

^1H NMR Signals from Urine Excreted Protein Are a Source of Bias in Probabilistic Quotient Normalization

Gonalo D. S. Correia,^{*,‡} Panteleimon G. Takis,^{*,‡} Caroline J. Sands, Anna M. Kowalka, Tricia Tan, Lance Turtle, Antonia Ho, Malcolm G. Semple, Peter J. M. Openshaw, J. Kenneth Baillie, Zoltán Takáts, and Matthew R. Lewis



Cite This: *Anal. Chem.* 2022, 94, 6919–6923



Read Online

ACCESS |



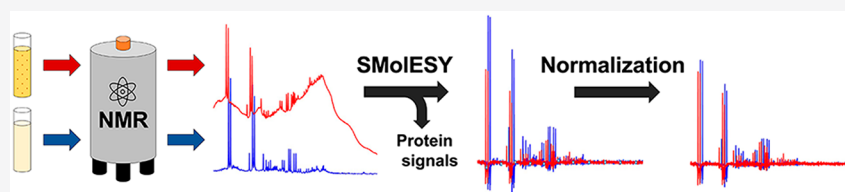
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Normalization to account for variation in urinary dilution is crucial for interpretation of urine metabolic profiles. Probabilistic quotient normalization (PQN) is used routinely in metabolomics but is sensitive to systematic variation shared across a large proportion of the spectral profile (>50%). Where ^1H nuclear magnetic resonance (NMR) spectroscopy is employed, the presence of urinary protein can elevate the spectral baseline and substantially impact the resulting profile. Using ^1H NMR profile measurements of spot urine samples collected from hospitalized COVID-19 patients in the ISARIC 4C study, we determined that PQN coefficients are significantly correlated with observed protein levels ($r^2 = 0.423$, $p < 2.2 \times 10^{-16}$). This correlation was significantly reduced ($r^2 = 0.163$, $p < 2.2 \times 10^{-16}$) when using a computational method for suppression of macromolecular signals known as small molecule enhancement spectroscopy (SMoLESY) for proteinic baseline removal prior to PQN. These results highlight proteinuria as a common yet overlooked source of bias in ^1H NMR metabolic profiling studies which can be effectively mitigated using SMoLESY or other macromolecular signal suppression methods before estimation of normalization coefficients.

Urine is a complex chemical mixture which contains metabolic end-products from host and associated microbiota, xenobiotics, and dietary compounds¹ in highly variable concentrations.² Urinalysis is routinely used in the clinic for noninvasive diagnosis of local conditions of urinary tract pathology and infection (e.g., via measurement of leukocytes and nitrite), systemic metabolic disease (e.g., diabetes via glucose), and to assess environmental and nutritional exposure. The ease of sample collection and diagnostic potential has made urine a focal point for biofluid-based metabolomics studies.³ ^1H nuclear magnetic resonance (NMR) spectroscopy emerged early in the evolution of the field⁴ as a suitable platform for such investigations owing to its broad linear dynamic range, excellent reproducibility, high throughput, and quantitative accuracy.^{5–7}

Urine exhibits strong intraday and interindividual variation in dilution owing to factors such as hydration status, kidney function, diet, medication, and voiding interval. This poses a fundamental challenge in urinalysis, especially of spot urine samples, that must be accounted for when accurately making or comparing chemical measurements⁸ in both clinical applications and metabolic profiling studies.⁹ Measurements of urinary creatinine, osmolality, specific gravity, or volume (in 24h collections) are routinely used in clinical biochemistry to

account for variable dilution (e.g., albumin-to-creatinine ratio) and allow comparison to normal reference ranges.¹⁰ However, in metabolomics studies, the use of profiling technologies such as ^1H NMR allows for the estimation of more robust normalization coefficients based on a broader view of the urinary metabolome. Probabilistic quotient normalization (PQN)¹¹ leverages the data captured in the metabolic profile as a whole and is the “gold standard” for computational normalization of ^1H NMR spectra in urinalysis studies. PQN uses the complete set of profile measurements to estimate a normalization coefficient for each sample that is representative of its dilution factor relative to a predefined reference (usually a median spectrum). Compared to other normalization methods (e.g., total area normalization), PQN provides robustness against bias from few signals whose intensity dominates the total profile integral. However, PQN is not

Received: January 27, 2022

Accepted: April 6, 2022

Published: May 3, 2022



suitable if a large proportion of variables (50% or more) covary systematically with factors other than the sample dilution.¹¹

