A Scalable Cellular Logic Technology Using Zinc-Finger Proteins

Christopher Batten, Ronny Krashinsky, and Thomas Knight, Jr.
MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA 02139
{cbatten|ronny|tk}@csail.mit.edu

Simple cellular logic circuits have been built by engineering the DNA of host cells. Similar to systems found in nature, these circuits use repressor protein concentrations as logic signals; a gate's input repressors interact with the cell's DNA to influence the production of the gate's output repressors. A limitation in building these circuits is the number of unique repressor proteins available to use as logic signals, and previous designs have consisted of only a few gates.

In this paper, we propose a scalable cellular logic technology with zinc-finger proteins acting as the unique repressor logic signals. A zinc-finger protein binds to DNA at a specific target site determined by the nucleotide sequence, and zinc-finger proteins can readily be engineered to target almost any sequence. Our proposed technology uses engineered zinc-finger proteins and target DNA sequences as a scalable solution to implementing independent logic gates. The technology additionally attaches dimerization domains to the zinc-finger proteins to enable cooperativity and provide logic gates with non-linear gain. We use simulations analyze our proposed cellular logic technology, including the interference caused by interactions between gates, and conclude that building robust circuits with hundreds of gates seems feasible.

1 Introduction

Synthetic cellular logic circuits which directly control biological cells have the potential to transform natural prokaryotic cells into a novel nanoscale engineering substrate. Such synthetic biological systems could have far reaching impacts in a variety of fields such as nanoscale semiconductor fabrication, biomaterial manufacturing, autonomous biosensing, and programmed therapeutics.

Current cellular logic circuits represent signals between cellular gates with natural repressor proteins. This logic technology is fundamentally limited by the number of natural repressor proteins which have been extracted from other organisms, characterized, and tested in the circuit host organism. Currently there are just a handful of such proteins and this limits state-of-the-art circuits to less than a dozen logic gates.

We propose using zinc-finger proteins (ZFP) as the foundation for a novel cellular logic technology that is scalable to hundreds of gates. ZFPs are known to be relatively easy to engineer such that they recognize almost any DNA sequence. A cellular logic circuit based on a ZFP logic technology would use a ZFP with a unique DNA recognition sequence for each signal. Since all signal proteins are structurally similar, characterizing them would be significantly easier than characterizing the widely disparate natural repressor proteins used in current cellular logic circuits.

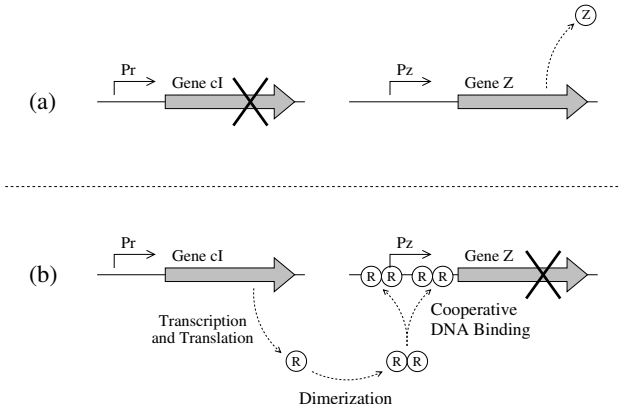

Figure 1: Schematic depiction of the *cI* repressor system found in the λ bacteriophage: (a) the repressor gene is not expressed and therefore the regulated gene *Z* is expressed, (b) the repressor gene is expressed and therefore the regulated gene *Z* is not expressed.

This paper first provides some background on cellular logic circuits and zinc-finger proteins before introducing one possible ZFP cellular logic technology. We introduce the concept of *inter-gate interference* in the proposed technology and demonstrate that careful engineering of various binding energies can significantly reduce the impact of such interference.

2 Background

This section provides some background and related work concerning cellular logic as well as zinc-finger proteins. Although the use of zinc-finger proteins in cellular logic has been proposed previously [21], this is the first work that we know of to examine a practical implementation.

2.1 Cellular Logic

Much of the previous work in synthetic cellular logic circuits has focused on engineering gene networks to implement the desired control system (see [9] for review). The majority of these approaches co-opt natural repressor proteins for use as the fundamental digital logic primitive. Repressor proteins are gene regulatory proteins which bind to DNA near the RNA polymerase promoter site and thereby inhibit gene

expression. Commonly used repressor proteins include the *lacI*, *tetR*, and *cI* proteins. Figure 1 illustrates the *cI* repressor system from the λ bacteriophage [14]. In Figure 1(a), the *cI* gene is not expressed which allows the RNA polymerase to bind to the P_z promoter and transcribe the regulated gene *Z*. In Figure 1(b), the *cI* gene is expressed and produces a large concentration of the repressor protein R. This repressor protein dimerizes and then cooperatively binds to an *operator* near the *Z* gene usually overlapping with the P_z promoter site. The bound R protein prevents the RNA polymerase from binding to the P_z promoter, effectively suppressing the expression of the regulated gene. The dimerization and cooperative DNA binding increase the system's *cooperativity* and introduce non-linearity into the system's transfer function.

If we consider the R protein concentration as the input signal and the Z protein concentration as the output signal, this simple repressor system is analogous to a basic digital inverter [18]. For a logic one input (large R concentration) the output is a logic zero (small Z concentration). For a logic zero input (small R concentration) the output is a logic one (large Z concentration). The non-linearity introduced by the system's cooperativity has the potential to make the system a regenerative inverter, meaning that degraded input signals that are still within specific noise margins will be restored to their full rail logic representation. This biochemical inverter is the basic primitive for most of the recent work in cellular logic circuits, and there has been quite a bit of work on experimentally characterizing as well as analytically modeling these logic gates [18, 20, 21, 22].

