LC
Preventing Pitfalls in LC–MS Proteomics Measurements

THEORY & FUNDAMENTALS
Alternative Workflows that Improve LC–MS Peptide Mapping of Antibodies

TECHNIQUE FOCUS
What's the Right Number of Replicates in Biopharmaceutical Analysis?

LC|GC®
north america
solutions for separation scientists

Volume 40 Number 11 November 2022
www.chromatographyonline.com

CELEBRATING 40 YEARS LCGC

GC
James Lovelock and His Electron Capture Detector (ECD)

APPLICATIONS
ANALYTICAL TOOLS FOR AN EXPANDING PFAS UNIVERSE

SAMPLES & METHODS
Test Your Extraction Knowledge for Preparing Samples

HAMILTON

HPLC Columns from Hamilton Company

Excelling in Separation and Purification:

PHARMACEUTICAL | FOOD AND BEVERAGE | INDUSTRIAL | ENVIRONMENTAL | FORENSICS | COSMETICS

Hamilton Polymer HPLC Column Advantages:

- Can be regenerated which saves the scientist time and money
- Have wide chemical capability
- Excellent sample recovery
- Can withstand pH from 1 to 13
- Can be operated at temperatures over 100°C

Hamilton Manufactures a Full Line of HPLC Columns for:

REVERSED-PHASE | **ANION EXCHANGE** | **CATION EXCHANGE** | **ION EXCLUSION**

Learn more: www.hamiltoncompany.com/HPLC

Lit. No. L80120 — 11/2022

Peak performance in every column.

TECAN SPE COLUMNS - FAST. EFFICIENT. SOLID PHASE EXTRACTION.

Modern biopharma workflows require speed and precision to ensure the first answer is the right answer. That's why Tecan SPE columns are designed to streamline your processes, increase your efficiency, and maximize your accuracy. Our automation-ready consumables enable cleaner extracts, lower solvent consumption, minimize errors, and improve reproducibility. Tecan SPE columns deliver peak performance so you can deliver trusted answers, every time.

www.tecan.com/biopharma-SPE

TECAN.

LC GC north america

MANUSCRIPTS: For manuscript preparation guidelines, see chromatographyonline.com/lcgc-author-guidelines, or call The Editor, (609) 716-7777. *LCGC* welcomes unsolicited articles, manuscripts, photographs, illustrations, and other materials but cannot be held responsible for their safekeeping or return. Every precaution is taken to ensure accuracy, but *LCGC* cannot accept responsibility for the accuracy of information supplied herein or for any opinion expressed.

SUBSCRIPTIONS: For subscription and circulation information: *LCGC*, P.O. Box 457, Cranbury, NJ 08512-0457, or email mmhinfo@mmhgroup.com. Delivery of *LCGC* outside the United States is 14 days after printing. (*LCGC Europe* and *LCGC Asia Pacific* are available free of charge to users and specifiers of chromatographic equipment in Western Europe and Asia and Australia, respectively.)

CHANGE OF ADDRESS: Send change of address to *LCGC*, P.O. Box 457, Cranbury, NJ 08512-0457; alternately, send change via e-mail to mmhinfo@mmhgroup.com. Allow four to six weeks for change. PUBLICATIONS MAIL AGREEMENT No. 40612608. Return all undeliverable Canadian addresses to: IMEX Global Solutions, P.O. Box 25542, London, ON, N6C 6B2, CANADA. Canadian GST number: R-124213133RT001.

C.A.S.T. DATA AND LIST INFORMATION: Contact Stephanie Shaffer, tel. (774) 249-1890, e-mail: SShaffer@mjhlifesciences.com.
REPRINTS: Contact Stephanie Shaffer, e-mail: SShaffer@mjhlifesciences.com
INTERNATIONAL LICENSING: Contact Kim Scaffidi, e-mail: kscaffidi@mjhassoc.com

CUSTOMER INQUIRIES: Customer inquiries can be forwarded directly to MJH Life Sciences, Attn: Subscriptions, 2 Clarke Drive, Suite 100, Cranbury, NJ 08512; e-mail: mmhinfo@mmhgroup.com

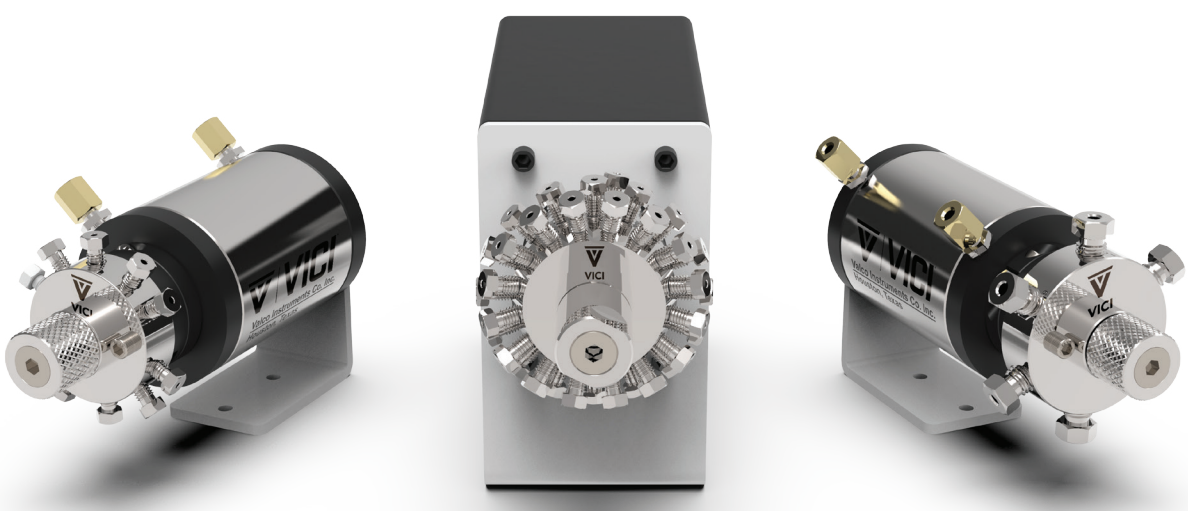
Alliance for Audited Media

50% Recycled Paper 10-20% Post Consumer Waste

485F US Highway One South,
Suite 210
Iselin, NJ 08830
(609) 716-7777

GAS TIGHT GC VALVES FROM THE COMPANY THAT INVENTED THEM

- For injection, stream selection and trapping
- Bores from 0.25mm (0.01") to 4mm (0.156")
- Zero dead volume fitting for 1/32", 1/16", 1/8" or 1/4" tubing
- Alloys and polymer composites to meet virtually any system requirement
- Manual, pneumatic or electric actuation



Scan the QR code to learn more about our GC valves

Learn More

VICI (800) 367-8424 sales_usa@vici.com

CONTENTS

COLUMNS

PEER-REVIEWED ARTICLE

LC GC north america
solutions for separation scientists

Cover Description: Investigating pollution from per- and polyfluoroalkyl substances (PFAS) around the world.

Cover image courtesy of stock.adobe.com/*Vastram*

DEPARTMENTS

LISTEN ONLINE

ANALYTICALLY SPEAKING PODCAST
LCGC | Spectroscopy

www.ChromatographyOnline.com/analytically-speaking-podcast

ALSO AVAILABLE ON
Apple Podcasts, SoundCloud, Google Podcasts

LCGC North America (ISSN 1527-5949 print) (ISSN 1939-1889 digital) is published monthly by MultiMedia Healthcare, LLC., 2 Clarke Drive, Suite 100, Cranbury, NJ 08512, and is distributed free of charge to users and specifiers of chromatographic equipment in the United States and Canada. LCGC is available on a paid subscription basis to nonqualified readers in the United States and its possessions at the rate of: 1 year (13 issues), $91.00; 2 years (26 issues), $163.00; in Canada and Mexico: 1 year (13 issues), $116.00; 2 years (26 issues), $182.00; in all other countries: 1 year (13 issues), $170.00; 2 years (26 issues), $304.00. Periodicals postage paid at Trenton, NJ 08650 and additional mailing offices. POSTMASTER: Please send address changes to LCGC, P.O. Box 457, Cranbury NJ 08512-0457. PUBLICATIONS MAIL AGREEMENT NO. 40612608, Return Undeliverable Canadian Addresses to: IMEX Global Solutions, P. O. Box 25542, London, ON N6C 6B2, CANADA Canadian GST number: R-124213133RT001. Printed in the USA.

Note from the CEO

Mike Hennessy, Jr.
President & CEO, MJH Life Sciences®

The analytical chemistry community is working hard to address the challenges of assessing contamination from per- and polyfluoralkyl substances (PFAS) in the environment, food, and our own blood. Regulatory bodies have lowered the acceptable concentrations for contaminated matrices; this reduction has resulted in the need for state-of-the-art instrumentation to detect sub-ng/L levels. In addition, suspect screening is becoming more popular, as the list of known PFAS grows in the thousands. In this month's cover story, Trever Schwichtenberg, a postdoctoral researcher at the Oregon Health and Science University, shares updates on some of the latest approaches, with the hope that they be adopted by laboratories in industry, government, and academia.

Improving peptide mapping of antibodies is the topic of another paper in this issue. The workflow presented here suggests employing pepsin as an alternative and complementary protease for digestion, along with using guanidine hydrochloride post-digestion. This approach can provide more comprehensive antibody peptide maps, to enable more thorough quality checking of biopharmaceutical products.

In our regular columns, "LC Troubleshooting" explains how to avoid common pitfalls that can lead to poor quality data when performing proteomics measurements by liquid chromatography-mass spectrometry (LC–MS). "Focus on Biopharmaceutical Analysis" considers the question of replication—asking how many replicates we need, and whether the answer depends on the application. "Sample Prep Perspectives" continues the refresher of fundamental principles of sample preparation, focusing on octanol-water partition coefficients as well as practical applications of the process. "GC Connections" looks at the electron capture detector (ECD), which was critical in the discovery of chlorofluorocarbons (CFCs) in the atmosphere and was invented by the recently departed James Lovelock. The ECD is still widely used today, but doing so effectively means understanding the principles of operation and tradeoffs in detection between selectivity, ease of use, and sensitivity. Finally, "Viewpoints" explains why electric field-assisted extraction can improve the sensitivity and selectivity, and is particularly useful with miniaturized and portable applications, and with biological samples.

We hope you enjoy this issue.

LC TROUBLESHOOTING

Pitfalls in Proteomics: Avoiding Problems That Can Occur Before Data Acquisition Begins

The analysis of peptides using liquid chromatography–mass spectrometry (LC–MS) for proteomics applications is powerful and can yield tremendously rich data sets. However, the exquisite sensitivity and global applicability of MS detection also makes it prone to contaminants that can easily and seriously compromise the quality of a data set. A short list of tips and tricks can increase the likelihood of consistently producing good LC–MS data in this context and streamline the troubleshooting process when problems do eventually occur.

Daniel Meston and Dwight R. Stoll

Some applications of liquid chromatography (LC) involve highly specific "tips and tricks" that are acquired by analysts over time, and they can make the difference between producing data that is either terrible or great. The analysis of peptides by LC coupled with mass spectrometry (LC–MS) is one such area where it can appear that highly skilled analysts possess superpowers or "golden hands" that enable them to acquire better data than anyone else in a laboratory. In the end though, careful attention to detail and a thorough understanding of the conditions and chemistry involved in these analyses can help avoid problems before they occur or accelerate troubleshooting problems when they do appear. For this month's installment of LC Troubleshooting, I invited Dr. Daniel Meston to share his short list of knowledge, tips, tricks, and advice related to analyzing peptides by LC–MS. There are many details that are important to consider in the preparation of samples and in the instrument prior to LC–MS data acquisition that can increase the likelihood of acquiring high quality data.

- Dwight Stoll

LC–MS has become the gold standard methodology used for *proteomics*, which is the identification and quantification of protein abundances in biological samples. This is because of the ability of LC-MS to provide sensitive detection for untargeted peptides and identify unknown peptides through tandem MS (MS/MS) fragmentation and comparison of the experimental spectra to theoretical spectra prepared in silico. Subsequently, LC–MS-based proteomics tools have become invaluable in diverse disciplines ranging from immunology to microbiology and food chemistry. However, analysis of biomolecules is generally susceptible to a number of unique and perilous sample preparation pitfalls that can severely degrade the quality of data acquired, even when the most sophisticated LC–MS systems are used, which is because of the unique chemistry and size of large biological macromolecules, such as proteins, peptides, and oligonucleotides to name a few. In this installment, we provide our views on some of the most common pitfalls in sample and instrument preparation relevant to proteomic analyses by LC–MS. We hope that increasing awareness of these potential problems can help users mitigate them and produce higher quality data in general.

Adding Contamination

All mass spectrometric methods require the ionization of the analytes of interest before they can be separated and ultimately detected. For the peptides encountered in proteomic workflows, the ionization step is generally straightforward because conventional tryptic digestion of proteins often results in amino acid chains with terminal amino groups. Indeed, the protonation of peptides is so favorable that they are often observed as multiply charged positive ions. Given the ease of ionization of peptides, MS can almost be regarded as a "universal detector" for peptide analysis. However, this sensitivity can also be a double-edged sword in the sense that the detector will also respond to very low concentrations of easy-to-ionize contaminants in our samples, thus degrading the overall quality of the acquired data.

Polymers

Polymers are perhaps the most frequently encountered type of contamination in proteomic samples, probably because they are present in most

laboratories at levels high enough to matter to LC–MS work. A few examples of polymer sources include skin creams and moisturizing products, pipette tips, chemical wipes that contain an abundance of different polyethylene glycols (PEGs), and siliconized surfaces that contain polysiloxanes (PSs). The presence of these contaminants can be readily recognized in MS spectra by their characteristically large numbers of regularly spaced peaks in the spectra (44 Da spacing for PEG, 77 Da spacing for PS).

Another source of PEG contaminants is the use of surfactant-based cell lysis methods that are routinely employed in most molecular biology laboratories and involve surfactants such as Tween, Nonident P-40, and Triton X-100. Residual surfactant in samples that are produced using these chemical lysis methods have the potential to produce MS signals that can largely obscure the MS signal of the target peptides, thus rendering the data useless. A dramatic example of such a result is shown in Figure 1. When such protocols are used, extreme care should be taken to selectively remove the surfactants from the sample prior to analysis. Although it is certainly possible to remove PEG contaminants using solid-phase extraction (SPE), it is far easier in the long run to avoid the problem altogether and not use the surfactants for cell lysis in the first place.

Urea

In addition to problems with surfactants, urea used in cell lysis buffers can also cause problems. It is known that urea can decompose to isocyanic acid, which can in turn covalently modify free amine groups in peptides through carbamylation reactions (1). This chemical modification of the peptide can be accounted for in peptide identification software, but only if the software is instructed to look for this.

Residual Salts

In general, residual salts can negatively impact chromatographic performance as well, and they can cause physical damage to fluidics and the MS-interface instrumentation by scratching the surfaces and clogging the emitter. As a result, residual salts should be removed from the sample prior to injection. Both urea and salts are most commonly removed using a reversed-phase (RP) clean-up step, such as SPE.

