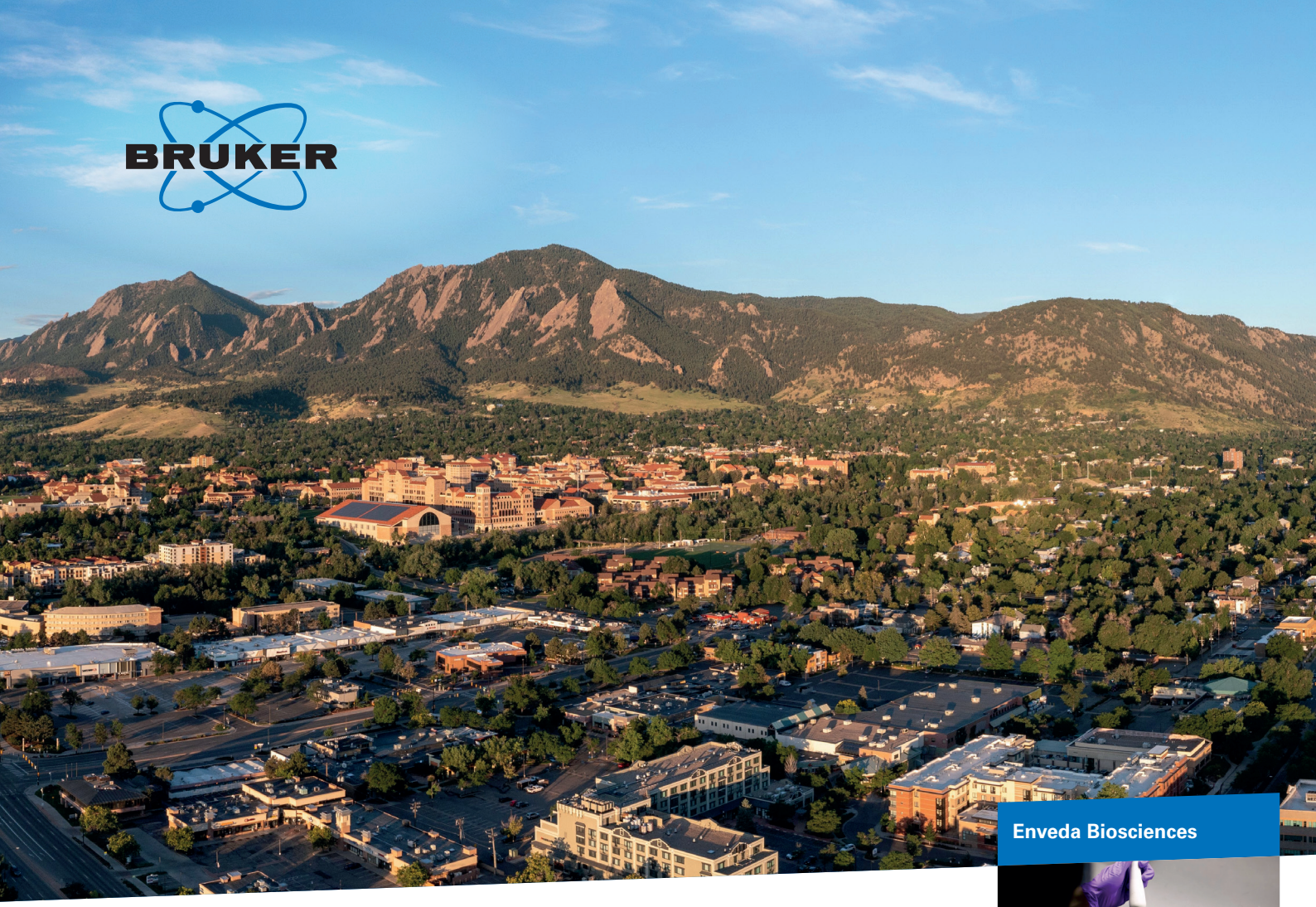


Enveda Biosciences

Natural products are the greatest untapped opportunity for drug discovery. Enveda Biosciences built the tools to capture it.

Customer Insights

TIMS-MS-driven machine learning enables Enveda's industrial-scale drug discovery platform

Researchers at Enveda Biosciences are translating nature into medicine with machine learning and trapped ion mobility tandem mass spectrometry (TIMS-MS/MS). Using Bruker's timsTOF Pro as a core element of their technology suite to identify, annotate, and utilize the structure and function of life's chemistry, the company's vision is to lead a shift from screening to informed database searching as the first step in drug discovery.

Working with Bruker

Enveda Biosciences has relied on Bruker timsTOF technology since 2021 to investigate the chemical composition of complex samples collected from natural sources. August Allen, Chief Technology Officer at Enveda Biosciences describes the relationship:

"Bruker's pace of innovation is astounding. As a business, it is exciting to partner with their cutting-edge technology for metabolomics research and we are proud to align ourselves with the innovation Bruker is driving. It enables us to better separate and identify the unique metabolites in our samples to prioritize new candidate drugs."

