# Implementing Genetically Encoded memory

Infinite Monkey Theorem group
Department of Bioengineering
Stanford University
{tons}@stanford.edu



**Abstract:** In this review, we outline the goals and possible directions of our project on implementing memory in a single-cell. What's memory? How a living cell stores information about its history? What're some possible ways to encode memory into DNA sequence in a living cell? We will also discuss some possible miniprojects to start exploring different constructions.

October 6, 2008

## Memory in biological system

(basic definition and outline of some natural examples of memory system) The system that can store information about its history. For example, cells that have received different signals (chemical, temperature, light, etc) in the past end-up behave differently, even though they are now in the exactly the same environment.

A single-living cell uses stored information from the past to cope with the selective pressure in the future. For instance, Multicellular organisms extensively use single-cell memory system memory system during development: stem cells that have gone through different chemical clues differentiate take different fates...

## Applications

A living cell needs memory, so do biologists and engineers.

Why do we want to a programmable memory system? We want to "record" something information that a natural living cell doesn't. Moreover, we would like to have a mechanism to help us retrieve the stored information. We can see application in the two areas.

- **tools for basic biology.**

  In many area of basic, it is very important the readout the history of a cell. For example, we may want to know how many time this particular cell has been divided.

  - Studies of aging: For example, the scenecence of budding yeast
  - Studies of development: Tracking cell-lineage
  - Cancer research: Cell division control
  - Directed evolution: mutagenize a specific location on the genome

- **components for engineering a cell.**

  - Reprograming a cell

    All computing machine require memory unit to temporary store information. For example, in development, a cell need to know when it should execute a particular genetic program.
  - Directed evolution: mutagenize a specific location on the genome

Memory is also extremely important in synthetic biology.

- Can we generalize our memory module?

  We might end-up spend lots of time building a memory system that can only store a particular information in a particular cell. At least, part of our memory system should be modular. For example, we could have the system that write/store/express genes. The writing part is coupled with a signal detector. We just need to change the signal detector, if we want to change the type of information we store.

- Studies of development: Tracking cell-lineage
- Cancer research: Cell division control
- Directed evolution: mutagenize a specific location on the genome

Our ultimate engineering goal is to create a memory module that can receive the signal of our choice, stably storage the information about that signal (e.g. have many time we shine the light to it? what glucose concentration it has been cultured in? how many time it has been divided, etc.), then we can easily readout the stored information.

USB drive for a living cell

## Memory system: abstract level

From engineering perspective, the memory system should have three major components: storage space, writing-machinery and reading machinery.

1. The storage-space can be any system that has multiple states. Some examples are different ON/OFF pattern of transistors, magnetic stripes of DVDs, different sequences of DNA, and different steady-states of gene expressions. The more states we have, the more information we can store.
2. The writing machine receives the signals and cause the storage-space to change its states. For example, DVDwriter gets signal from the computer rewrites magnetic pattern on DVD, signal-transduction machineries receives environmental clues (temperature, light, mechanical force or concentrations of some molecules inside or outside the cells) and bring the genetic circuits from one steady-state to another steady-state.
3. The reading machine extract information out from the storage, either for us to read or for coupling with the writing part to implement further computation (for instance, to add or subtract value from multiple-bits memory).

Note that our choice of storage mechanism will determine the possible scheme for writing and reading mechanism.

## How memory systems are implemented in a single-cell?

There are numerous examples of a living cell in nature with memory properties. (8), Almost every part of the cell can be used as a storage-space. However, given the limited knowledge we have about how a living cell functions, some implementation would be much more challenging than the other. In the following sections, we will enumerate possible mechanisms to store memory in a cell and discuss the advantages and challenges of each mechanism.

We roughly can divide the memory storage space of a cell into three levels (figure...)

1. At the bottom level, memory can be stored in the sequence of DNA (figure...). Different states corresponds to different DNA sequence (or module of the sequence encoded from particular function). Some natural examples are V(J)D recombination of lymphocyte and mating type switching in yeast. A synthetic biology example, as I have seen, is memory encoded invertase system by T. S Ham et al (5)

   - Advantages: DNA is a very stable storage-space and require only minimal energy and cell resource to maintain the information. The reading mechanism should be simple and natural: by coupling the information in DNA to the gene expression.