Water Quality

Not all high quality laboratory water is created equally, and there are many ways problems can arise with the water used for sample and mobile phase preparation. For example, in-line filters that are used to filter out deoxyribonucleic acid (DNA) can inadvertently lead to contamination of the water with PEG. Even the highest quality water produced on-site in the laboratory can begin to accumulate contaminants within a few days of production. Generally, we should avoid using water that has been

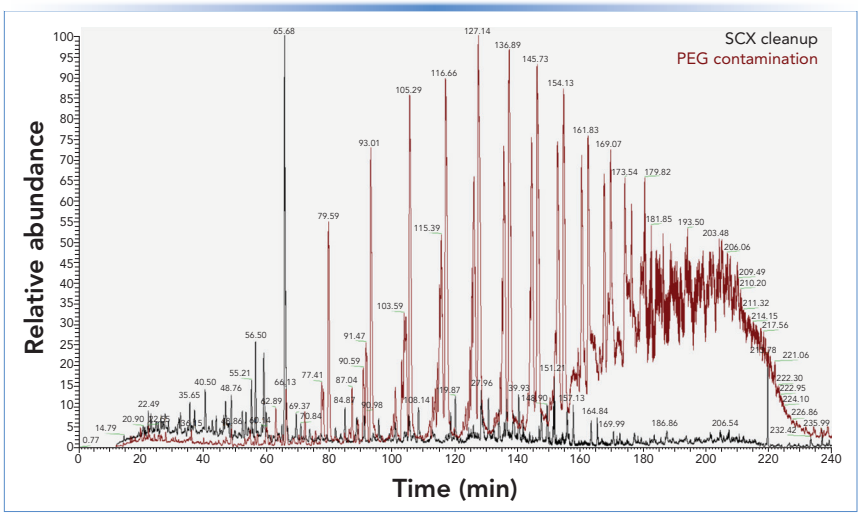


FIGURE 1: Comparison of LC–MS total ion chromatograms for the same peptide sample before (red) and after (black) cleanup by SPE to remove PEG contaminants. If the PEG is not removed, those species completely dominate the MS signal, making extraction of useful information about the target peptides extremely difficult.

sitting around for more than a few days after opening the bottle (if purchased) or production by a polishing system on-site. It is good practice to dedicate a specific subset of mobile phase bottles in the laboratory for LC–MS use only and to avoid washing them with any kind of detergent (2,3).

Keratins

The most abundant protein contamination found in proteomic samples are the keratin proteins that make up our skin, hair, and fingernails. It is not uncommon to observe more than 25% of the peptide content of a proteomics sample to originate from keratin-derived peptides. Any steps that can be taken to reduce this type of contamination will improve the ability to detect low-abundant proteins of interest. In general, natural fiber clothing, such as wool, should not be worn in proteomics laboratories. Ideally, all sample preparation steps should be performed in a laminar flow hood to prevent dust and skin from the analyst and laboratory air from entering samples. Additionally, gloves should be worn at all times and replaced after touching a contaminated surface, such as stopwatches, pens, and laboratory notebooks. However, once the proteins have been digested, one should consider not wearing gloves because these can be a further source of polymer contamination.

Protein and Peptide Adsorption

Peptide adsorption to the LC columns has been addressed in previous installments of "LC Troubleshooting" (4). A lesser appreciated pitfall in proteomics sample preparation is adsorption of the peptides of interest to the vessel used for sample pretreatment (for example, digestion with trypsin), as well as LC sample vials. Peptides are multifunctional molecules with tremendous diversity in properties, including hydrophobicity and the number of acidic and basic functional groups. These properties make them prone to adsorption onto materials with very different properties, such as glass and plastics. This observation is also true for proteins, but to an even greater degree (5). Adsorption of peptides from analytical samples onto such surfaces has been observed within timeframes as low as an hour after being placed in the LC vial, and they can result in a significant decrease in the apparent concentration of low-abundant peptides. A number of strategies to minimize sample adsorption to glass LC vials have been demonstrated, including the use of surfactants (which is not recommended for LC-MS, as discussed above). "Priming" vessels with a sacrificial protein, such as bovine serum albumin (BSA), is also done, similar to what is done with LC columns (4). The idea here is that the vessel is rinsed with a solution of a protein that is unrelated to the target proteins being analyzed (or synthetic peptides not found in biological samples); adsorption sites on the material are subsequently saturated and therefore unable to further adsorb peptides or proteins of interest from the analytical sample. Many vendors also offer "high-recovery" LC vials and other products that are engineered to minimize such undesirable analyte adsorption. An additional point to be aware of in this area is related to the removal of the sample solvent during the sample preparation process (for example, vacuum centrifugation to remove organic solvents). Completely removing all of the solvent promotes strong analyte adsorption onto surfaces; thus, it is helpful to avoid complete drying and to leave a small amount of liquid in the vial to increase analyte recovery.

Plastic micropipette tips present another opportunity to lose peptides to adsorption. With this in mind, it is important to limit the number of sample transfers during sample preparation, which has spurred the development of a number of so-called "one-pot" sample preparation methods that minimize contact between the sample and vessel, including nanoPOTS (6), SP3 (7), and FASP (8). From this research, it is becoming increasingly clear that single reactor vessel sample preparation protocols are superior to conventional protocols in the proteomics area.

Finally, when transferring peptide samples, it is best to avoid contact with metals. For example, most glass syringes are fitted with stainless steel needles by default. If a peptide sample is drawn up through a metal needle, peptides can be lost to adsorption on the metal surface. One way to avoid this scenario when using glass syringes

is to remove the plunger, fill the sample into the glass syringe barrel using a glass pipette, and then push the sample out of the syringe using the plunger, but through a PEEK capillary fitted to the syringe via a Luer lock fitting, rather than pushing it through the metal needle. Doing so is particularly important in the case of MS calibrations where many vendors utilize a peptide calibrant that can be depleted by metal syringe needles.

Sample Matrix Incompatibilities

Other aspects of the sample matrix are worthy of careful consideration. Although it is attractive to consider using trifluoroacetic acid (TFA) as a mobile-phase additive to improve chromatographic performance for peptides (that is, better peak shape, and increased retention, especially for hydrophilic peptides), this practice is generally frowned upon in the proteomics community because TFA can dramatically suppress peptide ionization, leading to lower overall detection sensitivity compared to using formic acid. An alternative to using TFA in the mobile phase is to add it to the sample and use formic acid to acidify the mobile phase. In this case, the TFA can enhance the retention of hydrophilic peptides on the precolumn, preventing them from eluting into the waste line. Meanwhile, the TFA exiting the precolumn is diverted to waste rather than entering the analytical column and MS interface.

Proteomics applications, which are outside conventional peptide sequencing, may also require specific buffers that preserve the structure or stability of proteins during analysis. It is not uncommon for these buffers to lead to salt precipitation on the electrospray emitter and the source cone of a MS system. In this way, one must optimize the conditions to use volatile buffers, such as ammonium acetate, or to physically clean the surfaces with water to prevent significant sensitivity loss. Recently, it has been demonstrated that post-column suppressor technology provides more flexibility in terms of the types of buffers that can be used, some of which are normally considered MS-incompatible (9).

Finally, the age of prepared mobile phases can be particularly problematic in proteomics applications where nano-flow systems are prevalent and mobile phase consumption is so low that a 100-mL batch of the mobile phase can last several weeks. Specifically, when buffer components with differing volatilities are mixed together, there can be selective evaporation of the more volatile component that will change the elution or ionic strength of the mobile phase, which can lead to changes in both absolute and relative peptide retention times. Primarily, this occurrence comes from incorrect viscosity

values that change the mixing behavior of modern nanoLC instruments. Additionally, changes in ionic strength will have a direct effect on the retention factor of the analytes, which will lead to variation in retention time. As a rule of thumb, mobile phases should be replaced every 1–2 weeks to ensure reproducible results. Additionally, in the case of pumping systems that rely on known mobile phase viscosities for accurate flow rates, the mobile phase viscosity should be determined when changing solvents. Instruments that require this feature build in software routines to facilitate these measurements. An additional point is to consider not using conventional membrane degassers when using nano-flow rate because the flow rate is so slow during this process. It has been noted in our laboratory that volatile buffer components, such as formic acid, are preferentially removed from the mobile phase during degassing. In these situations, it is better to utilize alternative degassing measures, such as sonication, that have a low risk for additional contamination of the mobile phase.

Optimizing the Electrospray Source Voltages in nanoLC-ESI-MS

Finally, as we reach the last step prior to the actual acquisition of LC–MS data, we must think about settings for the MS detector itself. Arguably, one of the most consequential missed opportunities in data acquisition for proteomics workflows is related to the optimization of electrospray source voltages. Once analytes are ionized, they are steered through the MS instrument in the direction of the detector using carefully manipulated electric fields. Therefore, analyte detection is highly dependent on the voltage settings for the ion optics that guide the motion of the ionized analytes. Failure to optimize these voltages can result in peptides over large regions of the mass range not making it to the detector, or significantly reduced numbers of detected ions.

Before analyzing any experimental samples, it is worth taking the time to incrementally change the ion source parameters step-wise (for a representative ion infused at the flow rate that will be used in the LC–MS method, we recommend the doubly charged ion of Glu-Fib or angiotensin II), while monitoring the signal change and the spectrum quality. For most modern MS instruments, this tuning step can be done automatically within the MS data acquisition software, but it can be beneficial to further manually fine-tune parameters to maximize signal for the ions of interest. Finally, once the detector settings have been adjusted, it is a good idea to analyze a well-characterized quality control (QC) sample (for example, a cytochrome C digest or HeLa cell digest) to enable a final check on the data quality before analyzing any novel samples.

Summary

Because of the size and varied chemistry of peptides found in proteomics samples, it is important to pay attention to details of each step leading to the acquisition of LC–MS data. This includes sample preparation chemistries, sample handling, optimization of instrument parameters, and control of mobile phases. Paying attention to these details from the start can help avoid data quality problems entirely and simplify troubleshooting when problems do arise. We hope that the tips and tricks discussed here are food for thought for analysts of all levels of experience, and ultimately improve data quality and minimize loss of precious samples and instrument time because of unanticipated problems.

References

(1) S. Sun, J.-Y. Zhou, W. Yang, and H. Zhang, *Anal. Biochem.* **446,** 76–81 (2014). https://doi.org/10.1016/j.ab.2013.10.024

(2) D.R. Stoll and T. Taylor, *LCGC North Am.* **38,** 320–324 (2020).

(3) D.R. Stoll and T. Taylor, *LCGC North Am.* **38,** 379–384 (2020).

(4) D.R. Stoll, *LCGC North Am.* **40**, 471–475, 483 (2022).

(5) Y. Van Wanseele, K. Maes, K. Lanckmans, J. Van Schoors, I. Smolders, and A. Van Eeckhaut, *Chromatographia* **81**, 65–72 (2018). https://doi.org/10.1007/s10337-017-3397-9

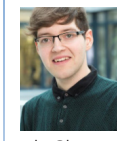
(6) J. Woo, S.M. Williams, L.M. Markillie, S. Feng, C.-F. Tsai, V. Aguilera-Vazquez, et al., *Nat. Commun.* **12,** 6246 (2021). https://doi.org/10.1038/s41467-021-26514-2

(7) C.S. Hughes, S. Moggridge, T. Müller, P.H. Sorensen, G.B. Morin, and J. Krijgsveld, *Nat. Protoc.* **14,** 68–85 (2019). https://doi.org/10.1038/s41596-018-0082-x

(8) J.R. Wiśniewski, A. Zougman, N. Nagaraj, and M. Mann, *Nat Methods* **6**, 359–362 (2009). https://doi.org/10.1038/nmeth.1322

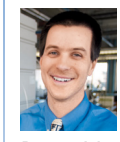
(9) S. Wouters, I. Pijpers, N. Vanden Haute, D. Meston, L. Dillen, F. Cuyckens, and S. Eeltink, *Front. Anal. Sci.* **2,** 1002935 (2022). https://doi.org/10.3389/frans.2022.1002935

ABOUT THE CO-AUTHOR



Daniel Meston is a postdoctoral researcher at the Free University of Brussels in Belgium. He also serves as the vice-president of the British Chromatographic Society. His research interests relate to the development of next-generation LC-MS methodologies for comprehensive proteome profiling, as well as novel workflows and technologies for biological sample preparation. His research results have been presented at numerous international symposia and have resulted in peer-reviewed publications. Direct correspondence to: Daniel.meston@vub.be

ABOUT THE COLUMN EDITOR



Dwight R. Stoll is the editor of "LC Troubleshooting." Stoll is a professor and the co-chair of chemistry at Gustavus Adolphus College in St. Peter, Minnesota. His primary research focus is on the development of 2D-LC for both targeted and untargeted analyses. He has authored or coauthored more than 75 peer-reviewed publications and four book chapters in separation science and more than 100 conference presentations. He is also a member of *LCGC*'s editorial advisory board. Direct correspondence to: LCGCedit@mmhgroup.com

SAMPLE PREP PERSPECTIVES

Class Is Back in Session: More Questions on Your Extraction Knowledge

In our September installment of "Sample Prep Perspectives," we shared sample exam questions from our graduate Chromatography and Separations course as a refresher of some fundamental principles and to give readers the opportunity to test their knowledge. We've dusted off our files and found additional questions for this month's column. These questions tend to focus on dealing with octanol-water partition coefficients, as well as more practical applications of sample preparation.

Douglas Raynie

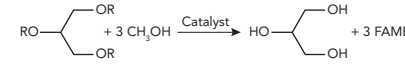
Question 1: Experimental measurements of K_{ow} are not necessarily straightforward. Describe how you would go about measuring the K_{ow} for a solute.

K_{ow}, or the octanol-water partition coefficient, is widely used to describe the distribution of solutes in biochemical, environmental, and analytical systems between nonpolar (lipophilic) and aqueous phases. It is defined as $K_{ow} = [S]_{oct}/[S]_{aq}$, though it is not simply the ratio of the solubility of the solute (S) in the octanol and water phases. One must consider all equilibria occurring in the system, including the mutual solubility of the two solvents. Prior to making the determination of K_{ow}, the octanol and water phases must be brought into contact with each other, thoroughly mixed to obtain saturation of each solvent in the other, and allowed to separate into two separate phases in equilibrium. The solute of interest should then be added until saturation is reached, and equilibrium between the octanol and water phases is achieved. Mixing can help obtain equilibrium more quickly. Solute concentration in each phase can be measured via chromatography, spectroscopy, or other suitable means. Specific procedures for measuring K_{ow} can be found in the methods of a number of regulatory and trade organizations.

Question 2: One common method for the extraction of fat from food samples is to use hexane. If you attempt to extract fat from foods using hexane in a microwave-assisted extractor, what will happen? Why?

