

Downloaded via UNIV OF LIVERPOOL on July 26, 2019 at 04:53:55 (UTC).
See <https://pubs.acs.org/sharingguidelines> for options on how to legitimately share published articles.

spectrometry assays,^{12–18} but these have largely been limited to one or a few standard matrices and lack a thorough investigation of interference as a potential confounding variable. This is especially worrisome as it is not commonplace yet in the field to use SIL peptides that aid in defining peptide retention times and help mitigate matrix effects.

An investigation into interference is critical as food ingredients each have a unique proteomic profile that, when combined into a food and modified through processing such as baking, equate to innumerable potential matrices. Consequently, although MRM transitions are meant to be specific to the peptide of interest, other molecular species in the matrix may unpredictably interfere in transition measurement through coelution and satisfaction of Q1 and Q3 transmission window resolution constraints.¹⁹ This challenge is well-known in the field of clinical proteomics^{19,20} and guidelines for MRM assay development dictate that transitions should be thoroughly screened for matrix-derived interferences.²¹ However, this ideal approach is not feasible for analyses of food, where each matrix contains a potentially unique interference profile. Here we describe a computational approach for addressing heterogeneous interference derived from diverse food matrices and apply it to understand patterns of allergen contamination in commercial foods with and without PAL.

■ EXPERIMENTAL SECTION/MATERIALS AND METHODS

Sample Preparation. Food products were selected predominantly from local grocery outlets with preference toward national brands and those with PAL. For each food, a 100 mg sample encompassing any heterogeneity in food content was homogenized with a 5 mm stainless steel bead using the TissueLyser system (Qiagen) in the presence of 1.6 mL of extraction buffer containing 7 M urea, 50 mM ammonium bicarbonate, and 10 mM dithiothreitol (DTT) for cysteine reduction. Samples were then centrifuged at 20,000g to separate the aqueous, lipid, and insoluble fractions. Protein, present in the aqueous fraction, was quantified using the Pierce 660 nm Protein Assay in microtiter plate format and then alkylated through the addition of iodoacetamide to a final concentration of 20 mM and incubated in the dark for 30 min. Samples were then diluted 1:10 with 50 mM ammonium bicarbonate to reduce urea concentrations to less than 1 M for overnight digestion at 37 °C with Trypsin Gold, mass spectrometry grade (Promega) at a ratio of 1 µg trypsin to 50 µg protein. Pierce LC–MS grade acetonitrile and sequencing-grade trifluoroacetic acid were then added to reach final concentrations of 5% and 0.5%, respectively, in order to quench digestion and retain polar peptides during subsequent desalting with C18 MacroSpin columns or 96-well MacroSpin plates (The Nest Group). Desalted samples were eluted with 120 µL twice and dried via speedvac before being reconstituted and transferred to Waters TruView LCMS Certified Clear Glass 12 × 32 mm Screw Neck Total Recovery autosampler vials such that after SIL peptide addition, targeted LC–MS/MS injections contained a total peptide mass of 20 µg.

LC–MS/MS Method Development. For each of the 14 selected allergens (almond, brazil nut, cashew, hazelnut, pistachio, walnut, egg, lupin, milk, mustard, peanut, sesame, soy, and wheat), we selected numerous tryptic peptides belonging to two or more clinically recognized allergenic proteins²² using the Allergen Peptide Browser²³ or via shotgun

proteomics of allergen protein extracts (see the [Supporting Information](#)). These shotgun proteomics data are available through the ProteomeXchange Consortium via the PRIDE²⁴ partner repository and should serve as a resource for the study of allergens and plant proteomics more broadly. Chosen peptides were then synthesized, and Skyline²⁵ was used to select optimal transitions and establish retention times for synthetic unlabeled peptides and stable isotope labeled (SIL) analogs, the latter of which contain ¹⁵N/¹³C-isotopically labeled C-terminal lysine or arginine. We then narrowed the list of candidate peptides with preference for those that lacked cysteine and methionine, ionized well, formed doubly charged precursor ions, belonged to multiple isoforms if an allergenic protein contained such variants, and were distributed along the elution gradient. We also attempted to avoid peptides sharing specificity to other species, but this was not possible in every case (see [Supporting Information](#)). The final method contained 265 transitions belonging to 53 peptides ([Table S1](#)).

For quantification, we used external calibrators with internal standardization by constructing an 8 point calibration curve from 1 to 500 ppm on a food mass basis using a serial dilution of target allergen extracts in a wheat matrix and constant addition of SIL peptides (see the [Supporting Information](#)). Allergen concentrations are therefore presented as ppm of allergen, for example soy flour; if desired, ppm of allergenic protein can be calculated using the protein fractions reported in [Table S2](#). Allergen detection limits in the wheat matrix, as measured by each allergen's most sensitive peptide(s), were 5 ppm for 9 allergens, 10 ppm for 1 allergen, and 25 ppm for 3 allergens ([Table S1](#)). The median R² value of all linear regressions was 0.994. For quantification, the most sensitive peptide “quantifier” was used for each allergen.

