Computing with living hardware

K.A. Haynes, M.L. Broderick, A.D. Brown, T.L. Butner, L. Harden, L. Heard, E. Jessen, K. Malloy, B. Ogden, S. Rosemond, S. Simpson, E. Zwack, A. Malcolm Campbell, T. Eckdahl, L.J. Heyer and J.L. Poet

Abstract: Our multi-institutional team of eleven undergraduates, one high school student, one postdoctoral fellow, and four faculty members explored the emerging field of synthetic biology and presented our results at the 2006 international Genetically Engineered Machine (iGEM) competition. Having had little or no previous research experience, biology, chemistry and mathematics students from four different institutions collaborated during the summer and fall semester of 2006. We identified the burnt pancake problem (sorting by reversals) as a mathematical puzzle ideal for solving with 'living computer hardware': *Escherichia coli* cells programmed to sort tandem fragments of DNA by reversals (DNA inversions or 'flipping'). Flipping is driven by a *Salmonella typhimurium* Hin/*hix* recombinase system that we reconstituted as a collection of BioBrick-compatible interchangeable parts. We tested functionality of these synthesised genetic parts and mathematically modeled the behaviour of pancake flipping. The living hardware system allowed us to consider future research applications such as regulating genetic element rearrangements *in vivo* and DNA computing. We found the field of synthetic biology to be ideal for learning, teaching, sharing, collaborating, and conducting integrative and original research with undergraduates.

1 Aims of the project: a biological approach to solving a mathematical puzzle

Our team set out to engineer bacteria in order to build living computer hardware that can compute solutions to a mathematical puzzle called the burnt pancake problem. The puzzle can be thought of as a stack of different sized pancakes, each having one burnt side and one golden side, arranged in an arbitrary order. The goal is to rearrange the pancakes by flipping individual pancakes or subsets of adjacent pancakes until the pancakes are sorted from largest to smallest with each pancake facing golden side up. In computer science this process is called sorting by reversals. As the pancake stack becomes larger, the number of possible arrangements increases and the problem becomes computationally intractable. To produce essentially unlimited computing power, we decided to harness the power of *Escherichia coli* DNA replication and cell division. We modified the Salmonella Hin/*hix* DNA recombinase system to perform DNA reversals on plasmid processors in a massively parallel processor (>200 plasmids per cell in a population of $\sim 1 \times 10^8$ cells/mL). Our system, affectionately called the *E. coli* House Of Pancakes (E.HOP) computer, is a proof of concept for computing *in vivo*. Mathematical modelling of random reversals helped us design the system and interpret the output of our E.HOP computer.

doi:10.1049/iet-stb:20070004

Paper first received 18th January 2007

K.A. Haynes, S. Rosemond, S. Simpson, E. Zwack and A. Malcolm Campbell are with the Biology Department, Davidson College, Davidson, NC 28036, USA

L. Harden and L.J. Heyer are with Math Department, Davidson College, Davidson, NC 28036, USA

A.D. Brown, T.L. Butner, L. Heard, E. Jessen, K. Malloy, B. Ogden and T. Eckdahl are with the Department of Biology, Missouri Western State University, St. Joseph, MO 64507, USA

M.L. Broderick and J.L. Poet are with the Department of Computer Science, Math and Physics, Missouri Western State University, St. Joseph, MO 64507, USA

S. Rosemond is also with the Biology Department, Hampton University, Hampton, VA 23668, USA

L. Heard is also with the Central High School, St. Joseph, MO 64506, USA

K.A. Haynes, A. Malcolm Campbell, T. Eckdahl and L.J. Heyer are also part of the Genome Consortium for Active Teaching, Biology Department, Davidson College, Davidson, NC 28036, USA