The presence of urinary proteins can exert a broad effect on ¹H NMR-based metabolic profiles by elevating the spectral baseline and contributing broad signals in a concentration-dependent manner (Figure 1a).¹² Because a large proportion

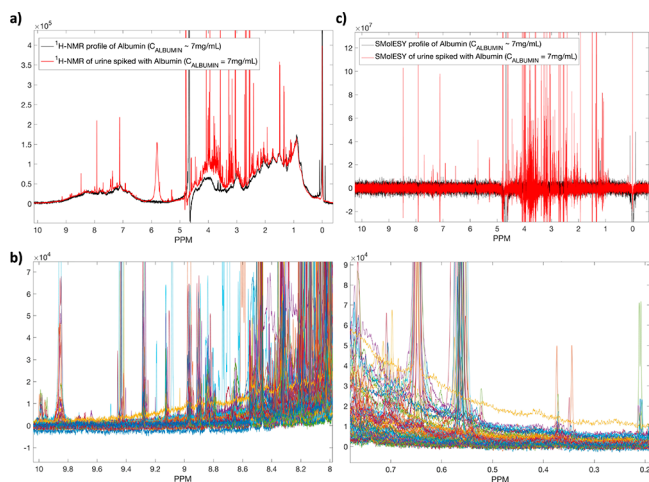


Figure 1. (a) The ¹H NMR spectrum of albumin (black line) compared to a real urine sample containing the same amount of albumin (~7 mg/mL). (b) Eighty-five urine ¹H NMR profiles from COVID-19 patients, focusing on the backbone -NH (left panel) and parts of methyl proteinic protons (right panel), showcasing the spectral baseline effect from the presence of proteinuria. (c) SMoESY application on the ¹H NMR spectrum of albumin (~7 mg/mL) and a real urine sample with the same amount of albumin. In both cases SMoESY succeeds in broad signals suppression as well as baseline homogenization, allowing the enhancement of sharp signals from small molecules (i.e., metabolites).

of signals across the profile are affected by the presence of substantial proteinuria, the fitness of PQN normalization for samples representative of many phenotypes may be challenged. Proteinuria is frequently encountered in chronic kidney disease,¹³ diabetes¹⁴ and obesity,¹⁵ and increases with age.¹⁶ Surprisingly, a thorough literature search did not reveal any critical assessment of the impact of proteinuria in the estimation of normalization coefficients. The ongoing Covid-19 pandemic and organized efforts to understand its underlying pathophysiological effects have provided both a need and the opportunity to evaluate this relationship, as proteinuria is prevalent in COVID-19 patients, independent of other comorbidities, and associated with disease severity and patient survival.^{17,18}

We used an ¹H NMR-based metabolic profiling approach to analyze urine samples from patients ($n = 1022$ spot urine samples from 711 patients) admitted to hospital with COVID-19, collected by the International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) following the WHO Clinical Characterization Protocol UK (CCP-UK). Further NMR experimental details are reported in the Supporting Information. Figure 1b shows several urine ¹H NMR profiles ($n = 85$) from COVID-19 patients, focusing on the backbone -NH and parts of methyl proteinic protons, clearly illustrating the effects of proteinuria on the small molecule profiles observed. The impact of macromolecular signals on the small molecule profile may be reduced when employing additional NMR experiments beyond the routine

one-dimensional general profile experiment (i.e., transverse relaxation (T_2) spectral editing experiments such as spin-echo pulse sequences¹⁹). However, the need for these additional experiments may not be anticipated at the outset of a urine profiling study, and their use both increases the experimental cost and decreases analysis throughput. We recently introduced small molecule enhancement spectroscopy (SMoESY)¹² as a computational alternative for removal of macromolecule-derived signals directly from routine one-dimensional NMR spectral profiles without the need of extra experiments, enabling the more specific and direct analysis of small molecule analytes (Figures 1c, S1, and S2). The approach can also be effectively reversed, removing the sharp small molecule-derived signals and providing an enhanced protein baseline for the purposes of urinary protein quantification (Figure 2a, S3).²⁰ Integration of the proteinic

