, to determine the optimal partitioning between hardware and software of the implementation of a system features. Here we propose to adapt co-design methodologies for synthetic biology. As a test case, we have designed an environmental sensing device that detects the presence of three chemicals, and returns an output only if at least two of the three chemicals are present. We show that the logical operations can be implemented in three different design domains: (1) the transcriptional domain using synthetically designed hybrid promoters, (2) the protein domain using bi-molecular fluorescence complementation, and (3) the fluorescence domain using spectral unmixing and relying on electronic processing. We discuss how these heterogeneous design strategies could be formalized to develop co-design algorithms capable of identifying optimal designs meeting user specifications.

1. Introduction

1.1. *The need for co-design of synthetic biology systems*

A major focus in the field of synthetic biology from the field's inception has been the design of biological "devices" [1]. To date, many biological analogs of common electronics parts have been developed. Some notable biological devices analogous to those commonly used for electronics system design include logic gates [2-4], asynchronous logic components [5], switches [6-8], oscillators [8-12], memory circuits [13] and most recently genetic counters [14]. While these devices have been excellent proofs-of-concept for the application of traditional engineering design to the design of biological systems, little progress has been made in the design of larger, more complex, biological constructs [15]. This was seen in 2004, when Blue Heron Biotechnology did not receive a single submission to their Big DNA Contest, which offered free synthesis of the most "interesting" DNA construct of over 40kb in length (http://tinyurl.com/bigdna).

The designs of early artificial gene networks were restricted to a single design domain: protein-DNA interactions regulating transcription [6,10,16]; intra-molecular interactions within RNA molecules [3,17-20], or even interaction between DNA molecules [21,22]. More recent publications however, report the combined use of multiple domains in a single design. The genetic counter [14] is a good example of a heterogeneous design combining circuitry operating in different domains. Two separate methodologies were used, resulting in counter devices that, while performing the same type of logic (being able to count to either two or three), were appropriate for different uses. The riboregulated transcriptional cascade (RTC) counter utilized a fast transcriptional cascade for counting. The DNA invertase cascade (DIC) counter used recombinases upstream of inverted promoters to count. Due to the dynamics of DNA recombination, these counters activate more slowly, and as such can be used only for the counting of low-frequency events. It is anticipated that future heterogeneous designs will also include non-biological elements, as well as biological circuits. One such example of this is an AND gate used to regulate protein folding, which utilized UV light and ATP as stimuli [23]. By including non-biological elements, the design space is increased, allowing for the design of more complex, and potentially better performing, constructs. In the AND gate example above, the wavelength and intensity of the UV light is an additional parameter that can be used to optimize the system. Similarly, one approach described here uses advanced detection to expand the list of viable reporters.

As synthetic biology matures, it becomes necessary to develop more sophisticated design strategies. The design of an industrial application calls for a systematic comparison of different possible designs meeting the user specification in order to identify an optimal design maximizing one or several figures of merit. Besides the design correctness (the design does what it is supposed to do), its performance, development cost, manufacturing cost, or reconfigurability are other criteria that an engineering team may need to optimize. When developing electronic systems, an important design decision is the partitioning of features implemented in hardware and those implemented in software, which is known as the hardware/software co-design problem [24,25]. Software ensures

† Corresponding author: peccoud@vt.edu
* These authors contributed equally to this work.

rapid time to market, design flexibility, and runs on inexpensive processors produced in large volumes. However, the development of application-specific circuits is needed when software running on generic processors cannot meet the required performance. The hardware/software co-design problem illustrates the issues arising when heterogeneous technologies are combined into a system. At a very high-level, the design of synthetic biology applications includes a wetware component (the living organisms), but also a hardware component represented by the instrument and software used to acquire and process signals generated by the biological component of the system. At a higher resolution, the wetware component is itself heterogeneous since the transcriptional, translational, and proteomic components of this machinery represent different design domains.. By leveraging methods acquired by electrical engineers to develop heterogeneous systems, the complexity, efficiency, and flexibility of synthetic biology applications will likely be dramatically increased, while reducing the production costs of these systems.

1.2. *An environmental sensor as test case for co-design analysis*

In the field of biosecurity there is an increasing recognition of the need for systems that can rapidly detect pollutants, contaminants, and biothreat agents (pathogenic bacteria, viruses and toxins) in food, agricultural products, pharmaceuticals, and environmental samples. For example, the safety of the US food supply is an ongoing concern because of potential impacts of contamination on both public health and the US economy. In addition to inadvertent contamination, concerns about a bioterrorism event such as the intentional introduction of pathogens and/or toxins, brings a new dimension to this problem.

Cells possess innate abilities that make them ideal for environmental sensing applications. Specifically, they are inherently able to detect small concentrations (parts per billion) of chemicals (or combinations of chemicals) in their environment and respond to it, usually with an amplified signal. Cells can be programmed by identifying three functional layers: an input layer, an information processing layer, and an output layer [26]. This abstract representation of the environmental sensing chain can help design environmental sensing devices relying on biological systems for transforming chemical information into electrical signals that can be recorded and processed by computer systems. For instance, a situational awareness monitoring system will rely on a network of geographically dispersed sensing units communicating chemical data to a central server or to personnel operating in their vicinity. In this scenario, the sensing units should be capable of some basic processing of chemical data in order to communicate informative data.

In many cases, the presence of individual molecules in the environment is not informative while the simultaneous presence of two molecules can provide valuable information worthy communicating. In the field of defense, many chemical agents and explosives are made from combinations of commonly available industrial chemicals. For example, mustard gas can be created with thiodiglycol, an industrial solvent used in dyes and other applications, and phosphorus trichloride, a common industrial chemical used to manufacture a wide range of organic phosphorous compounds. Rapid field detection of such combinations could be an important application of sensing devices [27]. In ecology, it is well established that levels of heavy metals can be below the safe threshold individually, but combine to be lethal in fish [28]. In human health, it was recently shown that the common herbicide Roundup is more toxic in the presence of its supposedly inert adjuvants, and has negative effects in pregnant women even at “safe” levels of the active ingredient [29]. These findings may lead to the elucidation of further chemicals that are safe alone, but unsafe together. Detection of such combinations could become important to assess environmental or security threats.