Only materials with a dipole moment will absorb microwave radiation. As an overly generalized statement, a material's dipole will align with the electromagnetic radiation and oscillate with the frequency of the radiation. The resulting molecular vibrations release energy, which serves to heat the surrounding material. Hence, hexane, which is nonpolar, will not absorb microwave radiation. In this case, any microwave absorption, and subsequent heating, will occur because of any residual water in the food. If one wishes to perform microwave-assisted extraction (MAE) with nonpolar solvents, there are a couple general ways this can be done. Addition of a polar solvent to the extracting solvents can be used. However, considering the "like dissolves like" principle, convenient systems combining polar and nonpolar solvent are somewhat difficult to obtain. Most modern, commercial microwave-assisted extraction systems have an available "susceptor," which will absorb the microwave irradiation and pass along the resulting heat via convection.

Question 3: Lipids samples, such as triglycerides, are generally hydrolyzed with an acid, then esterified to form fatty acid methyl esters (FAME), such as shown in equation 1 below, prior to analysis for total fats. The FAME are readily analyzed by gas chromatography–flame ionization detection (GC-FID). For example, this method is used for compliance with the Nutritional Labeling and Education Act, specifically in the food content labels we see in the United States.



[1]

Sucrose esters, such as shown in equation 2 below, are approved as fat substitutes when the R group is a fatty acid. Based on the chain length and saturation of the fatty

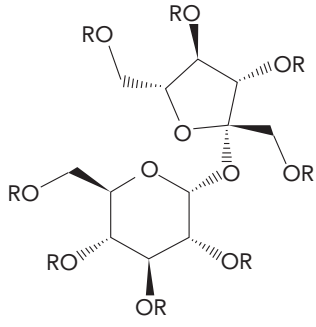
Icon image: Joe Zugcic, Zugcic Photographers, Inc.

TABLE I: Method results summary: spiked synthetic wastewater (3)

Method	Bases and Neutrals, Ave. Recovery (%)	Acids, Ave. Recovery (%)	Pesticides, Ave. Recovery (%)	Volume of Solvent Eluent (mL)	Time for Extraction, Dry, Conc (h)
SPE-100 mL	80.0	89.2	88.6	80	1.5
SPE-1000 mL	56.8	76.7	61.2	200	3.0
LLE-1000 mL	69.2	69.2	70.9	360	4–5

*: LLE time for extraction, dry and concentration estimated

acids used and the degree of esterification, the resulting sucrose ester will mimic a traditional lipid in terms of solubility properties, melting point, viscosity, and surface tension, among other traits.



[2]

Now, consider the situation of a food product that contains naturally occurring triglycerides, but is also processed with the sucrose ester. If acid hydrolysis and FAME analysis is used in the determination of lipids, the R groups in the sucrose ester would also be acid hydrolyzed, so the presence of the sucrose ester would give an artificially high value (such as a false positive) for the determination of triglycerides (analyzed as FAME). Suggest an extraction method with suitable selectivity to determine triglycerides in the presence of the sucrose ester.

This question can evaluate both an understanding of extraction selectivity and creative application of chemical knowledge. Because the properties of the sucrose ester are similar to traditional lipids, we cannot perform an extraction based on differential solubility. Perhaps the most common answer to this question is to perform some type of extraction which gets the FAME (and lipids) into the solution phase, followed by solid-phase extraction (SPE). This is certainly a promising method, and should work, though finding the appropriate combination of solvents and stationary phases is likely to take a great deal of research. Some students will suggest exclusion phases to perform the requisite separation of triglyceride from sucrose ester. Other forms of solvent extraction, such as pressurized solvent extraction (PSE), with the sorbent material added to the extraction vessel, may expedite the extraction, though the added effect of the heat may make obtaining suitable selectivity challenging.

The next most promising method would be supercritical fluid extraction. By manipulation of the applied temperature and pressure, one should reasonably expect to separate various classes of lipids, including the sucrose esters, from each other.

The method which was developed by the industrial creator of the sucrose ester (1) relies on the traditional method, with the exception that the hydrolysis step is performed with lipase. These enzymes offer selectivity as they look for the three-carbon glycerol backbone of mono-, di-, and triglycerides, and will not recognize the sugar backbone of the sucrose ester (the same principle by which the sucrose esters are "fat free").

To one trained in biochemistry or food science, this approach may seem straightforward. However, to many analysts, this approach requires some critical understanding of the chemistry of the system being studied, as well as out of the box thinking since this type of approach is not found in traditional analytical training.

Question 4: The Environmental Protection Agency (EPA) method 625 (2) is used for the determination of bases, neutrals, and acids in drinking water and wastewater via liquid–liquid extraction (LLE) of 1000 mL of sample. A solid-phase extraction alternative (3) has been developed that uses 100 mL of sample, significantly less solvent, and gives better analyte recovery. Table I provides a method results summary for spiked synthetic wastewater.

Discuss the advantages and disadvantages of SPE compared with LLE for this extraction.

For this question, we can assume that the student has not reviewed the application note with the new method or the EPA method, though this assumption may be incorrect for a take-home or similar exam. Based strictly on the information presented in the table (and devoid of other potentially relevant information), evaluation of the sample size indicates that the disk format of SPE must have been used. SPE cartridges are typically used of sample sizes smaller than tens of milliliters. Compared with LLE using samples of similar size (1000 mL), the SPE disk approach is approximately 1.3–1.7 times faster, and uses approximately 45% less solvent. With less solvent use, there is an associated green advantage of less waste, lower flammability hazard, and less use of potentially toxic solvents. Also, less solvent means lower cost and faster solvent evaporation steps. SPE does not deal with emulsion formation problems. Standard deviations are not shown, so we cannot comment on the reproducibility of the methods shown. Whether LLE or SPE is superior in terms of analyte recovery from one-liter samples is unclear. The recoveries shown, lower than the recoveries with SPE of a 100-mL sample, may be because of the

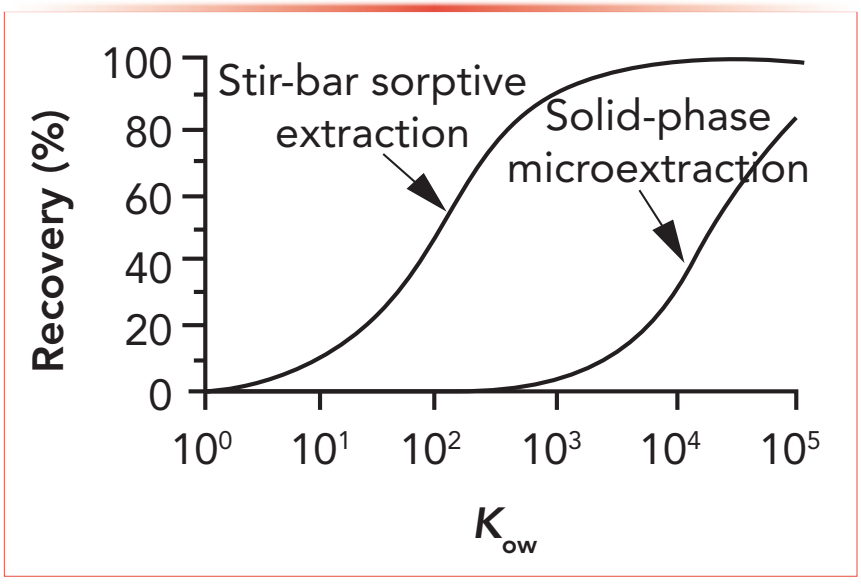


FIGURE 1: Recovery for solutes in function of the octanol-water partitioning coefficient K_{ow} for SPME (10-mL sample, 100-µm polydimethyl-siloxane fiber) and for stir-bar sorptive extraction (10-mL sample, 10 mm × 0.5 mm polydi-methylsiloxane-coated stir bar). (4)

difficulties of handling one-liter samples. However, closer inspection reveals that the recovery may be related to analyte acidity, so use of a mixed-mode SPE phase, perhaps by combining two or more disks, may improve the recovery via SPE. Use of 100-mL samples with SPE demonstrates several interesting features. First, the recoveries are consistently higher across all analyte types, indicating that perhaps the sorbent loading of the disks is more appropriate for the smaller samples. Another consideration is that, with the 100-mL samples and SPE, the ratio of solvent eluent volume to sample amount (0.8) is much larger than with the 1000-mL samples (0.2 for SPE and 0.36 for LLE). This increase in the relative amount of solvent used may explain the further reduction in extraction time and increased recoveries.

Question 5: Figure 1 (4) shows that, everything else being equal, stir-bar sorptive extraction (SBSE) can extract more solute (of equal K_{ow}) than solid-phase microextraction (SPME), and that solutes with a K_{ow} too low to be quantitatively extracted with SPME may be effectively extracted with the SBSE method. Explain why this is.

The stir bars used for SBSE have a much larger surface area than the fibers used in SPME. Consequently, they contain on the order of 50–250 times more extracting phase (sorbent). With such an advantage in terms of phase ratio, SBSE is expected to expected to separate various classes of solute with lower K_{ow} and greater amounts of solutes with similar K_{ow}.

However, this advantage of greater amount of solute extraction comes at a price. Both the sorption and desorption steps in SBSE are significantly slower than in SPME, though automation may help address the sorption and desorption steps. Although other practitioners may disagree, I contend that SPME is preferred when gas chromatography is the subsequent analysis step, and SBSE is preferred when liquid chromatography is used (though either approach can be used for any type of chromatography). However, as a result of this bias, it is important to remember that, with biological samples, macromolecules like sugars and proteins may need to be washed from the sorbent phase in SBSE, while this is less likely the case in SPME. Both SPME and SBSE are non-exhaustive equilibrium-based extractions, meaning that calibrations

must be determined during the method development phase to determine the extraction duration to achieve equilibrium. Finally, it should be noted that newer configurations of SPME, such as the SPME-Arrow, have increased sorbent amounts, though still less than found in SBSE.

Question 6: Consider an ionic compound, MX, composed of the cation M^+ and the anion X^-, which dissociates in water. Write the equilibrium equation for each equilibria in an extraction system.

Consider an organic extracting phase (org) and an aqueous sample phase (aq). The equilibria between each species in each phase must be considered:

1. org ⇄ aq
2. $MX_{(aq)} \rightleftarrows MX_{(org)}$
3. $MX_{(aq)} \rightleftarrows M^+_{(aq)} + X^-_{(aq)}$
4. $MX_{(org)} \rightleftarrows M^+_{(org)} + X^-_{(org)}$
5. $M^+_{(aq)} \rightleftarrows M^+_{(org)}$
6. $X^-_{(aq)} \rightleftarrows X^-_{(org)}$

Depending on the dissociation constant of $MX_{(aq)}$ and the sizes of the relevant equilibria constants, it is likely that only equilibria 2 and 3, and perhaps 1, will be important in an analytical extraction. For some polyatomic systems like the bicarbonate-carbonate system, all dissociations ($H_2CO_3 \rightleftarrows HCO_{3-} \rightleftarrows CO_3^{2-}$) must be considered.

Conclusion

I hope these exam questions over the past couple of columns has been fun and educational. I enjoyed putting them together. Now, with our next installment, we'll return to more focused treatments of specific sample preparation techniques and applications.

References

(1) D. Schul, D. Tallmadge, D. Burress, D. Ewald, B. Berger, and D. Henry. *J. AOAC Intl.* **81**(4), 848–868 (1998).

(2) US EPA Method 625, Revision A, *Federal Register* **80**(33), 8956 (February 19, 2015).

(3) W.R. Jones, A. Cannon, D. Gallagher, M. Ebitson, and Z. Grossner, *LCGC North Am.* **33**(12s), 66 (2015). http://www.chromatographyonline.com/new-method-us-epa-625-solid-phase-extraction-challenging-wastewaters

(4) F. David, B. Tienpont, and P. Sandra, *LCGC North Am.* **21**(2), 108–118 (2003).

ABOUT THE COLUMN EDITOR

Douglas E. Raynie

"Sample Prep Perspectives" editor Douglas E. Raynie is a Department Head and Associate Professor at South Dakota State University. His research interests include green chemistry, alternative solvents, sample preparation, high-resolution chromatography, and bioprocessing in supercritical fluids. He earned his PhD in 1990 at Brigham Young University under the direction of Milton L. Lee. Raynie is a member of *LCGC*'s editorial advisory board. Direct correspondence about this column via e-mail to LCGCedit@mjhlifesciences.com

GC CONNECTIONS

Selectivity and Sensitivity: The Electron Capture Detector (ECD), Its Unique Inventor James Lovelock (1919–2022), and GAIA

The electron capture detector (ECD) is among the original classical detectors for gas chromatography (GC). It is highly selective and sensitive for electron-absorbing compounds, especially those containing halogens. The ECD was developed in the 1960s by James Lovelock (1919–2022), who passed away earlier this year. The ECD is among the early detectors that is used in a relatively unmodified form today. The ECD was crucial in the discovery of chlorofluorocarbons (CFCs) in the upper atmosphere, ultimately leading to the international agreements limiting their use and reducing the ozone hole. In this installment, we review the ECD and its principles of operation and discuss the general tradeoffs in detection between selectivity, ease of use, and sensitivity. We also look at the unique life and work of Lovelock, both inventor of the ECD and one of the earliest scientists to bring attention to climate change through GAIA, which explained his theory that the Earth acts as a massive living organism.

Nicholas H. Snow

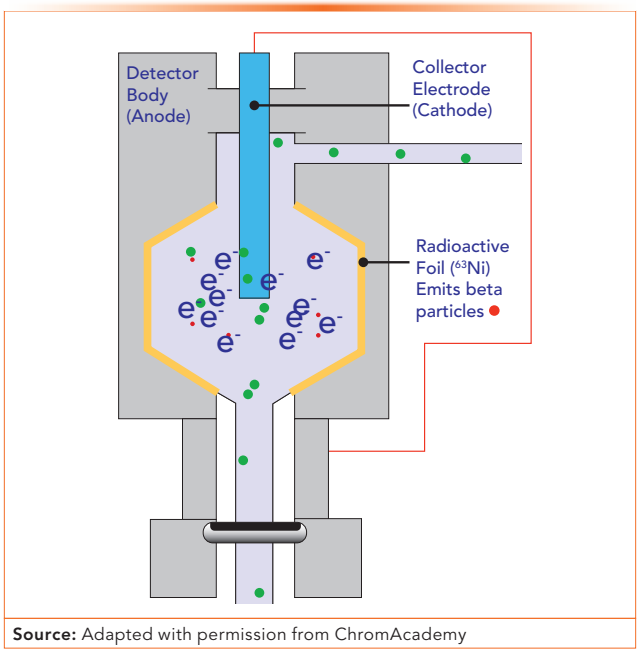
This summer, I was watching the news show "Sunday Today," and heard a name familiar to many gas chromatographers: James Lovelock. To chromatographers, Lovelock invented the electron capture detector (ECD) for gas chromatography (GC). He was featured in the weekly "Life Well Lived" segment, where the show features people who are not widely known to the public but made contributions to public life (1). The segment describes how the ECD and Lovelock's own work provided major influences on climate science today, nearly 60 years later.

