Volume 28 Fall 2014

Massachusetts Institute of Technology

Dark matter overestimated in the Milky Wayp. 4Editing genes with CRISPR (on cover)p. 10Facebook posts influence memoryp. 36





0 Ο 0 C Ծ C Φ 5 0 _ Φ U Ծ Φ S S Φ Ľ Φ μ $\boldsymbol{\omega}$ ų Л Ð S σ Б Ծ 0 U σ Ð S S σ σ

Contents

Introductory Letter

From MURJ Editors



Science News In Review

A look at the latest Science News.



Features



Editing Genes with CRISPR

Targeting and killing antibiotic-resistant bacteria, modeling lung cancer-causing genes, reversing liver disorders and modifying mice genomes in vivo — researchers have accomplished all this and more in the last year alone, through the use of a new gene editing technique.



MURJ Spotlight: Professor Douglas Lauffenburger MIT Department of Biological Engineering

This issue's spotlight features Professor Douglas Lauffenburger, Ford Professor of Biological Engineering, Chemical Engineering, and Biology and the Head of the Department of Biological Engineering. 10-13

14-18

MURJ

UROP Summaries

The Role of SIRT1 in Modulating Circadian Rhythms Through High-fat Diet

Jiapei Chen, Caitlin Ondracek, Leonard Guarente

Circadian rhythms are largely regulated by the suprachiasmatic nucleus (SCN) located in the hypothalamus. The SCN responds to light-dark (LD) cycles from the environment and generates an oscillation of 12-hour light and 12-hour dark cycles accordingly...

How Integration Mechanisms Impact the Performance of the Adjoined Mental and Physical Health Care in the Clalit Health System

Kyle Yuan, Wiljeana Glover, John Carrol

On May 1, 2012, the Socio-Economic Cabinet of the Israeli Government approved a reform in the mental health services in Israel, which transfers the responsibility **22-23** for providing mental health services from the Israeli Ministry of Health to the four non-governmental Health Maintenance Organizations (HMO) in the nation...

Determining the Effectiveness of a Carbonaceous Iron/Sand (CarFeSand) Column Filter System in Removing Heavy Metals from Polluted Water

Justin Yuan, Evangeline Ng, Reginald Thio

According to a study conducted by the World Health Organization (WHO), millions of people die each year from waterrelated illnesses or sanitation and hygienic problems due to lack of availability of water purification techniques and equipment. The objective of this project is to design and build an innovative and sustainable carbonaceous iron (FeO)/sand (CarFeSand) filter system...

mRNA Delivery and Expression Using a Vector-free Microfluidic Intracellular Delivery Platform

Derek Jang, Camilo Ruiz, Tushar Kamath, George Hartoularos, Jasdave Chahal, Armon Sharei, Klavs Jensen, Robert Langer

One of the greatest obstacles to clinical application of nucleic acids is successful delivery of the nucleic acids to the target cells. Although there are currently existing delivery methods, they have unique disadvantages that limit their efficacy...

26-28

Reports

A Reimplementation of Python's Dictionary Using Simple Tabulation Hashing and Linear Probing

Pedram Razavi, Thomas Georgiou, Russell Cohen, Erik Demaine

Here we analyze the performance improvements obtained by reimplementing the standard Python's dictionary data structure by replacing the built-in hash function for strings with one based on tabulation hashing, and its collision resolution strategy with linear probing. These choices are motivated by the recent theoretical results of Pătrașcu and Thorup who showed that in this hashing scheme operations take constant expected time...

30-35

Inferring Social and Linguistic Influences on Memory from Facebook Posts: An Exploratory Analysis

Brianna Jones, Rebecca Saxe

Why are some things easier to remember than others? To answer this question, we tested the short-term memorability of a variety of Facebook posts and book sentences by presenting participants with a large number of items then immediately testing their recognition memory. The stimuli were coded for several social and linguistic features...

36-40

Letters

MURJ

Volume 28, Fall 2014

of Technology

Massachusetts Institute

Massachusetts Institute of Technology

UNDERGRADUATE RESEARCH JOURNAL Volume 28, Fall 2014

Editors-In-Chief

Elliot Akama-Garren Tatyana Gubin **Reuben Saunders**

Research Editor Lakshmi Subbaraj

News & Features Editor Linda Jiang

Assistant Features Editor Eta Atolia

Copy Editors Carolyn Lanzkron Daphne Superville Winter Guerra

Features Contributing Staff Elizabeth Berg Puwanat Sangkhapreecha Amelia Trainer

News Contributing Staff Jeff Chen Pratheek Nagaraj Puwanat Sangkhapreecha Sebasthian Santiago

Research Staff Daphne Superville Do Yeon Park Kaustav Gopinathan Mahmoud Ghulman

Web Developer Ryan Alexander

Production Advisor DSGraphics

Printed at Upper Valley Printing

Massachusetts Institute of Technology 77 Massachusetts Avenue Cambridge, MA 02139

MURJ Staff MIT Undergraduate Research Journal

December 2014

Dear MIT community,

We are thrilled to publish the 28th issue of the MIT Undergraduate Research Journal. In this issue we have the privilege of presenting outstanding research conducted by MIT undergraduates from across campus and across disciplines. Here you will read about an approach of tabular hashing to increase efficiency of the Python programming language, a study of the linguistic influences on recognition memory, and a MISTI Singapore project to design a novel water purification column to increase much-needed access to safe drinking water.

We also present two features articles that explore cutting-edge scientific advances. In this issue, our features explore the recent explosion of CRISPR/ Cas9 technologies that are changing the field of biomedical research. In addition, we present an interview with Professor Douglas Lauffenburger, head of the MIT Department of Biological Engineering.

Biannual publication of this journal is only possible due to an extraordinary team of dedicated students, and we would like to thank all of our editorial board and contributors for their time and effort this semester. We would also like to thank all the undergraduates who took their time to share their research with us, and the greater MIT community. If you would like to contribute to future issues of the MIT Undergraduate Research Journal, we invite you to join our team of authors and editors or submit your research for our Spring 2015 issue. Please contact murj-officers@mit.edu if you have any questions or comments.

Best,

Elliot Akama-Garren (Co-Editor-in-Chief)

Tatyana Gubin (Co-Editor-in-Chief)

Reuben Saunders (Co-Editor-in-Chief)

lgarren Tatyano Gli Ruben A.SM

No material appearing in this publication may be reproduced without written permission of the publisher. The opinions expressed in this magazine are those of the contributors and are not necessarily shared by the editors. All editorial rights are reserved.

Science News In Review

NEUROSCIENCE

Optogenetics: a Powerful Tool to Understand and Control the Brain

In a landmark paper published in 2005 in Nature Neuroscience, researchers at Stanford showcased an ingenious system with which to activate neurons and cause them to fire action potentials by shining bursts of blue light on them. Known as "optogenetics," the approach is now utilized by scientists around the world to study and control—various parts of the brain.

The optogenetic system combines the ectopic expression of lightreceptive proteins called "opsins" in neurons and the installment of a penny-sized device containing numerous optical fibers. Once light is delivered to these neurons, opsins respond by allowing the influx or efflux of charged ions, which then lead to the activation or inactivation of the neurons. imaging (fMRI) have allowed researchers to investigate the average activity of broad brain regions, optogenetics enables scientists to turn on or off specific sets of neurons with exquisite precision. The potential of optogenetics is virtually boundless and its implications far-reaching.

For example, Prof. Edward Boyden-the first author of the 2005 paper who now has his own lab at MIT-has collaborated with colleagues to cure mouse analogs of blindness and post-traumatic stress disorder (PTSD) by applying optogenetics. Earlier this year, researchers at UC Davis employed optogenetics to erase specific negative memories in mice by inactivating neurons associated with the retrieval of these memories-a result that points directly to novel potential treatments of psychiatric disorders. -J. Chen

Sources:

http://www.nature.com/neuro/journal/v8/n9/ abs/nn1525.html

http://www.nature.com/mt/journal/v19/n7/ full/mt201169a.html

http://www.cell.com/neuron/abstract/S0896-6273(14)00895-2

ASTROPHYSICS

Dark Matter Overestimated in the Milky Way

One of the biggest enigmas in astronomy is why the Milky Way and other galaxies weigh as much as they do. Our current understanding of matter falls short in explaining the true weight of these galactic bodies. To compensate for our lack of knowledge we hypothesize the existence of dark matter - matter that has gravitational effects but seemingly snubs any interaction with electromagnetism.



A artistic depiction of the Milky Way galaxy. White regions represent ordinary and detectable matter, while the blue halo represents dark matter. Credit: European Space Organization/L. Calcada

To put it into perspective, we only know about 4 percent of matter in the universe, the rest is either dark matter or dark energy (a hypothetical energy that helps explain the expansion of the universe). Further, we believe that our own galaxy, the Milky Way, has 10 times more dark matter than ordinary matter.

Researchers now believe that the amount of dark matter in the Milky Way is about 8x10¹¹ times the mass of the Sun, half that which was previously thought. In order to determine this new measurement Astrophysicist Dr. Prajwal Kafle, from the University of Western Australia, and his team examined the fringes of the galaxy at a distance of 5x10^15 kilometers from Earth. Then, using the speed of stars as determined by electromagnetic radiation, they could determine the amount of mass needed to keep the stellar objects rotating at the measured velocities.

Current dark matter theories suggest the existence of satellite galaxies around the Milky Way. Based on the measured mass, there should be three satellite galaxies,



Neuroscientists insert genes of lightsensitive proteins in neurons, which they can then trigger on demand with pulses of light via optical fibers. Courtesey of Deisseroth lab at Stanford. Image credit: http://www.osa-opn.org/home/articles/volume_22/ issue 7/features/

While other advanced technologies like functional magnetic resonance

MURJ

Volume 28, Fall 2014

which is precisely what one can see. Thus, Kafle's work reconciles some of the cosmological problems associated with dark matter but still leaves many problems open.

-P. Nagaraj Source: http://www.sciencedaily.com/ releases/2014/10/141009091600.htm

Perfect Fluid of the Early Universe

The early stage of the universe was not unlike that of soup - fluid and hot. Particles existed in their most elementary forms. Quarks and gluons (basic matter and force particles) filled an ultra-hot environment, which we refer to as quarkgluon plasma. This exotic state of matter lacks "stickiness" (viscosity, heat conduction, and shear stress) making it quite nearly a perfect fluid.

Using particle accelerators to recreate conditions of the primordial The ALICE experiment showing simulated collisions of lead ions. By accelerating particles to near light-speed, scientists can explore an exotic form of matter known as quarkgluon plasma. Credit: ALICE Experiment at CERN



universe, experimental physicists have discovered new properties of this perfect fluid that have puzzled researchers for decades. The multiinstitutional team, called the JET Collaboration, combined the data from the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven National Laboratory in New York and Large Hadron Collider (LHC) at CERN in Switzerland. They were specifically looking for energy loss of high-energy particles, or "jets," inside the quark-gluon plasma. Knowing the energy loss provides clues to the density and interaction strength of the plasma.

The theoretical underpinnings of the work date back to Einstein's theory of special relativity and incorporate the motion of fluids into what is known as relativistic hydrodynamics. An important quantity in this field is the jet transport coefficient, which describes the level of interac-

MATERIALS SCIENCE

Super-Repellant Coating Repels Blood and Bacteria from Medical Devices

Medical devices are used in a great deal of crucial operations, from cardiac implants to joint replacements. However, once they are introduced into our bodies, they are treated as foreign substances, causing immediate blood clotting and a chance of bacterial infection. Now, using materials previously approved by the FDA, a research group from the Wyss Institute for **Biologically Inspired Engineering at** Harvard has developed a new surface coating that can help prevent blood from clotting when it comes into contact with over twenty substances made from glass, plastic, and metal that are commonly used in medical devices.