MADIC Algorithm. MADIC is a Python package for identifying and correcting interference in targeted mass spectrometry data that may arise from complex and heterogeneous sample matrices. MADIC accepts data exported as a Report from open-source software Skyline and thus is compatible with any common instrument vendor whose data can be loaded into Skyline. MADIC can be used from the command line or interactively and contains functionality for preprocessing, quality control, and interference detection and correction. The software, along with a visual tutorial and detailed instructions for installation and use are available on github: <https://github.com/dcroote/madic>.

MADIC implements four quality control filters. First, a transition passes the transition ratio quality control filter if its ratio is within a user-defined tolerance from its reference transition ratio as defined by synthetic standard injections. The tolerance is an adjustable parameter that can be specified in absolute or relative terms compared to the reference. These data have been evaluated with an absolute tolerance of 15%. Second, a transition passes the retention time quality control filter if its retention time is within a user-defined tolerance from the retention time of the SIL peptide transitions. As this tolerance will differ based on chromatographic method and consistency of elution, this tolerance is also an adjustable parameter. These data were evaluated with a tolerance of 4.2 s. Third, a peptide passes a signal-to-noise quality control filter if its peak intensity within retention time boundaries is greater than a user-designed multiple of the median background intensity outside of the retention time boundaries. These data were evaluated with a signal-to-noise threshold of 3. Fourth, a

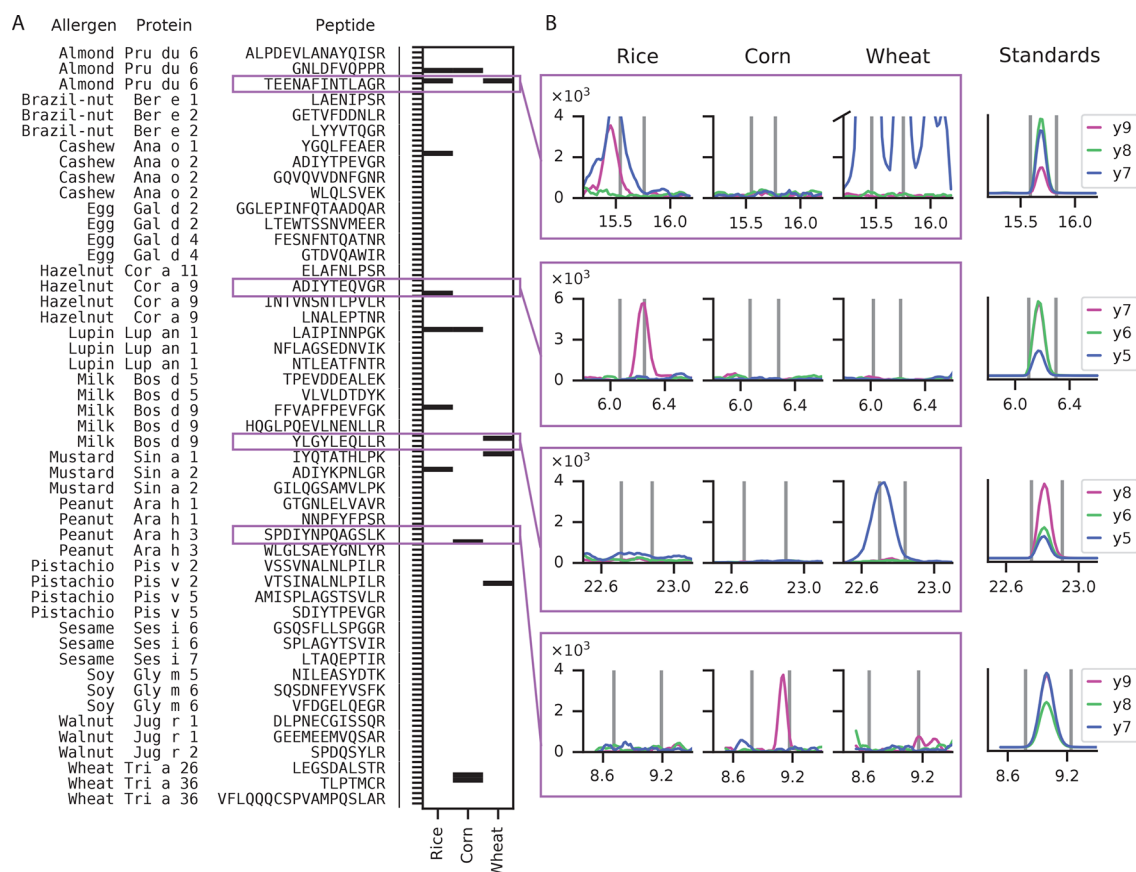


Figure 1. Heterogeneous matrix-specific interference in MRM transitions. (A) Binary heatmap of interference among 147 MRM transitions for common food ingredients rice (sample S8), corn (sample 68), and wheat (sample A10). (B) Example chromatograms illustrating interferences in (A). For comparison, synthetic standards depicting expected transition ratios are included on the right. Peak boundaries defined by SIL peptides are shown as vertical gray lines.

peptide passes the replicate quality control filter if each replicate passes all of the three aforementioned filters.