Detection in GC remains unique in analytical chemistry because of the large number of customized detectors. For early workers, detection was especially challenging because the detector was required to be sensitive, selective, and able to detect a low concentration or mass of analytes in a rapidly moving stream of carrier gas. Interestingly, many of these early detectors, including flame ionization detectors (FIDs), thermal conductivity detectors (TCDs), and ECDs remain widely used today, along with myriad additional ionization, physical property, and spectrometric detectors. Schug and McNair provided a summary and "grade sheet" for commonly used GC detectors (2). The ECD was graded A+ for limit of detection (LOD) and specificity, but it was given a C grade for qualitative speciation, linear range, and robustness, and a D grade for universal response.

A summary of the basic principles and operation for the ECD is provided in textbooks on GC and available in ChromAcademy, *LCGC North America*'s online learning platform (3–5). Overall, the ECD provides a unique combination of selectivity and sensitivity. Implied by its name, the ECD offers selectivity for compounds that include atoms that are favorable for attracting or capturing an electron. Recalling general chemistry, note that elements on the right and up in the Periodic Table tend to attract electrons. These include the most common heteroatoms, especially halogens. The ECD is highly selective for compounds containing these elements, and by contrast, generates little or no signal for common hydrocarbons. The ECD presents a useful complement and contrast to the FID, which gives strong signals for hydrocarbons.

Highly selective detectors, in general, offer greater sensitivity for those analytes for which they are selective. The ECD is noted to be the most sensitive of the classical detectors, with detection limits easily in the part-per-billion (ppb) concentration or picogram mass range. Finally, highly sensitive detectors, including the ECD, have very specific requirements for obtaining the expected high sensitivity. In the case of the ECD, the carrier and makeup gases must have scrupulous high purity, samples should be especially clean, and the operating environment should be free of contaminants containing electron absorbing atoms. Together, these operational factors lead to the grades shown for the ECD by Schug and McNair.

Figure 1 shows a schematic diagram of a typical ECD. The operating principle, while not fully understood, first involves the production of electrons from a radioactive source, typically ^{63}Ni. For safety, a sealed source is used and usually the detector manufacturer will have a blanket radioactive isotope license in case the end user does not. Radioisotope licensing was an early barrier for ECD users. If you ever need to dispose of



Source: Adapted with permission from ChromAcademy

FIGURE 1: Schematic of electron capture detector (ECD).

an ECD, you should consult with the manufacturer, a licensed health physicist, or radiation safety officer for proper disposal procedures.

Unique among ionization detectors, the ECD measures a reduction in the electrical current rather than production of a current. Electrons produced in the radioactive source ionize carrier gas, usually nitrogen, molecules flowing through the detector cell. If the carrier gas is helium or hydrogen, nitrogen makeup gas is used. Electrons produced by the ionized nitrogen migrate to the anode under a fixed voltage, generating a steady electrical current. If electron-absorbing species pass into the detector cell, they absorb electrons, resulting in a reduction in the standing current.

Figure 2 shows an early example of ECD selectivity from Lovelock's original work in 1960 (6). The top chromatogram shows the separation of eight compounds using a classical argon detector (not commonly used today). The bottom chromatogram shows the separation of the same eight compounds using the ECD. Note the stronger argon detector signals for the hydrocarbon and fluorinated compound, for the singly-chlorinated benzene, and the strong ECD signals for the multiply chlorinated compounds.

In 2007, Ettre and Morris provided a thorough discussion "The Saga of the Electron Capture Detector" in *LCGC North America* (7). They trace the developments that led to the ECD beginning in 1948, showing a fascinating scientific story that began with studying a different problem (whether drafts of cold air in a room really do cause illnesses) for which Lovelock developed an anemometer that worked by disturbing the slow flow of positive ions from a radon source. This anemometer was so sensitive that the signal was disturbed by small amounts of cigarette smoke in the atmosphere, capturing electrons, although at the time this was seen as a problem and not of any interest. Recall that smoking was much more prevalent in the 1950s than it is today. In my own career working with modern ECDs, I have observed deflections in ECD signals because of the presence of cigarette smoke odors. The full story, presented by Ettre and Morris, of how this observation ultimately led to today's ECD is well worth the read.

The Science Museum of London also has an early ECD and other equipment, including Lovelock's own home-made gas chromatograph on display and an extensive treatment of Lovelock's life and career, along with Lovelock's entire personal archive of documents and letters (8).

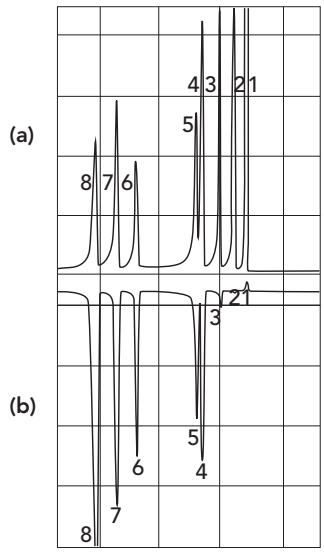
Lovelock, Atmospheric Chemistry, and GAIA

Lovelock's prolific scientific career, spanning seven decades, included many more contributions than inventing the ECD for use in GC. As with many of the early gas chromatographers, his career started in other areas. His first publication in 1941 related to the use of hypochlorite spraying for disinfectants (9). Then, he had a prolific publication record in the general area of disinfectants, sports training, prevention of the common cold, and other medical topics through the remainder of the 1940s and early 1950s. After earning his PhD in medicine in 1948, a shift toward instrumentation and measuring devices was seen throughout the 1950s. Through the 1950s, his medical research shifted to many aspects of preservation and chemical effects on red blood cells and some interestingly titled articles related to the re-animation of rats and mice frozen to 0 °C.

As Ettre and Martin describe, in the 1950s, Lovelock's laboratory was in close proximity to A.J.P Martin and A.T. James, who invented GC and they became collaborators, with Martin suggesting to Lovelock that he develop a highly sensitive detector that was operable by average chemists. They noted that the most sensitive detectors of the day were not easily operated by most chemists. The first mention of GC in the title of an article written by Lovelock occurred in 1958. In the article, he discussed an early detector, not the ECD but the argon ionization detector, which like the ECD used a radioactive source to produce electrons that ultimately ionized analytes. It was a nearly universal detector, used for many applications that were ultimately supplanted by the FID, which did not require a radioactive source (10). Using the argon ionization detector, Lovelock then published numerous papers on the ionization mechanisms of organic compounds in the vapor phase, most often using GC with the argon ionization detector.

Following the initial invention and publication of the ECD (6), Lovelock worked with Zlatkis to commercialize the detector, founding a company called Ionics Research, which provided ECDs to PerkinElmer and the Wilkens Instrument and Research Company, which later became the Chromatography Division of Varian bought in 2010 by Agilent Technologies. In 1962, coincidentally with the publication of the classic book *Silent Spring* by Rachel Carson, the first papers demonstrating ECD of pesticides at the picogram level were being published; although, as pointed out by Ettre and Morris, the ECD is often incorrectly cited, including in the "Sunday Today" segment, as one of the techniques on which Carson's findings were based (11). However, there is clear connection between Lovelock's work, the ECD, the early environmental movement, and today's challenges with climate change.

Using the ECD, Lovelock published the first work demonstrating the presence of chlorofluorocarbons (CFCs) in the atmosphere over the Atlantic Ocean between England and Antarctica (12). This work



Source: Reprinted with permission from reference (6). (Copyright 1962, American Chemical Society.)

FIGURE 2: (a) Argon detector and (b) ECD chromatograms of a mixture of hydrocarbons and halocarbons. 1. cyclohexane; 2. fluorobenzene; 3. chlorobenzene; 4. m- and p-dichlorobenzene; 5. o-dichlorobenzene; 6. 1,3,5-trichlorobenzene; 7. 1,2,4-trichlorobenzene; and 8. 1,2,3-trichlorobenzene.

demonstrated the presence of Freon-11 and other compounds resulting from the use of aerosol cans and refrigerants. Roland and Molina then used this work in developing the theory that halocarbons in the atmosphere were reacting with stratospheric ozone in a chain reaction, causing the depletion of ozone and the famous "ozone hole" (13). Roland, Molina, and Crutzen shared the Nobel Prize in 1995, and this work resulted in international treaties limiting the use of halocarbons.

We close with GAIA, a fundamental aspect of climate science today, that was mocked by most when introduced by Lovelock in the 1970s. Much like the seeming lack of interest in the 1950s in a detector that was affected by cigarette smoke in the room, GAIA, named for the Greek goddess of the personification of Earth, proposed that the Earth is a single self-sustaining organism, and this idea was universally rejected by the scientific community of the day (14). The basic principles and problems described by Lovelock in his original work remain debated in climate discussions today. With the invention of the ECD, its use in discovering CFCs in the atmosphere, work in planetary, medical, space and climate science, James Lovelock was not just a chromatographer but a true Renaissance man of science.

References

(1) Sunday Today, *Life Well Lived*. https://www.today.com/video/dr-james-lovelock-scientist-who-created-gaia-theory-dies-at-103-145164869912 (Accessed October, 2022).

• Continued on Page 542

FOCUS ON BIOPHARMACEUTICAL ANALYSIS

On Replication in Biopharmaceutical Analysis

The landscape of biopharmaceutical analysis features different analytical requirements in terms of instrumentation and type of analysis (QC, RTRT, and in-process). One of the questions that often crops up for routine analysis is "How much replication is enough?" Should all the samples be run in triplicate irrespective of the type of analysis, or does the type of analysis (such as charge or size variants) have any bearing on the number of optimal replicates?

Anurag S. Rathore, Srishti Joshi, Akshdeep Ahluwalia, and Jared Auclair

There are various factors that the user needs to consider in the matter of biopharmaceutical analysis—first, the objective of replication, and second, the trade-off with respect to cost and time. Ideally, the number of replicates should be large enough to adequately represent the variability present, but not so large that the information gained is unjustifiably expensive with respect to the costs (such as instrument time, analyst time, consumables, and samples). The optimal replication applicable to each type of analysis would therefore lie somewhere in between. For any given sample, one must also consider the sources of variability at play. Although some of these sources would be generally applicable to all types of analytical assessments (including analyst-dependent and system-dependent variabilities), some may be unique to the dynamics of a specific biomolecule, such as changes occurring during analysis (stability related). Traditional data accumulated for a specific analysis and molecule over time can help understand and evaluate effective replication numbers.

In this article, we use data generated in our laboratory for a published biosimilarity exercise on marketed trastuzumab biosimilars in order to understand and illustrate how the optimum replication number might be influenced by a) the analysis type; and b) devised strategies for efficient utilization of resources to estimate replication number for a given analysis type (Figure 1) (1).

Standard error: Often confused with standard deviation, it is the measure of the magnitude of the experimental error of an estimated statistic measured statistic (3).

Standard deviation: It is the measure of variability of the population, indicating the estimate of the variability of the population (3).

Blocking: It tackles the problem of pseudo replication, and thus serves as means of reducing experimental error by arranging experimental units into groups (blocks) that are as homogeneous as possible (2).

Randomizing: It helps to assure the validity of the estimate of the experimental error, and gives unbiased estimate of variance as well as treatment differences (2).

Region of Interest (ROI): A region of responses within a spectra that captures the most important or relevant information within the data set.

Outliers: These can be identified using principal component analysis and subsequently performing a Hotelling's T-squared test.

Errors

Blocking and randomization are used in statistical experimental design to balance variability and measure size. After calibrating the equipment and validating the method to reduce variation due to nuisance variables, replication is required to increase the experiment's informative power. The size of the experiment can be influenced by increasing or decreasing the number of replicates (2).

Increased replication helps to reduce the standard error, which is a measure of the magnitude of the experimental error of an estimated statistic. In contrast to the standard deviation, which relates to the inherent variety of observations within individual experimental units, the standard error refers to random fluctuations of a total estimate. By increasing replication, the standard error can be made arbitrarily small. This implies we want a standard error that is small enough to draw reasonable inferences, but not too small.

Types of Error

Type I Error

Type I error is when you declare an effect to be real, and, in reality, it is zero. Setting this to a certain value is usually represented by selecting a value for α (alpha), where α limits this kind of error to a related percentage of

probability. Therefore, type I error is also sometimes called *alpha (α) error*. Similarly, *type II* error, which will be discussed in the following section, is regarded as *beta (β) error* (4).

It bears mentioning that simply protecting against type I errors may not be sufficient. Samples of sufficient size may nearly ensure statistical significance, regardless of how little the difference is, as long as it is non-zero. Assuming we only want to discover differences that are important in practice, we add an extra precaution against a type I error by avoiding employing sample sizes greater than those required to defend against the second kind of error—the so-called type II error.

Type II Error

Type II errors occur when an effect is genuine, but is neglected. When the replicates vary slightly, such a failure may not be significant. Only when the disparity is significant does it become problematic. The second step is specifying the probability of actually detecting it. This probability (1-β) is called the *power* of the test. The quantity β is the probability of failing to detect the specified difference to be statistically significant, and therefore this error is also sometimes called a *beta (β) error*.

Calculating the Size

Mathematically, the size of the experiment depends on the estimate of variance (σ^2); the true difference (δ) that we are set out to detect; the effect of assurance with which it is desired to detect the difference (the power of the test); the significance level that will be utilized in the experiment; and the type of test being necessary (as in whether a one-tail or two-tail test). The required sample size, *n*, is given by the following formula:

$$n = \frac{2\sigma^2(Z_{a/2} + Z_{\beta})}{\delta^2} \qquad [1]$$

By rearranging equation 1, we can get the true difference, δ, in terms of size *n*, for an estimated certain variance as:

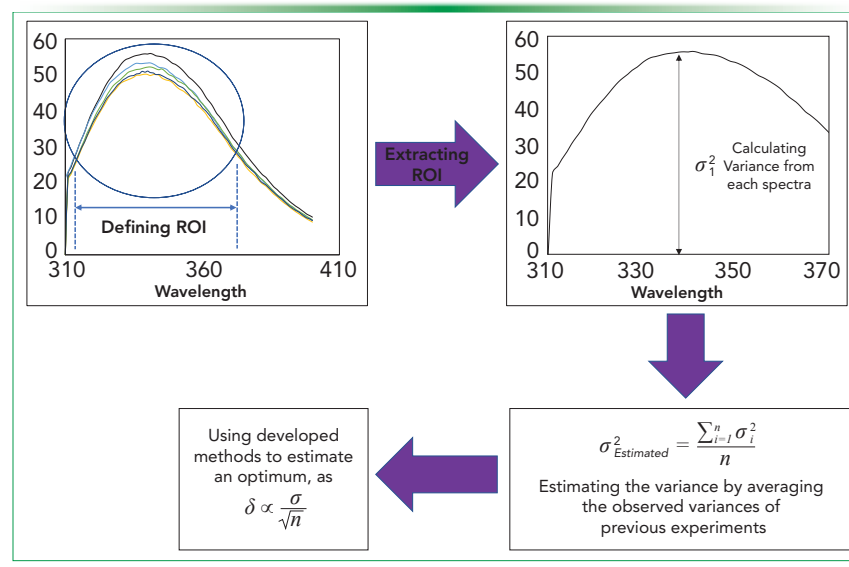


FIGURE 1: Strategy for the estimation of the optimum replication number.