Introducing Enveda Biosciences

Founded in 2019 by Viswa Colluru, Ph.D., Enveda Biosciences is using a combination of breakthroughs in metabolomics, machine learning, automation, and biological assays to revolutionize natural product drug discovery. Currently with 230 employees, the drug discovery team of biologists, chemists, software engineers, and data scientists is harnessing the world's largest and most diverse biologically annotated dataset of plant chemistry to discover the next generation of impactful small molecule therapeutics.

Natural product drug discovery has traditionally been hampered by a requirement to isolate and study potentially therapeutic molecules one at a time, making it a slow, expensive, and failure-prone endeavor. Enveda's cutting-edge technologies and methods enable the rapid assessment of biological activity, chemical structure, and drug-like properties directly from mixtures in parallel, massively speeding the characterization step and advancing the most advantageous molecules for further study.

Using this novel approach, Enveda is demonstrating that drug discovery can be initiated by an informed database search – querying chemical structure, biological activity, and organ distribution – to rapidly search for new leads.

Plant powered drug discovery

Natural products, particularly phytochemicals – molecules derived from plants – have historically played a key role in drug discovery due to their incredible diversity of chemical properties and biological activity. They are considered a valuable source of lead compounds for drug development, especially for cancer and infectious diseases (1). Despite several decades of declining interest from biopharma – where the emphasis has shifted to screening enormous libraries of synthetic molecules with predefined chemistries – natural products still represented 30% of the new small molecule drugs that came to market between 2000 and 2020, showing significantly larger success rate relative to investment (2).

Enveda recognizes that molecules from plants represent an enormous untapped library of drug-like chemicals that can be developed into novel small molecule drugs. Currently, however, most phytochemicals are unknown to science and the methods to identify novel molecules from natural sources are extremely arduous. Enveda's approach is to build and utilize a unique database of natural compounds, starting with plants, leading to rapid discovery of previously unidentified natural products to generate new medicines for patients suffering from a range of diseases.

Why natural products?

An emphasis on natural products as a source of medicines is well supported in the literature. For example, a recent review (March 2022) looked at the case for a renewed focus on nature in drug discovery, highlighting that drugs that more closely mimic natural products have conserved recognition elements and properties associated with them – the hypothesis being that they are working in an evolutionarily optimized manner (3).

In other work, researchers observed that there is a greater translatability of natural products. For example, chemical space, defined by high fraction sp^3 (Fsp^3), high stereochemical content, high oxygen content, high ring content and low aromaticity (properties enriched in natural products) correlated with increased progression through clinical trials. Natural products occupy areas of chemical space unpopulated by commercial, synthetic molecules and drugs as defined by high Fsp^3 (4).

Importantly, when complex chemical samples collected from natural sources are subjected to the best analytical techniques for recall of known structures contained in the sample, less than 10% return a match (5). This implies the existence of a large, untapped pool of chemistry derived from billions of years of evolution.

The Enveda workflow

Despite the evidence that natural products are a successful source of potential drug leads, the process to discover such molecules poses several challenges. The first is the scope of unknown chemistry. As mentioned above, most natural products are unknown to science, meaning that we know nothing about their identity, including their chemical structure. The second is the scope of unknown biological function. Enveda's platform routinely uncovers ten thousand distinct molecules per plant sample, each of which can interact with human biology in a range of ways.

To overcome these challenges and to find the very best molecules to progress as drug candidates, the Enveda platform first uses liquid chromatography to separate complex mixtures into fractions. A portion of each fraction is sent to the timsTOF instrument for structure prediction, while the other portion of the fraction is tested in dozens of different biological assays to determine activity. Statistical models are then used to deconvolute the data and determine which molecules are responsible for which biological activities. The molecules with the right combination of beneficial activity, low toxicity, and amenable structures are then prioritized for isolation.

Allen says: *"Isolating individual compounds from complex mixtures is a slow and expensive process that has plagued natural product drug discovery historically. Oftentimes you may end up re-isolating a compound that is already known or incorrectly isolating a compound that does not cause the underlying bioactivity.*

The power of Enveda's platform lies in our ability to prioritize the most interesting compounds to isolate based on structure and function annotations of complex mixtures."