- Challenges: The major challenge is in writing machinery: how can we specifically modify a specific location in DNA? There are many tools that have long been exploited by biologists to specifically modify genetic codes. Some examples are restriction enzymes, viral-integrase, transposons and DNA recombination. We will extensively discuss about this kind of implementation later in this article.

2. In the next level, memory can be stored epigenetically in the different configuration of chromatin, via, for example the methylation of bases or chromatin remodeling. Different states would be, for instance, the methylation patterns or the coiling configuration around histone. The most obvious natural examples is cell differentiation during development. Memory storage by modifying chromatin or methylation hasn't been exploited extensively by synthetic biologists.

    - Advantages: Similar to the information storage in DNA sequence, the information here is also stored in the stable "structures" of molecule. Minimal energy and other cell resources are required. Reading mechanism can be implemented by reading out gene-expression level. However, the relations between methylation or chromatin structures and the level gene-expression are less understood.
    - Challenges: Similar to information storage in DNA sequence, we need mechanisms to methylate or coil-up/loosen chromatin at specific locations. It seems like we have fewer biotechnological tools to do the job than in the case of DNA sequences.

3. At the top level, memory can be stored as the dynamics of gene products. Different states would be, for example, different steady-state of gene product concentration. There are so many example or multistabilities of gene expression in nature like. The synthetic example of this is a bistable toggle switch by Gardner et al (3).

    - Advantages: This is an example of dynamic memory. We could see the rapid change in the readout once new information get in. Moreover, nature seems to give us wide range of tool to coupling the signal detection to the change of the concentration or activities of some compound inside the cells.
    - Challenges: First the readout of the information is not simple and very diverse, depending one what compounds we use in the circuit. Second, this storage mechanism constantly consume energy and cellular resources as some gene must be ON all the storage time.Third, it's not trivial to expand the size of memory storage aka to engineer a circuit with multiple steady-state. At the current state of system biology, the predictive quantitative models for large genetic circuits have yet to be developed.

## How to re-write DNA sequence in a living cell?

How could we change the DNA sequence at the specific site in vivo? Nature has provide us several tools: restriction/ligation enzymes, invertase, integrase, transposase, terminal transferase, homologous/nonhomologous recombination etc. To pick to tools and further develop into writing machinery, we should consider the following factors:

- **Site-specificity** We probably don't want to change DNA at random places and screw-up cellular functions. Most of enzymatic tools such as integrase, transposase and invertase act one specific sequence of target DNA. The one question is how specific are they?Another question is whether we can modify the mechanism to change its sequence specificities

- **Efficiency and leakiness** Ideally, we want the sequence to change if and only if a cell get a signal for change. However, many natural mechanisms have pretty low efficiency: when the writing mechanism is triggered, only small fraction of cells actually do the writing. For instance, double-inversion recombination switches by Han et al (5) seems to have problems with the low inversion rate by invertase (note the reported efficiency).

- **Reversibility** We don't want our writing machinery to write the information and randomly erase the information at the same time. This is actually the case for many natural systems like transposon or phage-integration: the mechanisms catalyze both forward and backward rate of the sequence change at the same time. While we would like the memory to be erasable, we would like to control the forward and the backward change separately.

- **Complexity** the last but possibly the most important question is whether the tool is easy or hard to engineer, given the limited knowledge we have. Some natural writing mechanisms involve only few know enzymes, some need many, including poorly characterized factors. One question to ask is whether we can reconstruct the whole mechanism in the second organism.

Let's consider some natural writing mechanisms and their specificity, efficiency, reversibility and complexity (to do).