TABLE I: Z-values for sample size calculation (5)

Value	Variance	
α-value	$Z_{a/2}$ (two-sided)	$Z_{a/2}$ (one-sided)
0.01 (level of significance 1%) 0.05 (level of significance 5%)	2.576 1.960	1.282 1.645
β-value	Z_{β}	
0.01 (power 99%) 0.05 (power 95%)	2.326 1.645	

TABLE II: Summary of results for approach 1

R_o	5% significance for α, 95% power for β	1% significance for α, 99% power for β
FT-IR	4 (3.514)	4 (3.514)
CEX	4 (3.514)	4 (3.514)
Fluorescence	4 (3.514)	4 (3.514)

$$\delta = \frac{\sqrt{2\sigma^2(Z_{a/2} + Z_{\beta})}}{\sqrt{n}} \qquad [2]$$

where $Z_{a/2}$ and Z_{β} are the type I and type II errors, respectively, the values for which can be found in the Z tables from most standard books on statistics (see Table I).

In the following, we illustrate this approach using examples from three routine analytical techniques used in biopharmaceutical analysis—namely, Fourier transform infrared spectroscopy (FT-IR), cation exchange (CEX) chromatography, and fluorescence spectroscopy (FLR). For each of these methods, the considered readings were taken in respective region of interests (ROIs) while estimating the variance, and the final estimate was made by averaging the obtained variances for replicates in each batch for all the originator and biosimilars being considered here in this study. For FT-IR, the original data measured over 400–4000 cm^{-1} was taken in the ROI of 1400–1900 cm^{-1}. In CEX, the readings were over 0–20 min; the considered ROI is 9–17 min. In the fluorescence data, the ROI was taken in just the Gaussian curve, and it was between

TABLE III: Summary of results for approach 2

R_o	5% significance for α, 95% power for β	1% significance for α, 99% power for β
FT-IR	5 (5.496)	5 (5.496)
CEX	3 (2.658)	3 (2.568)
Fluorescence	3 (2.829)	3 (2.694)

TABLE IV: Summary of results for approach 3

R_o	5% significance for α, 95% power for β	1% significance for α, 99% power for β
FT-IR	2 (2.478)	3 (3.378)
CEX	3 (3.279)	4 (4.460)
Fluorescence	3 (2.865)	4 (3.900)

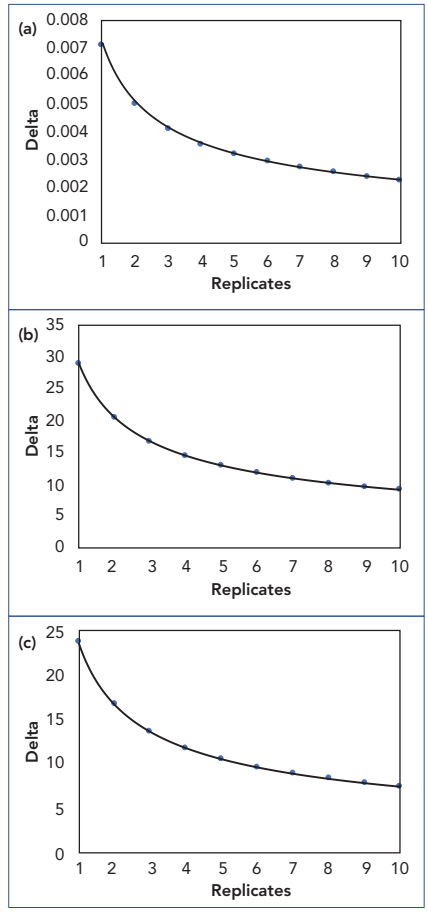


FIGURE 2: Curve showing increase in the impact of higher replication on captured differences. Ordinate data scales are different for (a) FT-IR, (b) CEX, and (c) FLR.

310–370 nanometers. After estimating the averaged variances, we can find the relation of how the δ changes with the replicate number. Here, we study two confidence intervals of 95% and 99% for each of the errors.

With the Z values and estimate of the variances, we can obtain a curve for δ versus the replication number (Figure 2). Starting with FT-IR data (Figure 2a), we obtained an average variance of $7.052e^{-06}$ in the observed ROI. This helps us visualize the increase in the impact of higher replication on the ability of capture differences between the given data sets. Similarly, the observed variances (averaged) in CEX (Figure 2b) and fluorescence (Figure 2c) data were 117.088 and 77.153, respectively, for the considered ROIs in each. As the relationship between the two axes is the same (inverse root of replication), the order of values on the vertical axes corresponds to the order of response measured for the experiments, from which the variances were calculated and averaged. The effect of increasing the power of errors (moving from 95% confidence to 99% confidence) reflects an upwards shift of the curve along the *y*-axis, implying that the overall trend of the curve remains the same, and that the true difference that could be captured is more at each particular replication number.

The relationship between delta and replication is non-linear and exponential in nature, meaning that, beyond a certain number of replicates (thereafter referred to as optimum replication R_o), the cost associated with replication is greater than that of the value of the information gained. It would then help to have a mechanism to decipher the value of R_o for a given analyses. To achieve this, we need methods that inherently consider the estimated variance and the power of tests used.

Estimating an Optimum (R_o)

While estimating an optimum, it is important that the devised methods capture the effect of the estimated variance for the different data sets, and yield a number accordingly. For example, R_o may differ among the different chromatographic and spectroscopic techniques. Keeping this in mind, we explored three approaches to yield an estimate for R_o. The methods essentially exploit

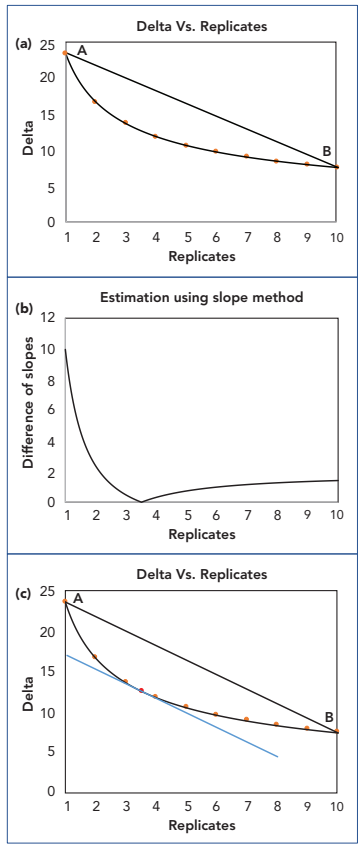


FIGURE 3: Estimating R_o using the slope method; (a) highlighted values at integral points; (b) absolute difference of slope of the curve w.r.t. the slope of line AB; and (c) red dot indicating the estimated point to be rounded to nearest integer.

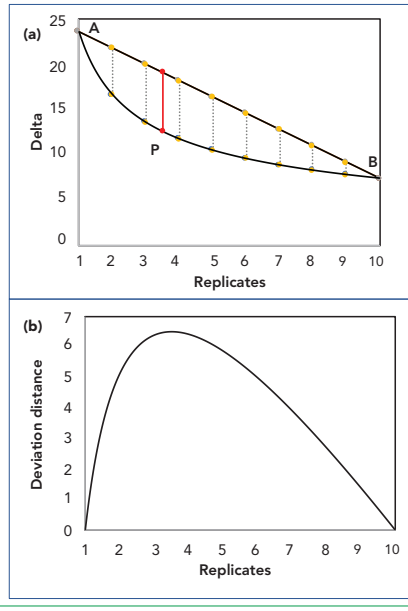


FIGURE 4: Estimating R_o using the slope method, alternatively seen as deviation distance; (a) the observed dips at integral points seen as parallel lines with optimum at P and (b) plot depicting the trend of the dip.

the shape of the curve obtained as a function of the relationship between delta and replication number, and are based on exploration of geometrical features (such as slope, center, and deviation from linear relation). We have applied this to three types of data sets commonly obtained in biopharmaceutical analysis—ion exchange chromatography, FT-IR spectroscopy, and fluorescence spectroscopy.

Estimation Using Slope Approach

Physically, the replicate number will be integral, and equation 1 should yield discrete values as the number goes from 1 to 10 (10 has been defined as a maximum here based on general laboratory practices). However, to leverage the curve mathematically, we have considered it to be continuous for this and the subsequent estimation below.

In the slope method, we first look at the overall decrease in the variance over the maximum replication that we could go to (defined here as 10). In graphic depiction, this number corresponds to the line AB in Figure 3a. The plots considered here are for the fluorescence data at 5% significance and 95% power for type I and II errors, respectively.

With that line, we find the point on the replicate axis that has a slope equal to that of line AB; one can easily argue that there would indeed exist such a point. Using equation 2, upon differentiating, we get a mathematical expression for the slope in continuous domain as:

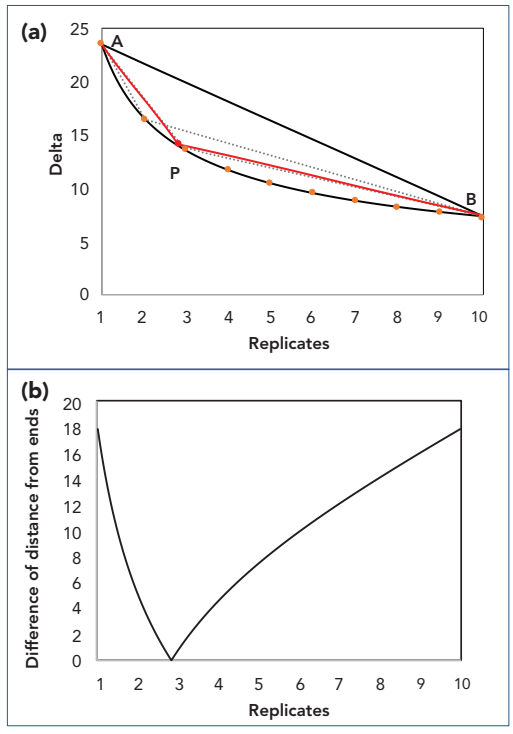


FIGURE 5: Estimating R_o using center method; (a) estimated point P subtending equal length or angle with the endpoints of line AB; (b) plot depicting the difference of sides from the ends

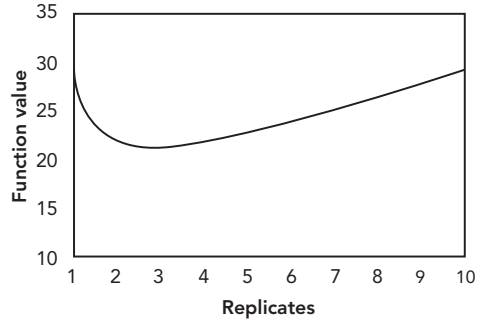


FIGURE 6: Estimating R_o using cost function.

$$\frac{d\delta}{dn} = \frac{-1}{2} \cdot \frac{\sqrt{2\sigma^2(Z_{a/2} + Z_{\beta})}}{\sqrt[3/2]{n}} \quad [3]$$

Using equation 3, we compare the slopes at each point, and get a number having the same overall decrease of detectable delta over replicate. We plot the difference between the slopes to visualize the trend, and then the minima of it. As depicted in Figure 3b, the minima is found to be at 3.514. Graphically, we will have a tangent with an equal slope as that of AB, as shown in the Figure 3c.

Because we are working in the continuous domain, we will get a positive real number as the estimate that needs to be rounded off to the closest integer for physical consideration. Having done that, we have estimated a number that we can say to be optimum. Here, the closest integral number to 3.514 will be 4, so that is our predicted optimum.

This method of slopes is based upon finding the maximum deviation of the delta values, if the relation was to be linear between them and the replicates with a slope the same as that of a line joining the end points of our considered replication range. It can also be said to be looking at the maximum dip present graphically, as depicted in Figure 4a. The deviation distances should naturally go to a maximum value as it is zero at both ends of the curve. Thus, the continuous deviation values can be plotted, as shown in Figure 4b. Mathematically, it so happens that the numerical optimization function becomes the same upon differentiating while maximizing the deviation distance, resulting in an expression identical to that obtained when the slope was compared. This is basically because of the nature of the function that we are using. The summary of the estimated optimum is as given in Table II.

We see that the predicted optimum number using this method is the same for all spectroscopic methods, thus essentially suggesting that this method has failed to recognize the variance in it, and it predicts a generalized R_o value.

Estimation Using Center Method

Through this approach, we find the center of the curve (the point on the curve that is equidistant from both the ends of the curve). As the point being equidistant from the two extreme ends (as in Figure 5a) are competing with the detectable difference and the replication done, the middle point can be an optimum balancing the two conflicting aspects of the features of the curve.

We plot the difference of the distance of the points on the curve and expect a minimum to be there, highlighting the point that is equidistant. Figure 5b shows the variation of the difference of distances for points on the curve from the ends. The minima occur at 2.829. By rounding off, the integral estimate that we get is 3. Graphically, we see that the point of the minima subtends equal

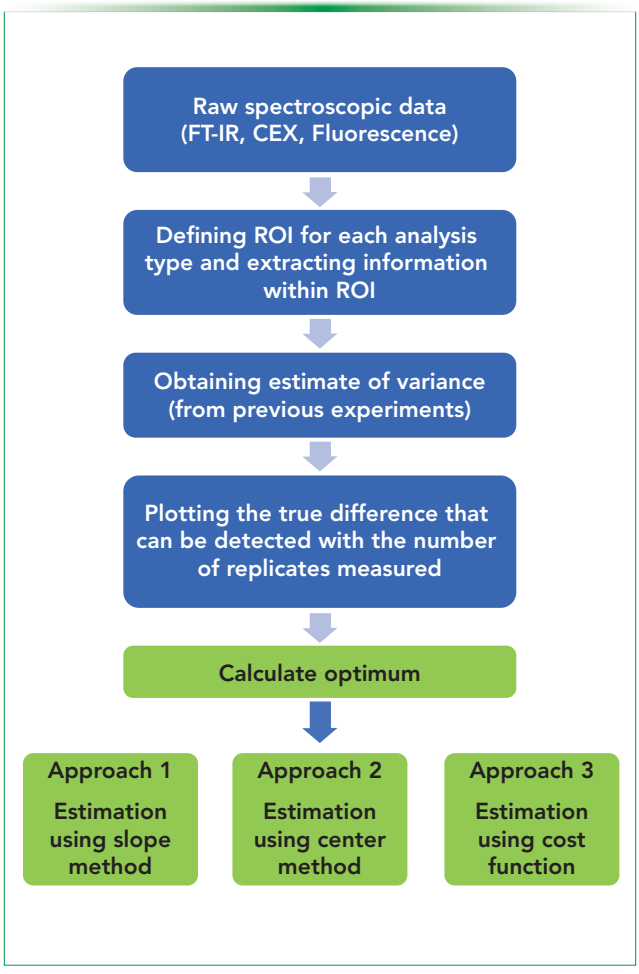


FIGURE 7: General scheme for R_o estimation.

angle with the line AB, and will essentially make an isosceles triangle, including itself and the triangle PAB, with PA and PB being the two equal sides (Figure 5a). The summary of the results obtained using this method is shown in Table III.