Table 1. Logic table of the environmental sensor

Input 1	Input 2	Input 3	Output 1	Output 2	Output 3
-	-	-	-	-	-
-	-	+	-	-	-
-	+	-	-	-	-
-	+	+	-	+	-
+	-	-	-	-	-
+	-	+	-	-	+
+	+	-	+	-	-
+	+	+	+	+	+

The following sections describe three possible methods for implementing a cell based system designed to be able to detect the presence of each pair of three different chemical inputs and produce three different electronic outputs in response. Formally, the system can be specified by a truth table (Table 1). These designs differ in where the logic is implemented in the system (Figure 1). In the first option, hybrid promoters that contain binding sites for transcription factors responsive to the inputs are used to control the expression of fluorescent proteins. Only when the proper inputs are present will reporter genes be expressed, thus the logic occurs at the transcriptional level. The second option is to implement the logic at the protein level. This is accomplished by coupling each input to the expression of a non-fluorescent fragment of a fluorescent protein. Only when the two proper inputs are present will the fragments associate to generate a fluorescent signal. A final option is to embed the logic in the electronic layer. In this case each input directly activates the expression of one of three different fluorescent proteins and the inputs present are determined by processing the pattern of fluorescence that is obtained. For all three scenarios, the fluorescent proteins used were cyan (CFP), yellow (YFP), and red (RFP), which are all easily separated from each other.

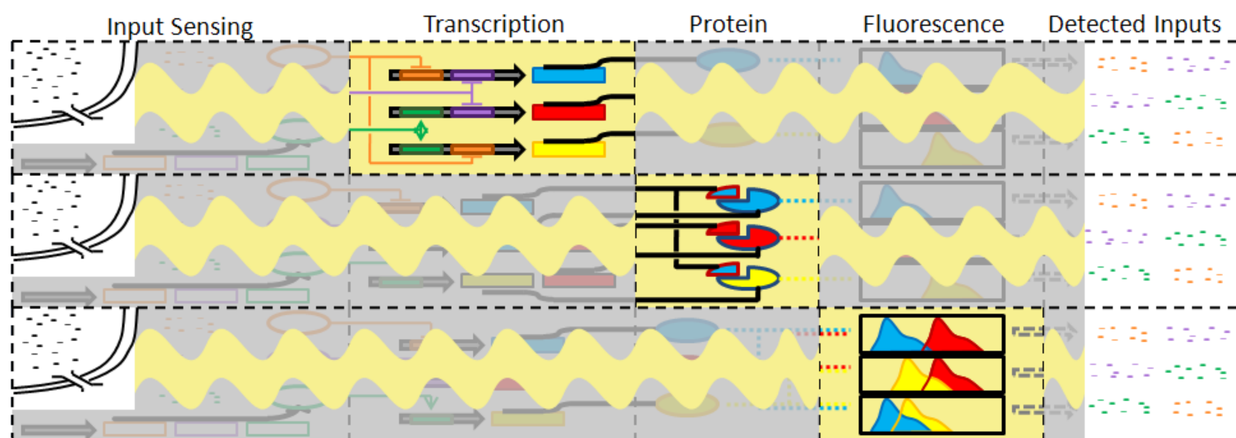

Figure 1. Implementation of logic in different design domains. The figure gives an overview of how each approach processes the environmental inputs. Wavy yellow lines indicate signal transduction, and yellow boxes highlight where the logic occurs. The details of each design are presented in Figure 2, Figure 3, and Figure 4

2. Solution 1: Hybrid Promoters

2.1. *Theoretical foundation*

One approach is to embed the design logic into transcriptional control with hybrid promoters. Since the input to the system is a set of small molecules, the first step is to sense the presence of the small molecule in the environment. In the described situation, these small molecules are ligands capable of binding to specific transcription factors. The sensor function is therefore accomplished by constitutive expression of the corresponding transcription factors. This sensing mechanism is conserved through each approach. Once bound by its ligand, the behavior of each transcription factor is altered. For example, some ligand-bound repressors can no longer bind to their corresponding promoters and exert their repressive properties. Thus, each transcription factor can be thought to switch on or off depending on the presence or absence of ligand.

In order to implement the design logic, the on/off state of these transcription factors must be processed. The hybrid promoter approach accomplishes the logic by controlling expression of a reporter gene through a promoter that responds to the state of pairs of transcription factors. That is, promoter A responds to the on/off state of transcription factors 1 and 2, promoter B to transcription factors 2 and 3, and promoter C to transcription factors 1 and 3. Each promoter responds only when both transcription factors have been toggled by the presence of the ligand molecule. The name “hybrid promoters” derives from the fact that they are engineered to respond to multiple transcription factors.

As a more detailed explanation, let us assume that inputs 1 and 2 are repressor proteins that repress only in the absence of their respective ligand. With no ligand present, both transcription factors effectively repress transcription of the reporter gene controlled by promoter A. In the presence of ligand 1 only, transcription factor 1 loses its ability to bind to promoter A, but transcription is still blocked by transcription factor 2. Likewise, if only ligand 2 is

present, transcription factor 1 continues to block transcription of the reporter gene. However, in the presence of both ligand 1 and ligand 2, neither transcription factor can bind to promoter A, and the reporter gene is freely expressed.