More sophisticated gates have been designed which use multiple inverters with a common output signal protein or which use externally generated inducer molecules to inhibit repressor proteins [20]. Others have explored the logic gate characteristics of randomly interconnected repressor systems and have identified NAND, NOR, and NOT IF gates [8].

These basic cellular logic gates have been composed into larger circuits including a three inverter ring oscillator [5], and a simple digital flip-flop [6]. Most of the previous work in cellular logic gates has used a small number of unique repressor proteins, and as a consequence such systems are limited to an equally small number of digital gates. Before researches can investigate larger and more complex circuits, a more scalable solution is needed to enable hundred gate systems.

2.2 Zinc-Finger Proteins (ZFPs)

Considered abstractly, a zinc-finger protein (ZFP) is a sequence-specific DNA "clamp", and ZFPs can provide a handhold to arbitrary locations on a piece of DNA. By far the most abundant DNA binding domain in eukaryotes [11, 15], ZFPs also have the potential to serve as an invaluable component in synthetic biological systems. With the basic ability to bind to any DNA site, ZFPs can directly function as repressors by blocking RNA polymerase. Importantly, a ZFP DNA binding domain can also be linked with an effector domain to enable a diverse range of applications [1, 11, 15]. Among many others, some effector domains include transcription factors for gene activation or repression, restriction enzymes for DNA cleavage, and dimerization domains for cooperative binding.

A single Cys_2-His_2 zinc-finger is a sequence of 30 amino acids with two conserved cystines and two conserved histidines which interact with a zinc ion to form a stable ββα fold [11, 15]. The α helix fits into the major groove of a DNA double-helix and its N terminus typically recognizes a 3 base-pair sequence of DNA nucleotides.

Zinc-finger proteins are typically composed of multiple fingers fused together to recognize longer DNA sequences. For example, the Zif268 protein (the first used to study ZFP-DNA binding) consists of three fingers which recognize a 9 bp DNA site. Just a few fingers in a ZFP go a long way in specifying a unique DNA location: for random DNA sequences, a 9 bp sequence would occur once every 260 thousand base-pairs, and an 18 bp sequence constructed out of six ZFPs would occur once every 69 billion base-pairs.

2.2.1 Engineering ZFPs

A crucial feature of ZFPs is their versatility in targeting arbitrary DNA sequences. When the key residues in a zinc-finger are modified, the ZFP is changed to target a different DNA sequence. Early hopes for a code [2] to fully predict ZFP–DNA interactions based on the key zinc-finger residues and the DNA bases have not materialized. In actuality, the ZFP–DNA interaction is quite complex and each finger typically recognizes 4 bp (or more) with a large variety of contacts between side chains and bases [12]. The most straightforward solution to constructing a poly-finger ZFP, combining individual fingers which each recognize a 3 bp sub-site, does not usually succeed because the target sites of neighboring fingers overlap and arbitrary finger combinations may conflict.

Fortunately, the target-site overlap problem can be easily avoided for a subset of DNA target sequences. A library of 16 individual zinc-fingers can be constructed which recognize 4 bp sites of the form 5'-GNNG-3' (where N can be any of the four bases, A, C, G, or T). These domains can then be composed into ZFPs in which each finger targets a 5'-GNN-3' triplet (e.g. to target sites of the form 5'-GNNGNNGNN-3'), and, by design, the target-site overlap problem is mitigated [4, 16]. This technique is commonly referred to as parallel selection, but we believe that *direct composition* is a more useful description of the construction process. Experimental selection must only be used to build and optimize the initial library of fingers, after which new ZFPs can be directly constructed by combining fingers from the library; indeed, new ZFPs can be produced in a matter of hours using standard PCR methods [15]. The parallel selection technique has recently been enhanced with domains that can recognize 5'-ANN-3' triplets [3], allowing any sequence which conforms to a repeating 5'-RNN-3' pattern (where R is G or A) to be targeted; in this way, a small library of finger domains can be used to target half of all possible DNA sequences.