Therefore, we see that each method yields a unique number based on the variance obtained from the respective technique for a given analysis, providing an appropriate method to make an estimation.

Estimation Using Cost Function

Having seen a couple of methods based on using geometrical aspects of the curve, we now try to develop a cost function that penalizes if the replication number is too high, and incentivizes for the low delta values, so that we anticipate a minimum in the function as we increase replication.

Initially, we proposed a function as equation 4. As *n* increases, the function value increases. However, at the same time, because of the corresponding decrease of delta values, the function value decreases as well.

$$J(\delta,n) = e^{\delta \cdot n} \quad [4]$$

Considering this, we can expect a minimum value, but that minima can possibly fall well above our maximum number of replicates taken. This function being explicit in terms of replicate number, upon solving this function mathematically, we get the R_o as equation 5:

$$R_o = \left(\frac{\sqrt{2\sigma^2(Z_{\frac{\alpha}{2}} + Z_\beta)}}{2}\right)^2 \quad [5]$$

We see here that optimum is dependent on the variance that is essentially desired, but the order of variance can be very different for experiments. To handle this uncertainty, we can further modify our function by dividing δ by σ_{min}, where σ_{min} is the lowest variance observed of all the previous experiments; for fluorescence data, the observed minimum in the data set was 48.472. The function then can be written as equation 6:

$$J(\delta,n) = e^{\left(\frac{\delta}{\sigma_{min}}\right)}(n) \quad [6]$$

Mathematically, the optimum for this function can be written to be at:

$$n_{\text{optimum}} = \left(\frac{\sqrt{2\frac{\sigma^2}{\sigma^2_{\min}}(Z_{\frac{\alpha}{2}} + Z_\beta)}}{2}\right)^2 \quad [7]$$

Although we have a mathematical expression, we plot the function and try to depict the minima that must be there. Figure 6 shows the plot of the cost function for fluorescence data (at 5% significance for α, and 95% power for β). We can see that the minimum is in our limit to our maximum range. The summary of the results obtained using this method are presented in Table IV.

Here again, we get a unique number for each experiment, thus being a consistent approach to estimate an optimum with.

Conclusions

Irrespective of the analysis, the general scheme to be followed for R_o estimation is shown in Figure 7, starting from the raw data to the proposed approaches.

R_o estimation approaches can be compared based on effect and the range of relation they share with the estimated variance of a given technique and analysis type, it being the defining feature for the relevance of R_o obtained. While the first approach is insensitive to estimated variance, it could be used as an estimation method where the innate mathematics inherently incorporates the impact of variance (and other such related parameters) over the estimation.

The second approach, while being sensitive to individual technique variance, does not measurably distinguish between confidence for both type I and II errors, as can be seen in Table III.

The last approach based on the developed cost function is both sensitive to technique specific variance as well as the error confidence interval, and thus can be applied across analytical characterization platforms to get a realistic and relevant understanding of the replication required, which is technique and analysis specific based on variances derived from existing analytical data, thus facilitating a more rational approach to defining replication.

Although the work discussed here is in specific context of biopharmaceutical analysis, it would be equally applicable across different domains where chemical data is generated in the form of *x and y* line plots.

References

(1) S. Joshi and A.S. Rathore, *BioDrugs* **34,** 209–223 (2020).

(2) P. Berthouex and L. Brown, *Statistics for Environmental Engineers* (CRC Press, Boca Raton, FL, 2nd ed., 2002).

(3) A. Banerjee, U.B. Chitnis, S.L. Jadhav, J.S. Bhawalkar, and S. Chaudhury, *Ind. Psychiatry J.* **18**(2), 127–131 (2009).

(4) D.G. Altman and J.M. Bland, *BMJ [Br. Med. J.],* **331**(7521), 903 (2005).

(5) J. Charan, R. Kaur, P. Bhardwaj, K. singh, S.R. Ambwani, and S. Misra, *Ann. Natl. Acad. Med. Sci. (India)* **57,** 74–80 (2021).

ABOUT THE CO-AUTHORS

Anurag S. Rathore is with the DBT Center of Excellence for Biopharmaceutical Technology at the Indian Institute of Technology, in Delhi, India. Direct correspondence to: asrathore@biotechcmz.com.

Srishti Joshi serves as the Coordinator of Analytical Characterization Division at the DBT-Centre of Excellence for Biopharmaceutical Technology, Indian Institute of Technology, under the guidance of Professor Anurag. S. Rathore, Department of Chemical Engineering, Indian Institute of Technology, Delhi.

Akshdeep Ahluwalia is a senior-undergraduate pursuing a Bachelor's plus Master's, Dual Degree program in Chemical Engineering at Indian Institute of Technology in Delhi, India.

Jared Auclair is an Associate Dean of Professional Programs and Graduate Affairs at the College of Science at Northeastern University, in Boston, Massachusetts. He is also the Director of Biotechnology and Informatics, as well as the Director of the Biopharmaceutical Analysis Training Laboratory.

• Continued from Page 535

(2) K.A. Schug and H.M. McNair, *LCGC North Am.* **33**(1), 24–33 (2015).

(3) H.M. McNair, J.M. Miller, and N.H. Snow, *Basic Gas Chromatography* (John Wiley and Sons, Hoboken, NJ< 2019), pp. 131–134.

(4) C.F. Poole, Ed., *Gas Chromatography* (Elsevier, Amsterdam, The Netherlands, 2nd ed., 2021), pp. 343–358.

(5) ChromAcademy, *GC Detectors*. https://www.chromacademy.com/channels/gc-training-courses/instrumentation/gc-detectors/ (accessed October 2022).

(6) J.E. Lovelock and S.R. Lipsky, *J. Am. Chem. Soc.* **82**, 431–433 (1960).

(7) L.S. Ettre and P.J.T. Morris, *LCGC North Am.* **25**(2), 164–178 (2007).

(8) Science Museum, *Something in the Air: James Lovelock and Atmospheric Pollution*. https://www.sciencemuseum.org.uk/objects-and-stories/chemistry/something-air-james-lovelock-and-atmospheric-pollution (accessed October 2022).

(9) J.E. Lovelock and O.M. Lidwell, The Lancet, **238**(6172), 746 (1941).

(10) J.E. Lovelock, *J. Chromatogr. A.* **1**, 35–46 (1958).

(11) R. Carson, *Silent Spring* (Houghton-Mifflin, Boston, MA, 1962).

(12) J.E. Lovelock, R.J. Maggs, and R.J. Waade, *Nature (London)* **241**, 194–196 (1973).

(13) F.S. Roland and M.J. Molina, *Rev. Geophys. Space Phys.* **13**, 1–36 (1975).

(14) J.E. Lovelock, *GAIA: A New Look at Life on Earth* (Oxford University Press, Oxford, United Kingdom, 1979).

ABOUT THE AUTHOR

Nicholas H. Snow is the Founding Endowed Professor in the Department of Chemistry and Biochemistry at Seton Hall University, and an Adjunct Professor of Medical Science. During his 30 years as a chromatographer, he has published more than 70 refereed articles and book chapters and has given more than 200 presentations and short courses. He is interested in the fundamentals and applications of separation science, especially gas chromatography, sampling, and sample preparation for chemical analysis. His research group is very active, with ongoing projects using GC, GC–MS, two-dimensional GC, and extraction methods including headspace, liquid–liquid extraction, and solid-phase microextraction. Direct correspondence to: LCGCedit@mmhgroup.com

Overcoming Incomplete Peptide Mapping of Antibody Complementarity-Determining Regions with Alternate Digestion Workflows

Peptide mapping of antibodies is an essential method to monitor peptide modifications in antibody lots that could affect the safety and efficacy of the product. Conventional protocols rely on protein digestion using proteases, such as trypsin, before mapping with mass spectrometry (MS). However, trypsin digestion may cause incomplete mapping of peptides, especially those that include highly hydrophobic peptides. Here, we show how pepsin can be used as an alternative and complementary protease for digestion that allows for improved sequence coverage, especially in proteins with highly hydrophobic regions. We also show that using guanidine hydrochloride (GuHCl) post-digestion improves peptide mapping results. Overall, these two methods—pepsin digestion and GuHCl post-digestion—can be used to provide more comprehensive antibody peptide maps, thereby enabling more thorough quality checking of biopharmaceutical products.

Ken Cook, Dan Bach Kristensen, Martin Ørgaard, and Trine Meiborg Sloth

Peptide mapping by liquid chromatography–mass spectrometry (LC–MS) has become the method of choice for in-depth characterization of biotherapeutic proteins, such as monoclonal antibodies (mAbs). The method is now commonly referred to as multi-attribute monitoring (MAM) because it enables simultaneous analysis and monitoring of multiple quality attributes in a single experiment, which traditionally would require multiple independent analytical experiments (1). The MAM approach allows scientists to confirm the primary structure of peptides, identify possible post-translational modifications (PTMs), and provide valuable quantitative information for critical quality attributes (CQAs) at the amino acid level (2,3). Biopharmaceutical companies are now increasingly using MAM for current Good Manufacturing Practice (cGMP) release testing (4), and there is an increasing regulatory awareness and expectation that MAM is employed as an analytical tool in biopharmaceutical development (5–7). Generally, the therapeutic protein is digested by a site-specific protease, and the resulting peptides are separated with LC and detected by LC–MS. The most common protease used for this purpose is trypsin, which cleaves after the basic amino acids, arginine and lysine, and several optimized MAM workflows based on trypsin, have been published (8). Lately, automation of MAM workflows has been gaining traction and recent publications have demonstrated good interlaboratory robustness and consistent results compared to the more conventional and laborious MAM workflows (9,10).

Although trypsin works well for the majority of antibodies and other biopharmaceuticals, trypsin digestion, in some cases, results in small peptides, which are too hydrophilic to be retained on the C18 column. Likewise, in this study, we demonstrated that trypsin digestion may result in large hydrophobic peptides, which absorb to surfaces and return no detectable peptides in LC–MS analysis. In particular, antibody sequences that contain several aromatic amino acids may result in highly hydrophobic peptides after trypsin digestion. These hydrophobic peptides can be difficult to recover, making trypsin-based peptide mapping impractical for those antibodies, which can be problematic if the elusive peptide is part of critical sequences of the antibody, such as the complementarity-determining regions (CDRs) that define antibody epitope specificity. CDRs are important for antibody function because this is where the antibody binds to its specific antigen. Therefore, it is crucial to monitor for peptide modifications in the CDR regions of antibodies.

Here, we demonstrate the use of pepsin as an alternative protease to trypsin and the use of guanidine hydrochloride (GuHCl) post-digestion to access difficult and otherwise elusive hydrophobic sequence regions. Both measures can mitigate against challenges in digestion strategies and provide comprehensive peptide maps, including elusive sequences in the CDR. The MAM workflow presented here builds on the simple and automated digestion workflow published previously (9).

Materials and Methods

We utilized an analytical method that involved automated sample preparation using the Smart Digest Pepsin kit (Thermo Fisher Scientific) combined with LC tandem MS (LC–MS/MS)-based peptide mapping. For all peptide mapping analyses, prepared peptides were analyzed using the Vanquish Horizon ultrahigh-pressure liquid chromatography (UHPLC) system and the Acclaim Vanquish C18 column. Mass spectrometric detection was performed using an Q-Exactive plus or Orbitrap (orbital

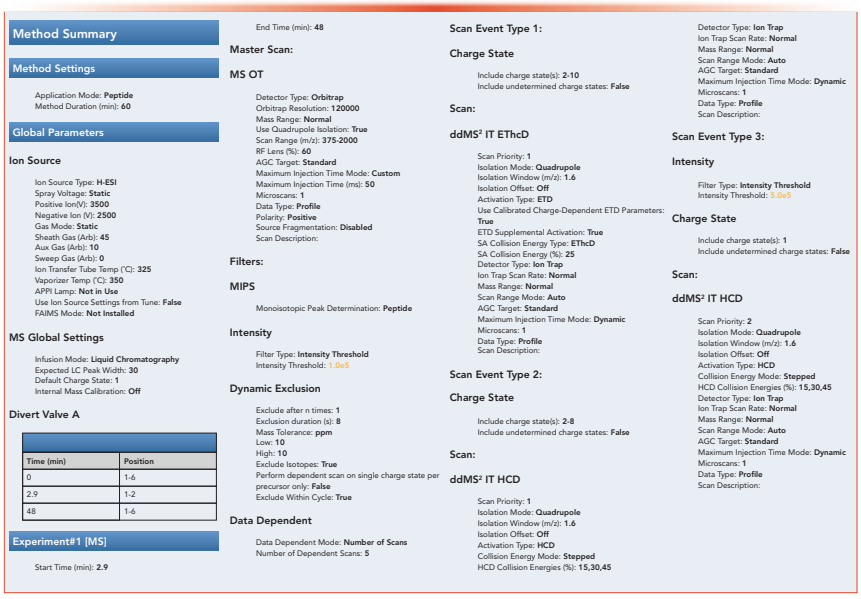


FIGURE 1: Detailed description of LC–MS method.

Instrument Parameters	
Precursor Mass Tolerance	3,00 ppm
Fragmentation Type	Use Thermo scan headers
Fragment Mass Tolerance 1	0,30 Da
Fragment Mass Tolerance 2	0,30 Da
Recalibration (lock mass)	None
Digestion	
Cleavage Site(s)	LWFYCM
Cleavage Side	C-terminal
Digestion Specificity	Non specific (slowest)
Missed Cleavages	-1
Modifications	
Modifications	Oxidation/ +15.994915 @ M,W | rare 1 Deamidated/ +0.984016 @ N,Q | rare 1 Gln->pyro-Glu/ -17.026549 @ NTerm | common1
Total Common Max	1
Total Rare Max	1

FIGURE 2: Details for software search parameters.

ion trap) Fusion Tribrid high-resolution accurate mass (HRAM) spectrometer (all instrumentation from Thermo Scientific).

Sample preparation and digestion was automated using the KingFisher Duo Prime purification system with 96 deep-well plates, controlled by Thermo Scientific's BindIt 4.0 software. Digestions were performed using magnetic Smart digestion kits, which use heat (>70 °C) to denature proteins. Next, 200 µg of the sample was added, and a Smart digest buffer was added to a final volume of 200 µL in each KingFisher 96 deep-well plate, which was digested in the KingFisher Duo for 30 min at 75 °C. After digestion, the samples were actively cooled to 5 °C, and 10 µL of 20% trifluoroacetic acid (TFA) and 70 µL of 8 M guanidine hydrochloride (GuHCl) were added to each sample.

A rapid slit seal cover was placed on the 96 deep-well plates, which were then mixed using a thermomixer and transferred directly to the Vanquish Horizon autosampler for peptide mapping analysis. Peptide mapping by LC–MS/MS was done using data-dependent acquisition (see Figure 1).

LC–MS/MS data were processed using Protein Metrics BYOS software, using the search parameters shown in Figure 2.

