

Vanderbilt University <http://www.vanderbilt.edu>

Creation of the largest human-designed protein boosts protein engineering efforts

by David Salisbury <http://news.vanderbilt.edu/author/davidsalisbury/> | Posted on Tuesday, Nov. 15, 2011 — 12:59 PM



<http://news.vanderbilt.edu/files/barrellprotein-JMCF.jpg>

Professor Jens Meiler, right, and research assistant Carrie Fortenberry, who is handling a model of the largest human-designed protein. (John Russell/Vanderbilt University)

If Guinness World Records <http://www.guinnessworldrecords.com/> had a category for the largest human-designed protein, then a team of Vanderbilt chemists would have just claimed it.

They have designed and successfully synthesized a variant of a protein that nature uses to manufacture the essential amino acid histidine. It is more than twice the size of the previous record holder, a protein created by researchers at the University of Washington <http://www.washington.edu/> in 2003.

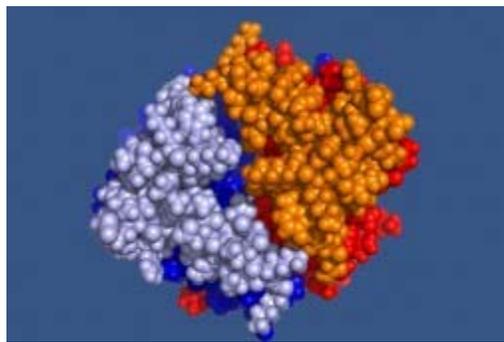
The synthetic protein, designated FLR, validates a new approach which the Vanderbilt scientists have developed that allows them to design functional artificial proteins substantially larger than previously possible.

We now have the algorithms we need to engineer large proteins with shapes that you don't see in nature. "We now have the algorithms we need to engineer large proteins with shapes that you don't see in nature. This gives us the tools we need to create new, more effective antibodies and other beneficial proteins," said Jens Meiler <http://www.vanderbilt.edu>

[/chemistry/faculty/meiler.php](#)), the associate professor of chemistry at Vanderbilt who led the effort.

Recently, protein engineers have verified a potential treatment strategy for HIV by using designed protein vaccines in mice and have designed artificial proteins that mimic antibodies in broadly neutralizing flu infections. The technique developed at Vanderbilt promises to expand the scope of these efforts substantially.

That is important because proteins are the most important molecules in living cells. They perform most of the vital tasks that take place within a living organism. There are hundreds of thousands of different proteins. They come in a variety of shapes and sizes. They can be round or long and thin, rigid or flexible. But they are all made out of linear chains of 20 amino acids encoded in the genome of the organism.



(<http://news.vanderbilt.edu/files/barrellprotein-model.jpg>)

Space-filling molecular model of the FLR protein clearly shows its barrel structure: the same structure that is found in 10 percent of all proteins. (Courtesy of the Meiler Lab)

Proteins assume this variety of shapes and sizes by the manner in which they bunch and fold. This complex process takes two steps. First, small numbers of adjacent amino acids form what scientists call secondary structures: the most common of which are a rod-like spiral shape called the alpha-helix and a flat, pleated shape called the beta-sheet. These secondary structures, in turn, interact, fold and coil to form the protein's three-dimensional shape, which is the key to its function.

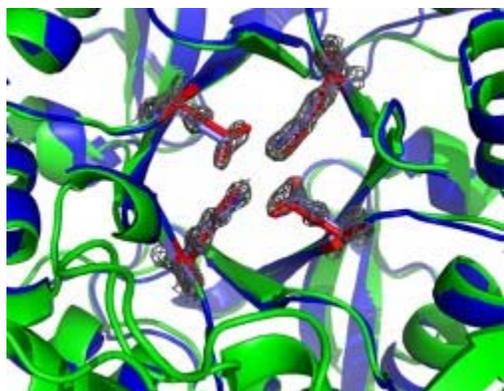
Over the past 10 years an increasing number of proteins that don't exist in nature have been designed "in silico" (in a computer). Scientists use sophisticated protein modeling software that incorporates the relevant laws of physics and chemistry to find amino acid sequences that fold into stable forms and have specific functions.