The transcription factor does not have to be an inducible repressor to accomplish the appropriate logic. There are 4 possible transcription factor responses: (1) the transcription factor *represses* only in the *absence* of its ligand, (2) the transcription factor *represses* only in the *presence* of its ligand, (3) the transcription factor *activates* only in the *absence* of its ligand, and (4) the transcription factor *activates* only in the *presence* of its ligand. Only options (1) and (4) are viable for implementing the design logic because they do not invert the signal. For example, a positive signal from the presence of a ligand would become a negative signal if the induced transcription factor transitioned from an inactive repressor (therefore allowing transcription) to an active repressor (therefore blocking transcription). Put another way, the positive (+) input signal would be inverted by *activation of the repressor* (-) to block expression from the promoter (+/- = -). On the other hand, the positive (+) input signal would be conserved by *inactivation of the repressor* (+) to allow expression from the promoter (+/+ = +), which preserves the signal.

Figure 2 illustrates an implementation of the logic using specific transcription factors and ligands. The rationale behind the choice of specific parts is described below.

2.2. *Proposed design*

A number of features need to be considered in the selection of appropriate inducible transcription factors. These features include: high range of control, compatibility with other transcription factors, and prior use in other applications. High range of control is necessary to ensure that the final signal is detected over the cellular noise. Compatibility with other transcription factors refers to the design of the actual promoter. For example, activators frequently must bind to specific promoter regions and thus may prevent the use of a second transcription factor that must bind an overlapping site. Last, well characterized transcription factors that have been widely used in other designs are preferred.

In synthetic biology, the list of commonly used genes is small and thus selection of appropriate parts is restricted. The first two appropriate transcription factors that match the criteria are LacI, which is inducible by isopropyl β-D-1-thiogalactopyranoside (IPTG), and TetR which is inducible by anhydrotetracycline (aTc). Both have operator sites that can be effectively placed in multiple locations to prevent interference with other transcription factors and both have been shown to have a high range of control [30,31]. Furthermore, hybrid promoters under the simultaneous control of LacI and TetR were described previously [31] which matched our logic criteria.

The selection of the third transcription factor is not as straightforward. LuxR is an attractive candidate. While LacI and TetR induce transcription by derepression in the presence of their respective ligand, LuxR activates transcription in the presence of its ligand acyl-homoserine lactone (AHL). LuxR has a high range of control for the wild type promoter [32]; it is described extensively in the literature, and it is commonly used in synthetic biology applications [26,33-35]. Although previous attempts to design hybrid promoter responding to LuxR/LacI or LuxR/TetR proved unsuccessful [31], a careful investigation of these LuxR hybrid promoters shows that none of them had spacing between the -10 box and the LuxR binding site that was identical to the wild type promoter. Given that this spacing has been shown to be important to ensure proper regulation of gene expression [32], we predict LuxR hybrid promoters can be redesigned if the suitable spacing is used.

The first promoter for implementing the environmental sensing device uses the sequence of the promoter A90 responding to LacI/TetR [31]. For the promoters responding to LuxR, we modified the wild type promoter for *luxI* and added *lac* and *tet* operators, respectively, downstream of the -10 box. Specifically, we replaced the sequence downstream of the -10 sequence in the wild type promoter with sequence downstream of the -10 box taken from promoters successfully responding to LacI and TetR [31]. Since the designed promoters already contain sufficient spacing downstream of +1, the 3 full sequences were assembled by simply adding a ribosome binding site sequence, a coding sequence for three different fluorescent proteins (CFP, YFP, RFP), and a terminator. Figure 1 shows the specific combinations of hybrid promoter and fluorescent reporter gene.

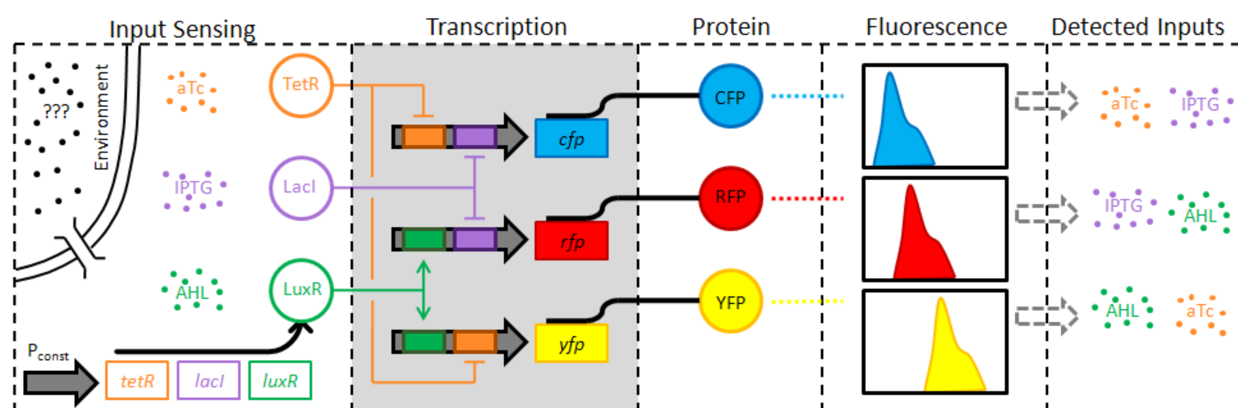

Figure 2. Hybrid Promoter Approach. The system logic is implemented in the control of transcription of 3 reporter genes. See text for detailed explanation of how the logic is processed. Orange indicates the aTc signal, purple the IPTG signal, green the AHL signal, blue the CFP signal, red the RFP signal, and yellow the YFP signal. Dots indicate small molecule inducers, thick solid arrows indicate promoters with color-coded operator boxes inside and adjacent boxes indicating genes, thick black lines indicate production of protein from a gene, circles indicate proteins, dotted lines indicate fluorescence measurement, and the gray box indicates where the logic occurs.