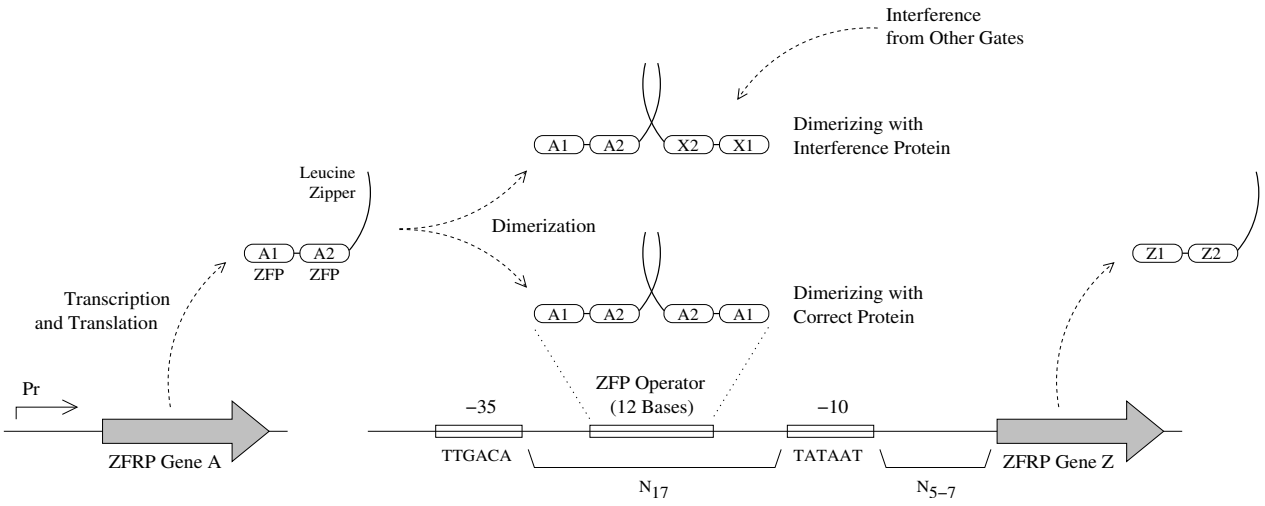

Figure 2: Schematic depiction of the proposed zinc-finger repressor system

Although other techniques for creating ZFPs exisit [7, 10], we believe that directly composing poly-finger ZFPs from a small pre-designed library of zinc-finger domains will probably prove to be the most useful construction method in synthetic biological systems. Researchers claim that careful use of a pre-designed library makes target-site overlap a non-issue when directly composing ZFPs [1], and combined with the convenience of this method it has become the standard for building ZFPs in many research labs and companies [15].

2.2.2 Engineering ZFP Dimers

Creating ZFP dimers is an important way to enhance their target site affinity and specificity, as well as to enable cooperative binding [12]. Effector domains that provide dimerization can readily be attached to ZFP DNA binding domains. A simple example system attaches the dimerization domain of Gal4 to two Zif268 fingers [13]; this construct recognizes two 6 bp symmetry-related subsites separated by a 13 bp spacer, and achieves a dissociation constant of $7.8 \times 10^{-19} M^2$ corresponding to half-maximal binding at a monomer concentration of 0.9 nM.

An improved system attaches the leucine zipper dimerization domain of GCN4 directly to the ZFP α helix C-terminus to give a more rigid dimer interface and to allow recognition of a contiguous DNA sequence [23]. This Zif23-GCN4 protein achieves a dissociation constant of $9.18 \times 10^{-18} M^2$. However, ZFP dimers bind opposite strands of the DNA helix, and thus provide the unique opportunity for the target sites to overlap; a 2 bp overlap improves the dissociation constant to $3.50 \times 10^{-19} M^2$. Furthermore, by randomizing and optimizing the linker between the domains, the dissociation constant was improved to $1.41 \times 10^{-21} M^2$. Additionally, the work describes how alternative leucine zippers (cJun and cFos) can be used to allow ZFP heterodimers to recognize asymmetrical DNA binding sites.

Another scheme uses peptide sequences as the dimerization domain for ZFPs [19]. A two-finger domain from Zif268 was extended with random 15-residue peptide sequences, then selected and optimized to evolve sequences that mediate dimerization.

3 Proposal

We propose a new class of proteins called zinc-finger repressor proteins (ZFRPs) as the basis for a scalable cellular logic technology which will enable hundreds of gates within a single cell (see Figure 2). ZFRPs are very similar to the engineered Zif23-GCN4 protein in that they are composed of two domains: a two-finger ZFP DNA binding domain which is engineered to recognize a specific operator and a leucine zipper dimerization domain [23]. ZFRPs differ from the Zif23-GCN4 protein since the dimerization energy for ZFRPs may need to be engineered as discussed below.

A circuit can contain many gates each with its own unique ZFRP and matching operator. The maximum number of unique ZFRPs is limited by the number of unique DNA sequences a ZFRP dimer can recognize (termed the ZFRP's *encoding space*). Although a ZFRP dimer recognizes 12 base pairs, our proposal uses ZFRP homodimers which reduces the encoding space to 4^6. Additionally, we envision using direct composition to engineer the ZFRP recognition sequences which further reduces the encoding space by a factor of two (see Section 2.2). Thus, the final encoding space for our design is 2048. A robust design will probably use less than the maximum encoding space to both increase specificity and remain within the metabolic capabilities of the cell. Thus a reasonable design could have on the order of one or two hundred unique ZFRPs.

Figure 2 illustrates how the ZFRPs dimerize and then bind to the promoter region of the regulated gene Z. The 17 base spacer between the -35 and -10 regions typical in σ^{70} bacterial promoters is engineered to match the corresponding ZFRP. Although all gates have ZFRPs with unique DNA binding domains, all ZFRPs have a common dimerization domain.