Once quality control is performed, MADIC identifies transitions with interference if they fail the transition ratio quality control filter while also exceeding a transition ratio threshold defined by a scaling function that takes into account the transition's expected transition ratio. This criterion is necessary to identify interference even within peptides that naturally have ratios skewed toward one high intensity transition. To avoid flagging low intensity noise as interference, a transition's area must also exceed an area threshold. As instruments differ in background intensity and intensity reporting schemes, these cutoffs are configurable parameters. These data were evaluated with a ratio of 10 and minimum area of 2000.

RESULTS AND DISCUSSION

Identifying Matrix-Dependent Transition Interference. Using our scheduled MRM method to quantify food allergen contamination, we encountered heterogeneous and matrix-specific MRM transition interference. For example, as illustrated in Figure 1, common food ingredients rice, corn, and wheat each have unique transition interferences that can inhibit quantification of trace amounts of allergen contamination. Moreover, each instance of interference is unique in peak intensity, peak width, and product ion affected.

While interference varied significantly by food product, we observed consistent interference within foods composed of a

dominant ingredient. As shown in Figure 2A, all 12 breads made with wheat contained wheat-specific interference in the milk *Bos d 9* (α -s1-casein) peptide YLGYLEQLLR, which is often-targeted in LC-MS/MS allergen detection assays.²³ Similarly, we found consistent interference in the peanut *Ara h 3* peptide SPDIYNPQAGSLK in all seven corn flours tested (Figure 2B). Importantly, the challenge of interference is not unique to our method, instrument, or instrument brand. For example, matrix-specific interference has previously affected allergen detection.²⁶ Moreover, as shown in Figure 2C–E, using our method of interference identification, we found consistent interference in data from an interlaboratory plasma protein MRM assay validation study.²⁷ The fact that this study included transition screening for interference prior to assay deployment emphasizes the challenge of simply avoiding interference even under well-controlled conditions with a single sample matrix.

Computational Approach for Addressing MRM Interference. We implemented an algorithm-termed MADIC that identifies interference and enables confident quantification of allergens from targeted LC-MS/MS data acquired using a single instrument method and processed with platform-independent and open source Skyline software.²⁵ Interference detection operates under the assumptions that interference will dominate the transition ratio and greatly exceed the intensity of noise surrounding the SIL-defined peptide retention time window. These aspects are illustrated using the aforementioned milk peptide YLGYLEQLLR in a

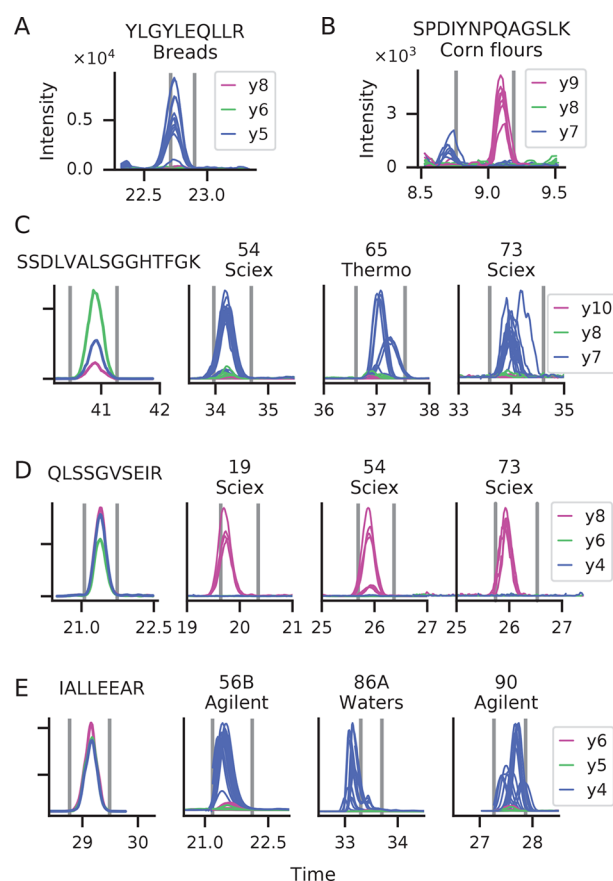


Figure 2. Ubiquity of interference. (A) Overlaid chromatograms illustrating consistent YLGYLEQLLR y5 transition interference in 12 breads. (B) Overlaid chromatograms illustrating consistent SPDIYNPQAGSLK y9 transition interference in 7 corn flours. (C–E) Examples of peptides exhibiting reproducible MRM interference across instruments and sites. Data were generated by a large-scale interlaboratory study²⁷ and interferences were identified using MADIC. QC samples in the first column of each row provide a reference for the expected transition ratios. For each set of chromatograms, identified by site and instrument vendor, between 4 and 25 injections have been overlaid after retention time correction.