3. Solution 2: Fluorescence complementation

3.1. *Theoretical foundation*

While transcriptional logic is prevalent in the majority of synthetic biology constructs, logic can also be performed in the protein domain. As previously reported, the anti-parallel Leucine zipper mediated direction of protein reassembly allows for the reconstitution of intact and functional GFP [36]. Further research has shown the ability to adapt protein reassembly of fluorescent proteins to visualize protein-protein interactions *in vivo* [37,38]. These methods describe “Bimolecular Fluorescence Complementation” (BiFC) which uses non-fluorescent fragments of fluorescent proteins bound to separate functional proteins; when interaction between these proteins takes place the non-fluorescent pairs combine to produce a fluorescent complex. In this second solution, inputs are sensed in the same manner as in Solution 1. However, here the promoters respond to single transcription factors and therefore a single environmental input. Hence, they simply transmit the signal to the protein domain. The information is processed by coupling non-fluorescent halves of fluorescent proteins. This makes an “AND” gate from the pairing of non-fluorescent halves to produce a final product of fluorescence dependent on the promoters engaged. The logic comes not only from which fragments are produced but from the fact that only certain combinations of fragments will produce a detectable fluorescence output. Fragment 1 (N-terminal) of GFP combines with fragment 2 (C-terminal) of GFP to form functional GFP, in contrast fragment 1 (N-terminal) of GFP cannot combine with fragment 1 (N-terminal) of CFP.

3.2. *Design*

In order to implement this method of logic processing, several factors must be considered. Dissection sites of CFP, YFP, and a monomeric form of DsRed, a Red Fluorescent Protein variant that yield two non-fluorescent halves capable of reassembly into fluorescent proteins have been previously reported [37]. CFP is split into two fragments at amino acid 155 yielding a CFP 1-155 (N-terminal) fragment and a CFP 155-239 (C-terminal) fragment, subsequently referred to as CFP-N and CFP-C respectively. Similarly DsRed and YFP are split into the following fragments: RFP-N (residues 1-168), RFP-C (residues 169-225), and YFP-N (residues 1-154). The C-terminal of YFP is not required, because YFP-N can combine with CFP-C to form a species that produces yellow fluorescence [37,38].

These fragments will be cloned downstream of three inducible promoters. Fragments CFP-N and RFP-C are placed under the control of LacI (inducible by IPTG), Fragments YFP-N and RFP-N are placed under the control of LuxR (inducible by AHL), and the final fragment CFP-C is placed under TetR (inducible by aTc) control. As shown in Figure 3, the expected outputs from this system are dependent on which fragments are expressed. Some factors which play a role in total fluorescence are: the concentration of the inducer, the relative strength of promoter, and the fragment complementation.

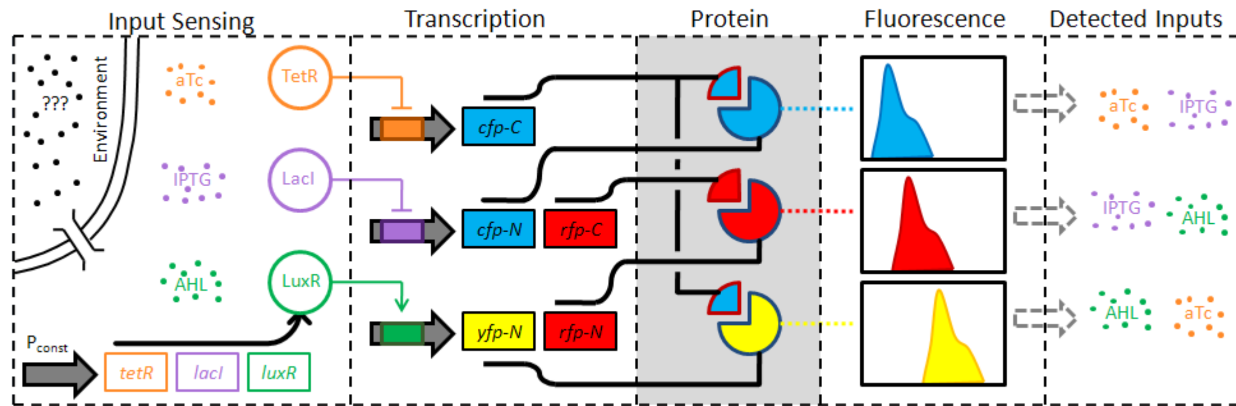

Figure 3. Fluorescence Complementation Approach. The system logic is implemented in the complementation of 3 fluorescent proteins. See text for detailed explanation of how the logic is processed. Orange indicates the aTc signal, purple the IPTG signal, green the AHL signal, blue the CFP signal, red the RFP signal, and yellow the YFP signal. Dots indicate small molecule inducers, thick solid arrows indicate promoters with color-coded operator boxes inside and adjacent boxes indicating genes, thick black lines indicate production of protein from a gene, circles indicate proteins, dotted lines indicate fluorescence measurement, and the gray box indicates where the logic occurs.

Logic circuit processing may be carried out by protein reassembly mediated by inducible promoters. For future experimentation, synthetically designed proteins for binding novel small and macro molecules could be incorporated into the system outlined here for logical processing.

Table 2. Previously reported fluorescent outputs generated by complementation of the different non-fluorescent protein fragments resulting in stable fluorescent protein complexes (given in color and name) or unstable/incompatible fragments (given by N/C "no color") [37].