E-mail: kahaynes@davidson.edu

2 Description of the work

2.1 Building the E.HOP computer

The burnt pancake problem is ideal for demonstrating the computational capabilities of living hardware. The biological equivalent to a burnt pancake is a functional unit of DNA such as a promoter or coding region. Similar to burnt pancakes, expressed DNA elements have directionality (5′ to 3′), require a specific order of the units (e.g., promoter followed by coding region) and can be flipped (cut, inverted, and spliced *in vivo* by cellular machinery). To flip units of DNA, we have reconstituted the Hin/*hix* recombinase system from *S. typhimurium* as a BioBrick-compatible set of components for use in *E. coli*. In *S. typhimurium*, native Hin recombinase activity is required for the inversion of a ~1 kb chromosomal segment that mediates the expression of the H1 and H2 flagellin genes during phase variation [1].

Hin recombinase was cloned from the *S. typhimurium* (Ames strain TA100) genome and tagged with the LVA degradation signal (part BBa_M0040 [2]) using PCR amplification. Hin-mediated DNA inversion requires the recombinational enhancer (RE), a cis-acting DNA element, and *hixC*, a symmetrical 26 bp sequence that is recognised by the Hin homodimer [3]. To construct these parts, we used the publicly available genomic sequence of *S. typhimurium* and a double-stranded DNA (dsDNA) assembly program we created for gene synthesis from overlapping oligos [4].

Any segment of DNA capable of being inverted (i.e., a DNA pancake) must be flanked by a pair of *hixC* sites. Hin recombinase recognises these *hixC* sites and cleaves both strands of DNA. With the help of the Fis protein bound to the RE, the invertasome complex inverts the *hixC*-flanked DNA fragment [5, 6]. In our system, selectable phenotypes of antibiotic resistance depend upon the proper arrangement of a series of *hixC*-flanked DNA segments in a plasmid. A selectable marker allows us to detect which cells have successfully solved the puzzle.

Using several modes of communication (Fig. 1), our iGEM teams at Missouri Western State University (MWSU) and Davidson College cooperatively addressed specific design considerations. We selected segments of DNA that could be sorted to yield a phenotypically unique solution and we built several construct intermediates to optimise positioning of the *hixC* sites. Our final design is a two-pancake stack (Fig. 1); pancake 1 is a pBAD promoter (part BBa_I13453 [2]) from the arabinose operon and pancake 2 is a tetracycline resistance coding region with an upstream ribosomal binding site (RBS-Tet^R, part BBa_S03562 [2]). We designed an insulated vector called pSB1A7 that effectively prevents read-through from the vector backbone into the pancake stack [2]. An additional plasmid encodes the AraC and Hin invertase proteins. This AraC/Hin generator (part BBa_J3108 [2]) was designed to express Hin-LVA in the presence of IPTG and pause transcription (via AraC binding to pBAD) during Hin-mediated DNA inversion. After co-transformed cells have undergone random flipping, samples are grown in the presence of tetracycline to obtain colonies that carry a correctly sorted pancake stack. Before running the living hardware system, we mathematically modelled its behaviour to help us interpret the results.

2.2 Mathematical modelling of pancake sorting

Our mathematical representation of a burnt pancake stack is a signed permutation, in which each integer represents the pancake size and the sign of the integer represents the orientation. The permutation $(1, 2, \ldots, n)$ denotes a sorted stack of n pancakes in order from smallest to largest, all golden side up. A negative sign denotes a pancake facing burnt side up. Note that for n burnt pancakes, there are $(2^n \times n!)$ possible arrangements. To visualise how many flips are required to sort each arrangement, we generated graphs in which the signed permutations comprise the vertices and a flip of a single or multiple adjacent pancakes is represented by an edge connecting two vertices (Fig. 2).