serial dilution of milk in a wheat matrix. The consistently high y5 transition area originating from interference skews the transition ratio at lower milk concentrations and is overcome only at high milk concentrations (Figure 3A,C–D). Once interference is identified, the peptide can be excluded from analysis, avoiding a false positive result. Alternatively, the affected transition can be selectively excluded, optionally rescuing the ability to confirm presence of the peptide if the remaining transitions pass the quality control metrics discussed below (Figure 3B,E). Importantly, the matrix-specificity of interference and the selective exclusion of transitions in one matrix but not others enables a broadly applicable scheduled MRM method capable of detecting and quantifying allergens in a variety of foods, each with a unique interference profile.

Due to the complexity of interference in food, quantification of trace amounts of allergen requires additional quality control measures in addition to interference detection. Three established quantitative aspects of an interference-free chromatogram form the basis of the MADIC implementation: relative transition area ratios similar to synthetic standards, coelution of all individual target transitions with respective SIL

transitions, and a distinguished peak relative to the background. Additionally, the requirement that replicate injections satisfy each of these previous criteria serves as a fourth quality control criterion. Chromatograms shown in Figure 3F–H provide examples of these criteria. This work extends previous efforts limited to transition ratio and replicate filtering^{28,29} by also implementing signal-to-noise and transition retention time matching criteria and packaging it in freely available, tested, and documented open-source software.

The computational approach we developed to address heterogeneous matrix interference offers advantages over alternative strategies for improving assay selectivity. For example, in our hands the occasional reduction in interference using differential mobility spectrometry³⁰ did not outweigh the loss in signal intensity due to decreased transition dwell times (Figure S2). Another approach, MRM³, has successfully been applied to a panel of tree nut allergens,³¹ but its adoption is challenged by instrumentation requirements, nonspecific secondary product ions, manual scheduling windows, and long dwell times incompatible with a highly multiplexed method containing SIL peptides. Additionally, while high-resolution mass spectrometers offer an increased ability to discriminate between target analytes and interfering ions of similar mass in complex matrices, these benefits come at the cost of dynamic range, sensitivity, and measurement variation compared to MRM.³² Our approach, on the other hand, can be implemented without additional hardware or modification to existing MRM data acquisition strategies.

Quantifying Allergen Contamination in Commercial Foods. We developed a scheduled MRM method on a triple quadrupole mass spectrometer capable of detecting allergenic proteins belonging to the following 14 allergens: almond, brazil nut, cashew, hazelnut, pistachio, walnut, egg, lupin, milk, mustard, peanut, sesame, soy, and wheat. While we did not generate incurred samples due to the impossibility of reproducing the wide variety of processing observed in commercial foods, we dedicated significant time toward method development. Optimizations aimed at increasing recovery and sensitivity included narrowing the scheduling window duration to 1 min, using Skyline for selection of collision energies and declustering potentials, increasing the chromatographic flow rate to narrow peak width without unduly decreasing points across the peak, modifying the elution gradient length and shape as well as the distribution of peptides along the gradient to maximize transition dwell time (Figure S1B–D), increasing the peptide injection mass (Figure S1E), performing overnight trypsin digestion (Figure S1F), maximizing peptide desalting elution volume (Figure S1G), and optimizing the mass spectrometer cell exit potential (Figure S1H). Quantification of allergens relied on internal standardization using SIL peptides spiked into each sample paired with external calibration based on a serial dilution of allergens in a wheat matrix (see **Materials and Methods** and **Supporting Information**).

We applied this scheduled MRM method with interference detection to 84 common food products in order to survey patterns of allergen contamination (Figure 4). Most foods belonged to the broad categories of grains, dairy, or snacks. Overall, we observed 23 instances of contamination. Soy was the mostly commonly contaminating allergen, followed by tree nuts, wheat, and milk. Walnut and hazelnut contaminated 1 and 2 products with tree nut PAL, respectively, at concentrations of 5 ppm or less and were the only two tree