Fragment	CFP-N	CFP-C	RFP-N	RFP-C	YFP-N
CFP-N	N/C	BLUE	N/C	N/C	N/C
CFP-C	BLUE	N/C	N/C	N/C	YELLOW
RFP-N	N/C	N/C	N/C	RED	N/C
RFP-C	N/C	N/C	RED	N/C	N/C
YFP-N	N/C	YELLOW	N/C	N/C	N/C

4. Solution 3: Unmixing of fluorescence spectra

4.1. *Theoretical foundation*

For the simultaneous detection of 3 (or fewer) fluorescent proteins it is possible to find fluorescent proteins with suitably separated excitation and emission spectra, such that the 3 colors can be distinguished by the use of optical band-pass filters [39]. However, this becomes difficult, if not impossible, as the number of fluorescent proteins increases. Therefore, a spectral unmixing approach was developed for the detection of multiple fluorescent signals [40-42]. The use of spectral unmixing also allows the use of a wider range of fluorescent proteins, which is of great value when it is necessary to use fluorophores with similar properties such as maturation and degradation times.

Spectral unmixing relies on the *a priori* collected emission spectra of the individual fluorescent proteins in the system to determine which fluorophores have contributed to the observed signal. The experimental output spectrum, F, can be described by the system of linear equations:

$$F = \mathbf{X}A, \quad (1)$$

where the m data points in the output spectrum, F, and the weights of the n individual fluorophores, A are column vectors and $\mathbf{X}_{\mathbf{ij}}$ is the **i**th point in the spectrum of fluorophore **j**,

$$F = \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{pmatrix}, \; A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix}, \; \mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{pmatrix}. \quad (2)$$

If $m > n$, Eq. 1 describes an overdetermined system, which can be solved for A by a least squares fitting algorithm to minimize $F - \mathbf{X}A$.

4.2. *Design*

Biologically, spectral unmixing offers the simplest approach for implementing the logic required to detect combinations of environmental species, as it does not require any interactions between the various promoters or protein products. As illustrated in Figure 4, each input ligand triggers the production of a single fluorescent protein. The sensing mechanism is again the same as Solution 1. In this case, the input signal is transmitted all the way to the spectral detection via single-operator promoters controlling production of single fluorescent proteins. The various fluorescence components can be extracted by use of Eq, (1), and the concentrations of each chemical can be determined.

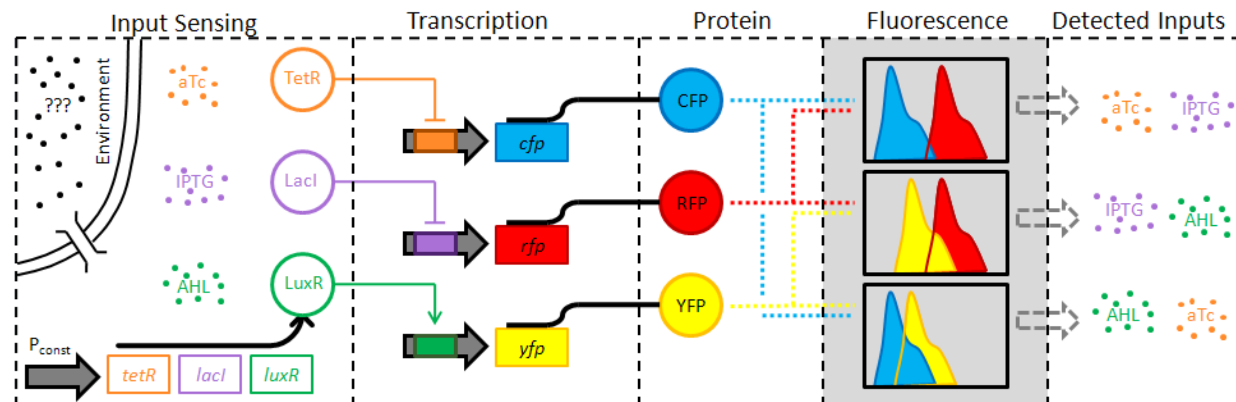

Figure 4. Spectral Detection Approach. The system logic is implemented in the unmixing of the measured spectra. See text for detailed explanation of how the logic is processed. Orange indicates the aTc signal, purple the IPTG signal, green the AHL signal, blue the CFP signal, red the RFP signal, and yellow the YFP signal. Dots indicate small molecule inducers, thick solid arrows indicate promoters with color-coded operator boxes inside and adjacent boxes indicating genes, thick black lines indicate production of protein from a gene, circles indicate proteins, dotted lines indicate fluorescence measurement, and the gray box indicates where the logic occurs.

4.3. *Preliminary data*

A spectrofluorimeter is required to collect the needed data to unmix the contributions of multiple fluorescent proteins. For the experiments reported here, a NanoDrop 3300 (Thermo Scientific, Wilmington DE) was used. The NanoDrop System uses a small sample volume of 2 μL. It contains three light-emitting diodes (LEDs) for fluorescence excitation: UV (peak emission at 365 nm), blue (470 nm) and white (460-650 nm). The collected fluorescence is dispersed over a 1024-pixel linear CCD, and allows the collection of wavelengths from 400-750 nm with a 4 nm resolution. Because the system does not include any optical bandpass filters, it is necessary to obtain spectra for blank samples, cells similar to those being used in the experiment but lacking any fluorophores.

Preliminary experiments were performed with four constitutively expressed fluorescent proteins: EGFP [43], acGFP [44], vYFP [45], and Citrine [46], each in a separate cell-line, to test the ability of the spectral unmixing algorithm to separate fluorescent proteins with similar emission spectra. Figure 5 presents the reference spectra for these four fluorophores. The spectra show that this combination of fluorescent proteins can be regarded as a worst-case scenario, as none of the four could be distinguished with the use of optical band-pass filters.

