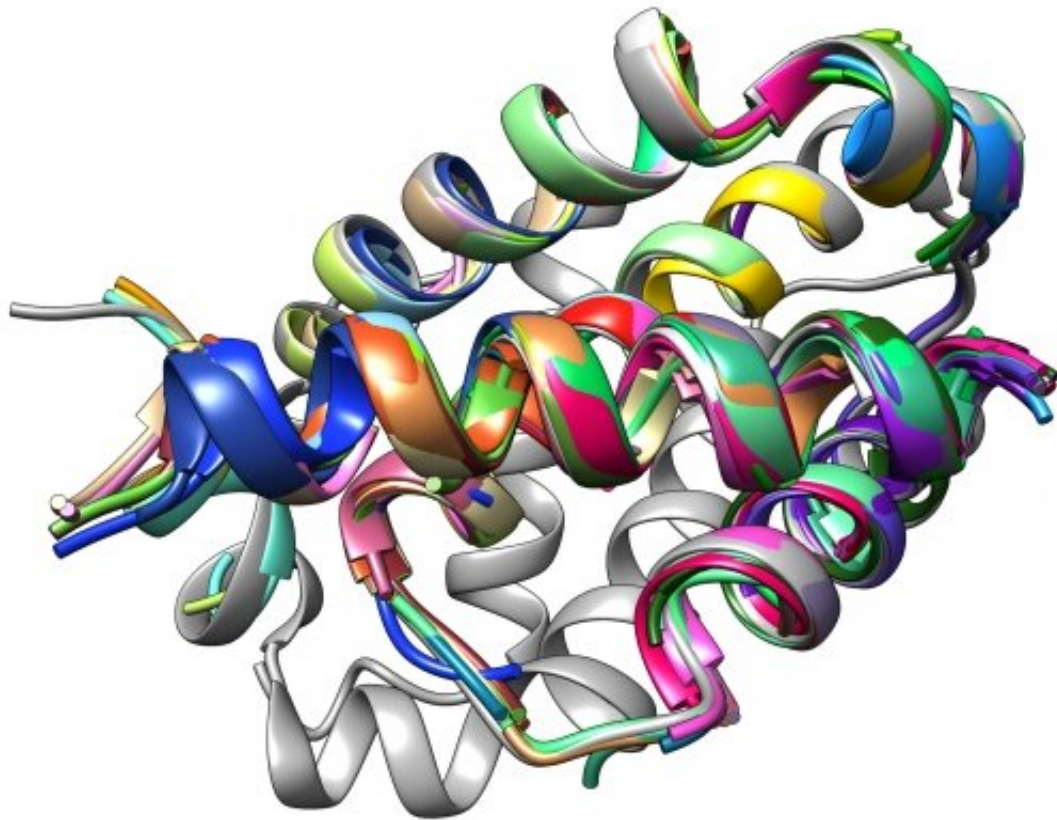# Predicting sequence from structure

February 18 2019, by Raleigh Mcelvery

The binding interface between a peptide and its Bcl-2 protein target is composed of common structural motifs known as TERMs. Credit: Sebastian Swanson and Avi Singer

One way to probe intricate biological systems is to block their

components from interacting and see what happens. This method allows researchers to better understand cellular processes and functions, augmenting everyday laboratory experiments, diagnostic assays, and therapeutic interventions. As a result, reagents that impede interactions between proteins are in high demand. But before scientists can rapidly generate their own custom molecules capable of doing so, they must first parse the complicated relationship between sequence and structure.

Small molecules can enter cells easily, but the interface where two proteins bind to one another is often too large or lacks the tiny cavities required for these molecules to target. Antibodies and nanobodies bind to longer stretches of protein, which makes them better suited to hinder protein-protein interactions, but their large size and complex structure render them difficult to deliver and unstable in the cytoplasm. By contrast, short stretches of amino acids, known as peptides, are large enough to bind long stretches of protein while still being small enough to enter cells.