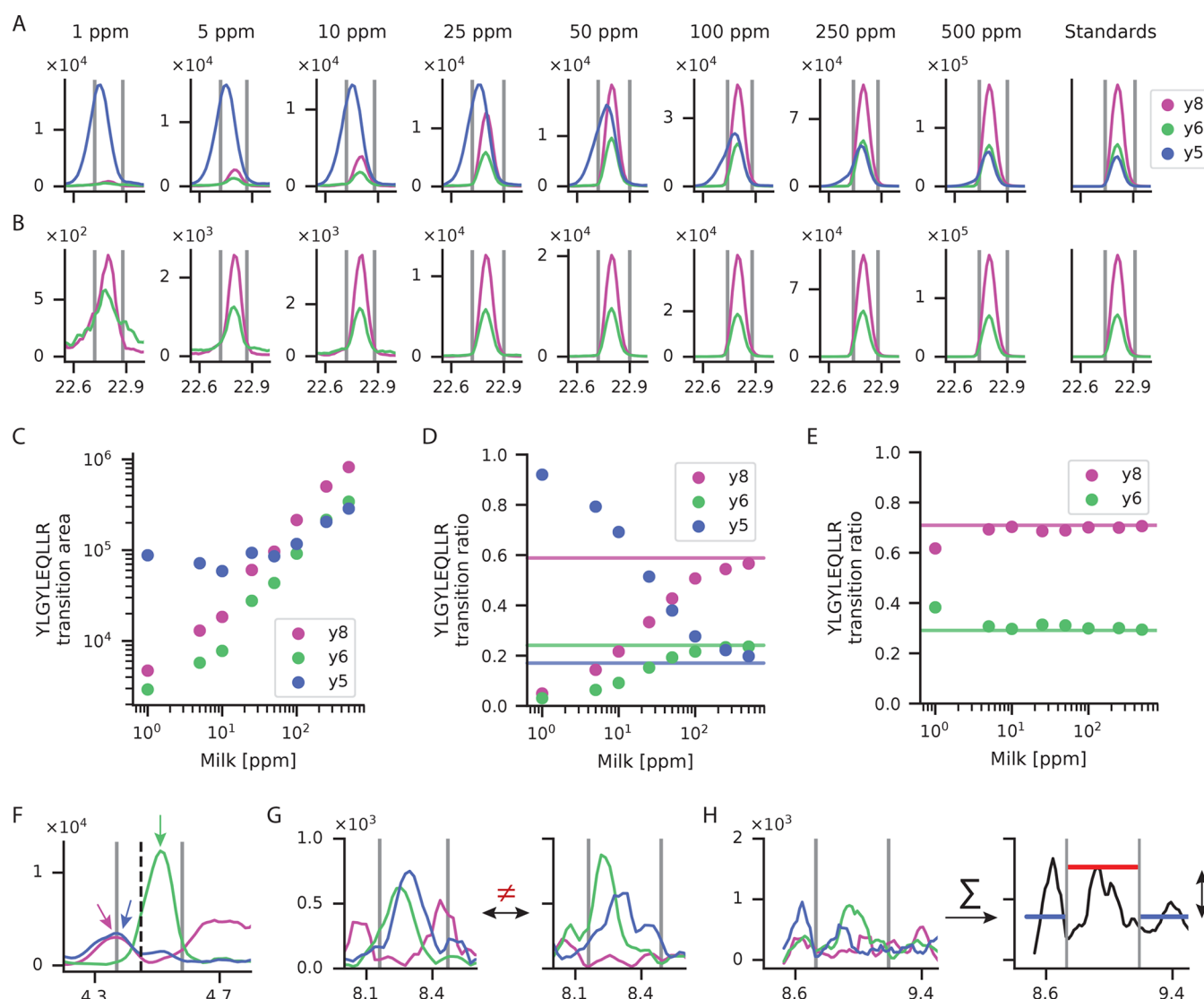


Figure 3. Transition interference identification and quality control criteria enable high confidence peptide detection. (A) Serial milk dilution in a wheat matrix highlights interference in the y5 transition of milk peptide YLGYLEQLLR. (B) Same as (A) but excluding the y5 transition. (C) Transition areas for chromatograms in (A). (D, E) Transition ratios for chromatograms in (A) and (B), respectively. Expected ratios derived from synthetic standard injections are represented by horizontal lines. (F–H) MADIC avoids false positives with quality control filters. (F) Transitions do not coelute and individual transition peak retention times (arrows) do not match the SIL peptide peak retention time (vertical dotted line). (G) Disagreement in transition retention times and transition ratios between duplicate injections. (H) Insufficient signal-to-noise determined by the ratio of peak apex within the SIL peptide-defined peak boundaries (red line) to median background outside of the peak boundaries (blue line).

nuts of six found to contaminate any of the products tested, despite similar or better detection limits for the other tree nuts. In products without PAL, hazelnut was found in two cookies (samples 16, 20) and a basil pesto sauce (sample 17) at less than 5 ppm. Egg was found in one penne pasta product (sample 55) with PAL at 3 ppm. Lupin, a flowering plant in the legume family commonly ground into a gluten-free alternative to wheat flour, was not found in any sample, which is consistent with its infrequent use in the U.S. Mustard similarly did not contaminate any product. Within the products sampled, peanut was the second-most commonly included allergen in PAL, and, despite being detected in all products that listed peanut as an ingredient, was not found in any other products. While this result may relate to our detection limit or use of a roasted peanut flour as a calibrator (Table S1), low peanut contamination rates between 0 and 4.4% have been previously reported in a number of studies across a variety of

food products and geographic locations,^{33–36} although rates reported for products with PAL may be higher.^{33,37}

PAL was featured on 58 products and individual labels listed anywhere from one to six allergens. The phrasing of PAL had numerous slight variations, but could be grouped into the following three general categories: “may contain,” “same facility,” or “same equipment” (Figure 4). Although the number of products with contamination was small, we did not find that phrasing had any strong relationship with risk of contamination. Studies evaluating the contamination of peanut³⁷ and/or hazelnut,³⁸ with larger numbers of products similarly concluded that PAL phrasing has little bearing on actual risk.