To account for differences in expression levels of the four different fluorescent proteins, the fluorescence spectra of the four cell-lines were normalized by their optical absorbance at 600 nm. The normalized spectra were

then used as the inputs to the unmixing algorithm, so that the extracted coefficients, such as those recorded in Table 3, are also in units of absorbance.
The four cell cultures were then mixed at known concentrations, and the resulting spectra shown in Figure 5b-f were analyzed with the spectral unmixing algorithm in order to extract the components of the mixtures. The measured optical densities at 600 nm for the existing cells in each mixture along with the extracted values are given Table 3. In all cases, there were no false-negatives. There are some false positive results, however, but their coefficients remained small (value less than 0.006).

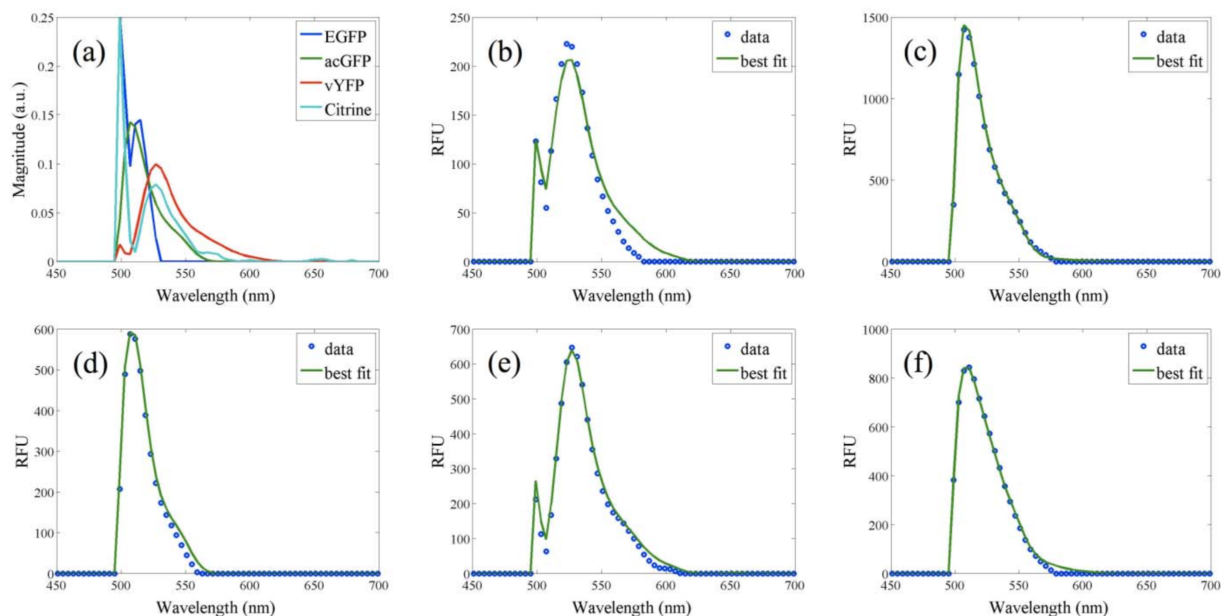

Figure 5: Experimentally measured emission spectra. (a) Emission profiles of cell cultures expressing 1 of the 4 fluorescent proteins. (b-f) Collected fluorescence spectra for mixtures of cell cultures expressing (b) EGFP and vYFP, (c) acGFP and Citrine, (d) acGFP and EGFP, (e) Citrine and vYFP, and (f) acGFP, Citrine, EGFP, and vYFP.

There are several possible explanations for the observed discrepancies between the measured and extracted optical densities of the true positives listed in Table 3. First, the small sample size used (2 μL) could mean that the cell concentrations in the measured sample did not completely reflect the concentrations of the stock solution. Also, as can be seen in Figure 5a, there is a sharp peak visible in the fluorescence spectrum of all of the fluorophores except acGFP at ~500 nm. This peak appears to be part of the spectrum of the blue LED used for exciting all of the fluorescent proteins, and most likely contributes to some of the errors in the extracted contributions.

Table 3: Measured and Extracted optical densities of mixtures of cells expressing a single fluorescent protein. Gray boxes indicate measured values of 0.

OD_{600}	Mix1		Mix 2		Mix 3		Mix 4		Mix 5	
	Meas.	Fit	Meas.	Fit	Meas.	Fit	Meas.	Fit	Meas.	Fit
acGFP	0	0.003	0.050	0.099	0.046	0.040	0	0.004	0.030	0.062
Citrine	0	0.006	0.070	0.039	0	0.002	0.081	0.014	0.027	0.017
EGFP	0.053	0.040	0	0	0.048	0.060	0	0	0.032	0.011
vYFP	0.048	0.029	0	0.006	0	0	0.082	0.100	0.027	0.037

5. Discussion

5.1. *Heterogeneous solutions to a synthetic biology design problem*

We have proposed three distinct solutions to a specification, namely the detection of distinct combinations of chemical signals. The first proposed solution uses a set of three hybrid promoters, performing the logic at the transcriptional level. This transcriptional solution would quickly become intractable if the number of input chemical

signals is increased. While we have shown a potential solution to the problem for three input signals, the number of parts needed to detect combinations from four possible inputs would require more parts than are currently available.

The second proposed solution still uses biology for the entirety of the logic, but does so in the protein domain. By moving the logic into the protein domain, the genetic circuitry is simplified. This allows the circuit to more easily be combined into a larger system, since the potential for cross-talk is decreased. Even though strategies to split fluorescent proteins still remain to be explored in a more systematic way, there are likely a limited number of fluorescent proteins that can be used in this context. So, the issue of the limited number of parts available faced in the transcriptional domain is also relevant in the protein domain. The assembly of the fluorescent proteins will increase the activation time of the system which can be an advantage or an inconvenience depending on the specific application of the environmental sensing device. In some cases, a fast response will be needed. In other cases, a longer maturation time can be used to time average the device response.