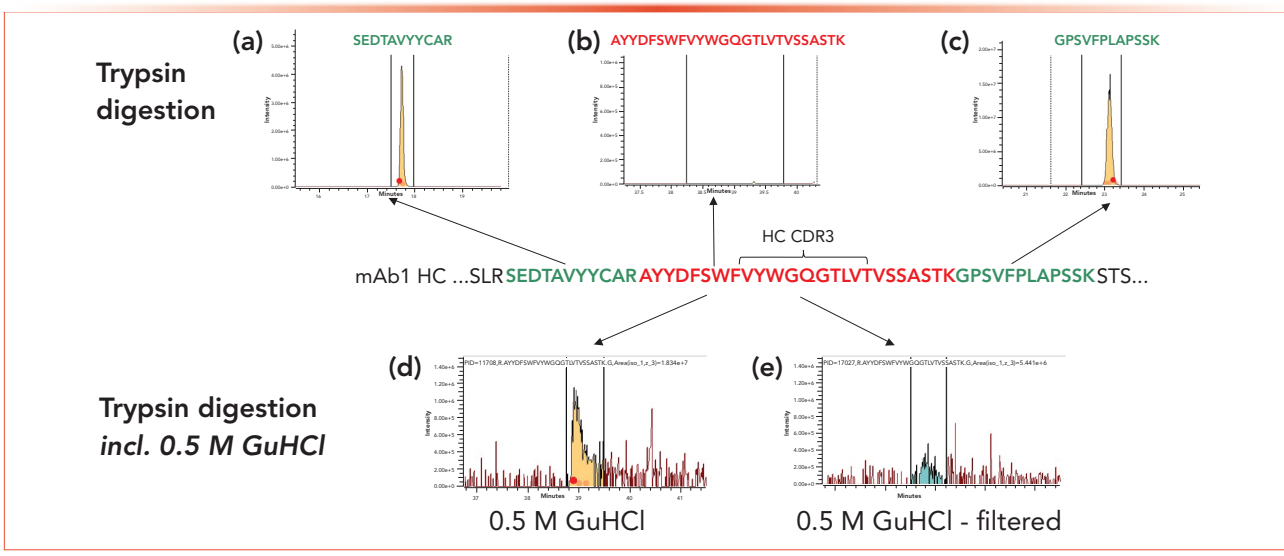


FIGURE 3: (a,c) Flanking peptides (green amino acids) confirm good trypsin activity; however, the (b) CDR region (red amino acids) was poorly digested. (d and e) Additives like GuHCl (0.5 M) during and after digestion did not solve the issue. For all figures, abscissa label is *minutes* and the ordinate label is *intensity*.

Results

Part One: Trypsin Versus Pepsin Digestion

Complete peptide mapping of antibodies is not always feasible with trypsin when the antibody sequence results in small hydrophilic peptides that are not retained on the C18 column or if the peptides contain several aromatic amino acids, which results in the production of highly hydrophobic peptides. As such, mapping of some antibodies misses coverage of the CDR if hydrophobic regions are present when using trypsin digestion methods. This inability to reliably map CDRs is a major analytical challenge for the characterization and QC of antibodies because this region is critical for antibody function and binding.

An example is shown in Figure 3, where results of tryptic digestion of mAb-1 showed incomplete mapping of the CDR. It was clear that trypsin was successfully digesting the region (that is, flanking peptides were detected, as seen in Figure 3a), but the peptide spanning the CDR3 region was not detected (Figure 3b). Additives such as urea, GuHCl, dimethyl sulfoxide (DMSO), isopropanol, and acetonitrile that were added during and after digestion did not solve the issue (Figure 3d and 3e).

Investigating this issue further, we identified other mAbs with the missing sequence coverage in CDRs containing numerous aromatic, highly hydrophobic amino acids. From these findings, we hypothesized that using an alternative protease that digests around the aromatics could create more accessible fragments.

Therefore, we focused on pepsin as an alternative to trypsin. Pepsin digestion resulted in improved sequence coverage in CDRs of all mAbs (see one example in Figure 4b) compared to trypsin digestion (Figure 4a). Figure 4c shows extracted ion currents for peptides from the pepsin digest covering the CDR-3 region, a region which is not covered by the trypsin digestion (Figure 4a).

Overall, these results show that pepsin is a viable alternative and complementary digestion method to trypsin for the peptide mapping of antibodies with challenging hydrophobic regions, including those around key regions, like CDRs.

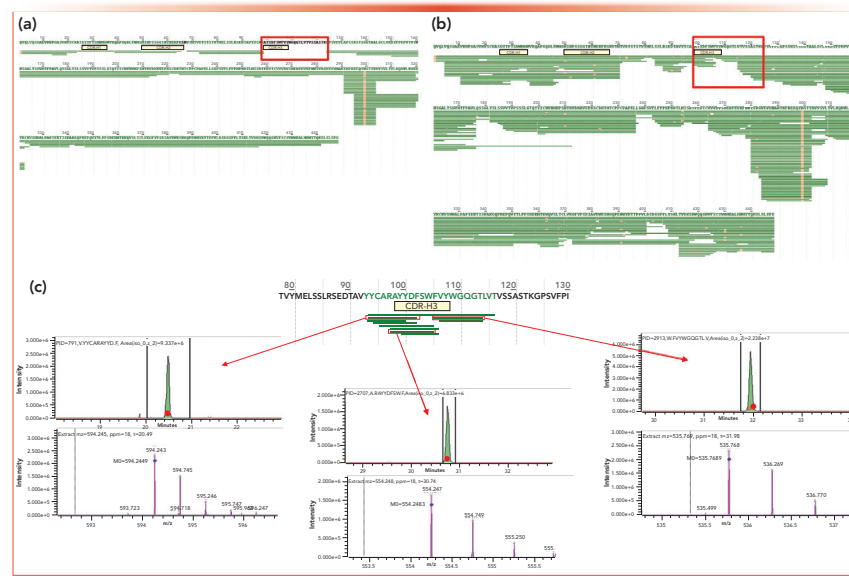


FIGURE 4: (a) Peptide map showing absence of peptide mapping in the CDR region (marked red) with trypsin digestion, (b) peptide map showing successful detection with pepsin digestion. (c) Example peptide sequences with validated pepsin-based sequences of CDR region. (c) For upper set of three figures, abscissa is labelled *minutes*, and ordinate is labeled *intensity*. For lower set of three figures, abscissa is labeled *m/z*, and ordinate is labeled *intensity*.

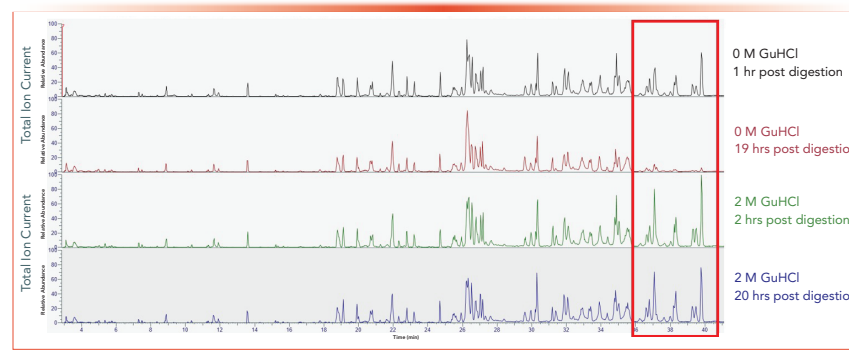


FIGURE 5: Time dependent loss of hydrophobic peptides post-digestion in autosampler; *time (minutes)* versus *total ion current*.

Part Two: Guanidine Hydrochloride Post-Digestion

During our experiments, we noticed a tendency for the hydrophobic peptide sample signal intensity to drop when left for elongated times in the autosampler, which led to the idea that adding GuHCl after digestion could be an effective way of maintaining peptides in a solution and prevent loss of the hydrophobic peptides to the wall of the sample vials. Here, we tested whether post-digestion treatment of samples with GuHCl improved peptide recovery after prolonged storage in the autosampler.

We performed Smart Trypsin digestion of a mAb, subsequently adding 8 M GuHCl to a final concentration of 2 M GuHCl in one of the sample wells directly after digestion. No GuHCl was added to the other mAb digest. The samples were then placed in the autosampler at 5 °C until time of analysis. Samples with 0 M and 2 M GuHCl were analyzed between 1–20 h post-digestion (see Figure 5).

• Continued on Page 549

THE NEXT GENERATION

A special article series highlighting the work of young analytical chemists.

Tools for Investigating the Expanding Per- and Polyfluoroalkyl Substances (PFAS) Universe

Per- and polyfluoralkyl substances (PFAS) continue to need novel analytical tools for their proper characterization. Regulatory bodies are lowering the acceptable concentrations for contaminated matrices, requiring state-of-the-art instrumentation to detect sub-ng/L levels. Suspect screening (for known PFAS without certified analytical standards) is becoming more popular as the list of known PFAS grows to thousands of compounds. The methods that currently exist for suspect screening require updating to move them from boutique, study-by-study methods to those that can be adopted more universally to generate transferable data sets. Some of those updates are presented here with the hope that they are adopted by laboratories in industry, government, and academia.

Trever Schwichtenberg

Per- and polyfluoroalkyl substances (PFAS) are a class of compounds characterized by the presence of highly inert carbon–fluorine bonds that impart both oleophobic and hydrophobic properties. These properties have made PFAS useful ingredients in a variety of manufacturing processes, including food packaging, papers and textiles, industrial nonstick coatings, and aqueous film-forming foams (AFFFs) for extinguishing hydrocarbon-based fuel fires. These compounds are sometimes called "forever chemicals," because they do not break down readily in the environment and those that do (precursors) degrade to more stable end products. The U.S. Environmental Protection Agency (EPA) published a strategic roadmap in 2021 outlining plans to deal with PFAS contamination (1).

The world of PFAS is an exciting avenue of exploration for analytical chemists, because advanced studies involving fate and transport in the environment, ecotoxicity, and treatment can be performed. However, fundamental questions about the detection, quantification, and communication of PFAS results still need to be answered. These challenges bring together separations science, mass spectrometry (MS), data science, environmental engineering, and even public policy.

What constitutes a PFAS molecule has been an ongoing debate. Common among definitions is the presence of a carbon–fluorine bond. However, many molecules can contain a single carbon–fluorine bond and would not be considered PFAS. The main definitions have been built into the PubChem Classification Browser (2), where analysts can generate lists of PFAS according to each definition. Various lists of PFAS have been curated; as a result, there is a lack of consensus because of the varying definitions across lists. The list we chose to use in our studies is one curated by the National Institute of Standards and Technology (NIST) of the United States (3). All lists contain varying amounts of data for each entry, with each having at least the molecular formula. The NIST list contains the molecular structure in a machine-readable format (international chemical identifier [InChI] in this case), and this structure becomes the identifier for each compound. Acronyms and the International Union of Pure and Applied Chemistry (IUPAC) names have been inconsistent across lists; InChIs are a way to unify communication of the results (4). The conversion of InChI keys to draw molecular structures presents an approachable and time-saving exercise in computer coding for the intrepid chemist.

Trever Schwichtenberg

MS has been the choice detector, coupled with gas chromatography (GC) or liquid chromatography (LC), for volatile or non-volatile PFAS analysis, respectively. Most PFAS receiving regulatory scrutiny have been non-volatile, and LC–MS has been the most often used tool. Targeted analysis of PFAS is done using

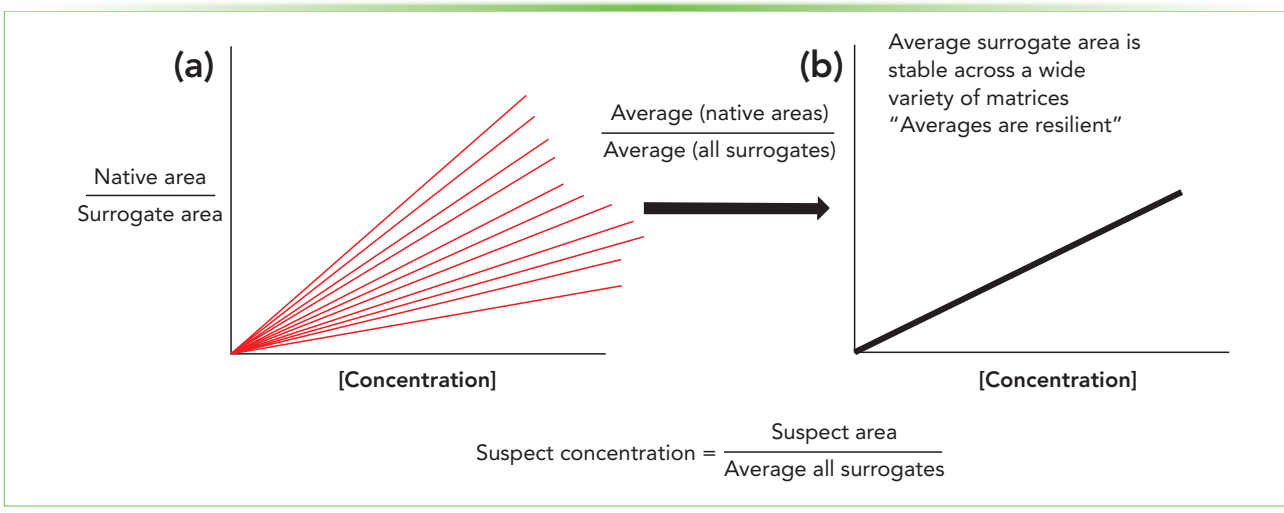


FIGURE 1: Graphic depicting (a) multiple calibration curves and (b) the single PFAS curve for estimating suspect concentrations. One-to-one matching involved estimating which calibration curve would most closely mimic the suspect PFAS. (b) The new technique uses one curve for all suspect PFAS, and would unify semi-quantitative techniques across studies and laboratories.

triple-quadrupole detectors to monitor ion transitions for quantification. The availability of analytical standards has increased greatly over the past 20 years, from almost no standards to mixes where dozens of PFAS can be included in a targeted run. Not all native standards have an exact-matched isotopically labeled surrogate, so some surrogates are borrowed by natives within a PFAS class. It is best to avoid recruiting surrogates from natives where high concentrations are expected. High concentrations will suppress the surrogate signal, which is fine for the exact-matched

native but not for a native borrowing the surrogate. For low concentrations, advances in instrumentation and preconcentration steps have led to sub-ng/L levels of detection (LOD) for certain PFAS. Such low levels make contamination an issue not to be underestimated, even in routine targeted analysis.

After targeted analysis, suspect screening and non-targeted analysis are the next frontier in PFAS identification. Suspect screening requires the use of a high-resolution detector, such as a quadrupole time-of-flight (QTOF), ion mobility (IMS), or an orbital trap MS instrument. Mass-to-charge (*m/z*) ratios are screened against a list of PFAS to look for similar masses. Then, confidence in the identified compound is increased through isotope ratio differences, the presence of diagnostic fragments, and other data acquired. These confidence levels were initially communicated using the scale developed by Schymanski and others for all non-targeted MS work (5). This scale was the framework for a new scale that our team created to communicate confidence of PFAS identification using features specific to PFAS (6).