(a)	Dimerization	$R+R \rightleftharpoons R_2$	K_{R+R}	$=$	$(R)^2/(R_2)$	$=$	$e^{E_{dim}/RT}$
(b)	Dimer Binding	$O+R_2 \rightleftharpoons R_2O$	K_{R_2+O}	$=$	$(O)(R_2)/(R_2O)$	$=$	$e^{2E_{op}/RT}$
(c)	Monomer Binding	$O+R \rightleftharpoons OR$	K_{O+R}	$=$	$(O)(R)/(OR)$	$=$	$e^{E_{op}/RT}$
(d)	Monomer Binding	$R+O \rightleftharpoons RO$	K_{R+O}	$=$	$(O)(R)/(RO)$	$=$	$e^{E_{op}/RT}$
(e)	Cooperative Binding	$OR+R \rightleftharpoons R_2O$	K_{OR+R}	$=$	$(OR)(R)/(R_2O)$	$=$	$e^{(E_{op}+E_{dim})/RT}$
(f)	Cooperative Binding	$RO+R \rightleftharpoons R_2O$	K_{RO+R}	$=$	$(RO)(R)/(R_2O)$	$=$	$e^{(E_{op}+E_{dim})/RT}$
(g)	Protein Synthesis	$O \rightarrow O+Z$	k_x				
(h)	Protein Decay	$Z \rightarrow$	k_{deg}				
(i)	Dimerization	$X+X \rightleftharpoons X_2$	K_{X+X}	$=$	$(X)^2/(X_2)$	$=$	$e^{E_{dim}/RT}$
(j)	Inter-Gate Interference	$X+R \rightleftharpoons XR$	K_{X+R}	$=$	$(X)(R)/(XR)$	$=$	$e^{E_{dim}/RT}$

Table 1: Chemical equations used in the ZFRP analytical model

This means that a ZFRP for a specific *victim* gate (for example, ZFRP A in Figure 2) might dimerize with the ZFRP from a different *attacker* gate (for example, ZFRP X in Figure 2). This *inter-gate interference* decreases the concentration of ZFRP A dimers and at high ZFRP interference concentrations, could cause the victim inverter to incorrectly change its output. The next section will use analytical models to illustrate that careful engineering of the dimerization energy can reduce inter-gate interference while still providing the cooperativity needed for a regenerative logic technology.

4 Analysis

In this section we develop an analytical model for the ZFRP system proposed in the previous section. This model is based on the set of chemical reaction equations listed in Table 1. Equations (a) and (b) model ZFRP repressors (denoted with the symbol R) dimerizing and then binding to the appropriate operator (denoted with the symbol O). Equations (g) and (h) model protein synthesis and decay (the output protein is denoted with the symbol Z). These four equations are sufficient to model systems with strong dimerization energies, but for weaker dimerization energies these equations fail to capture the cooperative binding of monomers to the operator. Equations (c) through (f) broaden the model to account for both dimerization off the DNA and also the cooperative binding of monomers to the DNA. Equations (i) and (j) complete the model by including the affects of inter-gate interference (the interference protein is denoted with the symbol X). Table 1 illustrates how the equilibrium dissociation constants can be derived for all but Equations (g) and (h) from a given dimerization and operator energy.

Previous researchers have measured engineered three finger ZFPs as having nanomolar dissociation constants [2, 23]. This corresponds to a DNA binding energy of approximately -12 kcal. If we assume that DNA binding energy scales linearly with the number of recognized bases, we can estimate the binding energy for a single finger ZFP to be -4 kcal. Therefore, for this work we assume our two-finger ZFRP monomers have a DNA binding energy (E_{op}) of -8 kcal to -9 kcal. The Zif23-GCN4 protein has a macroscopic dissociation constant on the order of 10^{-18} M^2 which corresponds to a binding energy of -24 kcal [23]. Since the Zif23-GCN4 protein also uses two-finger ZFP DNA binding domains we can approximate the isolated dimerization energy (E_{dim}) to be approximately -6 kcal to -8 kcal. For this work we further assume that it is possible to create ZFRP variants with phage display which decrease E_{dim} as desired.

The rest of this section discusses cooperativity in the model before investigating inter-gate interference and its influence on an appropriate operating regime. The section finishes by examining a representative transfer curve for a ZFRP inverter, and the effects of inter-gate interference on this transfer curve.

4.1 Basic Cooperativity

We first analyze cooperativity in a basic repressor system (without interference). In addition to the equations in Table 1, we note that if (O_T) is the total operator concentration, then:

$$(O_T) = (O) + (RO) + (OR) + (R_2O) \tag{1}$$

and if (R_T) is the total repressor concentration, then:

$$\begin{aligned} (R_T) &= (R) + 2\cdot(R_2) + (RO) + (OR) + 2\cdot(R_2O) \\ (R_T) &\approx (R) + 2\cdot(R_2) \end{aligned} \tag{2}$$

where the simplification is based on the assumption that $(O_T) \ll (R_T)$.

The fraction of operator not bound (free) is:

$$\frac{(O)}{(O_T)} = \frac{(O)}{(O) + (RO) + (OR) + (R_2O)} \tag{3}$$

After dividing the top and bottom of this equation by (O), and substituting from Table 1, the equation simplifies to:

$$\frac{(O)}{(O_T)} = \frac{1}{1 + 2\frac{(R)}{K_{R+O}} + \frac{(R)^2}{J_{R+R+O}}} \tag{4}$$

where $J_{R+R+O} = e^{(2E_{op}+E_{dim})/RT}$.