Common Patterns of Allergen Contamination. Milk was found at 32 ppm (sample 49) and 20 ppm (sample 33) in two snacks containing chocolate chips. The latter, a chocolate chip cookie, elicited an adverse reaction in a milk-allergic

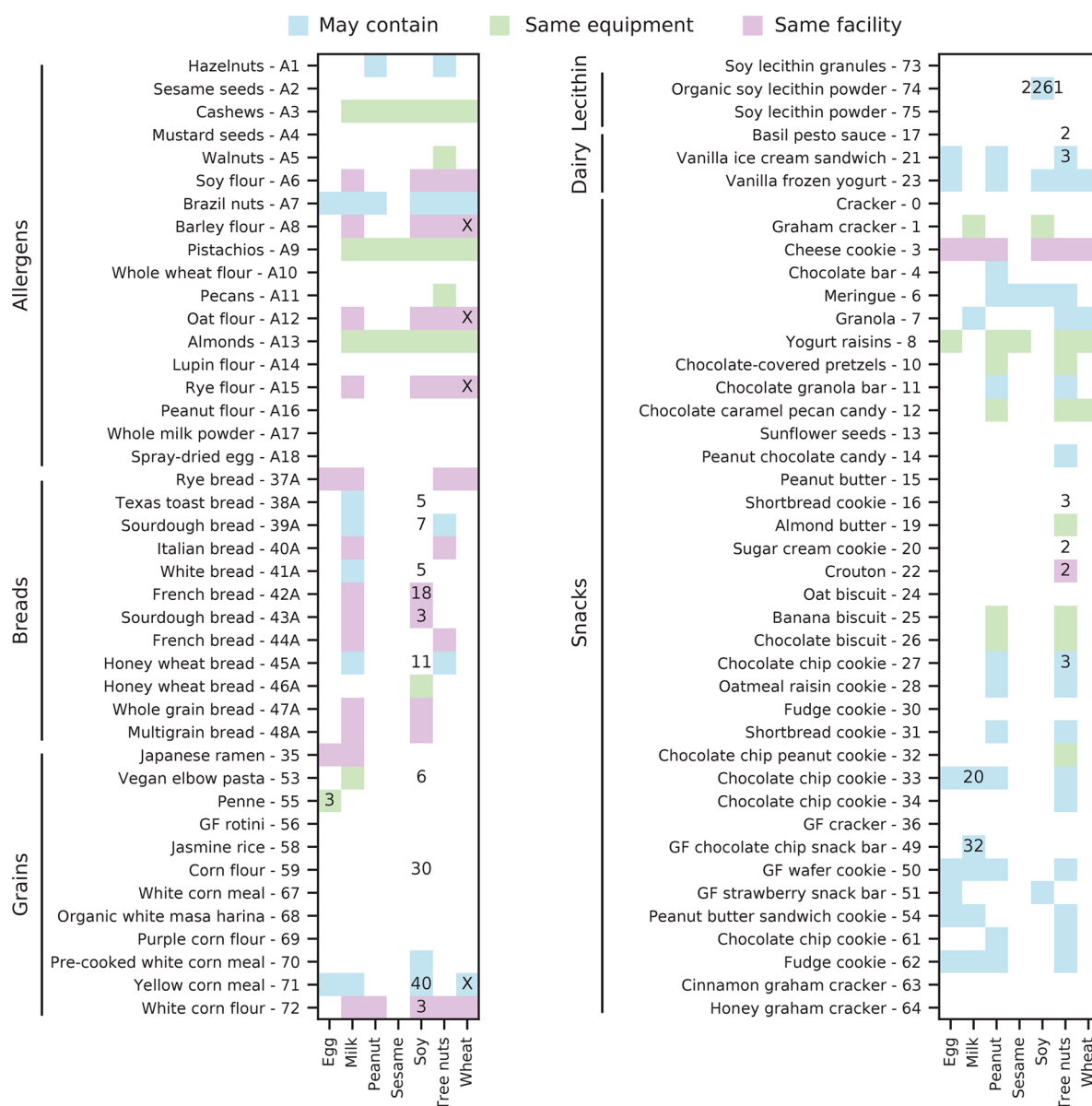


Figure 4. Allergen contamination in 84 unique commercial foods with and without PAL. Samples have been grouped by general category and are labeled with a description and sample ID. Values represent ppm food allergen, for example soy flour, except wheat, for which contamination is illustrated with an “X” due to the use of wheat as the calibration matrix. GF = gluten free. Almond, brazil nut, cashew, hazelnut, pistachio, and walnut have been aggregated as “Tree nuts”. Lupin and mustard have been omitted for visualization purposes as neither contaminated any products.

individual (private correspondence) and a second, independent package of the same product had nearly identical levels of contamination (23 ppm milk, sample 65). Interestingly, we found no evidence of milk contamination in a fudge cookie (sample 62) from the same manufacturer, which highlights the challenge that allergic consumers face in finding a trustworthy brand.