In these graphs, forward and reverse orientations of pancake stacks (i.e., (1, 2) vs. $(-2, -1)$) are considered distinct, whereas in a biological system they are functionally equivalent. Cells carrying either (*pBAD, RBS-Tet^R*) transcribed from left to right or (*RBS-Tet^R*reverse, *pBAD* reverse) transcribed from right to left will express tetracycline resistance. Since both of these arrangements can be thought of as correctly sorted stacks, the modelling would have to be modified to determine the number of flips necessary to reach either configuration and only half of the signed permutations could be considered. To discriminate between the two biologically equivalent correctly sorted stacks, the DNA construct includes a stationary promoterless red fluorescent protein reporter (*RBS-RFP* reverse, part BBa_J31011 [2]) that is expressed when *pBAD* is in the reverse orientation (Fig. 1).

Some additional biological factors must be considered in order to model the behaviour of the living hardware system. Flipping might be biased for DNA fragment size, proximity to the RE, or single vs. multiple pancakes. Furthermore, plasmid copy number influences the probability of randomly solving the problem in a single cell. Thus far, we have used MATLAB to simulate cell survival after completely random flipping on a single-copy plasmid; continued work will generate simulations that consider flipping bias and high plasmid copy number. Data from these simulations will help us to interpret the behaviour of Hin-mediated flipping *in vivo*.

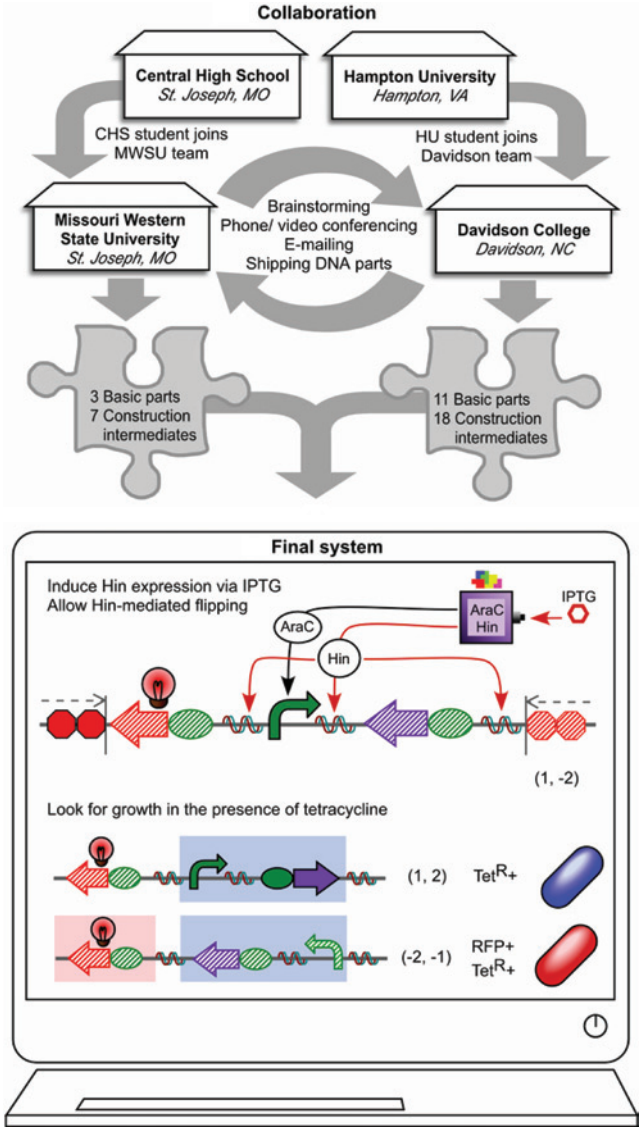

Fig. 1 *Engineering living hardware via multi-institutional collaboration*

During the summer of 2006, a senior from Central High School and an undergraduate summer research fellow from Hampton University were hosted by MWSU and Davidson College, respectively. Students and faculty at MWSU and Davidson College used a variety of media to develop project ideas and to execute research