The growth in awareness of the true number of PFAS has far outpaced the availability of analytical standards for true quantitative analysis. The demand for PFAS research led to the idea of reporting new or suspected compounds in a semi-quantitative way. Most new or suspected PFAS are either part of an existing homologous series, or have only small changes to headgroups. For new or suspected PFAS, one would borrow the calibration curve from a similar compound that does have an analytical standard, assume an equal molar response by the detector, adjust for molar mass differences, and call this a semi-quantitative concentration. This technique assumes that the ionization efficiencies between the suspect PFAS and the borrowed calibration PFAS are similar. Although this strategy worked initially, a number of problems have been found over time. The discovery of new PFAS compounds has outpaced the ability to match them to existing standards. As studies have become more complex, involving multiple environmental matrices with vastly different PFAS concentrations, the matching had to be altered for different matrices. Different matching schemes are used across the literature, making comparisons across studies difficult. Finally, different laboratories run slightly different analyte lists, making a universal matching scheme unattainable. To address this, we developed a new method to estimate concentrations that we believe can accommodate the rapid growth in demand for PFAS testing and information.

The PFAS curve replaces one-to-one matching with a single, average curve comprising all the PFAS standards run in a particular study (Figure 1). The same curve is used for all suspected PFAS, including non-targeted PFAS where the exact structure and location of functional groups (and thus estimated ionization efficiency) may not be known. This technique was partially inspired by an idea from data censoring, where traditional techniques, such as substitution of values below the limits of quantification (LOQ), presented a false sense of knowledge about the data (7). In the case of one-to-one matching, ionization efficiencies of suspect compounds can only be known using analytical standards, and without those, the ionization efficiencies are an educated guess at best. The single PFAS curve is likely wrong for most suspects, but it is wrong systematically, and if multiple studies use this technique, the estimated concentrations are comparable across data sets. Finally, it is a computationally simple approach compared to one-to-one matching and has a better chance of being adopted by laboratories across government, academia, and industry. We are currently working on a study using ^{19}F-NMR to measure the total organic fluorine to quantify the variability of this method.

The details of naming and quantifying PFAS are important when the scale of PFAS contamination is realized. A recent study proposed that most rainwater, even in remote areas, contained PFAS above the health advisory levels set by the U.S. EPA (8). Solutions to this problem will require new thinking and incorporating new methods of data analysis, such as machine learning. The more data can be generated and shared, the better these problems and solutions can be understood. Some of the techniques outlined here should help in streamlining and automating much of that work.

Acknowledgments

The author would like to thank the Strategic Environmental Research and Development Program (SERDP) for funding through grants ER19-1205 and ER20-1375.

References

(1) *PFAS Strategic Roadmap: EPA's Commitments to Action 2021-2024* (U.S. Environmental Protection Agency, Washington, D.C., EPA Report EPA-100-K-21-002, 2021).

(2) National Center for Biotechnology Information, *PubChem Classification Browser.* https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120 (accessed October 2022).

(3) B.J. Place, *Suspect List of Possible Per- and Polyfluoroalkyl Substances (PFAS), National Institute of Standards and Technology,* (2021). https://doi.org/10.18434/mds2-2387

(4) S.R. Heller, A. McNaught, I. Pletnev, S. Stein, and E. Tchekhovskoi, *J. Cheminform.* **7,** 23 (2015). https://doi.org/10.1186/s13321-015-0068-4

(5) E.L Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, and J. Hollender, *Environ. Sci. Technol.* **48**(4), 2097–2098 (2014). https://pubs.acs.org/doi/pdf/10.1021/es5002105

(6) J.A. Charbonnet, C.A. McDonough, F. Xiao, T. Schwichtenberg, D. Cao, S. Kaserzon, et al, *Environ. Sci. Technol. Lett.* **9**(6), 473–481 (2022). https://pubs.acs.org/doi/10.1021/acs.estlett.2c00206

(7) D.R. Helsel, *Statistics for Censored Environmental Data Using Minitab and R* (John Wiley & Sons, Hoboken, NJ, 2nd ed., 2012).

(8) I.T. Cousins, J.H. Johansson, M.E. Salter, B. Sha, and M. Scheringer, *Environ. Sci. Technol.* **56**(16), 11172–11179 (2022). https://pubs.acs.org/doi/pdf/10.1021/acs.est.2c02765

ABOUT THE AUTHOR

Trever Schwichtenberg is a Postdoctoral Scholar at the Oregon Health and Science University, in Portland, Oregon. Direct correspondence to: trever.schwichtenberg@gmail.com

United States Postal Service
STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION
Required by 39 USC 3685

1. ***Publication Title:*** *LCGC*
2. ***Publication Number:*** 698950
3. ***Filing Date:*** 9-28-22
4. ***Issue Frequency:*** Monthly
5. ***Number of Issues Published Annually:*** 12
6. ***Annual Subscription Price:*** $97.99
7. ***Complete Mailing Address of Known Office of Publication (Not Printer):***
MULTIMEDIA HEALTHCARE LLC, 2 Clarke Dr, Suite 100, Cranbury, NJ 08512-3619 228
8. **Complete Mailing Address of Headquarters or General Business Office of Publisher (Not Printer):**
MULTIMEDIA HEALTHCARE LLC, 2 Clarke Dr, Suite 100, Cranbury, NJ 08512-3619
9. ***Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor***
Publisher: Brian Haug, MULTIMEDIA HEALTHCARE LLC, 485F ROUTE 1 S STE 210, ISELIN NJ 08830-3072
Editor: Laura Bush, MULTIMEDIA HEALTHCARE LLC, 485F ROUTE 1 S STE 210, ISELIN NJ 08830-3072
Managing Editor: John Chasse, MULTIMEDIA HEALTHCARE LLC, 485F ROUTE 1 S STE 210, ISELIN NJ 08830-3072
10. ***Owner - Full Name:*** MULTIMEDIA HEALTHCARE LLC, LLC, 2 Clarke Dr, Suite 100, Cranbury, NJ 08512-3619
11. ***Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities:***
If none, check box. → ☑ None
12. ***Does Not Apply***
13. ***Publication Title:*** *LCGC*
14. ***Issue Date for Circulation Data Below:*** September 2022
15. ***Extent and nature of circulation***

	Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
A. Total Number of Copies (Net press run)	32,017	30,831
B. Legitimate Paid and/or Requested Distribution		
1. Outside County Paid/Requested Mail Subscriptions Stated on PS Form 3541	23,016	22,878
2. In-County Paid/Requested Mail Subscriptions Stated on PS Form 3541	0	0
3. Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid or Requested Distribution Outside USPS	0	0
4. Requested Copies Distributed by Other Classes Mailed Through the USPS	21	9
C. Total Paid and/or Requested Circulation (Sum of 15b (1), (2), (3), and (4))	23,037	22,887
D. Non-requested Distribution (By Mail and Outside the Mail)		
1. Outside County Non-requested Copies Stated on PS Form 3541	8,962	7,922
2. In-County Non-requested Copies Stated on PS Form 3541	0	0
3. Non-requested Copies Distributed Through the USPS by Other Classes of Mail	0	0
4. Non-requested Copies Distributed Outside the Mail	0	0
E. Total Non-requested Distribution (Sum of 15d (1), (2), and (3))	8,962	7,922
F. Total Distribution (Sum of 15c and 15e)	31,999	30,809
G. Copies not Distributed	19	22
H. Total (Sum of 15f and 15g)	32,017	30,831
I. Percent Paid and/or Requested Circulation	71.99%	74.29%

16. ***If total circulation includes electronic copies, report that circulation on lines below***

A. Requested and Paid Electronic Copies	1,909	560
B. Total Requested and Paid Print Copies + Requested/Paid Electronic Copies	24,946	23,447
C. Total Requested Copy Distribution + Requested/Paid Electronic Copies	33,908	31,369
D. Percent Paid and/or Requested Circulation (Both print & Electronic Copies)	73.57%	74.75%

17. ***Publication of Statement of Ownership*** - Will be printed in November 2022 issue of this publication.

18. ***I certify that all information on this form is true and complete. Signature and title of Editor, Publisher, Business Manager, or Owner***
Jonathan Severn, Circulation Director

Date: 9/28/2022

• Continued from Page 545

For the sample with 0 M, GuHCl, a distinct loss of hydrophobic peptides is observed for 19 h post-digestion (see red box in Figure 5). In contrast, no loss in signal intensity is observed 20 h post-digestion for the sample with 2 M GuHCl. These results confirm that 2 M GuHCl effectively prevented loss of hydrophobic peptides during storage at 5 °C in the autosampler.

Based on these results, the post-digestion addition of 2 M GuHCl is now a routine part of our peptide mapping workflows for antibodies and other biopharmaceutical products.

Conclusion

Complete sequence coverage by peptide mapping of antibodies is not always possible (for example, when antibodies contain highly hydrophobic regions rich in aromatic residues). There may consequently be poor coverage of CDRs of antibodies when using conventional trypsin digestion methods, which poses an analytical challenge for the characterization and QC of such antibodies.

Here, we show two methodologies that increase the coverage of antibody peptide mapping in chromatography and MS-based workflows: (i) substitution of trypsin for pepsin as a digestion protease increase the coverage of peptide mapping, and (ii) post-digestion addition of GuHCl to prevent time-dependent loss of hydrophobic peptides in the autosampler prior to analysis. These two methodologies improve MS-based workflows, enabling highly accurate and reproducible antibody peptide mapping results. Coupled with the right software tools, high-throughput and ease-of-use capabilities of modern MS systems, the improvements presented here enables laboratories to efficiently screen and characterize challenging mAbs.

References

(1) R.S. Rogers et al, *MAbs* **7,** 881–890 (2015).
(2) K. Diepold et al, *PLoS One* **7,** e30295 (2012).
(3) X.C. Yu et al., *Anal. Chem.* **83**(15), 5912–5919 (2011).
(4) D. Ren, *Trends Biotechnol.* **38,** 1051–1053 (2020).
(5) S.Rogstad et al., *Anal. Chem.* **91,** 14170–14177 (2019).
(6) I. Apostol et al., *Curr. Opin. Biotechnol.* **71,** 206–215 (2021).
(7) R.S. Rogers et al., *AAPS J.* **20,** 7 (2018).
(8) T. Mouchahoir and J.E. Schiel, *Anal. Bioanal. Chem.* **410,** 2111–2126 (2018).
(9) S. Millán-Martín et al., *Anal. Bioanal. Chem.* **412,** 6833–6848 (2020).
(10) C. Qian, B. Niu, R.B. Jimenez,J. Wang, and M. Albarghouthi, *J. Pharm. Biomed. Anal.* **198,** 113988 (2021).

Ken Cook is the Manager Expert Support Biopharma at Thermo Fisher Scientific, in Hempstead, Hertfordshire, United Kingdom. Dan Bach Kristensen is a Principal Scientist at Symphogen, in Copenhagen, Denmark. Martin Ørgaard and Trine Meiborg Sloth are Senior Technicians at Symphogen. Direct correspondence to: ken.cook@thermofisher.com

VIEWPOINTS

Sample Preparation by Electric Field–Assisted Extraction

Han Chen and Robbyn K. Anand

Extraction-based sample preparation has been recognized as a significant step in separation science. To improve the sensitivity and selectivity of the extraction, some new approaches have been introduced, including those that employ an electric field. These electric field–assisted extraction methods create another dimension for sample preparation, and are compatible with miniaturized, portable, and multifunctional analytical platlate proteins, organelles, viral capsids, and biological cells, among others.

Analytes of environmental or biological origin occur in complex matrices, which complicates their separation and detection. For this reason, sample preparation methods such as extraction are a crucial step prior to chromatographic analysis. The goals of sample preparation technologies are purification and pre-enrichment of target analytes to be separated and then detected by the instrument. Purification reduces the concentration of interfering matrix substances, which affects separation and measurement through a decrease in sensitivity and a significant increase in maintenance, loss in chromatographic efficiency, and distortion of peak shape. In addition, high-abundance species might overwhelm the column. Extraction accomplishes purification and pre-enrichment by partition of the analyte into or onto an extraction phase.

Advances in extraction seek to improve selectivity and to boost pre-enrichment of the analyte through its increased mass transport to and uptake into or onto the extraction phase. For systems in which analytes or interferents are charged, integration of electric fields normal to the phase boundary can be leveraged to improve these features of the extraction. Electric fields can be generated and modulated by a simple pair of electrodes, and are by nature a green, cost-effective, and user-friendly addition to an extraction protocol. For this reason, the use of electric fields in extraction is an important modern trend in analytical chemistry. Although many electric field–assisted separation techniques have been developed and studied in depth, such as electrodialysis, gel electrophoresis, capillary electrophoresis, and capillary electrochromatography, the use of electric fields in extraction is still an underdeveloped area.

In the context of extraction, there are three areas where an applied voltage can be employed. Electric fields can be used to drive more rapid mass transport of charged analytes to the extraction phase, as well as partition of a species with a certain charge to increase selectivity of the extraction. Further, electrosynthesis can be utilized to synthesize and prepare membranes and solid-phase sorbents for extraction. It is notable that use of an electric field to perform any of these three functions can be readily implemented in microfluidic devices to create highly integrated analytical platforms.

Using an electric field in extraction has several advantages, such as improved selectivity for transfer of ionized species, efficient matrix cleanup, and modulation of the sample preparation process for potentially exhaustive extraction of target species. Electric fields have been employed in both solid-phase and liquid–liquid extractions.

Electric fields have been employed to enhance solid-based extraction by either direct manipulation of extraction or by migration of charged species to the interface during the extraction phase or away from the interface during the elution stage. Compared to traditional solid-phase extraction, electric field-enhanced extraction provides a new dimension to separation art, making it more versatile and dynamic.

Techniques that employ non-linear (graded) electric fields for electrokinetic extraction are less developed and have the potential to access new classes of analytes and to more dramatically enhance extraction efficiency. These techniques include dielectrophoresis (DEP), for manipulation of polarizable particles (including those lacking a net charge), and electric field gradient focusing (EFGF) methods. Most prevalently, these electrokinetic methods have been integrated with biological analysis. For example, DEP has been used to selectively accumulate proteins, organelles, viral capsids, and biological cells, among others.

We have briefly summarized applications of electric fields as a means to modulate extraction. These methods enrich target analytes, improve sensitivity, and hold the potential to integrate separation and detection. These methods are broadly applicable to biological, environmental, and pharmaceutical samples. These advantages warrant further investigation into scaling and commercialization of these methods and development of new non-linear electrokinetic techniques for enrichment of polarizable particles and charged analyte prior to extraction.

Han Chen is a PhD student, and **Robbyn K. Anand** is an associate professor of chemistry, at Iowa State University in Ames, Iowa. Direct correspondence to: rkanand@iastate.edu

Read the full original version. **Scan code for link.**

Medical World News®

powered by MJH life sciences

A 24-hour streaming program

For Health Care Professionals, By Health Care Professionals

Season 7 is streaming now!

www.medicalworldnews.com

Be Bold

We're on your side

Advancing your IC science–45 years strong

Learn more at **thermofisher.com/ICBold**

thermo scientific