Equation 4 gives the fraction of operator not bound as a function of the free repressor concentration (R). To determine (R) as a function of (R_T), we substitute for (R_2) in Equation 2

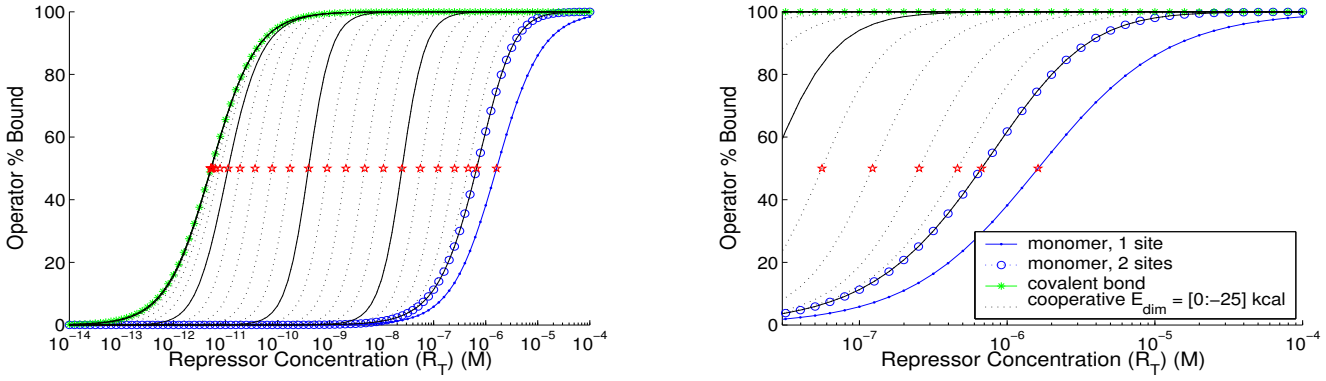

Figure 3: Cooperative binding with varying dimerization energy (for E_{op} = -8 kcal). The plot on the right zooms in on the region of higher repressor concentrations (and its legend applies to both plots). The cooperative binding curves vary E_{dim} from 0 kcal (for the rightmost curve) down to -25 kcal (for the leftmost curve) in increments of -1 kcal, with multiples of -5 kcal shown as solid curves. For each E_{dim} curve, the star indicates where half of the operator sites are bound.

based on the dimerization relation in Table 1, and then use the quadratic formula to derive:

$$(R) = \left(\sqrt{1 + \frac{8 \cdot (R_T)}{K_{R+R}}} - 1\right) \cdot \left(\frac{K_{R+R}}{4}\right) \qquad (5)$$

We note that when $(R_T) = K_{R+R}$, $(R) = \frac{(R_T)}{2}$ (meaning that half of the total repressors are free and half are bound as dimers) as expected. We term this point the *dimerization inflection point*.

Figure 3 shows how the bound operator percentage depends on the total input repressor concentration for varying dimerization energies (E_{dim}). When $E_{dim} = 0$ kcal, the repressor monomers bind to the two operator sites independently; the curve tracks that predicted by a simpler model with no dimerization (*monomer, 2 sites*). As the dimerization energy becomes stronger, a lower concentration of repressor is needed for operator binding to occur. More importantly, the cooperativity causes the slope of the binding curve to increase. Eventually, when the dimerization energy becomes strong enough, all of the repressors exist as dimers and the curve tracks that predicted by a simpler model with a covalently bound monomer.

To further analyze cooperativity, we create hill plots like those shown in Figure 4. This is an alternative view of essentially the same data as Figure 3. For these curves, r is the ratio of the bound repressor concentration to the total operator concentration:

$$r = \frac{(RO) + (OR) + 2 \cdot (R_2O)}{(O_T)}$$

and n is the number of operator binding sites (2 in this case). As demonstrated by the star markers in Figures 3 and 4, the hill plot positions the middle of the operator binding curves (where $\frac{r}{n} = \frac{1}{2}$) at $y = 0$. The slope of the curves at this point gives the *hill coefficient*, a measure of cooperativity where a value of 1 indicates no cooperativity and a value of 2 indicates maximum cooperativity (for dimerization of two monomers).

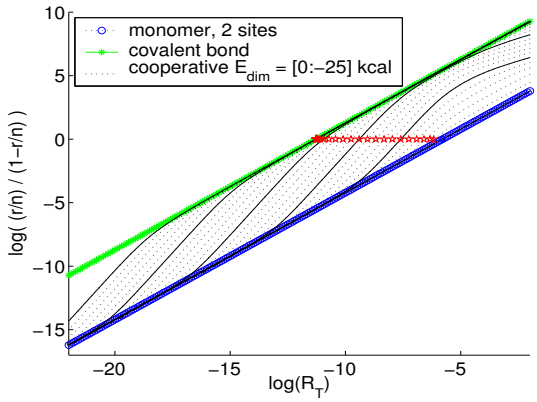

Figure 4: Hill plots with varying dimerization energy (for E_{op} = -8 kcal). The cooperative binding curves show a range of E_{dim} as in Figure 3, and for each E_{dim} curve the star indicates where half of the operator sites are bound.

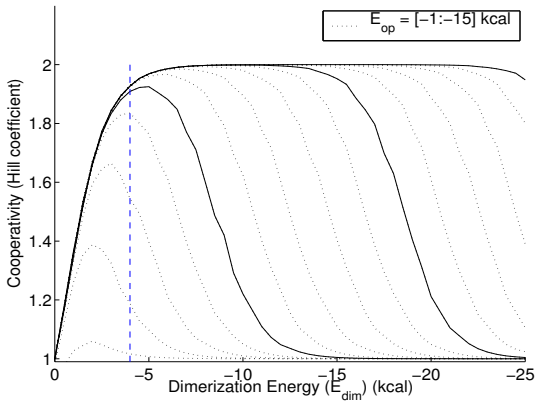

Figure 5: Hill coefficients as a function of dimerization energy with varying operator binding energy. E_{op} varies from -1 kcal (for the leftmost curve) down to -15 kcal (for the rightmost curve) in increments of -1 kcal, with multiples of -5 kcal shown as solid curves.

