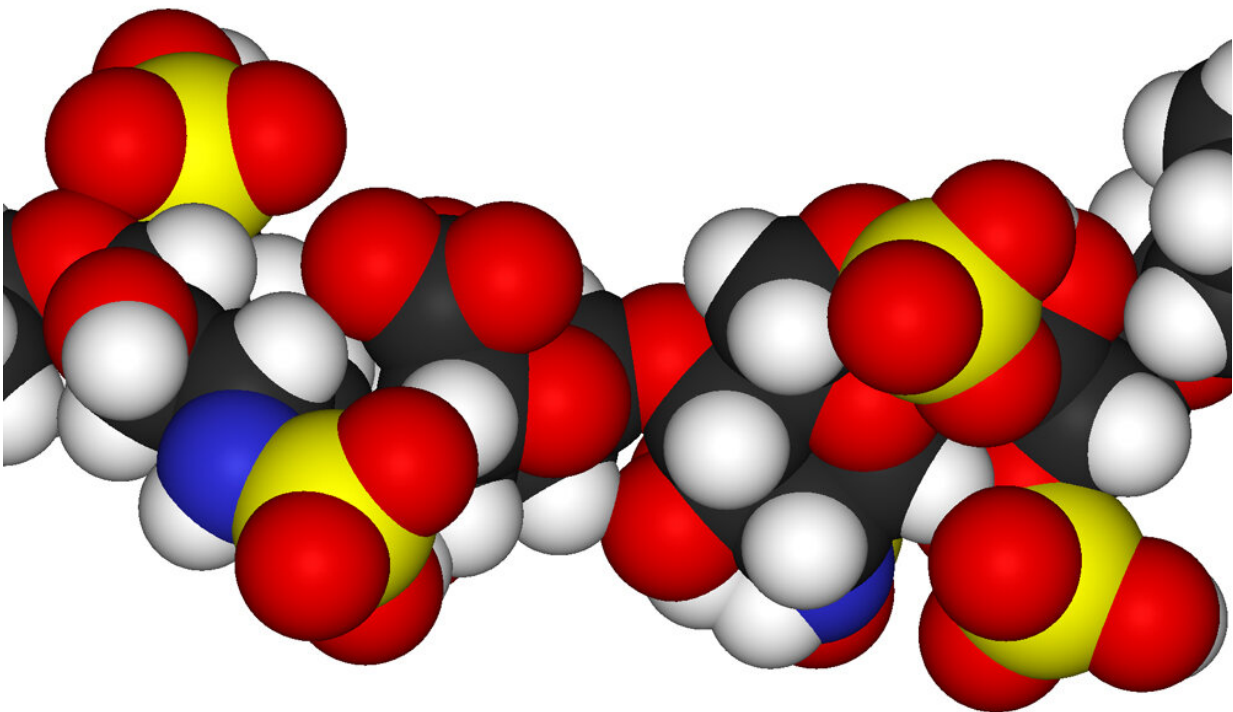


# Time needed to sequence key molecules could be reduced from years to minutes

March 8 2021

---



A nanopore and image recognition software can sequence a sulfated glycosaminoglycan in real time. Credit: Rensselaer Polytechnic Institute

Using a nanopore, researchers have demonstrated the potential to reduce the time required for sequencing a glycosaminoglycan—a class of long

chain-linked sugar molecules as important to our biology as DNA—from years to minutes.

As published this week in the *Proceedings of the National Academies of Sciences*, a team from Rensselaer Polytechnic Institute showed that [machine-learning](#) and image recognition software could be used to quickly and accurately identify sugar chains—specifically, four synthetic heparan sulfates—based on the [electrical signals](#) generated as they passed through a tiny hole in a crystal wafer.

"Glycosaminoglycans are a complex repertoire of sequences, as the work of Shakespeare or a poem by Yates is a complex collection of letters. It takes an expert to write them and an expert to read them," said Robert Linhardt, lead researcher and a professor of chemistry and chemical biology at Rensselaer Polytechnic Institute. "We've trained a machine to quickly read the equivalent of words with four letters like 'ababab' or 'bcbcbc.' These are simple sequences that don't have any meaning, but they show us that the machine can be taught to read. If we extend and develop this technology, it has the potential to sequence the glycans or even proteins in real time, eliminating years of effort."