The Keating lab at the MIT Department of Biology is hard at work developing ways to quickly design peptides that can disrupt protein-protein interactions involving Bcl-2 proteins, which promote cancer growth. Their most recent approach utilizes a computer program called dTERMen, developed by Keating lab alumnus, Gevorg Grigoryan Ph.D. '07, currently an associate professor of computer science and adjunct associate professor of biological sciences and chemistry at Dartmouth College. Researchers simply feed the program their desired structures, and it spits out amino acid sequences for peptides capable of disrupting specific protein-protein interactions.

"It's such a simple approach to use," says Keating, an MIT professor of biology and senior author on the study. "In theory, you could put in any structure and solve for a sequence. In our study, the program came up with new sequence combinations that aren't like anything found in

nature—it deduced a completely unique way to solve the problem. It's exciting to be uncovering new territories of the sequence universe."

Former postdoc Vincent Frappier and Justin Jenson Ph.D. '18 are co-first authors on the study, which appears in the latest issue of *Structure*.

Same problem, different approach

Jenson, for his part, has tackled the challenge of designing peptides that bind to Bcl-2 proteins using three distinct approaches. The dTERMen-based method, he says, is by far the most efficient and general one he's tried yet.

Standard approaches for discovering peptide inhibitors often involve modeling entire molecules down to the physics and chemistry behind individual atoms and their forces. Other methods require time-consuming screens for the best binding candidates. In both cases, the process is arduous and the success rate is low.

dTERMen, by contrast, necessitates neither physics nor experimental screening, and leverages common units of known protein structures, like alpha helices and beta strands—called tertiary structural motifs or "TERMs"—which are compiled in collections like the Protein Data Bank. dTERMen extracts these structural elements from the data bank and uses them to calculate which amino acid sequences can adopt a structure capable of binding to and interrupting specific protein-protein interactions. It takes a single day to build the model, and mere seconds to evaluate a thousand sequences or design a new peptide.

"dTERMen allows us to find sequences that are likely to have the binding properties we're looking for, in a robust, efficient, and general manner with a high rate of success," Jenson says. "Past approaches have taken years. But using dTERMen, we went from structures to validated

designs in a matter of weeks."

Of the 17 peptides they built using the designed sequences, 15 bound with native-like affinity, disrupting Bcl-2 protein-protein interactions that are notoriously difficult to target. In some cases, their designs were surprisingly selective and bound to a single Bcl-2 family member over the others. The designed sequences deviated from known sequences found in nature, which greatly increases the number of possible peptides.

"This method permits a certain level of flexibility," Frappier says. "dTERMen is more robust to structural change, which allows us to explore new types of structures and diversify our portfolio of potential binding candidates."

Probing the sequence universe

Given the therapeutic benefits of inhibiting Bcl-2 function and slowing tumor growth, the Keating lab has already begun extending their design calculations to other members of the Bcl-2 family. They intend to eventually develop new proteins that adopt structures that have never been seen before.

"We have now seen enough examples of various local protein structures that computational models of sequence-structure relationships can be inferred directly from structural data, rather than having to be rediscovered each time from atomistic interaction principles," says Grigoryan, dTERMen's creator. "It's immensely exciting that such structure-based inference works and is accurate enough to enable robust protein design. It provides a fundamentally different tool to help tackle the key problems of structural biology—from protein design to structure prediction."

Frappier hopes one day to be able to screen the entire human proteome

computationally, using methods like dTERMen to generate candidate binding peptides. Jenson suggests that using dTERMen in combination with more traditional approaches to sequence redesign could amplify an already powerful tool, empowering researchers to produce these targeted peptides. Ideally, he says, one day developing peptides that bind and inhibit your favorite protein could be as easy as running a computer program, or as routine as designing a DNA primer.

According to Keating, although that time is still in the future, "our study is the first step towards demonstrating this capacity on a problem of modest scope."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology