

# Proteins use a lock and key system to bind to DNA

January 16 2019, by Julie Langelier

---



Credit: CC0 Public Domain

You can think of DNA as a string of letters—As, Cs, Ts, and Gs—that together spell out the information needed for the construction and function of cells. Each cell in your body shares the same DNA. So, for cells to take on their differing roles, they must be able to turn on and off specific genes with precise control. The genes active in a brain cell, for instance, are different than those active in a skin cell.

This is achieved in part by the action of "DNA binding proteins" that latch onto the [human genome](#) at particular places to turn genes on or off. Now, researchers at the Gladstone Institutes led by Katherine Pollard, Ph.D., made a major discovery about how these proteins bind to DNA.

Scientists have traditionally thought that DNA binding proteins use patterns in the genome's code of As, Cs, Ts, and Gs to guide them to the right location, with a given protein only binding to a specific sequence of letters. However, many proteins bind to several different letter combinations, and two different proteins may recognize the same pattern.

Despite this multitude of overlapping patterns, proteins never seem confused about where they're supposed to bind. In the new study, published in *Cell Systems*, the Gladstone scientists discovered that proteins must rely on another clue to know where to bind: the DNA's three-dimensional shape.

"For decades, we've had difficulty explaining how proteins find the correct places to bind in the DNA, and how they do that in a specific way and without binding to the wrong places," said Pollard, a senior investigator and director of the Gladstone Institute of Data Science and Biotechnology. "We hypothesized this could be explained by the structural aspect of the genome."

That's because DNA's string of letters is also a physical, three-dimensional structure, twisted into the famous double-helix shape and wrapped up into a microscopic package. Within its ladder-like structure, a variety of twists, grooves, and gaps can be found between the rungs and sides. Pollard and her team realized these variations create a type of keyhole that select proteins slot into. If the grooves on the protein don't match those on the genome, the key won't fit.

"There's a rich scientific literature on how proteins interact with each other or bind to chemicals, and it's always through a kind of lock and key mechanism; why would proteins binding to DNA be any different?" said Md. Abul Hassan Samee, Ph.D., a postdoctoral fellow at Gladstone who is the first author of the study. "We think the proteins dock onto DNA as a 3-D structure, just like when they interact with other proteins or with chemicals."

Earlier work had raised the possibility that DNA shape provides additional information to proteins on where to bind, but it was unclear how influential these shapes were. To test their theory, the researchers adapted a common machine learning algorithm typically used to identify the letter sequences proteins bind to, except now they were looking for patterns in shape. They discovered that over 80 percent of proteins bind to a specific shape pattern in the genome.

The researchers say that although the proteins are frequently not reading the alphabetical code of the genome, the sequence of the letters is still vital to dictating where these proteins bind, but because it defines the genome's shape. Curiously, very different letter sequences can designate the same structure, while slightly different letter sequences can result in wildly different structures.

This fact helps explain the two biggest mysteries in protein binding to DNA. First, proteins that bind to multiple different letter sequences turn out to be homing in on the same spatial pattern, and second, proteins that appear to share letter sequences are in fact attaching to very different shapes. What's more, proteins that frequently bind to the genome as a pair are attracted to specific shapes that can differ from the shapes they recognize when they bind alone.

The current work was all done with computer modeling, so the researchers' next step is to prove their theory using molecular

experiments.

"It was accepted that a pattern of As, Cs, Ts, and Gs where a protein bound to DNA had a particular shape," said Pollard, who is also a professor at UC San Francisco and a Chan Zuckerberg Biohub investigator. "But nobody had looked to see whether other binding locations that couldn't be explained with that pattern of letters might have the same shape. If we can show in a dish that proteins can recognize a DNA location because of its shape, even when it doesn't contain the established letter sequence, I think it would be game changing."

In recent years, scientists have discovered that most genetic mutations that result in disease are not in the genes themselves. Instead, they occur in so-called "dark DNA"—the 99 percent of the human genome that influences how, when, and where genes are turned on or off. With their recent discovery, the researchers have opened the door to understanding a new way that mutations could affect gene expression and, as a result, the functioning of cells.

"There's a huge effort right now to understand how mutations in this dark DNA cause disease, and that's important because for most complex diseases, the majority of the [genetic mutations](#) are outside of genes," explains Samee. "Everyone has been looking at the letter sequences and asking whether the mutations disrupt those sequences, but our work shows that you also need to ask whether the mutation is changing the shape of the DNA. You could have a mutation that changes the letter sequence, but if it doesn't change the shape, it may not always change the [protein](#) binding."

The paper "A de novo shape-motif discovery algorithm reveals preferences of transcription factors for DNA shape beyond sequence-motifs" was published by the journal *Cell Systems* on January 16, 2019.

**More information:** Md. Abul Hassan Samee et al. A De Novo Shape Motif Discovery Algorithm Reveals Preferences of Transcription Factors for DNA Shape Beyond Sequence Motifs, *Cell Systems* (2019). DOI: [10.1016/j.cels.2018.12.001](https://doi.org/10.1016/j.cels.2018.12.001)

Provided by Gladstone Institutes

Citation: Proteins use a lock and key system to bind to DNA (2019, January 16) retrieved 25 April 2024 from <https://phys.org/news/2019-01-proteins-key-dna.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.