The team coated medical-grade tubing and catheters implanted in pigs' blood vessels and found that the devices prevent blood clotting up to 8 hours without anticoagulants like Heparin, which can have potentially



This Scanning Electron Microscope (SEM) illustrates how red blood cells form blood clots. Credit: James Weaver, Harvard's Wyss Institute

lethal side effects such as excessive bleeding. The idea of this technology came from Slippery Liquid-Infused Porous Surfaces (SLIPS), which is a novel surface technology created by Dr. Joanna Aizenberg, a Wyss Institute Core Faculty member at Harvard. The technology is inspired by the carnivorous pitcher plant, which has a slippery surface to catch insects that prevents everything from ice to crude oil from passing through.

In addition to preventing blood clotting, the paper reports that the coating can prevent bacteria from adhering to the medical device. If used in patients, this coating can prevent sepsis, which is caused by severe blood infections. The research team even tested the technology with a gecko (an animal with super sticky feet) and the gecko failed to adhere to the material. The research team asserts that, in the future, this technology can be applied to more complex medical devices, such as dialysis machines.

-P. Sangkhapreecha Source: http://www.sciencedaily.com/ releases/2014/10/141012134845.htm

Volume 28, Fall 2014

MURJ

Science News in Review

tion between the jet and the ultrahot matter. The collaboration determined a range of values for this coefficient, which further elucidates the perfect fluid nature of the ultrahot matter. Additionally, this work provides a great model for making precise measurements of this enigmatic form of matter.

Future work will adjust the collision energies to investigate how temperature affects the quark-gluon plasma's behavior. The team hopes that this will provide key insights for the phase transition between ordinary matter and the exotic ultra-hot matter.

–P. Nagaraj

Source: http://www.sciencedaily.com/ releases/2014/10/141002141858.htm

MEDICINE

From Embryonic Stem cells to Insulin-Producing Beta Cells

Diabetes, a disease characterized by either insulin deficiency or resistance, currently affects around 382 million people worldwide. The principal cause of diabetes is the deficiency or autoimmune destruction of beta cells. Beta cells are responsible for storing and releasing insulin, a crucial hormone in the metabolism of glucose.

Previous attempts to create insulin producing beta cells from human pluripotent stem cells have proven to be unsuccessful. This was mainly because, although these cells produced insulin, they were not fully functional when transplanted into mice. However, in a paper recently published in Cell titled "Generation of Functional Human Pancreatic β Cells In Vitro", a group of researchers in the Harvard Stem Cell Institute have managed to produce beta cells that are able to function correctly in vivo from human embryonic stem cells. By using published and new controlled differentiations, they



Human insulinmaking beta cells derived from stem cells. Credit: Harvard Institute Melton Laboratoru

were able to direct differentiation of embryonic stem cells to stem cell-derived beta cells. When stem cell-derived beta cells were transplanted into mice suffering from hyperglycemia, or high blood sugar, their blood glucose levels became much lower than those found in mice transplanted with control cells.

Furthermore, after being injected with glucose, mice transplanted with the stem cell-derived beta cells were able to clear the glucose from their blood. This constrasts against the results taken from mice transplanted with the control cells, showing once again the efficacy of these beta cells. These results show great promise in the field of diabetes treatment, bringing concepts such as the engineering of functional pancreatic organs and creation of new powerful therapeutic techniques closer to reality than to science fiction.

-S. Santiago Source: http://www.sciencedirect.com/science/ article/pii/S0092867414012288

Unexpected Benefits from PCV13 Vaccine

Research presented at the IDWeek2014TM, an annual meeting of professionals in the field of infectious diseases and epidemiology, appears to indicate that the 13-valent-pneumococcal vaccine (PCV13) reduces antibiotic pneumococcal resistant infections in children by 62 percent.

The PCV13 vaccine was introduced in 2010, and is recommended for a large demographic: Children; adults 19 years of age or older with HIV, certain cancers, and kidney failure; and adults 65 years of age or older. It is sometimes given in conjunction with other vaccines. For example, one dose of PCV13 is recommended for all 65+ adults, followed 6 to 12 months later by one dose of pneumococcal polysaccharide vaccine (PPSV23). It is able to prevent diseases such as meningitis, pneumonia, and blood stream infections.

These results were discovered when Healthy 2020, a nationwide program for disease prevention, was analyzing data and observed that



The PVC13 vaccine is produced by Wyeth and marketed by Pfizer under the name Prevenar 13. Credit: http://www.mims.com/Hongkong/drug/info/ Prevenar%2013/?type=brief

the rate of antibiotic resistant pneumococcal infections had decreased from 9.3 cases per one hundred

MURJ

Volume 28, Fall 2014

Glucosamine

supplements may increase

our longevity.

Credit: Creative

Commons/bitzi

thousand children to 3.5 cases per one hundred thousand children. Because antibiotic resistance is one of the biggest problems in modern medicine, these data highlight the importance of vaccines, which can prevent the use of antibiotics to treat an infection by preventing patients from contracting the infection in the first place. Thus, vaccines may be a key player in the fight against the growing global epidemic of antibiotic resistance.

-S. Santiago

Source: http://www.sciencedaily.com/ releases/2014/10/141010134344.htm

Glucosamine Supplement Increases Life Span of Roundworms and Mice

Scientists led by Dr. Michael Ristow at ETH Zurich and four German research institutions have found that D-glucosamine supplements can prolong the lives of roundworms and mice compared to their control counterparts. They found that the lifespan of roundworms treated with D-glucosamine supplements was extended by 5%, while the lifespan of mice treated with D-glucosamine supplements was extended by 10%.

For the mice experiments, Ristow and his colleagues used mice that were 100 weeks old, similar to humans at 65 years of age. Both the control and experimental groups received the same diet, except that the latter was treated with D-glucosamine. The researcher found that the life span of mice with D-glucosamine increased by nearly 10 percent, which is equivalent to approximately 8 years of human life.

Ristow argues that glucosamine feeding simulates a low-carb diet



because glucosamine stimulates amino acid breakdown. When there is a shortage of carbohydrates, the amino acids become preferentially metabolized. These results using mice models cannot definitively be applied to humans but Ristow would recommend that people take D-glucosamine supplements since there are no known relevant side effects of glucosamine supplementation.

-P. Sangkhapreecha Source: http://www.sciencedaily.com/ releases/2014/04/140408122135.htm

CNA Corporation is a non-profit organization that conducts objective, empirical research and analysis to help decision makers develop sound policies, make better-informed decisions, and manage programs more effectively.

CNA Corporation is looking for individuals with experience in engineering, mathematics, economics, national security studies, public policy, computer science, and many other scientific and professional fields of study to fill positions as Research Analysts and Research Scientists. We look for people with the training and ability to reason soundly and to apply scientific techniques imaginatively—people who can see beyond surface details to the core of a problem and devise logical, practical solutions. There are also positions offered in information technology, human resources, finance and accounting, and office management.



CNA's Center for Naval Analyses, the Federally Funded Research and Development Center (FFRDC) for the Navy and Marine Corps provides research and analysis services to the military to help improve the efficiency and effectiveness of our national defense efforts.

And through CNA's Institute for Public Research, we assist an array of federal, state, and local government agencies working in such areas as education, homeland security, air traffic management, and energy, water and climate.

CNA has a diverse and inclusive work environment that challenges all staff to excel. We reward excellence with competitive compensation and benefits, and offer opportunities for long-term career progression.

CNA Corporation is an Equal Opportunity Employer All job applicants are subject to a security investigation and must meet eligibility requirements for access to classified information.



Science News in Review

Volume 28, Fall 2014

CELL BIOLOGY

MURJ

Suicide, not murder, of helper T cells drives AIDS progression

It has been known for nearly three decades that the depletion of helper T cells, or CD4 T cells, is the main reason behind the opportunistic infections that eventually kill HIV-infected individuals. However, the exact mechanism of this cell death had been elusive-until two pivotal papers from Dr. Warner Greene's lab at UCSF were published in Nature and Science on the same day in December 2013. The researchers showed that, contrary to conventional wisdom-which attributes the CD4 T cell death to virus-induced apoptosis-the vast majority of helper T cells actually perish by a different kind of suicide termed "pyroptosis."

Derived from the Greek phrase for "fire falling," pyroptosis is an intensely inflammatory form of cell death where dying cells release proinflammatory cytokines to recruit more immune cells to the suicide zone, causing even more inflammation and death. Studying human lymphoid tissues propagated in the laboratory and spiked with HIV, the UCSF researchers noticed a striking pattern: those few CD4 T cells that are "productively infected" by HIV (meaning that the virus has integrated its genetic material into the host genome and can produce daughter virions) activate the conventional executioner Caspase-3 and die by apoptosis; conversely, 95% of the CD4 T cells are only "abortively infected" (meaning that incomplete viral mRNA is transcribed and no viral particles can be assembled), and that these helper T cells activate Caspase-1 via the assembly of a multiprotein complex termed the "inflammasome." The inflammasome response in turn results in the unnecessary vicious cycle of inflammation and cell death.



Mechanism of cell death in productively infected (a) and abortively infected (b) helper T cells. The Caspase-1-induced inflammatory cell death termed pyroptosis accounts for the vast majority of the cases. Credit: http://www.nature.com/nature/journal/v505/n7484/full/505492a.html

This discovery not only sheds light on the long-standing mystery of cell death mechanism in HIV infection but also suggests a novel strategy for therapeutic intervention. The researchers were able to stop the fiery chain reaction *in vitro* with a Caspase-1 inhibitor, which has already been proven safe in humans and is now actively pursued by pharmaceutical companies as the next-generation HIV drug. Unlike other antiviral medicines that target the virus itself, this potential treatment aims to blunt the overactive host immune response and may further increase the efficacy of current therapeutic cocktails.

Sources:

http://www.nature.com/nature/journal/v505/ n7484/full/nature12940.html http://www.sciencemag.org/content/343/6169/428 http://www.sciencemag.org/content/343/6169/383.full

-J. Chen



MURJ

NURJ Features

MURJ

Editing Genes with CRISPR

Targeting and killing antibiotic-resistant bacteria, modeling lung cancercausing genes, reversing liver disorders and modifying mice genomes in vivo — researchers have accomplished all this and more in the last year alone, through the use of a new gene editing technique.

BY ELIZABETH BERG

Gene editing is the process of manipulating an organism's DNA to alter gene expression. A technique that has become popular in recent years, gene editing offers the potential to completely cure genetic disorders. To accomplish this task, scientists work with nucleases, enzymes capable of cutting DNA at specific points and inducing the cell to glue this strand back together without maintaining the original sequence. Zinc-finger nucleases (ZFNs) and transcription activatorlike effector nucleases (TALENs) were two of the original nucleases used for this process. ZFNs are engineered to target specific sequences of DNA; however, scientists are concerned that ZFNs may splice the incorrect DNA region if they do not code for a unique region. This is an issue called "offtarget cleavage." TALENs contain a cleavage domain, which cuts the DNA, and transcription activatorlike effectors (TALEs), which are repeating chains of 33 to 35 amino acids. The 12th and 13th amino acid in each chain, the only two that do not repeat, are responsible for designating the DNA binding site. As with ZFNs, these effectors have the potential to lack specificity, while also being much more costly to produce. Luckily, a groundbreaking new gene editing technique lurks on the horizon with the potential to fix some of the issues found in previous methods.



Cas9 (CRISPR associated protein 9) is an RNA-guided DNA nuclease enzyme that is pushing the boundaries of gene editing and has the potential to completely cure genetic disorders.

Credit: http://commons.wikimedia.org/wiki/File:Crystal_Structure_of_Cas9_in_Complex_with_Guide_RNA_and_Target_DNA.jpg

Clustered regularly interspaced short palindromic repeats, also known as CRISPR, are short symmetric segments of DNA. In bacteria and archaea, these base sequences, along with RNA guide strands and their associated Cas genes, compose a system that provides protection against viruses by recognizing and targeting foreign viral DNA. This system provides prokaryotes similar protection to that which the immune system offers eukaryotes. Just as your own immune system keeps a record of past diseases you have been exposed to, these repetitive CRISPR sequences surround "spacers", or segments of DNA from pathogens that have attacked the bacteria in the past. If a similar pathogen invades again, the CRISPR system identifies that the foreign DNA matches one of the spacer segments and attacks it.