- **Homologous Recombination**
  - Specificity:
  - Efficiency:
  - Reversibility:
  - Complexity:
- **Invertase** Hams et al implemented the recombination switch using double inversion. Although invertase activity was optimized, only small fraction of cells successfully inverted their DNA.
- **Transposon**
- **Integrase/excisase**
  - Specificity:
  - Efficiency:
  - Reversibility:
  - Complexity:

  MP Calos group at Stanford has reported DNA integration in *E.coli* by $\phi C31$ at near 100% efficiency (4)
- **Terminal Transferase**
- **DSB and non-homologous recombination**

  By directed evolution Zn finger domain, endonuclease that introduce DSB at particular sequence can be engineered (7) . In fact, this system has been used for knocked out individual gene in Zebrafish (6)
- **Yeast mate-type switching**

  Yeast's DNA shuffling mechanisms has been used to mediate antibody class switching (9)
- **V(J)D recombination/ somatic hypermutation in immune system**

# Well, what should we do next?

Let's imagine that we have wet lab and everything all set, here are some possible mini-projects to start with

- **Turn on/off a given gene by DNA sequence level**

  With few exceptions, a living cell in nature control gene expression on epigenetic levels. In this project we will build the simple system for turning on/off genes at DNA sequence levels. Signal-1triggers a cell to change DNA sequence so that it can express something out; signal-2 trigger a cell change DNA sequence back to non-expressing state.

  Yeast mate-type switching occurs when the DNA sequence from the silence locus is copied into the active locus (figure). For example, haploid *C.cerevisiae* has either MATa or MAT$\alpha$ at the active locus and HMR and HML locus that contain the similar sequence but under the silencer. HO nuclease induce DSB on MAT. Then, the sequence from HML or HMR is transfered to the broken region to fix the damage. This could result in gene conversion.

Will such gene conversion mechanism still work,with high efficiency, if we change the sequence at MAT to something else that we want? The simple toy-experiment would be the switching of expressed fluorescence color like in figure (). If we cannot move the whole fluorescence protein into MAT, we could use the conversion mechanism to produce different kind of mRNAs sequence, that indirectly control something else.

Intergrase/excisase system is another strong candidate due to its efficiency, simplicities and controllable reversibility. We can use the integration of a particular sequence to turn some gene OFF (or ON) and use excision to convert DNA back to the original state.

(1) (2)

**Collaborators , co-adivisor**: Allan Cambell, MP Calos

- **Modifying site-specificity of enzyme writer**

  Natural DNA-manipulating enzymes have their own sequence specificity. Wouldn't it be nice to modify enzyme so that it can write at other sequence we want? or

- **The chain of DNA modification**

  The modification of DNA at one site can trigger the machinery for modifying DNA at the other site

.

# References


[1] Leander Blaas, Monica Musteanu, Rainer Zenz, Robert Eferl, and Emilio Casanova. Phic31-mediated cassette exchange into a bacterial artificial chromosome. *Biotechnique*, 43:659–664, 2007.

[2] Michele P. Calos. The $\phi c31$ integrase system for gene therapy. *Current Gene Therapy*, 6:633–645, 2006.

[3] T.S. Garner, Charles R. Canter, and James J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 404:339–342, 2000.

[4] Amy C. Groth, Eric C. Olivares, Bhaskar Thyagarajan, and Michele P. Calos. A phage integrase directs efficient site-specific integration in human cells. *Proc Nat Acad Sci USA*, (11)97:5995–6000, 2000.

[5] TS Ham, SK Lee, JD Keasling, and AP Arkin. Design and construction of a double inversion recombination switch for heritable sequential genetic. *PLOS One*, page 3(7): e2815. doi:10.1371/journal.pone.0002815, 2008.

[6] Xiangdong Meng, Marcus B Noyes, Lihua J Zhu1, Nathan D Lawson, and Scot A Wolfe. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotech*, 26:695–701, 2008.

[7] Helen Pearson. The fate of fingers. *Nature*, 455:x337–381, 2008.

[8] Denis M. Wolf, Lisa Fontaine-Bodin, Ilka Bischof, Gavin Price, Jay Keasling, and Adam P. Arkin. Memory in microbe: Quantifying history-dependent behavior in a bacterium. *PLOS One*, page doi:10.1371/journal.pone.0001700, 2008.

[9] Ali A Zarrin, Catherine Del Vecchio, Eva Tseng, Megan Gleason, Payam Zarin, Ming Tian, and Frederick W. Alt. Antibody class switching mediated by yeast endonuclease-generated dna breaks. *SCIENCE*, 315:337–381, 2007.