On Christmas day, 1801, Thomas Jefferson, then president of both the United States of America and the American Philosophical Society, received a letter from the society’s vice president, Robert Patterson, a frequent correspondent who was a mathematics professor at the University of Pennsylvania. Patterson began by defining four requirements of what he called a “perfect cypher.” It should work in any language, be easy to memorize, and be simple to perform. Most important, an ideal cipher should be “absolutely inscrutable to all unacquainted with the particular key or secret for decyphering.”

Patterson described a technique that met his criteria, and gave an example. “I shall conclude this paper with a specimen of such writing,” he boasted, “which I may safely defy the united ingenuity of the whole human race to decypher to the end of time….” Indeed, by all accounts, neither Jefferson nor anyone else could break Patterson’s challenge cipher for the next two centuries.

Until now. In a recent article in American Scientist, and in a talk at Harvard sponsored by the mathematics department, Lawrence Smithline ’94 explained how he decoded Patterson’s cryptic message in 2007. Smithline, a mathematician at the Center for Communications Research (in Princeton, New Jersey; a division of the Institute for Defense Analyses), used methods that were available in the early nineteenth century—if beyond that era’s mathematical intuitions—though he accelerated the computations with computers.

A coding technique like a simple substitution cipher consistently replaces one letter of the alphabet with another. The Caesar cipher, for example, shifts each letter a fixed number of places ahead in the alphabet—say, three—and XYZ becomes ABC. But since at least the fifteenth century, cryptographers have realized that simple substitution ciphers are vulnerable to frequency analysis. For example, “e” is the most frequently used letter in English. In a sufficiently long substitution cipher, whatever letter appears most often probably substitutes for “e.” Letter counts suggest a limited set of choices to try for the most commonly used letters.

In his more sophisticated code, Patterson wrote his message openly, without capitals or spaces, but vertically on ruled paper, “in the Chinese manner,” in columns from left to right. This produces a grid of lowercase letters that are gibberish when read left to right, but a perfectly clear message when read in columns. Next he broke this grid into sections of up to nine lines each, numbering each line 1, 2, 3, etc., and re-ordering them randomly within the section—though all sections would repeat the same reordered sequence of numbers. He also inserted up to nine arbitrary letters at the beginning of each line, which had no bearing on the message content but drastically increased the inscrutability factor. He filled vacant spaces at the end of the line with similarly random letters.

“It will be absolutely impossible, even for one perfectly acquainted with the general
Virtual Tumors

I f a single word could describe the diverse disorders collectively known as cancer, it might be unpredictable. Trying to judge the course and prognosis of this meandering, unforgiving, and frequently fatal disease is almost impossible.

But Thomas Deisboeck, as director of a pioneering project to “engineer” different kinds of tumors, is trying to make cancer more predictable—and hence, more treatable. His Center for the Development of a Virtual Tumor (CViT) doesn’t grow cells in a lab or study cancer in mice. Instead, it serves as a “virtual laboratory,” using computers rather than test tubes, and three-dimensional images instead of lab animals, to foster collaboration among researchers from around the world. “At CViT, we extract data from experiments and the scientific literature, and build models with it,” explains Deisboeck, an assistant professor of radiology at Harvard Medical School. “Basing our computational models on a cell’s interaction with its environment, we can simulate everything from a single cell to an organ, and make predictions from what we observe.”

The digraph table helped evaluate those guesses. “For instance, the letter pair ‘vj’ is impossible in English, so that excludes any alignment that creates that digraph,” Smithline wrote. “Alternatively, the letter pair ‘qu’ is rare, but when there is a ‘q,’ it must line up with a ‘u.’ When ‘q’ and ‘u’ do line up, that is strong evidence in favor of that alignment.” Lastly, he applied dynamic programming—a technique used today in computational biology to find, for example, similar regions in two DNA base sequences—to statistically identify top-scoring guesses on section size, row pairs, and extra letters. (The dynamic program works despite significant errors in transcribing the handwritten cipher to typed characters.) Certain constraints in Patterson’s cipher, Smithline wrote, “reduced the overall computational load to fewer than 100,000 simple sums—tedious in the nineteenth century, but doable.”

This analysis allowed Smithline to decrypt the challenge cipher that had held its message inviolate for more than two centuries. Had Jefferson cracked the code, he quite likely would have divined the entire message from its first few words: “In Congress, July Fourth,...” — the preamble to the Declaration of Independence, from Jefferson’s own hand. ~CRAIG LAMBERT

LIBRARY OF CONGRESS

Patterson’s challenge cipher (left) to President Jefferson. He believed it “absolutely impossible” to decode, “even for one perfectly acquainted with the general system,” as it had more than “ninety millions of millions” possible keys. In fact, neither Jefferson nor anyone else, during two centuries, was able to decrypt it.

Lawren Smithline E-mail address: lawren@idaccr.org

CANCER’S NEW REALITY

Virtual Tumors

system, ever to decipher the writing of another without his key,” Patterson wrote. He estimated the number of possible keys at more than “ninety millions of millions.” The cipher’s effectiveness so strongly impressed Jefferson that he forwarded it to his ambassador to France, Robert Livingston, who nonetheless persisted in an older nomenclator code, based on a catalog of numbers representing words or phrases.