The two components of this system that work together to destroy foreign DNA are "guide RNAs" and an enzyme called Cas9. When a CRISPR system encounters an invading plasmid or bacteriophage with genetic material that resembles one of its own spacer regions, it synthesizes

a strand of matching RNA. This strand of RNA "guides" the Cas9 enzyme to the foreign genome. Cas9, as the offensive member of this team, breaks the invading DNA at regions where it matches the guide RNA.

While the surprising phenomenon of palindromic DNA segments was first observed in the 1980s, scientists have only recently begun piecing together the structure of the CRISPR system to adapt it for new applications. Instead of using this enzyme to cut apart



Simplified diagram of a CRISPR locus. The three major components of a CRISPR locus are shown: cas genes, leader, and repeat-spacer array. For the repeat-spacer array, repeats are shown as grey boxes, and spacers are colored bars. Credit:http://commons.wikimedia.org/wiki/File:SimpleCRISPR.jpg

bacteriophages, modern research goals include using the CRISPR system to edit genes by modifying guide RNAs and Cas9 enzymes to remove genetic mutations.

MIT researchers are now at the forefront of using CRISPR to modify the genome of living organisms

> Just this past February, researchers from MIT Broad Institute, alongside scientists from the University of Tokyo, built the first detailed image of the Cas9 complex, consisting of the Cas9 enzyme, which cuts unwanted DNA out of the target genome, and RNA



Researchers in the Koch Institute for Integrative Cancer Research successfully reversed genetic liver disease in mice using CRISPR. Credit:http://commons. wikimedia.org/wiki/ Mus#mediaviewer/File:Lab_ mouse_mg_3263.jpg guide strands, which direct these enzymes to the foreign DNA. The ability of the Cas9 complex to target and destroy specific sequences of DNA offers exciting potential.

For these researchers, this newly developed understanding of how these CRISPR pieces are structured is the first step to understanding how to alter their function to fit their own desires.

While there is still much more to discover about how CRISPR works in bacterial cells, researchers have recently started studying potential applications of this system towards animal gene regulation. MIT is at the forefront of this new field of study; although examples of using CRISPR techniques with cultured cell lines and other *in vitro* applications are well documented, many MIT researchers are investigating ways to use CRISPR genome editing techniques to modify the genetic material of living organisms.

One group doing just this is the Anderson Lab, in the Koch Institute for Integrative Cancer Research. Working with mice that carried a mutated gene incapable of breaking down the amino acid tyrosine, researchers in the Anderson Lab investigated ways to reverse this mutation. They

Volume 28, Fall 2014

MURJ



CRISPR genome-editing enables researchers to defeat the antibiotic resistance genes of bacteria like the S. Aureus bacteria shown here in gold Credit: http://commons.wikimedia.org/wiki/File:Staphylococcus_aureus_bacteria_escape.jpg provenge.com

injected these infected mice with three things: an unmutated DNA template, RNA guide strands that were directed at the mutated DNA segment, and a CRSIPR Cas gene. Ultimately, they determined that the healthy gene was incorporated into approximately one half of one percent of the mice's liver cells. After enough time, these healthy cells replaced the infected cells, effectively curing the liver disorder and marking the first time that CRISPR was successfully used to reverse disease in living animals.

In addition to fixing the single mutations that cause metabolic disorders, this technique is being studied to help eliminate potentially more complex diseases. Earlier this fall, researchers from the Broad Institute and a handful of other MIT labs developed a new mouse model that should hopefully make it easier to test CRISPR techniques in living mammals. They inserted Caso, the enzyme that bacteria and archaea use to cut foreign DNA, into a live mouse. The easiest way the understand a gene's function is to study what happens in an organism when that gene is deleted or modified. The CRISPR/Cas9 modified organism now contains an enzyme that allows the researchers to quickly and easily edit the mouse's genome while it is still alive and functioning. This greatly facilitates the study of genetic functions. To perform similar experiments without this mouse model, scientists would need to extract cells and apply the CRISPR system to them in a petri dish. This is a much slower process, with a large risk that the cells would die before they could even be modified to express Cas9. to study externally, so there is no longer a race to induce Cas9 expression and complete experiments before the cells die. While the original engineered mouse can now be found at Jackson Laboratory in Bar Harbor, Maine, this method is available for use by anyone in the broader scientific community.

While we are grateful for bacteria as the original source of these CRISPR segments, scientists won't cut these bacteria any slack. Professor Timothy Lu and his biological

engineering lab have been hard at work to turn these bacterial defense mechanisms against their creators. Antibiotic resistant bacteria have become an urgent threat in recent years, with strains of bacteria evolving more quickly than scientists can develop new antibiotics to kill them. Timothy Lu's lab is studying ways to identify and target the specific genes in a given bacteria that provide resistance against antibiotics. The CRISPR genome-editing system is perfectly suited for this task. By designing RNA guide strands that

From livers to lungs, and from bacteria to mice to humans, the applications of CRISPR have seemingly limitless potential

Taking advantage of this new Cas9-expressing mouse, researchers have created a model of lung adenocarcinoma, a dangerous and complex form of lung cancer that results from mutations in a number of different genes. They have also been able to extract cells identified a particular antibiotic resistant gene, called NDM-1, these researchers were able to use CRISPR techniques to kill 99% of the bacteria carrying this resistance.

All these projects are just beginning, and certainly require more testing and experimentation.

Features

MURJ

As a field of research, CRISPR gene editing is still in its infancy, but it is already clear that the range of eventual benefits offered by these small, repetitive pieces of DNA is enormous. From livers to lungs, and from bacteria to mice to humans, researchers, armed with these mysterious new tools, have seemingly limitless potential. ■

References

- Baker, Monya. "Gene Editing at CRISPR Speed." Nature Biotechnology 32.4 (2014): 309-12. Nature. 8 Apr. 2014. Web. 7 Nov. 2014.
- "Broad, MIT Researchers Reveal Structure of Key CRISPR Complex." Broad Institute of MIT and Harvard. N.p., 13 Feb. 2014. Web. 07 Nov. 2014. Citorik, Robert J., Mark

Mimee, and Timothy K. Lu. "Sequence-specific Antimicrobials Using Efficiently Delivered RNAguided Nucleases." Nature Biotechnology (2014): n. pag. Nature. 21 Sept. 2014. Web. 7 Nov. 2014.

- Gaj Thomas, Charles A. Gersbach, and Carlos F. Barbas III. "ZFN, TALEN, and CRISPR/ Cas-based methods for genome engineering". Trends in Biotechnology (2013): 397-405. July 2013. Web. 11c Nov. 2014.
- Hsu, Patrick D., Eric S. Lander, and Feng Zhang. "Development and Applications of CRISPR-Cas9 for Genome Engineering". Cell (2014). 5 June 2014. Web. 11 Nov. 2014.
- Meade-Kelly, Veronica. "Research Reveals Structure of Key CRISPR Complex." MIT News. N.p., 13 Feb. 2014. Web. 07

Nov. 2014.

- Meade-Kelly, Veronica. "Researchers Engineer New Mouse Model to Study Disease." MIT News. N.p., 25 Sept. 2014. Web. 07 Nov. 2014.
- Pollack, Andrew. "A Powerful New Way to Edit DNA." The New York Times. The New York Times, 03 Mar. 2014. Web. 07 Nov. 2014.
- Sander, Jeffry D., and J. K. Joung. "CRISPR-Cas Systems for Editing, Regulating and Targeting Genomes." Nature Biotechnology 32.4 (2014): 347-55. Nature. 2 Mar. 2014. Web. 7 Nov. 2014.
- Trafton, Anne. "Battling Superbugs." MIT News. N.p., 21 Sept. 2014. Web. 06 Nov. 2014.



MURJ

MURJ Spotlight: **Professor Douglas Lauffenburger** MIT Department of Biological Engineering

This issue's spotlight features Professor Douglas Lauffenburger, Ford Professor of Biological Engineering, Chemical Engineering, and Biology and the Head of the Department of Biological Engineering

BY PUWANAT SANGKHAPREECHA & AMELIA TRAINER

In the 1970s, people could hardly imagine the direction in which biology was headed. They did not know that biotechnology would soon create a revolution in biology. Some scientists with a passion for biology wanted to bring engineering into the field. Among those scientists was Douglas A. Lauffenburger, a chemical engineering graduate who became fascinated with the world of biology. With a love for mathematics and physics, he perceived and explored biology through the eyes of an engineer. He became a professor in the MIT Chemical Engineering Department in 1995, and now he is the Ford Professor of Engineering and the Head of the MIT Biological Engineering Department. Recently, MURJ talked with him about his career path and how he traversed the decades of revolution in bioengineering.

MURJ: "Why did you decide to initially study chemical engineering?"

Lauffenburger: In high school, I liked math, chemistry, and physics, and I did not like biology. When I graduated from high school in 1971, biology was much more descriptive and less mechanistic. In college, I wanted to major in something for which I could take chemistry, math, and physics, which were the fields I was most interested in. I had no idea what a chemical engineer did. I wanted to study chemical engineering because I thought I would enjoy the classes, and I did. However, the problem was that I had a summer job with chemical and oil industry companies. It was very good, but I realized I did not want to work on plastic, gasoline, and so forth for my career. I thought about law school because I was interested in government. Then, I took a biology class in cellular biophysics which -covered enzyme kinetics, membrane transport. metabolism, and signaling

-- and I was fascinated by that! I thought I could study cells using math, physics, and chemistry, so I decided that I wanted to apply engineering to cell biology. I went off to graduate school, but there were no bioengineering graduate programs that were interested in cells and molecules. They all focused on physiological imaging, prosthetics, and so forth. They had nothing to do with cells and molecules because, after all,



Professor Douglas Lauffenburger, Ford Professor of Biological Engineering, Chemical Engineering, and Biology & Head of the Department of Biological Engineering Credit: Lauffenburger Research Group archives

it was still in the 1970s prior to the biotech revolution. I got into a chemical engineering program where I could do research that dealt with cell biology. My advisor had a friend who was a microbiologist, and he said that we could study chemotaxis, a cell biology function of how white blood cells, blood vessel cells, or bacteria find the right direction to move. That was my Ph.D. thesis, and I have been interested in how to bring together

MURJ

Volume 28, Fall 2014



Currently, Douglas Lauffenburger's research focuses on pushing the bounds of therapy in complex diseases, particularly in the field of cancer therapeutics Credit: Lauffenburger Research Groum archives

engineering and biology ever since then.

MURJ: "So you started with your interest in math, chemistry, and physics, but you later applied those fields to biology."

Exactly. When I went to graduate school, what I took was mostly biology classes: biophysics, microbiology, immunology, cell biology, biochemistry. They had nothing to do with my chemical engineering Ph.D. curriculum. However, I was just fascinated by biology, and I wanted to learn everything I could. As I got my Ph.D., I thought this would be a fabulous thing to do: to understand biology better in terms of engineering.

MURJ: "When you said that you did not want to work with plastic and oil, why was that?"

That is a really good question. The technical questions were interesting enough. However, in terms of what I wanted to do with my life, I just did not think that solving problems and making advances in that area would bring me much excitement or use my talents most beneficially for the world.

MURJ: "At that time, did you foresee the biological revolution?"

Not at all. What is not true is that somehow I was prescient and could just imagine the future decades later. That was not true. All I knew was what I was really interested in. Having found cell biology, I just got really excited about it. When I was in graduate school between 1975 and 1979, molecular biology and biotechnology really took off. That was when recombinant DNA was happening, and that was when monoclonal antibodies were happening. All of these experimental tools that people could now manipulate biology with were coming to the fore. Also, in the early 80s, biotechnology companies like Genzyme, Biogen, and Amgen were founded, and people started using biotechnology in very important ways. So, I was in the right place at the right time. It was solely because of my interest,

not because somehow I foresaw the future.