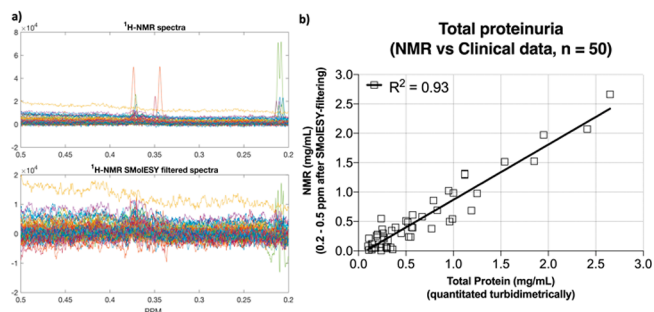


Figure 2. (a) Fifty ¹H NMR spectra from the COVID-19 cohort focusing on the proteinic methyl group-containing spectral region between 0.2 and 0.5 ppm (upper panel). After processed SMoESY filtering, the resulting spectra (bottom panel) are free from small metabolites sharp signals (or the remaining signals contain the same negative/positive part with almost zero integral) and their integration provides an estimate of the total protein. (b) The absolute quantification of total urinary protein via NMR highly correlates to the measured protein concentration by clinical methods (for both NMR and clinical methods see Supporting Information).

methyl group-containing spectral region between 0.2 and 0.5 ppm (Figure 2a) accurately represents the total amount of urinary protein as validated by comparison with turbidimetric measurement (Figure 2b). In the present study, we compared the PQN coefficients obtained from the standard one-dimensional (1D) ¹H NMR spectra and their SMoESY processed counterparts (Figure 3). Although there is a good correlation between both measures (Pearson's $\rho = 0.89$), there is also a visible trend in the deviations from the regression line associated with high levels of total protein quantified in each sample, confirming that proteinuria does influence the estimation of PQN coefficients. Urinary protein excretion alone explains approximately 42.3% ($r^2 = 0.423$) of the variance in PQN coefficients estimated from the 1D ¹H NMR spectra without macromolecular baseline removal (Figure 4a). This trend is greatly reduced ($r^2 = 0.163$) when the macromolecular signature is removed via SMoESY prior to PQN (Figure 4b), closer to the observed association between urinary creatinine and protein concentration ($r^2 = 0.063$, Figure S5).

A comparison between the PQN coefficients and creatinine concentrations is shown in Figure 5. While creatinine concentration can be affected by multiple biological factors, PQN coefficients should still correlate with creatinine levels in

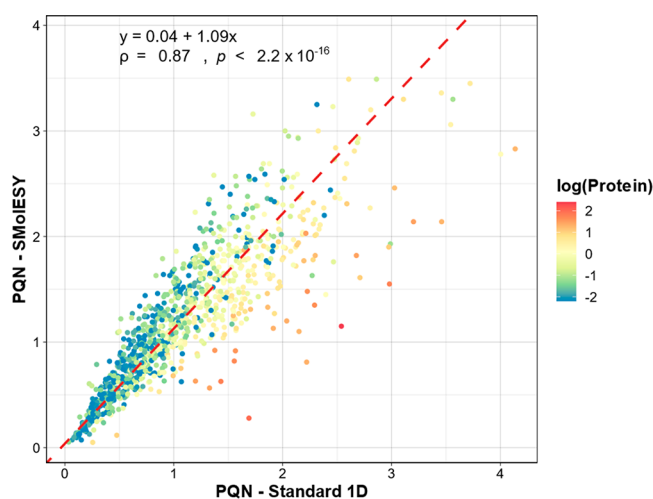


Figure 3. Agreement between PQN coefficients estimated from the standard ^1H NMR spectra and the corresponding SMoESY processed data. The linear regression trendline (dashed red line) was estimated with the orthogonal least-squares Passing–Bablok method. The regression coefficients, Pearson correlation coefficient (ρ) and the p -value from the two-sided correlation significance test are shown in the top left corner. Data points are colored by the natural logarithm of the estimated protein concentration (mg/mL).

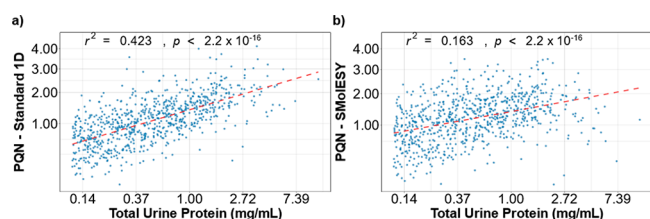


Figure 4. PQN coefficient variance explained (r^2 , estimated from the linear regression models plotted in red) by protein concentration, in the (a) standard and (b) SMoESY processed NMR spectra. Protein values equal or below the limit of detection (LOD) = 0.11 mg/mL were excluded for this analysis (final $n = 810$). PQN coefficients were square root transformed, and urine protein measurements were log-transformed.

spot urine samples, as observed in Figure 5a,b. Although the Pearson correlation between PQN coefficients and creatinine concentration is only marginally improved in the SMoESY data set ($\rho = 0.71$ vs $\rho = 0.69$, respectively) the residuals from the linear regression models (dashed red lines in Figure 5a,b) have a strong association with total urine protein ($r^2 = 0.43$, Figure 5c) in the original spectra, a trend which is reduced in SMoESY data ($r^2 = 0.1$, Figure 5d).

