Cambridge Isotope Laboratories, Inc.
isotope.com

Isotope Days 2024- Will Thompson, PhD
Utilization of Stable Isotope-Labeled Metabolites for Automation in Data Processing in Microchip CE-MS Metabolomics

Andrew Percy, PhD (00:00):

The final speaker is Dr. Will Thompson. Will is a principal scientist of 908 Devices in the life sciences r and d group. In his work at 908, he is predominantly focused on developing tools for automating data collection and analysis for metabolomics and bioprocessing applications using microchip CEMS platforms. In his career, he has amassed over a hundred peer reviewed articles across scientific areas, and today he will discuss some of the latest metabolomic developments and applications conducted at 908 using stable isotope labeled metabolite standards together with CEMS. Will the floor is yours.

Will Thompson, PhD (00:49):

Great. Thank you so much Andrew. Pleasure to be with everyone here today. That and clean up. So can you just confirm you can see my slides, Andrew?

Andrew Percy, PhD (01:03):

Yes, we can.

Will Thompson, PhD (01:04):

Okay, great. Thank you. Yep. So good morning everyone, or good afternoon, wherever you may be. So today I'm just going to be talking about how we've partnered with Cambridge isotopes to develop a custom stabilized isotope labeled mixture for use in our CEMS metabolite quantitation work. I'll give a quick intro to zip chipp and microchip CE ms, in case you're maybe not familiar, I'll talk about how we're using this custom stable isotope labeled mix and then do a little bit of a deep dive, an academic dive if you will, into automated data processing and how we're using this to save significant time in our data analysis and few application vignettes. So very quickly, many folks may not be familiar with how CEMS is maybe used or commercialized by 908, but the zip chip is a front end capillary electrophoresis system. It's on the top left, you can see the zip chip system and how it interfaces with the thermo explorers two 40 inside the box, which contains all the electronics and power supplies and everything fluidics needed to run the system.

(02:17):

There's a little chip and that glass microchip slide has incorporated a separation channel. CE drives the separation based on a voltage gradient, so a separation across the two channels and then we electro spray directly off the corner. So high resolution separation directly compatible with thermo Bruker and Sciex spectrometers and has some advantages from the perspective of separating particularly polar compounds because we don't need to retain the compounds and we separate, we do a nice job with CCE of separating salt and other matrix effects away from the sample. So we're able to do direct dilution and analysis from complex matrices such as plasma serum and biofluids media, those kinds of things. Okay, so what we're working to do here at 908 and the group that I work in is couple the hardware,

which we've been making some time with some novel new consumables and software in order to really round out the portfolio of end-to-end workflows.

(03:22):

And so the partnership with Cambridge has been very nice partnership so far and I've worked with them to develop a 36 components stable isotope mixture, which we've shown here on the slide. And this is a component of a new consumables package that we're working on, which includes system suitability standards, the stable isotopes, and then blanks, calibrators, and quality controls along with all the other consumables needed to drive the separation and analysis. Currently this is being mainly work done on high resolution mass spectrometers with thermo orbit straps, but there's no technical reason why it couldn't be done on other instruments as well.

(04:06):

So what's in the mix? So currently there are two or a number of main components, but two main uses. So we have the stable isotope labeled internal standards. 19 of them are used as indexing, what we call indexing compounds to correct for any migration time drift. We utilize a standard indexing approach common in GCMS and also somewhat in reverse phase chromatography. So we measure these 19 compounds and we calculate, we have an index calculated for them and all of the other unknown compounds are indexed associated with those compounds. So we take the index time or the measured time, we calculate an index and then we sort of know then where to look in real time for all of the unknown compounds. It gives us a nice handle for making sure we're integrating the correct peaks for quantitation. Included in the mix are 20 amino acids, eight non-canonical amino acids, and eight acylcarnitines.

(05:15):

So where do these stable isotope labeled standards play a role? So I mentioned that the sample preparation is quite simple and fast. So of course these stable isotopes are in a methodol ammonium acetate mixture and we simply do a precipitation, a protein precipitation using that on a 0.45 or one micron filter depending on the approach. And that 10 x dilution of most samples, we use a hundred x dilution for subculture media. The system is since enough to do that is then just filtered through a filter plate collected in the bottom of a 96 well plate for analysis. We also, so that corrects for matrix effects and iron suppression and those kinds of things. We of course also use it to help with calibration. So we use it to normalize the peak area and then combine with our external calibration materials, we can then build a proper calibration for quantifying the metabolites we measure. And then the final, which I'll spend some time on today, is really using those isotopes to improve the speed of our data analysis. So an example shown in the top left is isoleucine leucine and allo isoleucine, which are the same mass. So we have to utilize some very intelligent sort of algorithms to make sure that alo can pick up the correct peak for the purpose of quantitation.