Even in the early 1980s, people did not think there was a role for engineers in cell biology. Cell biologists said they didn't need engineers to help us understand biology. Engineers thought that, in building biotechnology, they could build bioreactors to make proteins from cells and separate proteins out of cells. However, they would leave the cell biology to the biologists. Because the biotechnology was just getting started, it was not clear then that the engineers needed to understand cells themselves. That came later.

MURJ: "With your great interests in math, physics, and chemistry, did you feel out of place in the biology world before biological engineers started becoming recognized?"

Absolutely. I had my own lab at the University of Pennsylvania with graduate students, and, when we would present posters at the American Society for Cell Biology meetings, almost everybody looked at the title and then walked past. We rarely got invited to give talks, and most people ignored our posters. I would also say the same thing for the chemical engineering world because, when I went to the American Institute of Chemical Engineers meetings, most of our talks were in the "Miscellaneous" sections, actually. I gave many talks where my audiences were only my fellow speakers. Sometimes, you may wonder why the world should care about your work, but, in the end, you just follow your own passions.

MURJ: "What is your research currently focused on?"

My current research is about improving therapy for complex diseases, especially cancer. I have an affiliation with the Koch Institute and do a lot of collaboration with folks there as well as clinicians. So, I work on a lot of cancer therapeutics. With cancer, there are a lot of mutations and metabolic pathways that are dysregulated. We

are figuring out the best pathways to aim drugs at. It is usually complicated, so that one drug with one target is not enough. A combination of therapies is actually needed. This is an engineering problem because it deals with multiple variables and pathways, feedbacks and feedforwards. It is so complicated that our intuition alone cannot handle predictions, computational analysis, and kinetics, and so forth come into play. It turns out to be perfect place for engineers to do research. Along with therapeutics for cancer, this is also the case for other complex diseases, such as immunological and inflammation-based diseases. For example, we work on HIV, diabetes. endometriosis, and colitis. In all of those, we look at multiple pathways and cell types, and we use systems biology to look at the complexity.

MURJ: "So bioengineers not only create tools used in clinics, but they also study



When Professor Lauffenburger began research, biological engineering was a nascent field. Now, its size and applications are growing at a progressively rapid pace. Credit: Lauffenburger Research Group archives

and understand how cells function though analytical approaches."

Yes, that is what makes biological engineering at MIT different. Most institutions have had biomedical engineering for decades, and that is mainly about engineers using physics- and chemistry-based devices for clinical applications for example, how to improve signal processing to get better images of organs or how to control drug delivery. Biological engineering at MIT is different because we do engineering based on biology, not physics or chemistry. For cancer therapeutics, we do experiments in genetics or molecular biology, and we can predict where a drug target might be. We are not building devices out of physics or chemistry: we are actually there inside a cell.

MURJ: "Because biological engineering is such a complex problem, how collaborative is the MIT biological engineering community?"

People are very collaborative within the department. Even in just our biological engineering department, we have experts in many different fields. Also, of course, we have a great deal of collaboration with the biology department, chemistry department, and other engineering departments. That is one thing that makes MIT such a great place. Undergraduate and graduate students and postdocs work as a team and may collaborate with other people from different departments.

MURJ: "What is your particular expertise in the biological engineering department?"

That is a very interesting question. I think our expertise is that we work on how to combine quantitative cell biology

Features

MURJ

Volume 28, Fall 2014

experiments with computational analysis. We are not solely a computational group, where some people do an experiment and give us their data, and we analyze it and see what insight we can give you back. Instead, our strength is

also how you may design and carry out the experiments in the first place to cover multiple variables so that the data you generate is more understandable.

We not only develop computational algorithms to analyze data, but we also design and do experiments to generate the data. This experiment - computation integration is our expertise.

MURJ: "What do you think the future of bioengineering will look like?"

What I would say is, all of engineering is trying to make technologies that are useful to society. That is really why engineers are educated and hired. The world has problems, and the world has opportunities. If you can make the right technology or the right product, you can address those problems and make people's lives better. However, you have to make those technologies out of something. If you're a computer scientist, you write software; you make a product using a certain branch of mathematics. If you're a mechanical engineer or an electrical engineer, you'll make products out of physics. If you're a nuclear engineer, you'll make products out of nuclear physics. If you're a chemical engineer or a materials scientist, vou'll make products out of organic chemistry and inorganic chemistry. All of these engineering disciplines have been around for decades, and

they all try to make technologies using branches of physics, math, and chemistry that are known to the world. Biological engineering is about something analogous. We are now educating people to make products and technologies

"We can do so much...and that's what's going to happen over the next few decades. Every realm of societal need is going to have products that have come from biology"

> that the world needs out of cells and biological molecules. We might be generating tissues, might be developing microbial ecosystems, might be evolving viruses and yeast that have certain properties.

> My point of going through all this is that I think, in the coming decades, biological engineering is going to become increasingly useful in making products that the world uses. Right now, the realm of biological technology might seem to apply to small niches. Maybe you make therapeutic antibodies or cell cultures, and maybe you do environmental remediation with some microbes. But we can do so much more than that. We can do so much more than all of that, and that's what's going to happen over the next few decades. Every realm of societal need is going to have products that have come from biology.

MURJ: "Do you think it's going to be a big revolution?"

Absolutely. I mean, it already is. However, it's still relatively small. You can see part of it in genetically modified organisms for agriculture, and we can do more there. For example, in the place of chemical fertilizers, we can figure out microbial routes for nitrogen fixation. With manufacturing, instead of making things with chemical reactions at high pressure and temperature, we see now that one can evolve viruses or even construct nucleic acids to be nanotechnology templates and structures that are useful.

> We can just do manufacturing in whole a different way. certainly. Also, medicine, reprogramming stem cells, and finding cancer drugs that are

more predictably useful – more can be done with all of these.

MURJ: "While commenting on the issues regarding putting biologically engineered products into society, you used genetically modified organisms as an example product for agricultural use. There is a lot of stigma attached to GMOs, and some people don't trust them. Can you comment on that?"

I'd say two things. Number one, the concern doesn't really match the problem because people have demonstrated for centuries that you can manipulate the genetics of plants and even animals to improve traits. People have done crop breeding and selected seeds that could bring out desired properties. That's already genetic engineering. Nowadays, it's just done in a different way - specifically, inserting genes and taking other genes out. Humankind has always manipulated agriculture genetically, just in a much less precise way.

Having said that, I think the crucial need for attention in biological engineering and really one of the important reasons that I think our educational mission is

Volume 28, Fall 2014

so important is any technology that's made needs to be predictive. You need to have some sense of the risk the risk-benefit ratio and cost-benefit ratio. These days, if you think of any other realm of society, products aren't made without computational models that will predict what's going to happen. You build an aircraft, an automobile engine, a car, a bridge, or a circuit – people don't put these things together randomly using trial and error and hope they work. They've discovered some of the laws governing what's going on - the physics and the chemistry - and turned it into equations that say, 'If I have these problems, then this will happen. Maybe I don't know these probabilities precisely, but here's a potential risk. I can calculate how bad it would be if that were to happen, how probable it could be, and what I could put into the system to design against it.' That's what's done in all realms of technology.

MURJ

It's this predictive capability of understanding how well you can predict that risk. That's what needs to be done for biotechnology, and that's what society needs to know.

MURJ: "This is just out of curiosity -- has there been a Nobel Laureate who is a bioengineer?"

Not yet. I would hope at some point in time there will be. Maybe in your lifetimes. I can imagine it happening because the problems being worked on are so hard and so important.

MURJ: "What kind of advice would you have for a biological engineering undergraduate student who is soon to enter the field?"

I would say that they have tremendous potential for doing a lot of good with their career. At the same time, what they are doing is so new; a lot of the world does not know what to make of it. When I started in this field thirty years ago, if I presented a poster and one person out of a hundred stopped and looked at it, that was a good thing. For the other 99, it did not look like the cell biology they were accustomed to. I would say our challenge is, when our students go out, most of the world is not going to recognize them. Some people will say, 'Well, you don't quite look like a biologist, but you don't look like a mechanical engineer or a chemical engineer or a physicist or a chemist. What is it that you can contribute to the world?' We're going to have to demonstrate what we can do to make people say, 'Oh, that's what biological engineers can contribute to the world alongside the other disciplines.' We're still in the very beginning, basic stages, and the world still needs to see demonstrations of what we can offer.



MURJ UROP Summaries

The Role of SIRT1 in Modulating Circadian Rhythms Through High-fat Diet

Jiapei Chen¹, Caitlin Ondracek², Leonard Guarente²

¹Student Contributor, Class of 2017, Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Department of Biology, MIT, Cambridge, MA 02139, USA

Introduction

Circadian rhythms are largely regulated by the suprachiasmatic nucleus (SCN) located in the hypothalamus. The SCN responds to light-dark (LD) cycles from the environment and generates an oscillation of 12-hour light and 12-hour dark cycles accordingly. This oscillation is auto-regulated by interacting transcriptional activators and repressors. Two central clock proteins are BMAL1 and CLOCK, which positively regulate the expression of target genes Cryptochrome (CRY) and Period (PER). When the protein levels of CRY and PER reach a threshold, they trigger a repression of BMAL1 and CLOCK, resulting in the self-controlled oscillation. The circadian clock is responsible for coordinating many biological processes, such as hormone secretion and body temperature maintenance. According to Bellet et al., 10-15% of all transcripts in different tissues display circadian oscillation. The disruption of circadian rhythm can lead to irregular cell metabolism and proliferation.

A recent study showed that circadian rhythms are also controlled by SIRT1 (Chang et al., June 20, 2013). SIRT1 regulates the circadian clock by activating the transcription of BMAL1 and alters the stability of PER2 through deacetylation. Aged wild-type (WT) mice have less SIRT1 in their SCN, and therefore less BMAL1 activity. This change in SIRT1 and other circadian regulators leads to a longer intrinsic circadian period, disrupted activity pattern, and a longer adjustment to changes in jet-lag type situations. Young brain-specific SIRT1 knockout (BSKO) mice show the same phenotypes as aged mice, whereas young mice overexpressing brain SIRT1 are more resistant to these age-associated behaviors.

It is also known that SIRT1 levels are affected by high-fat diet. Suter et al. demonstrated that *in utero*

exposure to maternal high-fat diet (MHFD) leads to decreased SIRT1 expression and activity in the offspring through observed epigenomic modifications.

Our goal is to determine whether SIRT1 is required for HFD-induced changes in the circadian rhythm. This will allow us to gain a deeper understanding of the role of SIRT1 in affecting the circadian rhythm and how high-fat diet, compounded with the influence of SIRT1, regulates the circadian clock.

Methods

We used sixteen C57BL/6 mice, half WT and half BSKO. Out of the eight WT and BSKO mice, four of each genotype were fed with regular chow (RC) and the remaining four were fed with high-fat diet (HFD) with 60% fat.

BSKO mice were generated by the Nestin/Cre method. Nestin is a brain-specific promoter that drives *cre* expression only in the brain, making it a good candidate to study SIRT1 genes in the hypothalamus. Mice with loxP sites flanking the SIRT1 gene were designated genotype SIRT1 (fl/fl). After Cre has attached to both loxP sites, the SIRT1 gene was enclosed in a loop and enzymatically cleaved from the original DNA strand. Using this process, we bred a Cre(+/-) mouse with a WT SIRT1(fl/fl) mouse and their offspring would have genotype SIRT1 (fl/fl) Cre (+/-), which was a BSKO. WT mice had genotype SIRT1(fl/fl) Cre(-/-).

To measure circadian rhythms, the mice ran on wheels and we recorded the daily activity as wheel revolutions were collected by a computer every 6 minutes. The mice were first acclimated to the running wheels under a 12-hour LD cycle. Then, they were switched to an all-dark cycle (DD). We analyzed the data collected in terms of the number of wheel revolutions per day and how long the active periods were. In addition, their weights were monitored throughout the experiment.