Figure 6 shows the principal component analysis (PCA) scores plots obtained when the SMoESY processed data set is normalized using PQN coefficients estimated from the standard 1D NMR spectra (Figure 6a) or from SMoESY data (Figure 6b). Despite the similarity in major trends in both plots, clustering based on protein concentration is more marked when spectra are normalized with the standard PQN procedure (Figure 6a).

Normalization procedures are crucial to correctly interpret urinary metabolic profiles. Here, we show that protein signals can confound probabilistic quotient normalization, and it is reasonable to assume this could also happen with other computational normalization methods. We recommend the removal of protein baseline signals prior to estimation of

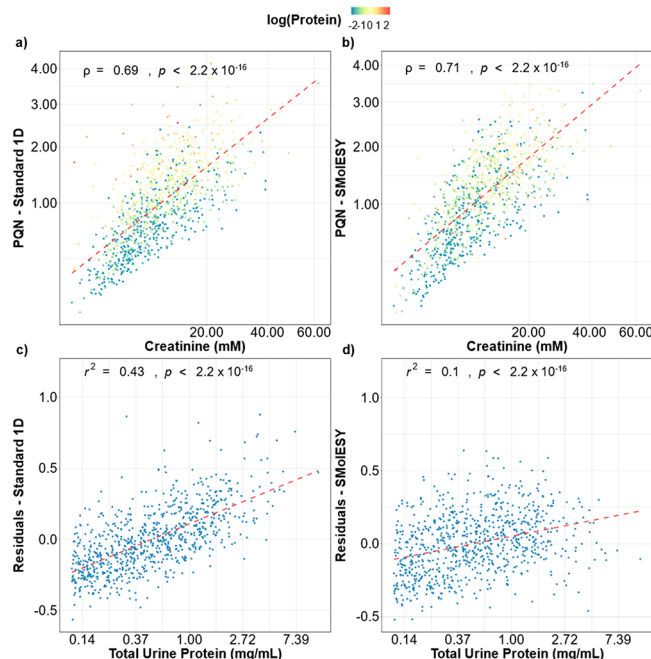


Figure 5. Correlation between PQN coefficients estimated from (a) the standard ^1H NMR or (b) SMoESY processed spectra and creatinine. Pearson correlation coefficient and p -value from the two-sided correlation significance test are shown in each figure. Data points are colored by the natural logarithm of the estimated protein concentration. PQN coefficients and creatinine measurements were square root transformed. (c,d) The residuals from the OLS regression trendlines (dashed red lines in panels a and b) and their residual association with total urine protein (log-transformed). Protein values equal or below the LOD = 0.11 mg/mL were excluded in panels c and d (final $n = 810$).

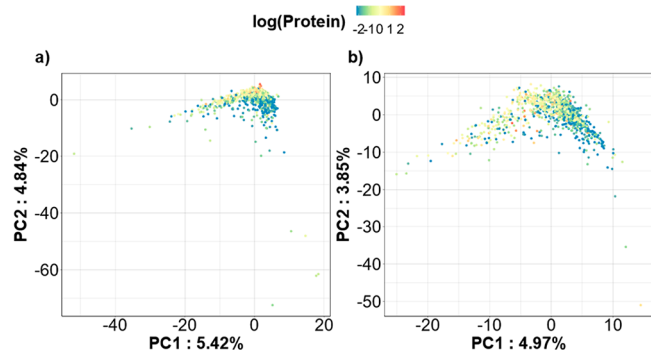


Figure 6. PCA scores plots for the SMoESY processed data set normalized with the PQN coefficients estimated from the (a) standard 1D and (b) SMoESY processed NMR spectra. NMR data was unit-variance scaled. Protein values were log-transformed and values equal or below LOD were imputed by replacement with the LOD value = 0.11 mg/mL.

normalization coefficients. This can be performed experimentally²¹ or with computational methods.^{22,23} However, we advocate the use of SMoESY, because of its ease of application to ^1H NMR spectra (including retrospective application where proteinuria is observed after the fact), being fast and highly effective method for removing broad baseline signals from protein and improving the estimation of normalization coefficients. Our observations and proposed methodology are of high importance for the accurate normalization of urine

biofluid ^1H NMR spectra, especially in the context of studies on diseases and phenotypes where proteinuria is likely to be present (e.g., diabetes, chronic kidney disease, pregnancy, infection, or protein rich diet).²⁴

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.2c00466>.