(06:43):

So really there's three time consuming areas, spending a lot of time on myself and the service area, providing data to customers in the past system, suitability, deciding a note, go no go before data collection, data quality checks. So deciding if the data's high enough quality and whether or not you need to recollect anything. And then quantitation are three sort of areas where you have to make time consuming decisions manually, often. And so today I only have really time to go over two of them. So I'm going to talk about the data quality check and the quantification piece. So the way I look at this is that typically when we're doing a mass spectrometry experiment, we go in and we will have some sort of manual check after the data is collected where we say, okay, do we need to do anything? Is the data high quality?

(07:37):

We'll click through and look at each file manually to make sure to do an evaluation of this disadvantages of this. It is slow, it is usually a spot check and it's not a quantitative assessment. So we make some subjective assessment of this. So we'd like to automate this and improve this process. So in this case we've just, I believe that the simplest set of criteria that works is the way we want to go. So in this case, we've utilized four of the compounds out of our 36 component mixture and we make a couple of basic assessments. So the software goes in, it looks at the raw data and it says, do I measure these four peaks? Do I see the relative migration time of these two in the middle or within some range we need so we can make sure that the automated data analysis can proceed if they're in a certain range.

(08:29):

And when we have a signal threshold cutoff to where we say, okay, can we measure them reproducibly? And so if we make that measurement, then what we want to do is we have assessment of a peak area, for instance, where we may have had a missed injection and then the software can go in and say, okay, if you've missed an injection, it looks like you missed one. Would you like to recollect? So you click a button, it can look at those four standards, they should be in every sample. And then we have the ability to recollect them in an automated manner.

(09:06):

We provide a skyline file for instance, that looks at those particular compounds that allows the user to go back in and do and do a sanity check if you'd like to look at it manually. The second most common and tedious problem that I spent literally lost, I feel like years of my life to do this, is to really do this peak assignment correctly. So on the left I've shown an example from our separation, but this also occurs of course in every mode of separation where we have Isobaric species. And most software will then look at this and particularly difficult to always assign the correct peak when you have a software solution. So for instance, if you've got isoleucine leucine and allo isoleucine, even if the migration time predicted is closest to say isoleucine in this case, software will often there's a little bit of error in the prediction and then software will often pick the highest, most abundant peak in that area.

(10:11):

And then you have to go in and manually correct this peak integration. This happens across lots of virtually ever every software package that I'm familiar with. And so in cases where it gets it incorrect many times and then it can get it correct sometimes, but often when it correctly assigns it, it's kind of by chance it would be in these cases these correct assignments occurred just because the peak on the left happens to be the most abundant. So we've done a lot of work to understand why this is, and really it happens for two reasons because the software most often has no knowledge of the relative evolution time of the analytes. And it also doesn't know whether or not the peak has already been assigned to another analyte because each compound is essentially assessed on its own. And so we're using the stable isotopes and IRT, we can come up with an intelligent approach to solve this problem.

(11:09):

So we started doing this back in 2021 quite some time ago with just buying the NSKA approaches and building these indexing. So the first step is just to correct any migration time variability you see across the study. So in CE we see some variability due to salt content. It doesn't change the order of the compounds alluding, it just may change how long it takes for those salts to be separated from the matrix itself. And so it delays the separation a little bit. So by using indexing as you see on the right, we can decrease the migration time variability across the study from a little bit of wiggle in these measurements to the bottom where basically no variability between the runs. We've now expanded this from the NSKA mix to now using this custom 36 component mixture. This is an example across four different laboratories where we might see some between matrices between samples.

(12:12):

These are kind of real experiments run across serum urine plasma and a calibration materials run across the six experiments in four different labs. We can see there is some variability which makes it pretty difficult to use absolute time as the measurement for whether or not a compound is correct. However, if we translate that to the index standard indexing standards, and on the bottom right you can see that trigon lean always eludes between prolene three at a specific location. And so if that time in the index space is quite reliable as far as a measurement of it, whether or not the compound is selected for integration correctly. So that can reduce the variability between laboratories, between chip devices, et cetera, to a spot where it's quite reliable predictive use, I'm sorry, predictive measure of identity. The second problem of course is even if we do that, there can be some slight wiggle between the prediction.

(13:16):

And so we need to fix this problem of which peak is assigned to the right matrix. So as a human viewing this, we can know that the small peak on the right is always alloy lucine, for instance. And so we go through and we just manually integrate the little one on the right if we need to do it correctly. So the challenge is how do we tell software to do that? And so in this case, this problem as we pitched it to our software colleagues is known as the linear sum assignment problem. So the problem basically states you've got a set of identities and a set of peaks and you need to assign them such that the minimum distance, the total distance between the library and the observe is minimized. And then you also need to make sure that no two compounds are assigned to the same peak.

(14:05):

And so this problem has a number of different solutions and we've adopted one called the Hungarian algorithm to do this. Of course, this top right is an oversimplified version of the scenario. So sometimes you can have one of these peaks is missing, so your mismatch between the number of compounds and the number of peaks, but we solve that, that's where the isotope labels come in because we can gain, even if this peak is not known to be iso leucine for instance, we can assign it correctly using also peaks that are out of the same M over Z axis. So using the stable isotopes, we have that handle on these specific compounds and helps us to add where the situation is more complex than I've shown. So how does this work? Just a few examples here where we show the separation on the left where beane has a difficulty oftentimes because of so much lower than the amino acid that shares its mass. And so we can correctly assign it on the top right homoserine, allo threonine and threonine all share a mass. And so we can make sure we assign homoserine correctly. And on the bottom right is how we can reassigning alloy lucine. So in this case, with an automated workflow, we're taking something that virtually is almost never assigned correctly using standard software and doing it nearly a hundred percent of the time correctly with automation.