Allen adds: "*Our function annotations rely on a mixture of industry standard functional assays alongside our in-house bioactivity assays, all formatted in 1536-well plates, so we can annotate bioactivity on hundreds of thousands of compounds."*

Bioactive components of natural products tend to be present in low abundance, making it difficult to isolate large enough quantities to perform structure analysis using the gold-standard technology, NMR. Structural analysis is essential for prioritizing drug candidates, though, as it determines synthesizability, amenability to medicinal chemistry, and many pharmacodynamic and pharmacokinetic properties. To overcome this challenge, Enveda has pioneered new ways of analyzing the data produced by high resolution mass spectrometry using machine learning. The key to identifying natural product structures is the combination of clean MS/MS fragmentation patterns and ion mobility data generated from Bruker's timsTOF Pro 2 mass spectrometers.

"Our structure annotations rely on our machine learning models, MS2Mol and MS2Prop, which learn the language of mass spectrometry to accurately predict the structure and properties of compounds in complex mixtures.

We use multiple timsTOFs to collect mass spectra on all our samples quickly. The Bruker timsTOF Pro 2 is central to our whole process."

MS/MS-based metabolomics can take a mixture of compounds extracted from a natural source, up to 1000s of molecules at a time, separate them using chromatography and pass them through a tandem mass spectrometer (5).

Bruker's timsTOF Pro 2 systems separate molecules based on their size and shape while enabling the measurement of collisional cross section (CCS) area. As CCS values do not change based on the ionization source, each measurement is a unique characteristic of that molecule. TIMS is then coupled with MS/MS fragmentation in a mode called parallel accumulation serial fragmentation (PASEF). As mobility separated ions are "eluted" from the TIMS cell, they are sent to the time of flight (TOF) mass analyzer either directly to collect accurate mass data for the precursor or alternatively fragmented to collect clean mobility separated MS/MS data.

The first stage (MS1) measures the mass of the individual compounds and their abundances, while the second stage (MS2) fragments the compounds into pieces and measures the mass and abundance of each fragment piece. Metabolites are mainly studied through MS/MS in which molecules are ionized and then broken apart in a collision cell. The masses of the resulting pieces are detected and compiled into an MS2 mass spectrum.

One of the main challenges with metabolomics is interpreting the MS2 spectrum. The mass of a compound and the mass of its component fragments can be detected, but this becomes complicated when combining them into an identifiable chemical structure.

Multiple species may be present in a chromatographic *m/z* peak so to further separate and characterize these metabolites, PASEF is used to produce mobility offset, mass aligned (MOMA) MS/MS fragmentation patterns, perfectly aligned in a horizontal plane.

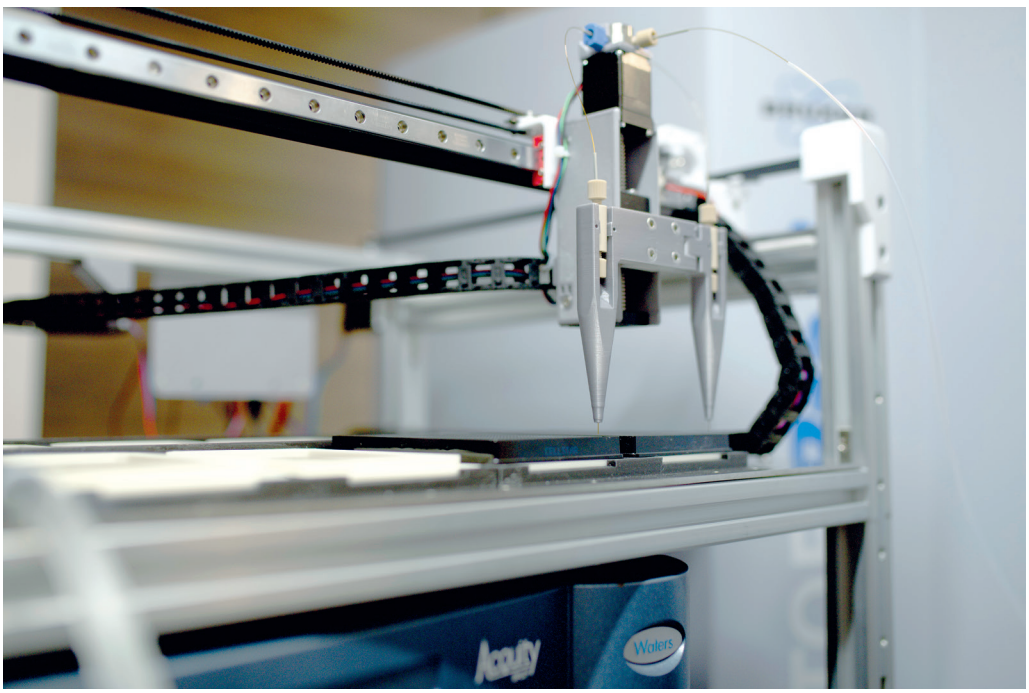
Enveda Senior Scientist Pelle Jan Simpson says: "*The challenge in the industry is no longer the acquisition of MS/MS spectra itself but the rate of acquisition and the quality of the spectra.*

The timsTOF platform performs well with both and provides a single-point encoding of a molecule's shape in the form of its collisional cross-section, which can be utilized to link molecules between acquisitions even when separations are varied."