Results

Our results are represented in Figure 1. We found that HFD shortens the intrinsic period of WT mice whereas HFD lengthens the intrinsic period of BSKO mice. When on HFD, WT and BSKO mice have similar periods, suggesting that SIRT1 may not play a role in HFD-induced changes in circadian rhythms. In general, we also found that SIRT1 mice have shorter intrinsic periods. In terms of weight gain on HFD, BSKO mice gain significantly less than WT mice (Figure 2).

Period Length (tau) after 7 Weeks RC/HFD Chi-Square Method



Figure 1. The average circadian period length of each group of mice.



Weight Gain after 8 Weeks of RC/HFD

Average of 4 Mice/Group

Figure 2. The average weight development of each group of mice.

Therefore, we propose the following question: could SIRT1 in the brain play a role in weight gain from a HFD?

Acknowledgements

I'd like to thank my direct supervisor, Caitlin Ondracek, for her time and efforts in guiding me in my UROP work and helping me make this poster. I'd also like to thank Professor Guarente for letting me become involved in his lab and learn so much about research. Finally, I'd like to thank the UROP program at MIT for giving me the funds and the opportunity to explore this project.

References

- Chang, H. C., & Guarente, L. (2013). SIRT1 Mediates Central Circadian Control in the SCN by a Mechanism that Decays with Aging. Cell, 153, 1448-60.
- Suter M. A., Chen A., Burdine M. S., Choudhury M., Harris R. A., Lane R. H., ... Aagaard K. M. (2012).
 A maternal high-fat diet modulates fetal SIRT1 histone and protein deacetylase activity in nonhuman primates. The FASEB Journal, 12, 5106-14.
- Bellet M. M., & Sassone-Corsi P. (2010). Mammalian circadian clock and metabolism—the epigenetic link. Journal of Cell Science, 123, 3837-48.
- Musiek E. S., Lim M. M., Yang G., Bauer A. Q., Qi L., Lee Y., ... FitzGerald G. A. (2013). Circadian clock proteins regulate neuronal redox homeostasis and neurodegeneration. The Journal of Clinical Investigation, 123, 5389–5400.

How Integration Mechanisms Impact the Performance of the Adjoined Mental and Physical Health Care in the Clalit Health System

Kyle Yuan¹, Wiljeana Glover², John Carrol³

¹Student Contributor, Class of 2016, Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Sociotechnical Systems Research Center, MIT, Cambridge, MA 02139, USA.

³Sloan School of Management, MIT, Cambridge, MA 02139, USA.

Background

On May 1, 2012, the Socio-Economic Cabinet of the Israeli Government approved a reform in the mental health services in Israel, which transfers the responsibility for providing mental health services from the Israeli Ministry of Health to the four non-governmental Health Maintenance Organizations (HMO) in the nation (Nirel et al., 2014). The reform bill establishes the right to receive mental healthcare, determines the basic package of services, defines who is responsible for providing services, and aims to integrate physical and mental health services in Israel by 2015 (Samuel et al., 2013). There are political, organizational, and professional challenges to such a reform and our research aims to outline strategies to improve access, quality, and efficiency of the new integrated healthcare system. (Gröne et al., 2002).

In light of the challenges posed by enacting this reform, the Israel Institute of Technology and the Massachusetts Institute of Technology partnered with Clalit Health Services, the largest HMO in Israel. Our proposed collaborative research project aimed to (1) describe the current state of mental health operations in the Clalit Health System and the Israel Ministry of Health and (2), identify best practices of integrated operations based on a different medical condition, diabetes mellitus.

Methods

Data collection took place with the participation of 66 individuals within clinics of a major Israeli health maintenance organization, Clalit Health Services (CHS). Over a period of 2 months, we conducted 8 meetings with Israeli physicians, healthcare officials, and other clinic workers. Through collaborative efforts with CHS, more than 60 individuals were interviewed through semistructured focus group meetings. We developed interview guides to gauge current state of diabetes & mental health management, current interactions between primary & secondary, and future state interactions between primary & secondary care. An abridged portion of the interview guide for a primary care clinic may be seen in Figure 1.

Responses in both Hebrew and English were recorded, translated, and transcribed into field notes from all of the meetings. We compiled the qualitative data from all 8 interview sessions and performed analysis with the assistance of MAXQDA. We analyzed the interviews by focusing on five management categories: Organizational, Operational, Technological, Knowledge-Oriented, and Social. Through this methodology, we developed a quantitative measure of satisfaction in diabetes and mental health care, as outlined by Figure 2.

Current State- Diabetes Care within Primary Care

Please describe how diabetes care is usually handled in your clinic (e.g., provide medications, referrals, etc.)

Current Interactions between Primary Care and Diabetes Clinics

What kinds of communication occur between diabetes centers and primary care clinics, and what is exchanged? (e.g., formal or informal interactions of patient information)

Future State- Interactions between Primary Care and Diabetes centers

In an ideal case, how should the interactions between primary care and diabetes centers work?



Conclusion

The focus group meetings provided valuable insight into espoused and enacted policies in both diabetes and mental health care within the HMO and throughout the nation. There are significant differences between how diabetes and mental health care is currently managed, most notably in terms of information flow, patient follow-up, and continuing physician education. Overall, physicians and healthcare workers are much more satisfied with the state of diabetes care than mental health care. Physicians are taught how to deal with "softer" psychiatric cases, but currently, there is suboptimal emphasis on education for primary care staff in mental health management. The best practices we identified for the upcoming healthcare reform include (1) standardize diagnosis procedures in mental health, (2) improve access to psychiatrists in primary care clinics, (3) improve communication between primary care physicians and specialists, and (4) develop continuing education programs for mental health workers. The data and conclusions collected will be instrumental in guiding the policies that will have significant impact on the mental health programs available to Israelis in 2015.

Figure 1. Abridged Semi-Structured Interview Guide. Depending on the attendants of the particular meetings, we developed interview guides to gauge current and future states of care. This particular interview guide was used for a primary care clinic.

Figure 2. Satisfaction Levels in Diabetes and Mental Health Management. 37 broad coding themes were developed and defined by our research team. These themes were split into 5 management categories: Organizational, Operational, Technological, Knowledge-Oriented, and Social. To gauge the level of satisfaction within diabetes and mental health treatment, we noted the number of positive codes in each category.

References

- Aviram U, Rosen H. (1998). Policy and services in the field of mental health in Israel – the planned reform in light of the national health insurance law. Hevra veRevaha [Society and Welfare], 18, 161–189.
- Gröne, O & Garcia-Barbero. (2002). Trends in Integrated Care – Reflections on Conceptual Issues. World Health Organization, Copenhagen, 1, 79-92.
- Nirel and Samuel. (2014). Work practices and the provision of mental-health care on the verge of reform: a national survey of Israeli psychiatrists and psychologists. Israel Journal of Health Policy Research, 3, 25.
- Samuel H, Rosen B., (2013). How are the Health Plans Preparing to Provide Mental Health Services? Jerusalem: Myers-JDC-Brookdale Institute, 636, 13.

Determining the Effectiveness of a Carbonaceous Iron/Sand (CarFeSand) Column Filter System in Removing Heavy Metals from Polluted Water

Justin Yuan¹, Evangeline Ng², Reginald Thio²

¹Student Contributor, Class of 2016, Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Department of Chemical Engineering, Engineering Product Development, Singapore University of Technology and Design, Singapore

Background

According to a study conducted by the World Health Organization (WHO), millions of people die each year from water-related illnesses or sanitation and hygienic problems due to lack of availability of water purification techniques and equipment (Prüss-Üstün et al., 2008). The objective of this project is to design and build an innovative and sustainable carbonaceous iron (FeO)/ sand (CarFeSand) filter system that demonstrates simple, yet effective, on-demand point-of-use water purification. Current leading companies in portable household water filtration, such as Brita, utilize silver-impregnated activated carbon as the main sorbent. However, silver can leach into the collected water and be toxic to human health.

Our CarFeSand filter is adapted from a centuriesold chemical remediation concept: water slowly percolates through layers of different filter media which efficiently separates harmful pathogens, hydrophobic organic contaminants (HOCs), and heavy metals from contaminated water via entrapment and adsorption onto the filter substrate, while avoiding unwanted leaching of byproducts back into the water (Chung et al., 2008). Widening the accessibility to safe drinking water will improve the quality of life for the communities of underdeveloped countries in need.

Materials and Methods

This project focuses on the removal of lead(II) (Pb^{2+}) ions from two different types of water—deionized (DI) water and artificial groundwater (AG)—by running it through a modular filtration system employing three different types of sorbents—fine charcoal, iron filings, and fine sand—to simulate environmental conditions (Zhao et al., 2009). By placing 40 g of iron filings in a PET bottle on top, followed by a PET bottle of 175 g of fine charcoal (2.8 mm diameter) in the middle, and completed with a PET bottle of 95 g of fine sand at the bottom, polluted water is forced to slowly percolate through these different layers of media (Figure 1).



Figure 1. Three PET bottles each filled with the pre-determined optimal mass of iron filings, fine charcoal, and fine sand are mounted on a ring stand.

Results and Discussion

The average sorption of lead ions into the filter substrates and out of the test water ranged from ~72%-99%, with average flow rates ranging from 0.57 mL/sec to 1.10 mL/sec. Our household-scaled CarFeSand filter shows significant efficacy in heavy metal removal from polluted water, economical practicality in the future consumer market, and will serve as an attractive, small adaptation of the traditional large, slow sand filters, such that they can be uniquely operated under rugged conditions. We subsequently see a decrease in heavy metal concentration due to prolonged periods of high surface area contact with these sorbents. Lead serves as a model for heavy metals and what interactions with other similar contaminants might result in. The final concentration of lead after the 0.5 ppm and 1.0 ppm tests falls within the WHO safety guidelines for drinking water, and the 2.0 ppm test only exceeds this guideline by 0.02 ppm (WHO, 2011). Furthermore, at an average flow rate of 1.66 mL/sec, it will only take approximately 10 minutes to obtain 1 liter of clean drinking water, a rate fast enough to sustain a family in need (WHO, 2011).

Conclusions

Ultimately, the beauty of the innovative CarFeSand filter lies in its modular design. Since each sorbent reaches saturation point at different periods of time, previous models of tightly packed substrate layer filtration systems were wasteful to discard each time any individual layer becomes saturated. Having modules in a column that contain different types of sorbent has proven to be time efficient and environmentally friendly. Instead of throwing away an entire filtration system, each time a sorbent become saturated, it is simple to remove that module and replace it with fresh sorbent. The future implications of a modular design can lead to beneficial overall water quality improvement for the general public. We hope that the results of this research will play an integral role and shed new light for future development of water purification techniques that better the many lives of those who lack access to a clean water supply.

References

- Prüss-Üstün, A., Bos, R., Gore, F., Bartram, J. (2008). Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health. World Health Organization, Geneva, Switzerland.
- Chung, S., Gandee, J., Suidan, M. T., & Venosa, A.D. (2008). Effect of sand particle size, oil contamination, and water table level on the effectiveness of sorbents in wicking oil from contaminated wetland, Proc. 2008 International Oil Spill Confer-

ence. Washington, D.C.: American Petroleum Institute.