The hill coefficients are shown in Figure 5 as a function of the dimerization energy for various operator binding energies. As expected, the coefficient is 1 when $E_{dim} = 0$, increases to a maximum of 2 as the dimerization energy increases, and eventually returns to 0 when the dimerization energy overpowers the operator binding energy. For operator binding energies greater than -5 kcal, the cooperativity begins to plateau at a dimerization energy of around -4 kcal. This energy corresponds to a K_{R+R} of $2.4 \cdot 10^{-4}$ M, which indicates the total repressor concentration for the dimerization inflection point (half the repressors are dimers at this concentration).

4.2 Inter-Gate Interference

To model interference between different gates, we define (Q_T) to be the total concentration of repressors, including both the input to the gate of interest (R_T) and the interfering signals (X_T):

$$(Q_T) = (R_T) + (X_T) \tag{6}$$

Then, we derive the total free repressor concentration (Q) as a function of (Q_T) as in Equation 5:

$$(Q) = \left(\sqrt{1 + \frac{8 \cdot (Q_T)}{K_{R+R}}} - 1 \right) \cdot \left(\frac{K_{R+R}}{4} \right) \tag{7}$$

This equation makes use of the fact that $K_{X+X} = K_{R+R}$. The free repressor concentration for the gate of interest (R) is simply computed as the appropriate fraction of the total free repressor concentration:

$$(R) = (Q) \frac{(R_T)}{(X_T)} \tag{8}$$

We use this value as the input to Equation 4 to determine the free operator fraction.

Figure 6 shows how interference affects the operator binding curves for various dimerization energies. The dimerization inflection point is denoted as K_{dim} in the figure. Inter-gate interference begins to have an impact on the operator binding curves when (X_T) is about an order of magnitude less than the dimerization inflection point. Once (X_T) reaches the inflection point, half of all ZFRPs in the cell are dimerized, and if $(X_T) \gg (R_T)$ this means that half of (R_T) is squandered. This effect can be seen in Figure 6; as (X_T) increases more input repressor (R_T) is required to transition the gate (i.e. the operator curves move to the right).

4.3 Operating Regime

As shown in Figure 6, interference can shift the operator binding curves. If these shift during operation, the variance reduces the noise margins of the gates, and in the worse case it could cause a "one" output of one gate to be interpreted as a "zero" input to another gate or vice-versa.

Since the amount of interference depends on the state of all gates in the cell, we can not rely on maintaining any particular interference concentration. Instead, we wish to characterize

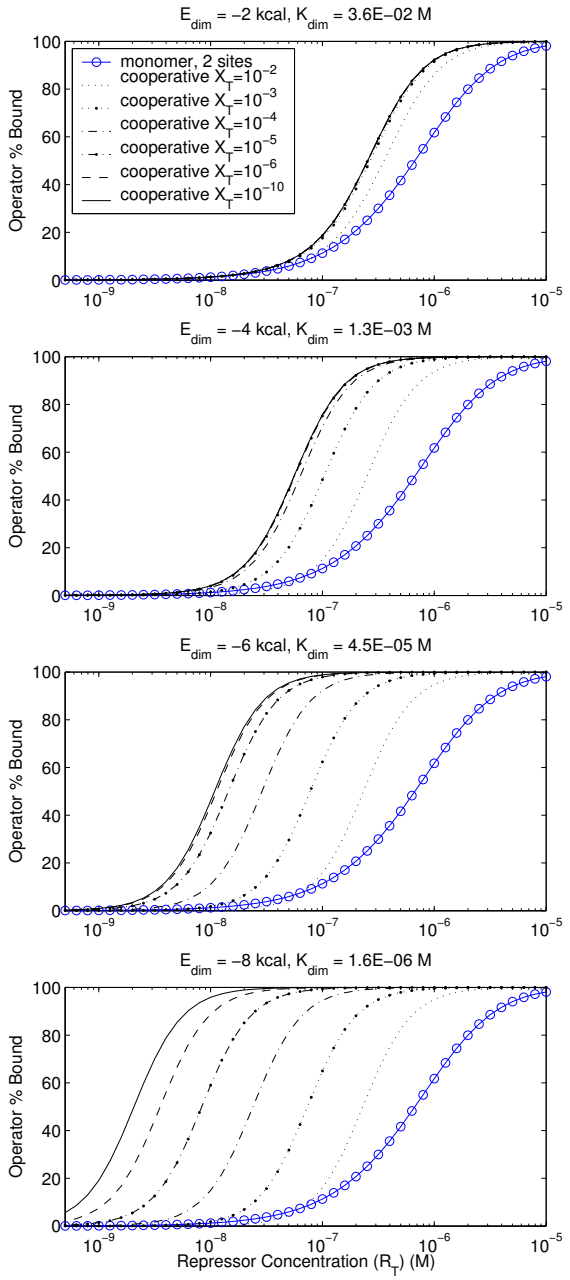

Figure 6: Cooperative binding with interference. The dimerization inflection point (K_{dim}) is shown in the title. (E_{op} = -8 kcal)

a system that is functional up to some maximum interference threshold, and then guarantee that this maximum is never exceeded. Figure 6 demonstrate that the dimerization energy determines the maximum interference concentration that can be tolerated before the operator binding curves begin to shift. To reduce interference, we want E_{dim} to be as weak as possible. However, E_{dim} must be strong enough to enable cooperativity; as shown in Figure 5 cooperativity plateaus at an E_{dim} of around -4 kcal. When E_{dim} = -4 kcal an interference concentration of up to around $X_T = 10^{-4}$ M can be tolerated (see Figure 6). We will show in the following section that it will be unlikely for this maximum to be exceeded.

