BIOCHEMISTRY

Sequencing for carbohydrates

A technique combining mass spectrometry and infrared spectroscopy shows promise for unambiguous sequencing of linear oligosaccharides.

DNA and protein sequencing methods, the first of which were developed decades ago, are crucial techniques for biologists. But, despite the essentiality of carbohydrates for cell structure and cell communication, the development of a comparable approach for carbohydrate sequencing has lagged far behind.

DNA is a chemically simple polymer, made up of just four nucleotides. In proteins, the 20 amino acid building blocks are strung together in linear chains by a single type of chemical bond. Carbohydrates, however, are another story. Monosaccharide building blocks (of which there are many) come in series of isomers that differ only in the arrangement of their atoms. Monosaccharides are linked together via glycosidic bonds, but such linkages can vary in both stereochemistry and regiochemistry. These variations are relevant to biological function and therefore are important to capture. But this chemical complexity—not to mention the branched structures of many carbohydrates—has made it challenging for researchers to develop a sequencing approach that can resolve carbohydrate structures without ambiguity.

"The development of all kinds of analytical strategies [for carbohydrate science] is slowed down in the absence of a robust, routine [sequencing] technology," notes Isabelle Compagnon of the Université de Lyon in France. Currently, the only method for complete carbohydrate structural analysis is nuclear magnetic resonance (NMR) spectroscopy, which is slow and requires specialized expertise. Many groups have been working on developing more practical and high-throughput methods for carbohydrate sequencing, several based on mass spectrometry (MS), but none of the methods developed to date have been demonstrated to resolve all possible isomerisms.

Compagnon and her colleagues recently reported a promising new method for oligosaccharide sequence analysis that combines the strengths of MS with infrared (IR) spectroscopy. "On one hand, spectroscopy provides direct and refined structural detail at the atomic level but requires a relatively large amount of thoroughly purified sample," says Compagnon.

GENOMICS

MULTISCALE 3D VIEW OF THE GENOME

Ultradeep sequencing of Hi-C data from mouse cells in different developmental stages can address outstanding questions about 3D chromatin function.

The last 10 years have seen powerful techniques to analyze a genome; not just its linear sequence of bases, but its organization as a meticulously organized, 3D structure. But how much this organization matters remains unclear.

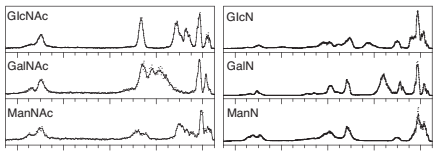
"We know that the 3D structure of the genome is complex, hierarchical, highly organized and it does affect gene expression to some extent," says Giacomo Cavalli from the CNRS and the University of Montpellier. But then he adds an essential caveat: "we don't know by how much." Whether most genes need 3D organization in order to function or whether most can work independently of conformation are still open questions, according to Cavalli.

Various renditions of chromosome conformation capture methods, such as Hi-C, yield DNA–DNA contact information at different resolution, from entire active and inactive compartments, megabases in length, to topologically associating domains (TADs) at 0.5–1 Mb, to smaller scale loops several kilobases long. Cavalli and his team wanted to reconcile these methods that operate at various scales using only Hi-C data to quantify the relative importance of the different contacts during neuronal differentiation.

The team used pure populations of mouse embryonic stem cells (ESCs), neural progenitor cells (NPCs) and fully differentiated cortical neurons and ensured that the ESCs and NPCs were in the G1 phase of the cell cycle to match the G1/G0 phase of the postmitotic neurons. To capture as many contacts as possible, they sequenced to a depth that Cavalli describes as mind boggling. "We sequenced about 50 billion paired end reads. But that is what you need to do if you want to have a really high resolution map in all stages."

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

"On the other hand, MS readily applies to complex samples in small amounts with limited sample preparation, but has a poor structural resolution for carbohydrates since it does not intrinsically disambiguate all types of isomers." By combining the two complementary technologies into a single instrument, the group developed a technique that reliably resolves carbohydrate isomerisms and may be developed into a robust sequencing approach.



Monosaccharide building blocks of carbohydrates can be distinguished by their IR signatures. Image reprinted with permission from Schindler *et al.* (Springer Nature).

Compagnon and her colleagues systematically analyzed monosaccharide and disaccharide standard compounds and found that these compounds can be distinguished by their unique, gas-phase IR spectral signatures. Starting from an oligosaccharide, tandem MS generates monosaccharide fragments in a sequential manner. By capturing IR spectra for each of these fragments in turn, an oligosaccharide sequence can be reconstructed by matching the fragment IR spectra to reference monosaccharide spectra. As a demonstration, the team applied the MS IR method to sequence a crude mixture of linear chito-oligosaccharides. Importantly, they found that the memory of the glycosidic bond stereochemistry was maintained throughout the MS manipulations. At the moment, it remains to be seen whether the technique could be further developed to resolve longer and more complex branching carbohydrate structures.

Compagnon's team is currently working with a prototype instrument, an ion-trap mass spectrometer modified with an IR laser to capture IR spectra immediately following molecular fragmentation. Using this instrument (currently available to external users at the OptoLYSE facility at Université de Lyon) requires an operator with both MS and laser technology expertise. Their goal for the next prototype is to automate spectroscopy data collection and analysis; such an instrument could be readily operated by a typical MS user.

Allison Doerr

RESEARCH PAPERS

Schindler, B. *et al.* Anomeric memory of the glycosidic bond upon fragmentation and its consequences for carbohydrate sequencing. *Nat. Commun.* **8**, 973 (2017).

As a result of their ultra-deep sequencing, the researchers saw a wide range of contacts, from chromosomal compartments to precise promoter–enhancer contacts in their Hi-C data. They also identified a type of contact not described before, "a contact that involves broad regions that are the whole coding parts of the gene and are more abundant for highly expressed genes," says Cavalli. Their role is not fully clear yet, but Cavalli speculates that they are involved in compartmentalizing splicing components.

By focusing on promoter–enhancer interactions, Cavalli attempts to tease out a connection between genome structure and function. "There was a big debate," says Cavalli, regarding this question: "is the opening of an enhancer separated in time and space from the opening of the promoter, or is it happening at the same time?" Their data show that once an enhancer is open for contact, it will not take long before it finds its target gene and activates it. Cavalli is quick to add that their temporal resolution is days, not minutes, and needs to be refined. Still, these results have important implications for deciphering the regulatory landscape of the genome. "Once you find a promoter contacting an open chromatin region you basically identified a very strong candidate for a regulatory region," explains Cavalli.

He sees 3D genomics as the next frontier—not just for its role in gene expression, but also for its contribution to DNA replication or repair, paired with an understanding of how much each layer, from compartment to domain to loops, is contributing. Cavalli predicts, "We will see a lot of this in the coming years."

Nicole Rusk

RESEARCH PAPERS

Bonev, B. *et al.* Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.