Tackling disease in three dimensions—supercomputers help decode RNA structure

July 13 2017, by Joan Koka

A cure for cancer, HIV and other stubborn diseases has evaded the brightest minds for generations. But with supercomputers – computing systems that can calculate, analyze and visualize extremely large amounts of data – researchers are gaining a leg up in the fight for better treatments and cures.

Researchers at the National Cancer Institute (NCI) are using supercomputers at the U.S. Department of Energy's (DOE) Argonne National Laboratory to advance disease studies by enhancing our understanding of RNA, biological polymers that are fundamentally involved in health and disease.

In collaboration with staff from the Argonne Leadership Computing Facility (ALCF), researchers have perfected a technique that accurately computes the 3-D structure of RNA sequences. This method, which relies on a computer program known as RS3D and Mira – the ninth fastest supercomputer in the world – gives researchers studying cancer and other diseases structural insights about relevant RNAs that can be used to advance computer-assisted drug design and development.

RNA not only functions as a DNA interpretation messenger for protein fabrication, but also plays a multifaceted role in regulating gene expression – such as when, where and how efficiently a gene is expressed. For this reason, researchers are actively seeking to understand

the functions of novel RNA sequences. And in order to get a complete picture, they need to know the biologically active forms of RNA, which are reflected in the complex 3-D structures that RNA sequences fold into after they're created.

"We already know the basic chemical groups for RNA and how they're composed, but what we don't know is what conformational structures they take," said Wei Jiang, a researcher at the Argonne Leadership Computing Facility who is one of the computational leads in the project.

"Getting the real functional structure, which is the 3-D structure, is very difficult to do experimentally, because the RNA polymer is too flexible," he said. "This is why we rely on computational simulation. Simulations can be used to explore hundreds or thousands of possible conformational states that would eventually lead us to the most likely 3-D structure."

The computer program RS3D was developed by a National Cancer Institute research team, led by researcher Yun-Xing Wang and postdoctoral fellow Yuba Bhandari, and optimized by ALCF researchers to run on Mira; Jiang played a central role in scaling the RS3D code to run on a large fraction of Mira, which improved its performance significantly.

As an input, RS3D uses known RNA sequence information and experimental data from small angle X-ray scattering, a technique that provides important structural information, such as particle size and shape, based on the scattering pattern that is generated when X-ray beams are applied to a target sample. With these inputs, RS3D outputs a low-resolution 3-D image of the topological structure of RNA that provides the most likely folding patterns.

"Since the biologically active form of RNA is a 3-D structure, going

from understanding the primary sequence and the two-dimensional layout of an RNA to understanding the 3-D form is a big stepping-stone that gives us a lot of useful information about biological functions," said Bhandari, one of the leaders of the project. "Understanding the structural basis provides a foundation for further investigating molecular interactions and biological pathways in various diseases."

The researchers validated their technique by using it to compute the 3-D structure of 18 RNA polymers whose structures are known. These select RNAs fold into a wide variety of structures that represent common folding architectures. Additionally, researchers used R3SD along with experimental data recorded at the synchrotron light source at Argonne, the Advanced Photon Source, to compute the structure of adenine riboswitch, an RNA structure known to regulate gene expression.

"One of the unique and advantageous features of this technique is the fact that it's fully automated, meaning it does not require the user to input an initial 3-D structural template to work. This sets it apart from other methods that perform similar calculations," Bhandari said. "This helps us eliminate any potential limitations or biases that could be introduced through a template, and make the whole approach easier to apply."

The researchers are now in the process of publishing their technique; the source code will be made available to researchers thereafter. A brief summary of their computational work, presented in an article titled "Modeling RNA topological structures using small angle X-ray scattering," is published in *Methods*.