Imagine making a necklace 10 beads long with beads that come in 20 different colors. There are more than 10 trillion different combinations to choose among. This provides an idea of the complexity involved in designing novel proteins. For a protein of a given size, the modeling software creates millions of versions by putting each amino acid in every position and evaluating the stability of the resulting molecule. This takes a tremendous amount of computing power which skyrockets as the length of the protein increases.

"The current limit of this approach, even using the fastest supercomputers, is about 120 amino acids," said Meiler. The previous record holder contained 106 amino acids. The newly designed protein contains 242 amino acids. The Vanderbilt group got around this limit by modifying the widely used protein engineering platform called [ROSETTA](#) (<http://boinc.bakerlab.org/>) so that it can incorporate symmetry in the design process.

Their success provides new support for a controversial theory about protein evolution called the gene duplication and

fusion hypothesis. The advantage of small proteins is that they can evolve rapidly in response to changing conditions, but larger proteins can perform more complex functions. Nature found a way to get both advantages by selecting small proteins that can interact with other copies of themselves to form larger proteins, which are called dimers. Once useful dimers have been created the gene that coded for the original protein is duplicated and fused to form a new gene that can directly produce the dimer. After it is created, the dimer gene is gradually modified by natural selection to make it more efficient or develop new functions.



http://news.vanderbilt.edu/files/barrellprotein-crystal_structure.jpg

An illustration that shows the accuracy of the computer model of the FLR protein by showing the computer model in blue and its experimental structure in green. The density of a structure called a salt bridge cluster is shown in grey superimposed on the computer model shown in red.

(Courtesy of the Meiler Lab)

Because they have two identical halves, dimers have a large degree of symmetry. By taking these symmetries into account, the Vanderbilt group was able to substantially reduce the amount of computing time required to create the FLR protein. Using 400 processors of the supercomputer at Vanderbilt's [Advanced Computing Center for Research and Education](http://www.accre.vanderbilt.edu/) (<http://www.accre.vanderbilt.edu/>), it took 10 days of continuous processing to find the most stable configuration.

To check the accuracy of their design, the researchers synthesized the DNA sequence that produces the protein, inserted it in *E.coli* bacteria and determined that they produced the protein and it folded properly.

The FLR protein assumes a 3-D shape called a TIM barrel, which is found in 10 percent of proteins and is particularly prevalent among enzymes. It is formed from eight beta strands that are surrounded by eight alpha helices arranged in a hexagonal shape like a tiny barrel.

The paper reporting this achievement appears in the Nov. 16 issue of the *Journal of American Chemical Society* (<http://pubs.acs.org/journal/jacsat>) and is available online (<http://pubs.acs.org/doi/abs/10.1021/ja2051217>). Members of Meiler's team are research assistant [Carie Fortenberry](http://www.meilerlab.org/index.php/users/show/u_id/6) (http://www.meilerlab.org/index.php/users/show/u_id/6), undergraduate students Elizabeth Bowman, Will Proffitt, and Brent Dorr and research assistant professors of biochemistry [Joel Harp](https://medschool.mc.vanderbilt.edu/facultydata/php_files/show_faculty.php?id3=11324) (https://medschool.mc.vanderbilt.edu/facultydata/php_files/show_faculty.php?id3=11324) and [Laura Mizoue](http://www.meilerlab.org/index.php/users/show/u_id/14) (http://www.meilerlab.org/index.php/users/show/u_id/14). The research was supported by grants from the Defense Advanced Research Projects Agency (<http://www.darpa.mil/>)'s protein design project and the National Science Foundation (<http://www.nsf.gov/>).

Contact:

David Salisbury, (615) 322-NEWS

david.salisbury@vanderbilt.edu (<mailto:david.salisbury@vanderbilt.edu>)