Health and Physiology

Floppy proteins and the hidden sequences they use to communicate

by **Theresa Hwang**[1] | PhD Student; **Amy Keating**[2] | Professor

[1]: Massachusetts Institute of Technology

This Break was edited by Ralph Bulanadi, *Senior Scientific Editor* - TheScienceBreaker

Our cells are filled with proteins. These proteins usually have a specific structure which help them to make the specific interactions we need to live. But how do short, floppy regions of proteins make such tight and specific interactions despite lacking any structure?

Image Credits: Pixabay

When the first protein structure was solved in 1958, scientists knew they had broken into a new frontier of biology. They felt that a protein's structure would reveal all the secrets of its function, and almost 70 years down the line, over 150,000 structures have now been solved. These data have proven enormously useful in expanding our knowledge of how proteins work.

However, as more proteins were discovered, scientists realized that a significant portion of all human proteins do not have a specific three-dimensional structure. In fact, 15-45% of proteins from eukaryotic cells are estimated to have large disordered regions. These disordered regions of proteins are mysterious, since, without a known structure, we cannot predict their function by comparing them to other proteins, according to what is known as the structure–function paradigm. The disordered regions can also exist as different shapes and are highly flexible. Many disordered regions have sections called short linear motifs (SLiMs). SLiMs can be thought of as small codes that recruit other, specific proteins to participate

in some function. SLiM-based interactions control behaviors ranging from cell movement to cell division and are so important that they are thought to make up 15-40% of all protein-protein interactions.

If the wrong proteins interact with each other, diseases such as cancer can occur, so making sure the right interactions happen is very important. One way that a protein can recognize the right partner is through their shape. Two proteins can interlock and "bind", much like pieces of a puzzle, resulting in a tight and specific interaction. However, SLiMs, which are found in flexible protein regions, do not have any defined 3D shape. They are also very short. The typical human protein is made up of about 400 small segments called amino acids. SLiMs are much smaller, consisting of just 4-6 amino acids. How can such a short and floppy stretch of protein recognize and bind to the right partner and no one else?

To understand how SLiMs make specific interactions, we must first find them. But tracking down SLiMs is difficult because they are short and bind weakly to their partners. To start with, we obtained a library of all human proteins chopped into 400,000 36-amino acid fragments. We compared this library against a folded protein called ENAH, which is known to bind to a SLiM that contains multiple amino acids known as proline. ENAH is interesting because it can help cells move throughout the body, but in cancerous cells, too much ENAH is made and all this movement can help cancer spread. This makes understanding ENAH binding particularly important for cancer treatment.

By searching in our library of protein fragments, we discovered over 100 unique sequences that bind to ENAH. When we closely examined the properties of their SLiMs, we found interesting features in the surrounding amino acids. For example, many of our sequences contained extra prolines at the end of their SliMs. We also found duplicated ENAH-binding SliMs spaced approximately 15 amino acids away from each other in our 36-amino acid long protein chunks.

Our next goal was to determine whether these patterns of amino acids made binding to ENAH more likely. We performed "truncation studies", where we cut out duplicated SLiMs or the extra prolines surrounding our SLiMs. We found that these regions did increase the binding of SLiMs to ENAH. To understand how the binding improved, we investigated the structure of ENAH bound to a SLiM with extra prolines. We found that these prolines bind to a new section on the surface of ENAH that enhances overall binding, revealing a new feature of ENAH-binding SLiM codes.

Our findings have important implications for SLiM-based interactions. For example, motivated by our finding that the regions that surround SLiMs can help binding, we can design molecules to bind to ENAH to stop its role in promoting cancer. This is something that we did in follow-up work, based on what we learned from this library-search approach. Because SLiMs are so common, we expect that we can use a similar approach to understand how these floppy protein sections can so reliably work with nearby amino acids, and therefore better understand the interactions that occur within our own bodies