- Zhao, D.H., Gao, H.W. (2009). Turning Calcium Carbonate into a Cost-effective Wastewater-sorbing Material by Occluding Waste Dye. Environ Sci Pollut Res. 17, 97–105.
- World Health Organization. (2011). Lead in Drinking-water – Background Document for Development of WHO Guidelines for Drinking-water Quality. WHO/SDE/WSH/03.04/09/Rev/1. World Health Organization, Geneva, Switzerland.
- World Health Organization. (2011). Guidelines for Drinking-water Quality. World Health Organization, Geneva, Switzerland.

mRNA Delivery and Expression Using a Vectorfree Microfluidic Intracellular Delivery Platform

Derek Jang¹, Camilo Ruiz², Tushar Kamath², George Hartoularos³, Jasdave Chahal⁴, Armon Sharei⁵, Klavs Jensen⁵, Robert Langer⁵

¹Student Contributor, Class of 2015, Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Student Contributor, Class of 2016, Department of Biological Engineering, MIT, Cambridge, MA 02139, USA

³Student Contributor, Class of 2015, Department of Chemical Engineering, MIT, Cambridge, MA 02139, USA

⁴Department of Biology, MIT, Cambridge, MA 02139, USA

⁵Department of Chemical Engineering, MIT, Cambridge, MA 02139, USA

Background

Current cancer treatment, such as chemotherapy, is very damaging to the patient. A growing trend in cancer treatment is to utilize the patient's own immune system to kill the tumor cells. In order to do so, the immune system must be able to remain active in the immunosuppressive environment that tumor cells create (Zou, 2005). We could also enhance the immune cells' ability to home to specific locations, such as tumors, for targeted therapies. This may involve delivering the right proteins to the immune cells, but proteins are transient and would require regular doses to keep the immune system activated. One possible solution is to encode the proteins in self-amplifying RNA (saRNA), which is mRNA that encodes RNA-dependent RNA-polymerase, in order to steadily activate an immune response.

One of the greatest obstacles to clinical application of nucleic acids is successful delivery of the nucleic acids to the target cells. Although there are currently existing delivery methods, they have unique disadvantages that limit their efficacy. Chemically modifying a deliverable to better facilitate cellular uptake, such as making it resemble substrates of active transport, requires highly specific modifications for each deliverable. Lipid vectors, which can carry deliverables and fuse with the cell membrane to release the deliverables into the cytosol, have low expression in human primary cells such as immune cells. Part of this may be explained by the deliverables being trapped in endosomes when endocytosed, hiding it from transcription or translation mechanisms. Viral vectors, which hijack a virus's ability to inject and express genetic material into a host cell, feature an inherent risk of unwanted genetic integration (Kim, 2010). Temporal membrane disruptive techniques, such as microinjection and electroporation, feature broader application because they physically move deliverables into cells by disrupting the membrane using forces. However, microinjection suffers from low deliver efficiency, and electroporation suffers in cell viability (Karra, 2010).

CellSqueeze

CellSqueeze is a vector-free microfluidic delivery platform (Figure 1).



Figure 1. A CellSqueeze chip.

By passing cells through channels smaller than their diameters, cells are mechanically deformed and induced to produce temporal membrane disruptions, allowing materials to diffuse into the cell. By optimizing delivery conditions such as channel length, diameter, and number of constrictions, as well as delivery medium

MURJ

and temperature, this device has been successfully used to deliver a wide range of materials, such as proteins, nucleic acids, and quantum dots, to a wide range of cells, from HeLa cells to T-cells, while maintaining high cell viabilities (Sharei, 2012).

Methods

mRNA Delivery and Expression

For each delivery, we co-delivered 3000 MW labeled dextran dye by Life Technologies with EGFP mRNA by TriLink, both at 0.1 mg/mL. Dextran dve was used to confirm successful membrane disruption because it easily diffuses into temporal membrane holes. Deliveries were done using chip and pressure conditions from previous experiments that optimized deliveries in the specific cell line. For example, HeLa cells were delivered using chips with constriction length of 10 µm and constriction widths of 6 or 7 µm, and at delivery pressures of 50 and 70 psi. For human T-cells, we used chips with constriction length of 10 or 30 µm and constriction width of 4 µm. and at delivery pressures of 80 and 120 psi. Each delivery cocktail was loaded into one reservoir of the device, pushed through the chip by a connected pressure system, and extracted from the other reservoir (Figure 2).

To assess CellSqueeze's nucleic acid delivery capabilities, we first delivered modified mRNA encoding GFP, known to be successfully delivered and expressed by traditional lipofection methods, into baby hamster kidney (BHK) cells, a cell line favored for transfections due to its



Figure 3. BHK cells expressing GFP encoded in modified mRNA.

limited innate defenses against foreign genomic material. We achieved comparable, and sometimes better, delivery efficiencies and expression (Figure 3) using CellSqueeze by optimizing delivery conditions. We then delivered the mRNA into HeLa cells, a robust human cell line also known to be transfected by traditional transfection methods. We again achieved comparable, and sometimes better, delivery efficiencies and expression by optimizing delivery conditions. Finally, we delivered the mRNA into human T-cells. We are currently in the process of



Figure 2. A diagram of parts of the delivery platform.

Figure 4. FACS analysis of mRNA delivery in human T-cells. The Pacific Blue channel monitors 3000 MW Dextran dye delivery as a measure of delivery efficiency, while the FITC channel measures GFP expression.

optimizing delivery conditions, but have seen over 30% GFP expression in T cells (Figure 4).

Discussion

Starting in the fall of 2013, we began using CellSqueeze to deliver mRNA into a range of cells, starting with the most basic BHK cells to the most clinically relevant human T-cells. We have been able to systematically optimize the delivery conditions, thanks to the modularity of the platform, in order to achieve comparable, if not better, mRNA delivery and GFP expression. So far, we have been delivering modified mRNA without any auxiliary materials to aid in delivery and expression, therefore the addition of drugs and substances to the delivery cocktail may boost delivery and expression.

Our next steps after successful optimization of our current expression of modified mRNA encoding GFP would be to test delivery of saRNA into human T-cells. If we can achieve a sustained GFP expression in the cells, we will move onto delivering saRNA encoding a desired protein that can keep T-cells activated in the presence of a tumor's immunosuppressive environment and evaluate the efficacy of this cancer treatment. Finally, we may evaluate delivery and expression of unmodified mRNA in order to understand the role of these modifications.

CellSqueeze's success in delivering and expressing nucleic acids will open many research and clinical doors. Its robustness against delivery material and cell type is a proven advantage over many traditional transfection methods, such as chemical modification and lipid vectors, which can be very specific for the delivery material and cell type. Its ease of use, low cost, competitive delivery and expression, and gentler treatment than other temporal membrane disruptive delivery techniques such as electroporation and microinjection make it friendlier to widespread adoption in academia, industry, and healthcare.

References

- Zou, Weiping. (2005). Immunosuppressive networks in the tumour environment and their therapeutic relevance. Nature Reviews Cancer, 5, 263-274.
- Kim, Tae Kyung, Eberwine, James H. (2010). Mammalian cell transfection: the present and the future. Analytical and Bioanalytical Chemistry, 8, 3173-3178.
- Karra, Daniela, Dahm, Ralf. (2010) Transfection Techniques for Neuronal Cells. The Journal of Neuroscience, 30, 6171-6177.
- Sharei, Armon, et al. (2012). A vector-free microfluidic platform for intracellular delivery. Proceedings of the National Academy of Sciences of the United States of America, 110, 2082-2087.

MURJ Reports

A Reimplementation of Python's Dictionary Using Simple Tabulation Hashing and Linear Probing

Pedram Razavi¹, Thomas Georgiou¹, Russell Cohen², Erik Demaine³

¹Student Contributor, Class of 2014, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Student Contributor, Class of 2013, Department of Electrical Engineering and Computer Science, MIT Cambridge, MA 02139, USA

³Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA 02139, USA

Here we analyze the performance improvements obtained by reimplementing the standard Python's dictionary data structure by replacing the built-in hash function for strings with one based on tabulation hashing, and its collision resolution strategy with linear probing. These choices are motivated by the recent theoretical results of Pătrașcu and Thorup (Pătrașcu, 2011) who showed that in this hashing scheme operations take constant expected time. Our new implementation of Python's dictionary attains significant performance gains on synthetic benchmarks and achieves an average of 4% improvement in execution time on a range of real-world programs. If the extra memory bandwidth and latency required to compute the tabulation hash can be further mitigated, our new implementation stands as a performant replacement to the one currently used.

1. Introduction

Python¹ (van Rossum, 1995) is a general-purpose, high-level programming language, and has been widely adopted because of its emphasis on code readability, and near-pseudocode syntax. With the advent of packages such as NumPy, SciPy, and Pandas, Python has begun to be extensively used (with questionable merit) in performance programming. However, since Python is an interpreted language, Python programs still lag behind their equivalents written in a compiled language, such as C, in terms of running time due to interpreter overhead.

1.1 Motivation

Python dictionary is a built-in associative array which is used to store and access (key, value) pairs. Python, as most interpreted languages, depends crucially on the efficiency of its internal dictionaries (hash tables), in particular when the keys are string-typed, because all the names in the code, for instance those of variables, classes, and methods, are all resolved at runtime. For instance, a dictionary can be instantiated to pass keyword arguments to a function. This means a dictionary could potentially be created and destroyed on every function call. Some other examples of internal usage of dictionaries in Python are: class method lookup, instance attribute lookup and global variables lookup (Python built-ins module), and membership testing.

This extensive internal use of dictionaries means that a running Python program has many dictionaries active simultaneously, even in cases where the user's code does not explicitly instantiate a dictionary. Take for example the following code:

¹ Python refers to the programming language and its technical specification. Based on context, Python can also refer to the official and most widely used implementation of this language, also referred to as CPython, distributed by the Python Software Foundation.

```
>> i = 2
>> b = i * 2
>> b
4
```

We instrumented Python's source code to gather various statistics on dictionary usage (e.g. number of collisions, probes, accesses, etc.) for a given invocation of the Python binary. Using this method, we found that official Python 2.7.3 interpreter performs 92 lookups in string dictionaries and 24 lookups in general purpose dictionaries to run this code snippet. This total of 116 dictionary hits is significant considering that this 3-line code never instantiate a dictionary explicitly. Hence, improvements in the hash function for strings entails performance improvements in every application, not just those that make heavy use of dictionaries. This observation and recent theoretical results in the field of hash tables motivates this work.

2. Current Python's Dictionary Implementation

In this section, we give a brief overview of the key implementation details of the dictionary data structure in CPython 2.7.3, the reference and most-widely used Python interpreter which is implemented in C.

Because string-only keys are dominantly used in various programs and are also extensively used to implement the internal language features, CPython takes several measures to optimize for the case when keys in a dictionary are exclusively strings. When a dictionary is instantiated, CPvthon defaults on using a custom string-only lookup function until a search for non-string is requested. In the latter case, CPython falls back to a more general lookup method and uses this general function onward. Two main optimizations are present in the string-only lookup function. First, since the string comparisons do not raise any exceptions by design, some of the error checking logic is bypassed. Second, the stringonly lookup function skips the rich set of comparisons ($\langle =, \rangle =, \langle , \rangle, ==, ! = \rangle$ which are provided for arbitrary data types but are not needed for string data type.

The current CPython dictionary implementation uses an iterative XORed polynomial terms for calculating the hashvalue of a string. In this implementation the hash value for sequential strings differs only in the least significant bits, for instance: hash("1001") = 659473584131336958and hash("1002") = 659473584131336957. This behavior is suitable for the common case of sequential strings. When the table size is 2^i taking the least significant *i* bits as the index can populate the table without collisions. Hence, this approach outperforms a hash function with more random output in this case, and is also simpler and faster to compute. On the other hand, this hashing strategy has the tendency to fill contiguous slots of the hash table. CPython tries to remedy this issue by utilizing a reliable collision resolution strategy of open addressing combined with a custom non-linear probing mechanism. In this scheme, CPython uses quadratic probing based on this function: $j = (5j+1) \mod 2^i$ where $0 < j < 2^i-1$. However, this recurrence alone is not sufficient to achieve a faster runtime. For a set of keys, this strategy behaves the same as linear-probing because it scans the table entries in a fixed order. To overcome this problem, CPython also uses the most significant bits of the hash value by using the following pseudocode snippet:

```
slot = hash; perturb = hash;
while (<slot is full> AND <item in slot
does not equal the key>) {
    slot = (5 * slot) + 1 + perturb;
    // Shift right by 5 bits
    perturb = (perturb >> 5);
}
```

Based on this new method the probe sequence will depend on nearly all the bits of a hash value. We should note that the value of perturb shift (currently set to 5) is crucial for a performant probing behavior. It should be small enough so that all the significant bits have an effect on the probe sequence but at the same time, it should be large enough so that in collision-prone cases the highorder hash bits can affect the earlier iterations.