Final living hardware system consists of two plasmids: (1) an AraC/Hin generator (purple box) that encodes the AraC and Hin invertase proteins (white circles); and (2) a two-pancake stack in which the *pBAD* promoter (green bent arrow) and a tetracycline resistance coding region (purple arrow) with an upstream ribosomal binding site (green oval) are flanked by *hixC* sites (wavy lines). Pancakes in the reverse orientation are hatched. IPTG induces *pLac*-driven Hin-LVA expression; Hin-LVA recognises the *hixC* sites (red vertical arrows). AraC binds to *pBAD* and pauses transcription during Hin-mediated flipping. Forward and reverse terminators (red octagons) in cloning vector *pSB1A7* block transcriptional read-through from the backbone into the pancake stack. *E. coli* cells are co-transformed with an unsorted pancake stack (i.e. (1, −2) shown here) and the AraC/Hin generator plasmids, then screened over time (increasing numbers of flips) for tetracycline resistance and RFP expression

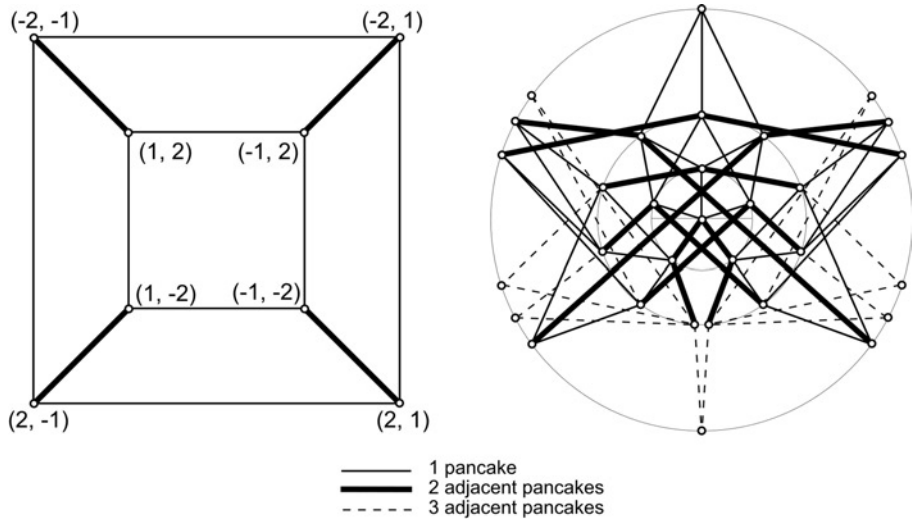

Fig. 2 *Mathematical modelling of burnt pancake flipping*

Graphs representing the relationships between signed permutations of pancake stacks. Each permutation is located at a vertex (open circle) and the edges denote flips of a single pancake or multiple adjacent pancakes
The graph for a two-pancake stack (left) has eight vertices
The graph for a three-pancake stack (right) with 48 vertices is plotted on a sphere. The 17 vertices in the Northern Hemisphere and 14 vertices on the equator are shown

3 Discussion

3.1 E.HOP computer is a proof of concept for computing in vivo, with implications for future data storage devices and transgenic systems

A two-pancake stack is easy to solve without the aid of a computer, but as the stack gets larger, the puzzle becomes more computationally challenging. The parallel processing capacity of the E.HOP computer should allow us to tackle larger problems with ease. The E.HOP computer also has potential to serve as a novel means for data encoding and storage. A series of DNA pancakes arranged in either the forward or backwards orientation is analogous to binary code (1's and 0's). Sorting by reversals generates $(2^n \times n!)$ configurations of n pancakes, a combinatorial explosion of data that could be written to plasmid 'hard drives' *in vivo*. Our BioBrick-compatible Hin/*hix* recombinase system may also prove useful for basic biological research. Controlled reversals *in vivo* would allow orientation-dependent function of DNA elements to be tested at a single locus. Large pancake stacks could serve as model systems to gain insights into gene rearrangements within syntenic chromosomes over evolutionary time. Furthermore, Hin-mediated genetic toggle switches could allow adjustable expression in transgenic organisms.