Urine samples collection details; NMR experiments/apparatus; NMR samples preparation SOPs; SMOESY data preparation; urinary creatinine quantification by ^1H NMR; urinary total protein quantification by clinical methods and ^1H NMR/SMOESY; statistical analyses details; PQN coefficients comparison with urinary creatinine concentration and osmolality measurements by clinical methods; correlation between urinary creatinine concentration and protein excretion; comparison between CPMG and SMOESY estimated PQN coefficients in human plasma heparin ^1H NMR spectra; supporting references (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Gonçalo D. S. Correia – Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom; National Phenome Centre, Imperial College London, London W12 0NN, United Kingdom; orcid.org/0000-0001-8271-9294; Email: gd2212@imperial.ac.uk

Panteleimon G. Takis – Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom; National Phenome Centre, Imperial College London, London W12 0NN, United Kingdom; orcid.org/0000-0002-7224-0412; Email: p.takis@imperial.ac.uk

Authors

Caroline J. Sands – Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom; National Phenome Centre, Imperial College London, London W12 0NN, United Kingdom

Anna M. Kowalka – Division of Diabetes, Endocrinology and Metabolism, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Clinical Biochemistry, Blood Sciences, North West London Pathology, Charing Cross Hospital, London W6 8RF, United Kingdom

Tricia Tan – Division of Diabetes, Endocrinology and Metabolism, Department of Metabolism, Digestion and Reproduction, Imperial College London, London W12 0NN, United Kingdom; Clinical Biochemistry, Blood Sciences, North West London Pathology, Charing Cross Hospital, London W6 8RF, United Kingdom

Lance Turtle – NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infection and Global Health, University of Liverpool, Liverpool L69 7BE, United Kingdom

Antonia Ho – MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow G61 1QH, United Kingdom

Malcolm G. Semple – NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool L69 7BE, United Kingdom; Respiratory Medicine, Alder Hey Children's Hospital, Liverpool L12 2AP, United Kingdom

Peter J. M. Openshaw – Faculty of Medicine, National Heart and Lung Institute, Imperial College London, London SW3 6LY, United Kingdom

J. Kenneth Baillie – Roslin Institute, University of Edinburgh, Edinburgh EH25 9RG, United Kingdom

Zoltán Takáts – Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom; National Phenome Centre, Imperial College London, London W12 0NN, United Kingdom; orcid.org/0000-0002-0795-3467

Matthew R. Lewis – Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London SW7 2AZ, United Kingdom; National Phenome Centre, Imperial College London, London W12 0NN, United Kingdom; orcid.org/0000-0001-5760-5359

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.analchem.2c00466>

Author Contributions

[‡]G.D.S.C. and P.G.T. contributed equally.

Notes

The authors declare no competing financial interest.

This work was done on behalf of and used human samples collected by the ISARIC4C consortium. The full membership list can be found in the following link: <https://isaric4c.net/about/authors/>

■ ACKNOWLEDGMENTS

This research is supported by grants from the National Institute for Health Research (NIHR) [award CO-CIN-01]; the Medical Research Council [grant MC_PC_19059, MC_PC_12025], the MRC UK Consortium for MetAbolic Phenotyping (MAP/UK) [grant number MR/S010483/1], and by the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool, in partnership with Public Health England (PHE), in collaboration with the Liverpool School of Tropical Medicine and the University of Oxford [award 200907]; NIHR HPRU in Respiratory Infections at Imperial College London with Public Health England (PHE) [award 200927]; infrastructure support was provided by the NIHR Imperial Biomedical Research Centre (BRC).

■ REFERENCES

- (1) Bouatra, S.; Aziat, F.; Mandal, R.; Guo, A. C.; Wilson, M. R.; Knox, C.; Bjorn Dahl, T. C.; Krishnamurthy, R.; Saleem, F.; Liu, P.; Dame, Z. T.; Poelzer, J.; Huynh, J.; Yallou, F. S.; Psychogios, N.; Dong, E.; Bogumil, R.; Roehring, C.; Wishart, D. S. *PLoS One* **2013**, *8* (9), No. e73076.
- (2) Takis, P. G.; Schäfer, H.; Spraul, M.; Luchinat, C. *Nat. Commun.* **2017**, *8*, 1662.