(15:44):

Just as a challenge to myself in this case, how much time can this save? So this is a study where we collected a to bioreactor study, a very common set of samples for us to collect 90 samples, 110 runs, including all the calibrators and quality controls, 300 targets and 142 measured metabolites. It took me about four hours to get through this with most of that being hands-on time for doing the curation. With the automated process, we feel like we can generate a 10 x improvement on the overall processing time, which is what we expect to be able to do with incorporating both the isotopes, the index alignment, and the automated algorithm approach. So two quick vignettes of application here. So we have, this is measurements from a urine study presented last year at ASMS from Corinne Moss and Josh Coon's labs. So the different colors on the plot on the left are all different individual patients.

(16:46):

Longitudinal urine study collected with different challenges that they were presented with diet and sleep deprivation in their laboratory. The black dots are the study pools that were run on each plate, and I've grouped them together at the beginning of each plate just so you can more easily visualize the difference between technical and biological variability. And so what we see is without any other corrections, so there's no post correction. This is just the raw micromolar quantitative values that come out of the approach. We're seeing quite high reproducibility between the measurements over time using the combination of the stable isotope labeled internal standards and the external calibration material separation and shown on the top right. And then of course, traditional principle components analysis, looking at the variability between the pools and the biology on the bottom. The last little application vignette is really one of something I didn't know very much about until a year ago or so.

(17:54):

But we've been doing a lot of work into CHO bioreactor media. So traditional kind of amino acids show these trends where amino acids are either utilized as food or in the case of glyc alanine, they may increase over the course of the bioreactor and somehow glutamate included in the feed. You may see this kind of spiking up and down of glutamate, but when we utilize this automated approach CEMS with a much larger panel, we can see much more interesting and complex trends as a function of time. So we have these compounds that increase over time with biomass or viable cell density. We have of course the fuel amino acids. We have some that show correlation with the lactate switch. And then we have others that have a peak either early what we call an early peak or a late peak, which are correlated to late exponential phase or stationary phase.

(18:48):

And then others which show more, much more complicated trends, which are quite interesting. As we kind of expand out more dense sampling, we start to see very complex trends in their quantitative expression as a function of bioreactor. So with this I'll close. So just to say that I'm excited about the work that we're doing with Cambridge isotopes to provide a solution with drastic reduction in time to result, mainly based on the automated data processing and data handling, the stabilized label internal standards. Working with Cambridge has been excellent to generate a nice custom mix that fits our purposes for the purpose of this new product. And then if you're interested in learning more, we are going to be releasing a product next year incorporating much of this work. And so if you're interested in hearing more, I encourage you to reach out to me directly and hopefully provide my information or information in the notes. Thank you.

Andrew Percy, PhD (19:54):

Thank you very much Will for that excellent presentation and shedding light into your analysis and interpretation routines for doing metabolomic CMS workflows. We're at the time cap for your session, but maybe a quick question on the data output format from your CMS workflow, if you could provide some indication on that before moving forward.

Will Thompson, PhD (20:24):

Yeah, sure. So currently the data is provided as a micromolar quantitative output file. So the data is just delivered as a matrix of quantitative values in each sample, but also we provide a skyline file on the backend for the purposes of digging into the data if you'd like to see what we've done to the data during our process.

Andrew Percy, PhD (20:50):

Excellent. Thank you again, Will. So if you can unshare. So that brings us to the close of this mass spec Isotope Day 2024 webinar. I'd like to thank all four speakers once again for their excellent presentations and all of you in attendance through your time and attention. Circling back to the goal for today's event, the session provided you with a platform to learn from scientists in the mass spec field who utilize stable

isotpe labeled standards or mixes for different application purposes. The hope was that you learned something new, strengthened your understanding of CIL's label products, and stayed engaged. At the close of the Zoom session, you'll be prompted to complete a short general survey. It would be greatly appreciated if you could take a few minutes to complete that. You should also expect to see in the near future a follow-up email with links to today's talks and other relevant information.

(21:57):

Should you have any questions or product requests, please don't hesitate to reach out via email and or visit the CIL website at isotope.com. This QR code provides a link to the webinar registration page. There you will find details on this and our future isotope day 2024 webinars. The next one in this three part series will be the NMR Isotope Day, which will feature presentations from five scientists. The webinar format will be akin to this Ms. Isotope Day and will take place on Thursday, October 10th at 10:00 AM Eastern Following suite is the environmental isotope day, and that'll take place ensuing week at Thursday, October 17th at 10:00 AM Eastern. So we look forward to seeing you there and thank you once again for all of your attendance and have a nice rest of your day.