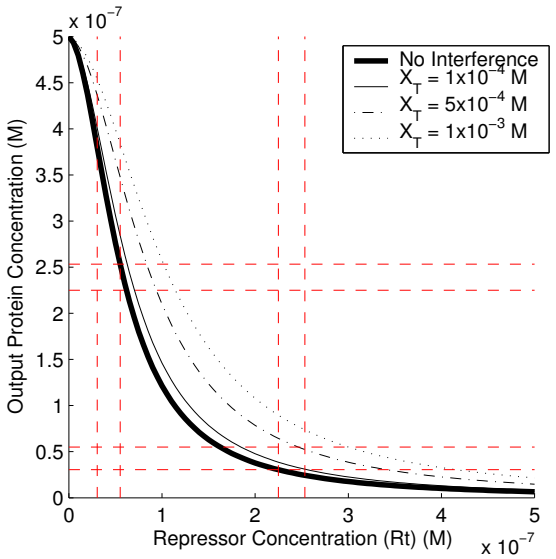

Figure 7: Transfer curve for increasing concentrations of inter-gate interference. Dashed lines denote optimal noise margins. (Parameters: k_x/k_{deg} = 500, O_T = 10^{-9} M, E_{op} = -8 kcal, E_{dim} = -4 kcal)

Parameter	Value
Operator Energy	-8 kcal
Dimerization Energy	-4 kcal
Synthesis to Decay Ratio	500
Operator Concentration	1×10^{-9} M
Max Protein Concentration	5×10^{-7} M
Gain at in = out	4.3
Max Output Logic Low Concentration	0.30×10^{-7} M
Min Output Logic High Concentration	2.51×10^{-7} M
Max Input Low Concentration	0.55×10^{-7} M
Min Input High Concentration	2.26×10^{-7} M
Low Noise Margin	0.25×10^{-7} M
High Noise Margin	0.28×10^{-7} M

Table 2: Typical ZFRP Gate Characterization

4.4 Transfer Curves

We now investigate the resulting transfer curves for ZFRP systems which operate in the regime suggested by the previous section. The following differential equation is derived from Equations (g) and (h) in Table 1.

$$\frac{d(Z)}{dt} = k_x(O) - k_{deg}(Z) \tag{9}$$

At equilibrium, $d(Z)/dt = 0$ and Equation 4 can be used to express (O) as a function of free repressor, yielding the following function for (Z).

$$(Z) = \frac{\frac{k_x}{k_{deg}} \cdot (O_T)}{1 + 2\frac{(R)}{K_{R+O}} + \frac{(R)^2}{J_{R+R+O}}} \tag{10}$$

This is the final transfer curve with Equation 5 providing (R) as a function of (R_T). For this work we assume $k_x/k_{deg} \approx 500$ [17] and $O_T \approx 10^{-9}$ M [14].

The solid dark line in Figure 7 is a representative transfer curve, and this curve is similar to the transfer curves in the literature [17, 20, 22]. Notice that this transfer curve is non-linear and has regions where the gain is greater than one and other regions where the gain is less than one. These are the hallmarks of a well-formed single-input digital logic device. Table 2 is a characterization of the ZFRP gate corresponding to the transfer curve in Figure 7. Assuming a single protein has a concentration on the order of 10^{-9} M, this ZFRP gate will produce approximately 500 output proteins when the input is a logic zero. This gate has a gain greater than four at the point on the curve where the input protein concentration equals the output protein concentration. Notice that this transfer curve can be used as an effective logic gate even though there is little noise attenuation for a logic zero input (low repressor concentrations). This is because any noise amplification from a logic zero input will be quickly attenuated when this signal is propagated as a logic one into the succeeding logic stage.

In characterizing the transfer curve, we maximized the high and low noise margins with the constraint that they were approximately equal. These noise margins are shown in Figure 7 with the horizontal and vertical dashed lines. The noise margins are approximately 2.6×10^{-8} M which is on the order of 25 proteins. These noise margins are tight, and a stochastic modeling approach is probably necessary to further investigate the implications of protein synthesis noise in such systems. This work, however, focuses on inter-gate interference and its affect on a victim ZFRP gate.

Figure 7 shows that increasing the interference concentration pushes the transfer curve up and to the right. This can cause faulty logic behavior amongst ZFRP gates as well as at the boundary between gates which use ZFRPs and gates which use a different logic technology.

Large interference concentrations on the order of 10^{-3} M, can cause the inverter in Figure 7 to output an incorrect logic value. If the input to a given inverter is 2.2×10^{-7} M (a logic one), then without inter-gate interference this inverter will have an output of 0.25×10^{-7} M (a logic zero). If this same inverter is now experiencing inter-gate interference, then its output will be an ambiguous 1.0×10^{-7} M. Subsequent gates which are not experiencing inter-gate interference can easily misinterpret this protein concentration as the wrong logic value. Even though at equilibrium all ZFRP gates within the same cell will experience similar amounts of interference (and thus have very similar transfer curves), transient effects could still cause serious problems for sequential logic. Since the inter-gate interference will vary across different cells, this type of noise can cause problems with inter-cell signaling as well. The unpredictability of logic value representations would make designing robust inter-cell signaling mechanisms very difficult.