Each MS/MS spectrum is coupled with Enveda's unique machine learning workflow that predicts the structure; these metabolites are further overlayed with their bioactivity so scientists can focus on molecules that have the greatest therapeutic potential.

The result is a map of a plant's natural product chemical space including bioactivity. To achieve this breakthrough in data processing, Enveda developed MS2Prop and MS2Mol – machine learning models that directly predict chemical properties and chemical structure, respectively, from mass spectrometry data.

MS2Prop and MS2Mol are built on a neural network architecture called transformers that was originally introduced to capture linguistic structure over entire passages of text. The team hypothesized that it could be ideal for MS/MS spectra, which lack a straightforward sequential or spatial dependency between the peaks. This makes them a poor fit for traditional deep learning via convolutional or recurrent neural networks (CNNs or RNNs). Transformers, on the other hand, could have application in deconvoluting MS/MS spectra. Their self-attention layers allow for learning complex dependencies based on the identity of the fragments alone without locality or ordering assumptions inappropriate for MS/MS data. Much as the transformers used to power ChatGPT were able to learn the rules of grammar, the transformers used in MS2Mol learned the grammar of the peaks in MS/MS data, recognizing patterns of peaks as corresponding to specific functional groups. In this way, MS2Mol can directly translate between the language of MS/MS and the language of chemical structure, providing Enveda's medicinal chemists with the information they need to make prioritization decisions.



From Screening to Searching

To tackle the unique challenge of working with mixtures of unknown and highly diverse natural compounds, Enveda redesigned the early-stage process, creating a powerful new approach, and a platform for industrial scale drug discovery.

Applying machine learning on the companion Bruker timsTOF Pro 2 data allows accurate prediction of compound structures.

Allen concludes: *"We are incredibly proud of our work, but it is just the beginning. We will continue building the largest metabolomics dataset, purpose-built for machine learning and applying active learning strategies to help us identify and characterize the mass spectra whose identity is most likely to improve our drug development programs."*

> *"What is also very exciting is that we see clearly how our approach could be extended beyond the world of plant biology to be applied more generally in drug discovery."*

References

[1] Atanasov AG, et al. (2015). *Discovery and resupply of pharmacologically active plant-derived natural products: a review.* Biotechnol. Adv. **33**,1582–1614.

[2] Newman D, Cragg GM. (2020). *Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019.* https://pubs.acs.org/doi/10.1021/acs.jnatprod.9b01285

[3] Young RJ, et al. (2022). *The time and place for nature in drug discovery.* https://pubs.acs.org/doi/full/10.1021/jacsau.2c00415

[4] Huffman BJ (2019). *Natural Products in the 'Marketplace': Interfacing Synthesis and Biology.* J Am Chem Soc. February 27; **141**(8): 3332–3346.

[5] Allen A (2022). *Transforming natural product drug discovery: machine learning for high-fidelity chemical property prediction from metabolomic data.* Medium. https://medium.com/@august.allen_51342/transforming-natural-product-drug-discovery-machine-learning-for-high-fidelity-chemical-property-94da4cd8ca73

Customer Insights

TIMS-MS-driven machine learning enables Enveda's industrial-scale drug discovery platform

Researchers at Enveda Biosciences are translating nature into medicine with machine learning and trapped ion mobility tandem mass spectrometry (TIMS-MS/MS). Using Bruker's timsTOF Pro as a core element of their technology suite to identify, annotate, and utilize the structure and function of life's chemistry, the company's vision is to lead a shift from screening to informed database searching as the first step in drug discovery.

About Enveda Biosciences

Enveda is harnessing the complexity of the natural world to tackle today's biggest healthcare challenges. With breakthrough advancements in machine learning, computational metabolomics and knowledge graphs, the company is discovering the next generation of small molecule therapeutics.

For more information, please visit: **www.envedabio.com**

About Bruker Corporation

Bruker is enabling scientists to make breakthrough discoveries and develop new applications that improve the quality of human life. Bruker's high-performance scientific instruments and high-value analytical and diagnostic solutions enable scientists to explore life and materials at molecular, cellular and microscopic levels. In close cooperation with our customers, Bruker is enabling innovation, improved productivity and customer success in life science molecular research, in applied and pharma applications, in microscopy and nanoanalysis, and in industrial applications, as well as in cell biology, preclinical imaging, clinical phenomics and proteomics research and clinical microbiology.

For more information, please visit: **www.bruker.com**

Bruker Switzerland AG

Fällanden · Switzerland
Phone +41 44 825 91 11

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com

Learn more at **www.bruker.com/timstofpro**