Finally, a third solution relies on the unmixing of fluorescence spectra for identification of molecules. In this case, biology is being used to generate the output signals, but the logic is being done outside of the biological domain. By implementing the logic outside of the biological system, the number of molecules possible to distinguish between is greatly increased, limited only by the number of transduction mechanisms and reporters. Also, the use of simpler biological circuits in this implementation circumvents any possible difficulties that may arise from incompatibilities between biological parts in the other two design schemes.

As we proceed with this project by physically implementing and characterizing each of these approaches, we will meticulously observe the differences in the difficulty of implementation and the performance of each design. As a result of unbalanced component behavior, each approach is likely to be improved by tweaking such elements as promoter strength, translational efficiency, degradation, etc. Thus, the cost and benefit of iterative designs will be evaluated as well. By further considering the difficulty of the design process for each solution, we will be able to holistically compare the strengths and weaknesses of the different approaches. Some comparative measures are discussed below. At this stage, it is important for the design team to acknowledge that there are multiple solutions to design problems and that the solutions can be implemented in different design domains. Just like the design of electronic systems is often a heterogeneous combination of hardware and software solutions, the design of a synthetic biology device can include multiple domains for the wetware component of the design as well as the hardware and software used to integrate the information originating from the design wetware component into a larger system.

A possibility that has not been considered in this paper is the combination of solutions implemented in different design domains. Are there solutions that could combine hybrid promoters and fluorescence complementation? Or could spectral unmixing be combined with fluorescence complementation to achieve better performance? Even though this manuscript proposes three distinct solutions, the universe of possible solutions is large and difficult to explore manually.

5.2. *Enabling co-design of synthetic biology application by design automation*

In order to compare different solutions to a design problem, it is necessary to define various figures of merits that can be used to quantitatively compare different solutions. Sensitivity, dynamic range, response time, robustness, or noise can be used to characterize the design performance. The development cost could be estimated by a function of the number of previously characterized components that can be reused in a new design. For instance, a solution requiring the development of a new promoter is expected to be slower and more expensive to implement than a solution relying on well characterized genetic parts such as the fluorescence unmixing approach. The manufacturing or production cost may also be a factor. The development of Solution 3 is the simplest but it relies on a more refined optical components that would increase the size and manufacturing cost of the device. This option may not be practical if millions of sensing devices needed to be distributed over large geographic regions to detect facilities manufacturing chemical weapons. Each specific application will require optimizing these metrics using multi-objective optimization algorithms [47,48].

Formalizing the representation of the design space is necessary to automate its exploration while searching for optimal designs. Fortunately, the wetware component of the system can be represented by the sequence of the synthetic DNA molecule implementing the design. Our group recently proposed to use formal languages to represent the structure of synthetic DNA sequences [49]. More recently, this original syntactic model was augmented with a semantic model used to predict the behavior encoded in a DNA sequence. By implementing this formalism in a logic programming language like Prolog [50], we were able to systematically explore a design space by generating structurally correct DNA sequences, compiling them into SBML files describing their behavior, and

simulating these files to identify solutions meeting a set of specification. Defining a distance in the design space, would make it possible to use optimization algorithms instead of a systematic exploration of all possible designs. In addition, it would be necessary to represent the non-DNA part of the designs by augmenting the language to represent detection systems and inputs.

The field of Synthetic Biology is growing by systematically adapting engineering practices to the design of biologically-inspired systems. The development of practical synthetic biology devices will require a system-level analysis and a co-design approach that have yet to be explored. In this paper, we have shown that the design space of a real-world device is large and may combine components developed in heterogeneous design domains. Finding optimal designs will require the use of design automation tools [51] like GenoCAD [52]. By adapting co-design methods used in more mature engineering fields [25,53,54], synthetic biology will fulfill its promise in the form of large, “interesting” circuits that were called for in the Big DNA contest .

Acknowledgments

This work was supported by the National Institutes of Health Grant 1 R01 GM078989, the National Science Foundation grant EF-0850100, the MITRE Innovation Program, and a generous donation from Science Applications International Corporation (SAIC) to the Virginia Bioinformatics Institute. We are indebted to James J. Valdes for indicating the possible application of this environmental sensor to the enforcement of the Chemical Weapons Convention and other nonproliferation treaties.

References

1. Endy D (2005) Foundations for engineering biology. Nature 438: 449-453.
2. Seelig G, Soloveichik D, Zhang DY, Winfree E (2006) Enzyme-free nucleic acid logic circuits. Science 314: 1585-1588.
3. Win MN, Smolke CD (2008) Higher-order cellular information processing with synthetic RNA devices. Science 322: 456-460.
4. Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A 100: 5136-5141.
5. Nguyen NPD, Kuwahara H, Myers CJ, Keener JP (2007) The design of a genetic muller C-element. ASYNC 2007: 13th IEEE International Symposium on Asynchronous Circuits and Systems: 95-104.
6. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403: 339-342.
7. Hasty J, Pradines J, Dolnik M, Collins JJ (2000) Noise-based switches and amplifiers for gene expression. Proceedings of the National Academy of Sciences of the United States of America 97: 2075-2080.
8. Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. Cell 113: 597-607.
9. Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. Nature 457: 309-312.
10. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403: 335-338.
11. Fung E, Wong WW, Suen JK, Bulter T, Lee SG, et al. (2005) A synthetic gene-metabolic oscillator. Nature 435: 118-122.
12. Chilov D, Fussenegger M (2004) Toward construction of a self-sustained clock-like expression system based on the mammalian circadian clock. Biotechnology And Bioengineering 87: 234-242.
13. Ajo-Franklin CM, Drubin DA, Eskin JA, Gee EP, Landgraf D, et al. (2007) Rational design of memory in eukaryotic cells. Genes Dev 21: 2271-2276.
14. Friedland AE, Lu TK, Wang X, Shi D, Church G, et al. (2009) Synthetic Gene Networks That Count. Science 324: 1199-1202.
15. Goler JA, Bramlett BW, Peccoud J (2008) Genetic design: rising above the sequence. Trends Biotechnol 26: 538-544.
16. Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. Science 296: 1466-1470.
17. Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nature Biotechnology 23: 337-343.