Commercial nanopore sequencing devices are used to sequence DNA, which is composed of four nucleic acid units, known by the letters A, C, G, and T, strung together in an endless variety of configurations. The device relies on an ionic current running through a hole only a few billionths of a meter wide in a membrane. Strands of DNA are placed on one side of the hole, and drawn through with the flow of the current. Each nucleic acid blocks the hole somewhat as it passes through, disrupting the current and yielding a particular signal associated with that nucleic acid. The devices, currently used for fieldwork, are only one of several relatively rapid and automated techniques for sequencing DNA.

Glycosaminoglycans, or GAGs, are a structurally complex class of

glycans—the essential sugars present in living organisms—found on the cell surfaces and extracellular matrix of all animals and perform many functions in cell growth and signaling, anticoagulation and wound repair, and maintaining cell adhesion. GAGs, currently extracted from slaughtered animals, are used as drugs and nutraceuticals.

Like DNA, GAGs can be subdivided into their constituent disaccharide sugar units. But whereas DNA is made of only four letters in a linear string, these glycans have dozens of basic units, some with attached sulfate groups, acid groups, and amide groups. For example, even a relatively small naturally occurring heparan sulfate molecule of six sugar units could have 32,768 possible sequences. Because of the challenge, glycan-sequencing remains onerous, relying on painstaking lab work and sophisticated analysis, involving techniques with names like liquid chromatography-tandem mass spectrometry and [nuclear magnetic resonance spectroscopy](#).

As part of his work, Linhardt, a glycans expert who developed a synthetic variant of the common blood-thinner heparin, sequences GAGs to understand naturally occurring forms and develop synthetic variants.

"Using standard analytical methods, it took us two years to sequence the first simple GAG," said Linhardt, a member of the Rensselaer Center for Biotechnology and Interdisciplinary Studies. "We have another one that we've worked most of the sequence out, and it's taken us five-plus years—and it's probably going to take us another five years to finish it,"

Reasoning that nanopore sequencing could be used to identify the disaccharide units in a GAG, the research team built its own nanopore device and synthesized four heparan sulfate GAG chains using the chemoenzymatic process developed by the Linhardt Lab. Importantly, these four heparan sulfates were very simple—made with combinations of only four different types of sugar units, assembled in a chain about 40

units long, and with a carefully controlled composition and sequence.

The team passed each heparan sulfate through the nanopore and producing a graph depicting voltage over time output of the device. Each of the four variants was run through the device more than 2,000 times, increasing the statistical likelihood of an accurate read given the rudimentary design of the experimental nanopore.

"The device sequenced the simplest heparan sulfate in real time and produced a pattern that our eyes could easily recognize right away for each of the four samples," Linhardt said. "You can immediately tell that they're different."

To ensure an unbiased analysis, the team fed the results into free machine-learning and [image recognition software](#) using Google's deep neural network, training the software to distinguish between the four different patterns and identify each variant of heparan sulfate. The most successful machine-learning model produced an analysis that was nearly 97% accurate.

"The information content in a GAG sequence can greatly surpass that of a similar quantity of DNA or RNA, meaning that the ability to rapidly read GAG sequences opens up a new window of understanding into the complex biochemistry of life" said Curt Breneman, dean of the Rensselaer School of Science. "This proof of concept study links innovative nano-detection methods with state of the art machine learning tools, and shows the power of interdisciplinary thinking to expand the frontiers of knowledge."

Reducing the speed at which the GAGs pass through the nanopore could increase accuracy, and the device can be trained on additional sugar units, and more complex sequences, all of which are future research goals. Linhardt said the machine would have to learn somewhere from

10 to 20 sugar units to fully sequence a GAG.

"This is a proof of concept; we've made it read two-letter words," Linhardt said. "Once we teach it the full alphabet, it will be able to read every different sequence. It'll be able to read all the words."

**More information:** Ke Xia et al., "Synthetic heparan sulfate standards and machine learning facilitate the development of solid-state nanopore analysis," *PNAS* (2021).

[www.pnas.org/cgi/doi/10.1073/pnas.2022806118](https://www.pnas.org/cgi/doi/10.1073/pnas.2022806118)

Provided by Rensselaer Polytechnic Institute

Citation: Time needed to sequence key molecules could be reduced from years to minutes (2021, March 8) retrieved 12 May 2024 from <https://phys.org/news/2021-03-sequence-key-molecules-years-minutes.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.