CPython initializes its table with 8 slots. The resizing strategy is based on maintaining a load factor invariant. Whenever there are n keys in the dictionary, there must be at least 3n/2 slots in the table. This load factor ensures that the table remains sparse and puts a bound on the number of potential collisions. CPython resizes a table whenever the load factor invariant gets violated. It quadruples the size when there are less than 50,000 keys in the table and doubles the size otherwise.

Overall, the design of CPython dictionary is simple and is optimized for more common use cases such as retrieval of sequential strings. Because it utilizes quadratic probing with perturbation, even if a data type does not implement a suitable hash function, the table is still populated with short chain lengths on average. However, there is room for improvements and optimizations. For instance, the current approach incurs suboptimal cache locality because probing a single slot might require jumping to the other slots which are not its immediate neighbors and consequently not necessarily stored in the cache.

3. Improvements

In this section we first study the theoretical motivation and justification for our implementation and later we look at some of the details of our implementation.

3.1 Theoretical Results

Knuth (Knuth, 1963) pioneered runtime analysis of hash tables by studying linear probing in 1963. Later, Wegman and Carter (Wegman, 1979) proposed different hashing functions including simple tabulation hashing. In simple tabulation hashing scheme, a key *x* is thought of as a vector of c characters: x_1, \dots, x_c . Initially, T_i tables are populated with random values and the final hash value of key x is calculated by $h(x) = T_1[x_1] \text{ XOR } T_2[x_2] \dots \text{ XOR } T_2[x_n]$ where XOR is the XOR operation. Recently, Pătrașcu and Thorup (Pătrașcu 2011) showed that simple tabulation hashing provides unexpectedly strong guarantees and is constant competitive with a truly random hash function for hashing when complemented with linear probing. They proved that for simple tabulation hashing with linear probing if the table size is $m = (1 + \varepsilon)n$, then the expected time per operation is $O(1/\varepsilon^2)$. This new guarantee asymptotically matches the bound for a truly random hashing function.

3.2 Implementation Details

Pătrașcu and Thorup (Pătrașcu, 2011) described simple tabulation hashing for fixed key lengths. For implementing this new hashing scheme, first we needed to figure out a variant of simple tabulation hashing for variable length strings that does not require a table for every possible index. Performing table lookups is potentially an expensive operation with many cache hits that can negate the superior cache behavior of linear probing. For the rest of this section, we explain the implementation details of our approach to these problems.

We modified the source code of CPython 2.7.3 to implement simple tabulation hashing with linear probing. We use a fixed number of tabulation tables and cycle between them for calculating the hash value of variable length strings. The method does not provide the same level of theoretical guarantees for strings with length longer than a set number of tables but empirically this approach performed well. We prepopulate a set of random tables in memory with random bits from random. org, which sources randomness from atmospheric noise in order to remove the reliance on pseudo-random number generators. Following is an excerpt from our hashing code. We use various bit tricks in order to prevent costly modulo instructions. For each byte in the string, we construct an index by adding the byte value and a table offset together. Then, we lookup the value in a large table of random 64 bit longs and then XOR it with the previous result in the chain of XORs.

while (--len >= 0) {

We also experimented with indexing using every pair of bytes in the string, treating them as 16-bit integers and using 2^{16} tabulation table entries but the results were worse due to increased memory thrashing.

On the engineering side, linear probing's caching behavior is superior in comparison to the probing scheme currently implemented in CPython. Consider an Intel Core 2 Duo processor with a 64-byte cache line. Each Python dictionary entry is 12 bytes so when Python probes a dictionary slot, it pulls in at least 4 more dictionary entries with it. However, when the current probing scheme probes another slot, it is not in the current cache line so it has to pull another cache line in. With linear probing, that next slot is most likely in the current cache line.

3.3 Optimizations

After implementation and profiling, we found that dictionary table lookups had over twice as many level 1 cache misses with tabulation hashing than without it. Hypothesizing that the cache misses were due to the tabulation tables not being in processor cache (16KB tables, 32KB L1 data, 3MB L2), we inserted x86 memory prefetch instructions into our hash code to tell the CPU's caching system that it should prioritize storing the tabulation tables in cache. We used GCC's builtin prefetch (void addr, char rw, char locality) to generate these instructions. However, this only yielded about a 1% performance improvement in synthetic benchmarks, negligible compared to the performance improvement of different hash schemes.

4. Results

In order to analyze the result of this new dictionary, we measured three main metrics. First, we compared the performance of the simple tabulation hashing with Python's current hashing scheme on random and sequential strings. Second, we measured the number of probes and collisions and overall the quality of our dictionary on structured and unstructured inputs. Finally, in this section we show some of the total time spent in dictionary operations by using a set of general Python benchmarks and some real-world programs. We ran the following on an Intel Core 2 Duo P8400 processor (3M cache, 2.26GHz) w/ 4GB of 1066MHz DDR3 RAM.



Figure 1. Eight tabulation tables with 1 byte indices provides the best performance for a dictionary with 6 million keys.

4.1 Table Number Tuning

For simple tabulation hashing with variable length strings, we decided to use a fixed number of N tables and round robin between them for every byte. We found the optimal value of N to be 8 empirically. We limited our choices to powers of 2 so that costly modulo operations could be avoided by using bit masks Figure 1 shows the performance of different numbers of tabulation tables with both 1 and 2 byte indices.

Eight hash tables indexing every byte separately resulted in the best performance. We believe this is due to the low entropy caused by fewer tables and too much memory bandwidth used by a higher number of tables. The optimal result may also be an artifact of the fact that we were using on the order of magnitude of thousands to millions of dictionary keys. Our optimal result resulted using 256 64-bit table entries in each of 8 tables using 16KB of memory.

4.2 Synthetic Benchmarks

To evaluate our implementation, we used Hash-tableshootout, a set of synthetic benchmarks which call the CPython dictionary code directly and hence bypass the interpreter overhead. We measured performance for both sequential strings, generated by converting sequential integers into their string representation, and for random strings, generated by the string representation of random integers. For each of the benchmarks, we inserted a given number of keys into the dictionary and measured the runtime. Figure 2 shows the performance for random and sequential strings side-by-side.



Figure 2. Random and sequential string performance comparison.

Both linear probing and simple tabulation result in improved performance over the current CPython dictionary implementation but combining both techniques results in over a 20% improvement. Figure 3 shows the average chain length, or dictionary cell probes per lookup for a run of the benchmark on sequential strings. Simple tabulation hashing with CPython's current collision resolution strategy results in the fewest number of probes. However, despite having the fewest number of probes, its real relative performance is worse as seen in Figure 3.

Linear probing performs faster than its average chain length suggests, showing the impact of looking at entries in the same cache line first rather than jumping around.



Figure 3. Simple tabulation hashing decreases the average chain length.

4.3 Real-world Benchmarks

Due to the heavy utilization of dictionaries in Python internals, it is valuable to see how dictionaries perform in a range of real-world programs. We used PyBench benchmark suite for this task. We saw significant improvements of around 20% in performance of some of the internals such as the special and normal class attribute lookups and string predicates. However, at the same time some of the other internals such as CPython exception handler were running around 12% slower compared to the current implementation. On average a 4.5% improvement in runtime was measured for the whole benchmark suite. For a more realistic view of the performance we also looked at a range of real-world programs.

The new implementation was tested on several standard real-world programs such as Tornado web server, Django web framework, html5lib and ten other real-world programs. For instance we wanted to measure if this new implementation can handle more requests compared to the current CPython implementation given a web server program. Different request rates (up to 2000 requests per second) were tested on a sample Tornado server and their corresponding response rates were recorded using httperf, a HTTP performance measurement tool. Considering the variance of request rates on different iterations of the test, we could not find any significant improvement or degradation of the response rate of the new implementation. However, the new implementation showed a smaller standard deviation in the response rates. This finding as well as those by other benchmarks showed that the new implementation gives more consistent runtime on different iterations of a same code compared to the current CPython implementation. This consistency is a desired behavior and we suspect that it is the result of more randomness in hash values for simple tabulation hashing and the smaller average chain lengths. The side-by-side performance comparison of more than ten applications is indicated in Figure 4. These results were consistent with the previous results of PyBench that showed about 4.5% improvement.

5. Discussion and Future Work

Simple tabulation yields faster and more consistent hash performance than Python's current implementation despite the increased number of memory accesses needed to compute the simple tabulation hash. Our results show that performance improvements of up to 19% are realizable in certain benchmarks with heavy lookup workloads such as class attribute lookups. Overall we measured 4% run time improvement on average for a range of real-world programs.



Figure 4. Percentage of overall improvements in some realworld Python applications. In a couple of the programs such as pure Python Pickle (an object serialization library) and html5lib (specification parser library) we calculated an improvement of 3% to 4%. In other cases we did not calculate a speedup but at the same time we did not find any real-world case which the new implementation performs significantly slower.

In terms of future work, we have 2 avenues of attack. First, we could use different dictionaries for different purposes. Many of the Python dictionaries are small, with a preset number of keys. New built-in functions are not added during the runtime of the program and thus they could use perfect hashing to improve lookups. Second, we could improve the performance bottlenecks illustrated by benchmarks. These mainly consist of the extra time and memory bandwidth and latency costs for computing simple tabulation hashing. If we optimize the hashing further through more profiling, memory layout, and instruction order optimizing, we may be able to achieve greater realworld performance improvements.

Given the strong performance of simple tabulation hashing and the only downside being the cost of memory lookups hampering performance, processor manufacturers should consider adding random tables into hardware as a ROM. We found that we only needed 16KB of randomness to achieve our 4% overall performance gains for Python. This would be a negligible cost in ROM given that processors have sizable higher-level caches nowadays. An x86 instruction set extension providing access to random tables would benefit numerous applications that use dictionaries as a core data structure. The changes needed to the Python interpreter are minimal, basically only changing a handful of lines of code in the hash functions of the "str" and "unicode" objects. In addition, if the tables are randomized at interpreter startup, the potential for malicious attacks against dictionary performance could be remedied.

References

- Carter, Lawrence; Wegman, Mark. (1979). Universal classes of hash functions. Journal of Computer and System Sciences, 18 (2), 143–154.
- Knuth, Donald. (1963). Notes on open addressing. Unpublished memorandum. See citeseer.ist.psu.edu/ knuth63notes.html.
- Patrascu, Mihai; Thorup, Mikkel. (2011). The power of simple tabulation hashing. Proceedings of the 43rd annual ACM symposium on Theory of computing, 1-10.
- van Rossum, Guido. (1995). Python Tutorial, CWI Technical Report CS-R9526.

Inferring Social and Linguistic Influences on Memory from Facebook Posts: An Exploratory Analysis

Brianna Jones¹, Rebecca Saxe²

¹Student Contributor, Class of 2015, Department of Brain and Cognitive Sciences, Department of Science, Technology and Society Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

²Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA

Why are some things easier to remember than others? To answer this question, we tested the short-term memorability of a variety of Facebook posts and book sentences by presenting participants with a large number of items then immediately testing their recognition memory. The stimuli were coded for several social and linguistic features and we examined whether any of these factors could predict how likely a sentences was to be remembered. We found Facebook posts were far more memorable than sentences from books. Moreover, sentences were variably memorable and memorability is consistent across people. Odd typography, indexical (i.e. a word whose meaning is dependent on the context in which it is used, such as "here" or "next Tuesday") count and sentence subject influenced the memorability of sentences. While these are the results of exploratory analysis, they suggest that our memory systems may be especially attuned to certain linguistic and social features of language.

