

With molecular data storage, cat videos could outlast us all

June 7 2019, by Caitlin Mcdermott-Murphy



Brian Cafferty, a postdoctoral scholar in the lab of George Whitesides, co-authored a paper on new advances in molecular storage. Credit: Kris Snibbe/Harvard Staff Photographer

From books to floppy disks to magnetic memory, technologies to store

information continue to improve. Yet threats as simple as water and as complex as cyberattacks can still corrupt our records.

As the data boom continues, more and more [information](#) is being stored in less and less space. Even the cloud—whose name promises opaque, endless space—will eventually reach its storage limit, can't thwart all hackers, and gobbles up energy. Now, a new way to store information lives outside the hackable internet, uses no energy once written, and, according to one of the researchers who developed it, "could allow information to be preserved for millions of years."

"Think storing the contents of the New York Public Library with a teaspoon of protein," said Brian Cafferty, a postdoctoral scholar in the lab of George Whitesides and author on a paper that describes the new technique. The work was performed in collaboration with Milan Mrksich and his group at Northwestern University.

"At least at this stage, we do not see this method competing with existing methods of data storage," Cafferty said. "We instead see it as complementary to those technologies and, as an initial objective, well-suited for long-term archival data storage."

Cafferty's chemical tool might not replace the cloud. But the [filing system](#) offers an enticing alternative to biological storage tools such as synthetic DNA, which scientists recently learned to manipulate to record any information, including GIFs, cooking tutorials, text, and music.

But while DNA is small compared with computer chips, it is large in the molecular world. And DNA synthesis requires skilled and often repetitive labor. If each message needs to be designed from scratch, macromolecule storage could be long and expensive work.

"We set out to explore a strategy that does not borrow directly from

biology," Cafferty said. "We instead relied on techniques common in organic and analytical chemistry, and developed an approach that uses small, low-molecular-weight molecules to encode information."

With just one synthesis, the team produced enough [small molecules](#) to encode multiple videos at once, making the approach less labor-intensive and cheaper than one based on DNA. For their low-weight molecules, the team selected oligopeptides (two or more peptides bonded together), which are common, stable, and smaller than DNA, RNA, or proteins.

Oligopeptides vary in mass, depending on their number and type of amino acids. Mixed together, they are distinguishable from one another, like letters in alphabet soup.

Making words from the letters is a bit more complicated: In a microwell—like a miniature version of a whack-a-mole, but with 384 holes—each well contains oligopeptides. When ink is absorbed on a page, the [oligopeptide](#) mixtures are assembled on a metal surface where they are stored. If the team wants to read back what they "wrote," they look at one of the wells through a [mass spectrometer](#), which sorts the molecules by mass. This tells them which oligopeptides were present or absent: Their mass gives them away.

To translate the jumble of molecules into letters and words, researchers borrowed the binary code. An M, for example, uses four of eight possible oligopeptides, each with a different mass. The four floating in the well receive a 1, while the missing four receive a zero. The molecular-binary code points to a corresponding letter or, if the information is an image, a corresponding pixel.

With this method, a mixture of eight oligopeptides could store one byte of information; 32 can store four bytes; and so on.

So far, Cafferty and his team have "written," stored, and "read" physicist Richard Feynman's famous lecture "There Is Plenty of Room at the Bottom," a photo of Claude Shannon (known as the father of information theory), and Hokusai's woodblock painting "The Great Wave off Kanagawa." Since it is estimated that the global digital archive will hit 44 trillion gigabytes by 2020 (10 times its size in 2013), an image of a tsunami seemed appropriate.

The team can retrieve their stored masterpieces with 99.9 percent accuracy. Their writing averages eight bits per second and their reading, 20. Because their writing speed far outpaces writing with synthetic DNA, at this stage reading could be both quicker and cheaper with the macromolecule. But with faster technology, the team's speeds will likely increase. An inkjet printer, for example, could generate drops at rates of 1,000 per second and cram more information into smaller areas. And improved mass spectrometers could take in even more information at a time.

The team could also improve the stability, cost, and capacity of their molecular storage with different classes of molecules. Their oligopeptides are custom-made and, therefore, more expensive. But future library builders could buy inexpensive molecules like alkanethiols, which could record 100,000,000 bits of information for just one cent. Unlike other molecular information storage systems, which rely on a specific molecule, this approach can use any malleable molecule as long as it can be manipulated into distinguishable bits.

Oligopeptides and similar choices are already resilient. "Oligopeptides have stabilities of hundreds or thousands of years under suitable conditions," according to the paper. The hardy [molecules](#) can endure without light or oxygen, in high heat and drought. And, unlike the cloud, which hackers can access from their favorite easy chair, the molecular storage can be accessed only in person. Even if a thief finds the data

stash, chemistry is needed to retrieve the code.

This story is published courtesy of the [Harvard Gazette](#), Harvard University's official newspaper. For additional university news, visit [Harvard.edu](#).

Provided by Harvard University

Citation: With molecular data storage, cat videos could outlast us all (2019, June 7) retrieved 23 April 2024 from <https://phys.org/news/2019-06-molecular-storage-cat-videos-outlast.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.