Fig. 3 *Team photo*

Front row (left to right): Trevor Butner, Brad Ogden, Eric Jessen; middle row (left to right): Todd Eckdahl, Laurie Heyer, Karmella Haynes, Lane Heard, Samantha Simpson, A. Malcolm Campbell, Jeff Poet; back row (left to right): Adam Brown, Marian Broderick, Sabriya Rosemond, Kelly Malloy, Lance Harden, Erin Zwack

3.2 Collaboration between underrepresented institutions yielded world-class research

In multi-institutional collaborations, efficient communication is the most significant limitation imposed by a lack of direct contact between team members. Creative thinking, teamwork, and electronic communication helped narrow the physical divide between the MWSU and Davidson College campuses. Overnight express shipping, e-mail, iGEM Wiki page editing [7, 8], online instant messaging, and conferences via phone and the internet allowed us to engineer our genetic devices in sync. Despite the lack of any face-to-face meetings until the 2006 iGEM Jamboree (Fig. 3), the MSWU and Davidson College teams produce two well-meshed presentations; one of these earned a 'Best Presentation' award (MWSU, first place) [9].

This year, only three primarily undergraduate institutions (PUIs) in the USA and two minority serving institutions (MSIs) were represented among many large research-driven universities located across the globe. Our team represented about half of the underrepresented institutions (MWSU (PUI), Davidson College (PUI) and Hampton University (MSI)). It is worth noting that compared to research-intensive institutions, our schools have no graduate students, small budgets and heavy faculty teaching loads. Team member cooperation and iGEM program resources mitigated these challenges and enabled us to conduct cutting edge research that blends biology and mathematics. Our collaborative research won five awards: 'Best Presentation' (MWSU, first place), 'Best Part' (Davidson,

second place), 'Best Cooperation and Collaboration' and 'Best Poster' (Davidson and MWSU, second place) and 'Best Conquest of Adversity' (Davidson and MWSU, third place) [9]. We have learned that multi-institutional collaboration on synthetic biology research can be a fun and rewarding experience.

4 Acknowledgments

K.A. Haynes, L. Harden, and S. Rosemond were supported by fellowships from the Howard Hughes Medical Institute (HHMI #52005120) at Davidson College. The Davidson team was supported by the James G. Martin Genomics Program and other campus resources. The MWSU team was supported by the Western Institute, the Western Summer Research Institute, the Student Excellence Fund, the Student Government Association, and Academic and Student Affairs.

5 References

1 Glasgow, A.C., Hughes, K.T., and Simon, M.I.: 'Bacterial DNA inversion systems,' in Berg, D., and Howe, M. (Eds.): 'Mobile DNA' (American Society for Microbiololgy, 1989), pp. 637–659
2 http://parts.mit.edu/registry/index.php/Main_Page, accessed January 2007
3 Lim, H.M., Hughes, K.T., and Simon, M.I.: 'The effects of symmetrical recombination site *hixC* on Hin recombinase function', *J. Biol. Chem.*, 1992, **267**, (16), pp. 11183–11190
4 http://gcat.davidson.edu/IGEM06/oligo.html, accessed January 2007
5 Johnson, R.C., Bruist, M.F., and Simon, M.I.: 'Host protein requirements for *in vitro* site-specific DNA inversion', *Cell*, 1986, **46**, (4), pp. 531–539
6 Bruist, M.F., Glasgow, A.C., Johnson, R.C., and Simon, M.I.: 'Fis binding to the recombinational enhancer of the Hin DNA inversion system', *Genes Dev.*, 1987, **1**, (8), pp. 762–772
7 http://parts.mit.edu/wiki/index.php/Davidson_2006, accessed January 2007
8 http://parts.mit.edu/wiki/index.php/Missouri_Western_State_University_2006, accessed January 2007
9 http://www.igem2006.com/results.htm, accessed January 2007