

New insight into how proteins find their DNA binding sites in the genome

April 3 2015, by Katharine Gammon



In 2014, Remo Rohs and his team summarized the progress toward cracking the transcription factor code in the Cell Press journal Trends in Biochemical Sciences. The journal cover, designed by USC Dornsife doctoral student Lin Yang, highlights the many variables that influence transcription factor-DNA binding.

Remo Rohs is looking for some deep connections. He is integrating genomics and structural biology to uncover some significant insights into how proteins recognize DNA.

While genomics deciphers DNA by studying the sequences of base pairs that encode genetic information, [structural biology](#) explores the impact of the actual 3-D structure of DNA. Rohs, however, aims to unite the two fields into something new—and hopefully more useful.

"Structural biology and genomics are big fields, but there is little interaction between these two worlds," said Rohs, assistant professor of biological sciences, chemistry and physics at USC Dornsife. "Genomics thinks of entire genomes in terms of sequence, and structural biology thinks of 3-D structures at high resolution but limited size."

In a March 9 paper in the *Proceedings of the National Academy of Sciences*, a team led by Rohs, which included researchers from Duke University and Columbia University, used a large data set of proteins to show that combining information on DNA shape and sequence resulted in a better understanding of protein-DNA recognition.

"Transcription factors are proteins that bind DNA to regulate genes, so knowing how and where they bind is of central importance in biology. This paper describes how modeling DNA shape can improve our understanding of transcription factor binding, with broad implications for many areas of research," said Steven Henikoff, a member of the National Academy of Sciences who edited Rohs' paper for PNAS.

Rohs' group used [machine learning](#) to train models that predict how well and where the transcription factors will bind to the genome. When thinking about how machine learning works, Rohs said, look no further than a search engine. "When Google tries to understand your consumer behavior by looking at the websites you visit—that is a feature," said Rohs, who holds a joint appointment in computer science at the USC Viterbi School of Engineering. "In the same way, we can use the features of the DNA—its sequence and shape—to predict whether a binding site is occupied by a protein or not."

Tianyin Zhou, a former graduate student in Rohs' lab who earned a Ph.D. in computational biology and bioinformatics from USC Dornsife in 2014, and is the lead author of the study, said that there are dual implications for the work.

"First, once we incorporate the DNA shape, we can get very good predictive models," he said. "And with this information, we can tell how gene expression is regulated. Second, when you know a mechanism, you can design or engineer a sequence to make it bind to the protein you want," said Zhou, who is now working as a software engineer at Google.

In another paper, published April 2 in the journal *Cell*, Rohs collaborated with Richard Mann, an experimental biologist at Columbia University, to tease apart the contributions of DNA shape and sequence. They took proteins that they knew require DNA shape for binding and mutated the amino acids that only recognize shape but not sequence.

The researchers looked at a group of proteins known as [Hox transcription factors](#), which are critical for early embryonic development. Rohs and colleagues found that introducing shape-recognizing amino acids from one transcription factor to another swapped binding specificities between Hox proteins.

Lin Yang, a doctoral student in computational biology and bioinformatics at USC Dornsife, said that understanding the fundamentals of binding specificities is a vital scientific goal. "When these proteins don't work properly—when they bind arbitrarily, or bind to incorrect sites—it might cause disease." Rohs credits Yang for the acceptance of the paper in *Cell* because Yang successfully used machine learning to identify the DNA shape features that are important for recognition.

A third paper, published March 11 in *Genome Research* in collaboration

with Eran Segal from the Weizmann Institute of Science, found that regions outside the [binding site](#) are important for binding.

In the future, Rohs hopes to continue his work on gene regulation in a more complex way. "When we talk about protein binding to DNA, we assume DNA is accessible, but in the cell it is folded up and covered by other proteins," he said. "So the next step is to integrate information about cooperative binding and the accessibility of binding sites, going from the in vitro to the more complex in vivo situation. This also includes epigenetic mechanisms such as DNA methylation, which is another interest of my team."

More information: "Quantitative modeling of transcription factor binding specificities using DNA shape." *PNAS* 2015 ; published ahead of print March 9, 2015, [DOI: 10.1073/pnas.1422023112](https://doi.org/10.1073/pnas.1422023112)

"Unraveling determinants of transcription factor binding outside the core binding site." *Genome Res.* gr.185033.114 Published in Advance March 11, 2015, [DOI: 10.1101/gr.185033.114](https://doi.org/10.1101/gr.185033.114)

"Deconvolving the Recognition of DNA Shape from Sequence." DOI: [dx.doi.org/10.1016/j.cell.2015.02.008](https://doi.org/10.1016/j.cell.2015.02.008)

Provided by University of Southern California

Citation: New insight into how proteins find their DNA binding sites in the genome (2015, April 3) retrieved 9 April 2024 from <https://phys.org/news/2015-04-insight-proteins-dna-sites-genome.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.