As discussed in Section 4.3, careful engineering of the dimerization energy can help reduce the affects of inter-gate ZFRP interference. Figure 7 shows that inter-gate interference has little influence on the transfer curve for interference concentrations less than 10^{-4} M. The maximum output protein concentration for a single gate is 5×10^{-7} M (Table 2), so a cell using the ZFRP logic technology could potentially have $10^{-4}/5 \times 10^{-7} = 200$ gates. Since the encoding space for the proposed ZFRP system and the metabolic capabilities of the cell are on the order of one or two hundred gates (see Section 3), this seems like a reasonable limitation.

5 Summary

Although the proposed ZFRP system exhibits good characteristics and allows for a reasonable number of gates per cell, there is still room for improvement both in terms of evaluating the system and in improving the system itself.

Extending the equilibrium model presented here to include non-equilibrium behavior would help quantify the speed of the system and allow one to investigate the affect of inter-gate interference on gate delay. As mentioned earlier, using a discrete model as opposed to a continuous model would allow one to include the effects of stochastic noise in protein synthesis. Of course actually implementing the proposed system in a live cell colony is the best way to know if the ZFRP logic technology is a viable alternative to traditional techniques.

An important way to improve the ZFRP proposal itself would be to increase the cooperativity of the system. Higher cooperativity would enable faster and more robust logic gates. Preliminary experiments with our analytical model show that increasing cooperativity creates transfer curves with greater gain in the switching region. Higher cooperativity also decreases the gain near the low and high logic values which results in better noise attenuation.

Synthetic cellular logic circuits promise to usher in a new era of innovation and discovery, but current approaches are stifled by logic technologies which cannot scale to more than a few tens of gates. We have introduced a novel cellular logic technology based on zinc-finger proteins, and using a simple continuous model we have shown that careful engineering of the dimerization energy can reduce inter-gate interference. Cellular logic technologies based on zinc-finger proteins are a promising technique for breaking through the current scalability wall.

References

[1] R. R. Beerli and C. F. Barbas, III. Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol*, 20(2):135–141, Feb 2002.

[2] Y. Choo and A. Klug. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci U S A*, 91(23):11163–7, Nov 1994.

[3] B. Dreier, R. R. Beerli, D. J. Segal, J. D. Flippin, R. R. Beerli, and C. F. Barbas, III. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem*, 276(31):29466–78, Aug 2001.

[4] B. Dreier, D. J. Segal, R. R. Beerli, and C. F. Barbas, III. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol*, 303(4):489–502, Nov 2000.

[5] M. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, January 2000.

[6] T. Gardner, C. Cantor, and J. Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403:339–342, Jan 2000.

[7] H. A. Greisman and C. O. Pabo. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, 275:657–661, 1997.

[8] C. Guet, M. Elowitz, W. Hsing, and S. Leibler. Combinatorial synthesis of genetic networks. *Science*, 296:1466–1470, May 2002.

[9] J. Hasty, D. McMillen, and J. Collins. Engineered gene circuits. *Nature*, 420(6912):224–30, Nov 2002.

[10] M. Isalan, A. Klug, and Y. Choo. A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat Biotechnol*, 19:656–660, 2001.

[11] A. C. Jamieson, J. C. Miller, and C. O. Pabo. Drug discovery with engineered zinc-finger proteins. *Nature Reviews Drug Discovery*, 2(5):361–368, May 2003.

[12] C. O. Pabo, E. Peisach, and R. A. Grant. Design and selection of novel cys(2)his(2) zinc finger proteins. *Annual Review of Biochemistry*, 70:313–340, Jul 2001.

[13] J. L. Pomerantz, S. A. Wolfe, and C. O. Pabo. Structure-based design of a dimeric zinc finger protein. *Biochemistry*, 37(4):965–970, 1998.

[14] M. Ptashne. *A Genetic Switch: Gene Control and Phage* λ. Blackwell Scientific Publications and Cell Press, 1986.

[15] D. J. Segal and C. F. Barbas, III. Custom DNA-binding proteins come of age: polydactyl zinc-finger proteins. *Current Opinion in Biotechnology*, 12(6):632–637, Dec 2001.

[16] D. J. Segal, B. Dreier, R. R. Beerli, and C. F. Barbas, III. 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*, 96(6):2758–63, Mar 1999.

[17] G. Sussman. Electrical models for genetic regulatory elements. Unpublished Document (swiss.csail.mit.edu/∼gjs/syntheticbiology/models.pdf), Jan 2003.

[18] Thomas Knight, Jr. and G. Sussman. Cellular gate techology. *Unconventional Models of Computation*, pages 257–272, 1997.

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

[20] R. Weiss. *Cellular computation and communications using engineered genetic regulatory networks*. PhD thesis, Massachusetts Institute of Technology, Oct 2001.

[21] R. Weiss and S. Basu. The device physics of cellular logic gates. In *NSC-1: The First Workshop on Non-Silicon Computing*, pages 54–61, 2002.

[22] R. Weiss, G. Homsy, and Thomas Knight, Jr. Toward in-vivo digital circuits. In *Dimacs Workshop on Evolution as Computation*, 1999.

[23] S. A. Wolfe, E. I. Ramm, and C. O. Pabo. Combining structure-based design with phage display to create new cys(2)his(2) zinc finger dimers. *Structure Fold Des*, 8(7):739–50, Jul 2000.