18. Win MN, Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A 104: 14283-14288.
19. Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. NatBiotechnol 24: 545-554.
20. Isaacs FJ, Collins JJ (2005) Plug-and-play with RNA. Nature Biotechnology 23: 306-307.
21. Kim J, White KS, Winfree E (2006) Construction of an in vitro bistable circuit from synthetic transcriptional switches. Molecular Systems Biology: -.
22. Okamoto A, Tanaka K, Saito I (2004) DNA logic gates. Journal of the American Chemical Society 126: 9458-9463.
23. Muramatsu S, Kinbara K, Taguchi H, Ishii N, Aida T (2006) Semibiological molecular machine with an implemented "AND" logic gate for regulation of protein folding. J Am Chem Soc 128: 3764-3769.
24. DeMicheli G, Gupta RK (1997) Hardware/software co-design. Proceedings of the Ieee 85: 349-365.
25. Wolf WH (1994) Hardware-software co-design of embedded systems. Proceedings of the IEEE 82: 967-989.
26. Kobayashi H, Kaern M, Araki M, Chung K, Gardner TS, et al. (2004) Programmable cells: interfacing natural and engineered gene networks. Proceedings of the National Academy of Sciences of the United States of America 101: 8414-8419.
27. Institute of Medicine (U.S.). Committee to Survey the Health Effects of Mustard Gas and Lewisite., Pechura CM, Rall DP (1993) Veterans at Risk : the health effects of mustard gas and Lewisite. Washington, D.C.: National Academy Press. xviii, 427 p. p.
28. Witeska M, Jezierska B. The effects of environmental factors on metal toxicity to fish; 2002 Oct 14-16; Brno, Czech Republic. Parlar Scientific Publications (P S P). pp. 824-829.
29. Richard S, Moslemi S, Sipahutar H, Benachour N, Seralini GE (2005) Differential effects of glyphosate and roundup on human placental cells and aromatase. Environ Health Perspect 113: 716-720.
30. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res 25: 1203-1210.
31. Cox RS, 3rd, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. Mol Syst Biol 3: 145.
32. Egland KA, Greenberg EP (1999) Quorum sensing in Vibrio fischeri: elements of the luxl promoter. Molecular Microbiology 31: 1197-1204.
33. Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. Nature 434: 1130-1134.
34. Karig D, Weiss R (2005) Signal-amplifying genetic enables in vivo observation circuit of weak promoter activation in the RhI quorum sensing system. Biotechnology And Bioengineering 89: 709-718.
35. Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proceedings of the National Academy of Sciences of the United States of America 101: 6355-6360.
36. Ghosh I, Hamilton AD, Regan L (2000) Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. Journal of the American Chemical Society 122: 5658-5659.
37. Kodama Y, Wada M (2009) Simultaneous visualization of two protein complexes in a single plant cell using multicolor fluorescence complementation analysis. Plant Mol Biol 70: 211-217.
38. Hu CD, Kerppola TK (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. Nature Biotechnology 21: 539-545.
39. Shaner NC, Steinbach PA, Tsien RY (2005) A guide to choosing fluorescent proteins. Nature Methods 2: 905-909.
40. Zimmermann T (2005) Spectral imaging and linear unmixing in light microscopy. Microscopy Techniques. pp. 245-265.
41. Dickinson ME, Bearman G, Tille S, Lansford R, Fraser SE (2001) Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. Biotechniques 31: 1272-+.
42. Lansford R, Bearman G, Fraser SE (2001) Resolution of multiple green fluorescent protein color variants and dyes using two-photon microscopy and imaging spectroscopy. Journal of Biomedical Optics 6: 311-318.
43. Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, et al. (2003) Global analysis of protein localization in budding yeast. Nature 425: 686-691.
44. Gurskaya NG, Fradkov AF, Pounkova NI, Staroverov DB, Bulina ME, et al. (2003) Colourless green fluorescent protein homologue from the non-fluorescent hydromedusa Aequorea coerulescens and its fluorescent mutants. Biochemical Journal 373: 403-408.

45. Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. Science 309: 2010-2013.
46. Heikal AA, Hess ST, Baird GS, Tsien RY, Webb WW (2000) Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: Coral red (dsRed) and yellow (Citrine). Proceedings of the National Academy of Sciences of the United States of America 97: 11996-12001.
47. Goh C-K (2009) Evolutionary multi-objective optimization in uncertain environments : issues and algorithms. New York: Springer.
48. Deb K (2001) Multi-objective optimization using evolutionary algorithms. Chichester ; New York: John Wiley & Sons. xix, 497 p. p.
49. Cai Y, Hartnett B, Gustafsson C, Peccoud J (2007) A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. Bioinformatics 23: 2760-2767.
50. Colmerauer A (1990) An Introduction to Prolog-Iii. Communications of the Acm 33: 69-90.
51. Sangiovanni-Vincentelli A (2003) The tides of EDA. Ieee Design & Test of Computers 20: 59-75.
52. Czar MJ, Cai Y, Peccoud J (2009) Writing DNA with GenoCAD. Nucleic Acids Res (in press).
53. Benini L, De Micheli G (2000) System-level power optimization: Techniques and tools. Acm Transactions on Design Automation of Electronic Systems 5: 115-192.
54. Bolsens I, DeMan HJ, Lin B, VanRompaey K, Vercauteren S, et al. (1997) Hardware/software co-design of digital telecommunication systems. Proceedings of the Ieee 85: 391-418.