Another pattern we observed was the contamination of grains into other grain-based products. For example, we detected wheat in a corn flour (sample 71), barley flour (sample A8), oat flour (sample A12), and rye flour (sample A15). Soy contaminated grain-based products even more frequently. Soy was found in one gluten-free pasta (sample 53) at 6 ppm as well as in numerous corn flours and breads, some with PAL and some without. Specifically, we found 20 ppm or less of soy in the crumb (inner portion) of two breads with PAL and four breads without PAL. In only two of these

contaminated breads we were also able to detect soy in the bread crust (Figure 5A). The fact that each pair of crumb and crust originated from the same dough enables this discrepancy to be attributed to temperature- and moisture-dependent effects of baking, which lowered extractable protein (Figure S3A) and reduced peptide intensity in a sample-dependent manner (Figure S3B). Soy was also found to contaminate corn flours. Among the seven corn flours tested, three were found to contain soy, and of these, two contained PAL (Figure 4). Two of the corn flours contained 30 ppm soy or more and a second package of sample 59 contained similar levels of soy contamination (45 ppm, sample 66).

We extended our analysis of soy presence to a subset of products that distinctly were labeled as containing soy but where soybean oil and/or the emulsifier soy lecithin were the only soy-contributing ingredients. We found variable amounts of soy in 23 of 30 unique products belonging to this category

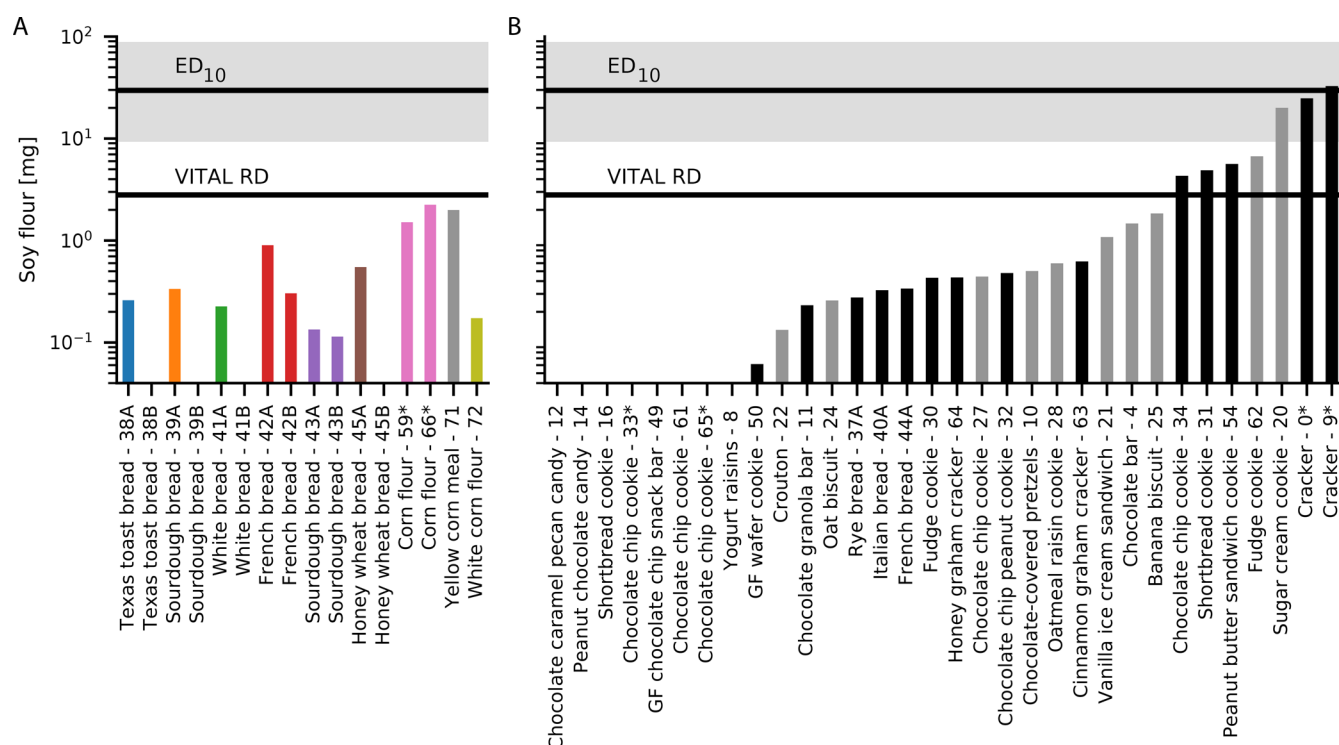


Figure 5. Soy contamination levels in commercial foods relative to the VITAL reference dose (RD) and ED₁₀ threshold. The 95% CI of the ED₁₀ has been shaded. Conversion from ppm soy for each food was made assuming a 50 g serving size. Pairs of samples marked with an asterisk are separate packages of the same product. (A) Soy contamination in breads and corn flours. Bread A/B designation refers to the crumb (inner) or crust (outer) portion of the bread, respectively. (B) Quantity of soy in 32 products that are labeled as containing soy, but where soy lecithin and/or soybean oil are the only soy-contributing ingredients. Gray bars indicate the presence of only soy lecithin, while black bars indicate the presence of soy lecithin and soybean oil.