- (3) Dinges, S. S.; Hohm, A.; Vandergrift, L. A.; Nowak, J.; Habbel, P.; Kaltashov, I. A.; Cheng, L. L. *Nat. Rev. Urol.* **2019**, *16* (6), 339–362.
- (4) Nicholson, J. K.; O’Flynn, M. P.; Sadler, P. J.; Macleod, A. F.; Juul, S. M.; Sönksen, P. H. *Biochem. J.* **1984**, *217* (2), 365–375.
- (5) Takis, P. G.; Ghini, V.; Tenori, L.; Turano, P.; Luchinat, C. *TrAC - Trends Anal. Chem.* **2019**, *120*, 115300.
- (6) Keun, H. C.; Ebbels, T. M. D.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Chem. Res. Toxicol.* **2002**, *15* (11), 1380–1386.
- (7) Beckonert, O.; Keun, H. C.; Ebbels, T. M. D.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. *Nat. Protoc.* **2007**, *2* (11), 2692–2703.
- (8) Edmands, W. M. B.; Ferrari, P.; Scalbert, A. *Anal. Chem.* **2014**, *86* (21), 10925–10931.
- (9) Blaise, B. J.; Correia, G. D. S.; Haggart, G. A.; Surowiec, I.; Sands, C.; Lewis, M. R.; Pearce, J. T. M.; Trygg, J.; Nicholson, J. K.; Holmes, E.; Ebbels, T. M. D. *Nat. Protoc.* **2021**, *16*, 4299–4326.
- (10) Helmersson-Karlqvist, J.; Årnlöv, J.; Larsson, A. *Clin. Biochem.* **2013**, *46* (1–2), 70–72.
- (11) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. *Anal. Chem.* **2006**, *78* (13), 4281–4290.
- (12) Takis, P. G.; Jiménez, B.; Sands, C. J.; Chekmeneva, E.; Lewis, M. R. *Chem. Sci.* **2020**, *11* (23), 6000–6011.
- (13) Garg, A. X.; Kiberd, B. A.; Clark, W. F.; Haynes, R. B.; Clase, C. M. *Kidney Int.* **2002**, *61* (6), 2165–2175.
- (14) Karalliedde, J.; Viberti, G. *J. Am. Soc. Nephrol.* **2010**, *21* (12), 2020–2027.
- (15) Rosenstock, J. L.; Pommier, M.; Stoffels, G.; Patel, S.; Michelis, M. F. *Frontiers in Medicine* **2018**, *5*, 122.
- (16) Sureshkumar, K. K.; Ray, T.; Clark, B. A. *Journals Gerontol. Ser. A* **2003**, *58* (4), M378–M381.
- (17) Chaudhri, I.; Moffitt, R.; Taub, E.; Annadi, R. R.; Hoai, M.; Bolotova, O.; Yoo, J.; Dhaliwal, S.; Sahib, H.; Daccueil, F.; Hajagos, J.; Saltz, M.; Saltz, J.; Mallipattu, S. K.; Koraishy, F. M. *Kidney Blood Press. Res.* **2020**, *45* (6), 1018–1032.
- (18) Ouahmi, H.; Courjon, J.; Morand, L.; François, J.; Bruckert, V.; Lombardi, R.; Esnault, V.; Seitz-Polski, B.; Demonchy, E.; Dellamonica, J.; Boyer-Suavet, S. *Front. Physiol.* **2021**, *12*, 12.
- (19) Carr, H. Y.; Purcell, E. M. *Phys. Rev.* **1954**, *94* (3), 630–638.
- (20) Vuckovic, I.; Denic, A.; Charlesworth, M. C.; Suvakov, M.; Bobart, S.; Lieske, J. C.; Fervenza, F. C.; Macura, S. *Anal. Chem.* **2021**, *93* (39), 13177–13186.
- (21) Nagana Gowda, G. A.; Raftery, D. *Anal. Chem.* **2014**, *86* (11), 5433–5440.
- (22) Worley, B.; Sisco, N. J.; Powers, R. J. *Biomol. NMR* **2015**, *63* (1), 53.
- (23) Brown, C. L. J.; Scott, H.; Mulik, C.; Freund, A. S.; Opyr, M. P.; Metz, G. A. S.; Douglas Inglis, G.; Montina, T. *Metabolites* **2022**, *12* (2), 148.
- (24) Carroll, M. F.; Temte, J. L. *Am. Fam. Physician* **2000**, *62* (6), 1333–1340.