Introduction

In a world where we are constantly bombarded with sensory stimulation, people have a remarkable capacity to retain massive amounts of information in long-term memory. While people's memory is very good overall, some kinds of information are more likely to be remembered than others. What accounts for these differences? While this question has been explored for certain kinds of sensory input, surprisingly little attention has been paid to memory for language, a potentially rich source of clues to what might make a thing more or less memorable in general. For example, language is especially useful for studying not only linguistic but also social factors that influence memorability.

In general, people have a remarkable ability to retain massive amounts of information in long-term memory even after brief exposure. In one study, participants were presented with 2,500 pictures for three seconds each. Surprisingly, recall rate reached as high as 92% (Brandy, 2008). People also have a remarkably good verbatim memory for sentences that are read out loud. In one study, participants heard texts that were 300 words long. After a period of six days, participants were asked to describe the scenes. Even after this delay of several days, participants were able to recognize and reproduce full sentences they had heard only once at above chance rates (Gurevich, 2010).

While people's memory is very good overall, some kinds of information are more likely to be remembered than others. An example of this is survival-relevant information. In one study, participants read either a "survival" story or a carefully matched control; they were then asked to rate common nouns for their survivalrelevance in the survival condition, and for pleasantness in the control condition. When they were later surprised with a retention test, participants recalled 49% of the words that were rated for survival-relevance, compared to only 36% of the words that were rated for pleasantness. Since social information is important for both survival and reproduction, it would not be surprising if it held a privileged place in memory (Kang, 2008).

Electrophysiology studies suggest that information conveyed through informal language can be easier to remember than that conveyed through formal language. One mark of informal communication is disfluency, illustrated for example by pauses in natural speech. In a study examining Event-Related Potentials (ERPs), also known as time-locked EEG signals, researchers compared participants' response to both predictable words (He likes long walks on the **beach**) and unpredictable words (I study in the toaster). An N400 effect was established for unpredictable words as compared to predictable words, meaning that "beach" in the above sentence, for example, would be more easily integrated into its context than "toaster" would be integrated into its context. This effect was reduced when the target words were preceded by a hesitation marked by the word "er," suggesting that "er" carried semantic meaning. Participants were also better able to remember words following this disfluency when later tested for recognition and recall. Both of these findings suggest that these words were processed differently as a result of hesitation, and that linguistic differences between formal and informal language can affect memory (Corley, 2007).

In sum, people have the capacity to retain massive amounts of information in long-term memory, and linguistic and social factors can increase the amount people are able to remember. I therefore hypothesize that Facebook posts will be better remembered than both less social and more formal writing and that certain social and linguistic factors will increase memorability. To test this hypothesis, I assigned participants to read either 100 Facebook posts or 100 sentences that were randomly drawn from books, coding the stimuli for a variety of social and linguistic features. I then measured participants' recognition memory of the stimuli.

Experiment 1

Participants and Materials

Sixty-two participants (29 male, 33 female) were recruited on the Massachusetts Institute of Technology (MIT) campus in Cambridge, Massachusetts. Participants were assigned to one of two conditions, with forty in the Facebook condition and twenty-two in the Amazon condition. Condition assignment was random for the first forty-four participants tested, while the next eighteen participants were all assigned to the Facebook condition.

We randomly selected 210 individuals from our own Facebook accounts, which together included over 580 friends. The most recent post meeting our criteria was then taken from each person's Facebook profile. Acceptable posts were less than 25 words long and were not connected with games, links, photos, or videos.

A total of 210 sentences from books were randomly collected from Amazon.com. Sentences with quotations, single-word sentences, and sentences that contained more than 25 words were not used. Examples are provided in the Appendix.

For each condition, 10 items were reserved for practice sessions and the remaining 200 items were used during

the test sessions. For each participant, 100 items were randomly selected as targets to be presented during the reading phase of the test session. The other 100 were reserved as lures to be randomly intermixed with the target sentences during the recall phase. A different combination of targets and lures was used for each participant.

Procedure

Participants were informed that they would be given a memory test after reading a series of sentences. During the first phase of the experiment ("reading phase"), 100 target sentences were individually presented for 3,000 ms each followed by a blank-screen for 250 ms (Figure 1).



Figure 1. Experimental design. Each participant saw either Facebook posts (left) or Amazon sentences (right), and each item was presented for 3000 ms. In the recall phase, items were presented in the same visual layout but remained on the screen until the participant responded.

Participants then took a self-paced recognition test that included 100 targets from the reading phase randomly intermixed with 100 lures. Participants indicated their confidence that each sentence had been seen (old) or not seen (new) during the reading phase using the 20-point rating scale shown in Figure 2. A "1" indicated that they were 100% certain the sentence had not appeared while "20" indicated that they were 100% certain it had appeared, with the intermediate range corresponding to lower certainties. Thus, a correct response for a lure would be between 1 and 10, and a correct response for a target would be between 11 and 20.



Figure 2. Confidence rating scale.

Results

Accuracy was measured in terms of confidence ratings. Responses were recorded on a 20-point scale, where 1 corresponds to high certainty that an item is a lure, 20 corresponds to a high certainty than an item is a target, and the intermediate range corresponding to lower certainties. A correct response for a lure would fall in the 1-10 range, while a correct response for a target item would be between 11-20. In order to compare confidence ratings, ratings (1-20) were transformed by subtracting the value from 10 and taking the absolute value of this number so that high values correspond to higher certainty.

Participants in the Facebook condition responded with higher accuracy and confidence for both target sentences and lure sentences compared to participants in the Amazon condition. This is illustrated in Figure 3, in which the receiver-operating characteristic (ROC) curves were generated starting at the rating 20 and continuing in decreasing order; the cumulate number of target and lure items given that rating or higher was computed, providing target and lure pairs for each level of confidence. Compared to the Amazon curve, the Facebook curve is shifted towards the area of the graph corresponding to higher hit rate and lower false alarm rate.



Figure 3. Receiver-operating characteristic (ROC) curves for Facebook and Amazon conditions. Compared to the Amazon curve, the Facebook curve is shifted toward higher accuracy.

Experiment 2

The methods for Experiment 2 were identical to those in Experiment 1, except only the Facebook stimuli were used.

Results

To test the variability and reliability of memorability of sentences, a split-half correlation was done using participants in the Facebook condition. Participants were randomly assigned to either Group 1 or Group 2 and the memorability of each sentence was compared between the two groups. While some sentences were remembered at barely above chance rates, others were recalled with 100% accuracy. This variability appears to be reliable. As shown in Figure 4, there is a strong positive correlation between the two groups for average accuracy. This means that a sentence's memorability was similar for both groups. This finding is mirrored by a similarly strong positive correlation between the two groups for average confidence rating z(198) = 0.49, p < 0.001.



Figure 4. The memorability (in terms of accuracy) of each Facebook post was compared across two randomly assigned groups. The strong positive correlation between accuracy of Group 1 and Group 2 shows that sentences were reliably memorable across different participants.

Experiment 3

The methods for Experiment 3 were identical to those in Experiment 2. The stimuli were coded for linguistic and social features.

Results

None of the linguistic features analyzed could alone account for the differences in memorability between sentences, however when non-grammatical capitalizations, repeated letters, ellipses and emoticons were grouped together, sentences tended to be better remembered (M = 15.92, SD = 2.2).

Sentences were also coded for social features including first, second and third person pronouns, and indexicals. While no difference was observed between first, second and third person pronouns, the presence of at least one indexical (M = 15.74, SD = 2.34) tended to increase a sentence's memorability over sentences without indexicals (M = 15.1, SD = 2.21), t(198) = -1.66, p =

0.09. Moreover, the number of indexical was positively correlated with memorability r(198) = 0.15, p = 0.02.

A third category of social feature analyzed was the subject of the sentence. The three categories were Reader, Writer, and Other. For example, the sentence "Thank you everyone for the kind birthday wishes!" refers to the reader, "YAY. I am finally friends with Angelina Jolie!" refers to the writer, and "snowww" refers to neither (called "Other"). There was no difference in memorability according to confidence ratings between Reader sentences and Other sentences, or between Reader sentences and Writer Sentences. The only effect found was between Writer and Other sentences, with the former being significantly more memorable than the latter t(175) = 2.12, p = 0.035.

Discussion

These results reveal three interesting findings. First, people have remarkable memory for written sentences and memory for Facebook posts is much stronger than for sentences drawn from books. Second, sentences are variably memorable and how memorable they are is consistent across people. Lastly, social and linguistic features including the number of indexicals a sentence contains can be predictive of its memorability. Odd typography and whether a sentence refers to the person who wrote it also appears to make sentences easier to remember. While these are the results of exploratory analysis, they suggest that our memory systems are especially attuned to certain linguistic and social features of language.

The fact that a sentence's memorability is stable across different people suggests that there must be certain features about the sentence that can account for this quality. However, the findings presented here do not seem to align with what is known about memory for isolated words. Specifically, it has been found that highly emotional words are better remembered than neutral words (Kensinger, 2003). However, I found that these features that signal emotion such as exclamation points, question marks, and emoticons are not predictive of memorability at the level of sentences. This may be because sentences are processed differently from individual words. Sentences are more than the sum of their component parts, with the presence or absence of words influencing the meaning of others, among numerous other linguistic factors. It would not be surprising if more complex levels of processing were needed to interpret these interactions. This is especially interesting because it suggests that it might not be possible to break down other types of complex stimuli into component parts in order to examine memory or other types of processing. That is, they must be considered as a whole.

This then raises the question of whether memory for sentences is primarily a function of linguistic features such as typography and semantics or of content and meaning. Somewhat surprisingly, the former appears to be true (Barclay, 1972). Even still, certain types of content such as social information appear to significantly impact memory (Holtgraves, 2008).

These results have important implications for business, education, cognitive science and numerous other areas. Yet many central questions remain open, and this area of research is highly deserving of further attention.

References

- Barclay, J.R. (1972). Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 2(3), 193-209.
- Brandy, T. F., Konkle, T., Alvarez, G. A., & Olivia, A. (2008). Visual long-term memory has a massive storage capacity for object details. PNAS, 105(38), 14325–14329.
- Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. Cognition,105(3), 658-668.
- Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. Language and Cognition, 2(1), 45-78.
- Holtgraves, T. (2008). Conversation, speech acts, and memory. Memory & Cognition, 36(2), 361-374.
- Kang, S. H. K., McDermott, K. B., & Cohen, S. M. (2008). The mnemonic advantage of processing fitnessrelevant information. Memory & Cognition, 36(6), 1151-1156.
- Kensinger, E.A. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? Memory & cognition, 31(8), 1169-1180.
- Masson, M.E.J. (1984). Memory for the surface structure of sentences: Remembering with and without awareness. Journal of Verbal Learning and Verbal Behavior, 5(23), 579-592.
- Mikes, L., Darby, R.S., Hwe, V., Bajic, D., Warker, J.A. (2013). Major Memory for Microblogs. Memory & Cognition, 4(41), 481-489.

Volume 28, Fall 2014

MURJ

Reports

Skowronski, J.J. (1991). Social Memory in Everyday Life: Recall of Self-Events and Other-Events. Journal of Personality and Social Psychology. 6(60), 831-843.

Appendix

Example Facebook posts:

1. My British silver fox of a professor named Nigel just said "and now for something completely different." Happy Valentines Day to me.

2. Thanks to everyone who made this summer so unforgettably perfect. I miss you all already.

3. So jealous of everyone up north enjoying the snow storm...

4. 7 years later...visited Ankara, Turkiye...got to see some buddies...

Example sentences from Amazon books:

1. He wondered how it could help but strike them.

2. He walked to the front of the cave and looked out.

3. Getting students to support the plan was not as easy as it might sound.

4. It's not cheating or copping out - it's smart.

WANT WORK THAT MATTERS?

bbn.com/careers Innovative R&D Raytheon BBN Technologies



We're hiring. You do the math.