of food (Figure 5B), which was surprising as allergenic soy protein is unlikely to originate from either soybean oil or soy lecithin. In fact, soybean oil, which is a highly refined oil, is exempt from allergen labeling requirements due to its negligible protein content.^{39–41} Soy lecithin, on the other hand, has been shown to sometimes contain small amounts of allergenic protein^{41–43} and, in agreement, we found soy protein only in 1 of 3 soy lecithin products tested (Figure 4). Taking into account the small contribution of soy lecithin by mass to foods, the amount of soy protein contributed by soy lecithin is generally considered insufficient to elicit a reaction in allergic individuals^{44,45} and some clinicians even suggest that allergic individuals ignore soy lecithin on labels.⁴⁶

Clinical Reactivity and Allergen Origins. To assess whether the measured soy concentrations in breads, corn flours, and products with soy lecithin and/or soybean oil could be clinically reactive, we first calculated the amount of allergen that would be consumed under the assumption of a 50 g food serving. Next, we compared these amounts with predicted population eliciting dose (ED) values compiled from statistical modeling of oral food challenge data. For example, ED₀₅ represents the dose predicted to elicit a response in 5% of the population. The Voluntary Incidental Trace Allergen Labeling (VITAL) initiative has recommended a reference dose (RD) for most major allergens and for soy is 1 mg based on the lower 95% confidence interval of the ED₀₅.^{47,48} On the basis of this threshold, the amounts of soy ingested from contaminated breads and corn flours would be unreactive (Figure 5A). For products with soy lecithin and/or soybean oil, on the other hand, seven were above the RD, one of which

exceeded even the estimated ED₁₀ (Figure 5B).⁴⁹ While these results should be confirmed with larger numbers of products, the practice of ignoring PAL when products contain only soy lecithin and/or soybean oil is potentially dangerous for individuals with a severe soy allergy.

For all products, we have only tested the final output of an increasingly complex globalized process involving myriad combinations of sourcing, transport, storage, manufacturing, and packaging steps. Consequently, while it is unknown how or when allergen introduction may have occurred, there are potential explanations for the patterns of contamination we observed. For example, milk often contaminates chocolate,^{50–52} which may be unsurprising given the difficulty of cleaning shared equipment used to produce both dark and milk chocolate. Cross-contamination of grains, on the other hand, can occur through the standard practice of agricultural comingling, which features shared use of harvesting, transport, and storage equipment. Historical precedent has exempted this type of contamination from labeling requirements, but contamination is common. One study reported soy contamination in 62.8% of wheat flours,⁴ a striking finding that, when considered with the common use of wheat in bakery products and snacks, may explain our soy lecithin findings as well as findings by the United States Food and Drug Administration in which nearly 25% and 11% of products within these categories were contaminated with soy, respectively.⁵³ Issues with comingling are not restricted to soy; one study found 88% of Canadian oat samples contained above 20 ppm of wheat gluten,⁵⁴ while others reported dramatically variable contamination frequencies depending on geographical location, food

category, and presence of specific labeling such as “gluten-free”.⁵⁵

While the prevalence of contamination may be concerning, the quantity of allergen must be considered with respect to predicted eliciting doses. For soy, our results and those of the aforementioned studies suggest that few products are predicted to elicit reactions. On the other hand, quantities of wheat gluten exceeding the “gluten-free” 20 ppm threshold have been found frequently in grains but less commonly among products explicitly labeled gluten-free.^{55–57} The patterns of contamination discussed here are likely a subset of allergen and food category combinations that represent high-risk scenarios. For these scenarios there is an outstanding opportunity for more rigorous testing along the supply chain and for labeling that is more informative than the unregulated, ubiquitous, and unhelpful PAL currently featured on commercial food products.

CONCLUSIONS

Through a survey of commercial foods, we have highlighted how a single multiplexed scheduled MRM method can overcome complex matrix interference and capture trends in allergen contamination such as milk in chocolate-containing products, soy in breads and flours, wheat in grains, and variable soy content in products with soy lecithin. Although our sample size was small relative to the product market, we found an overall low incidence of contamination and generally low amounts of allergen when contamination was found. To improve communication of actual risk to the allergic consumer, better characterizing scenarios that present high risks of contamination using a modality such as LC–MS/MS should be part of a comprehensive approach involving PAL standardization,^{58,59} incentives to avoid contamination,⁶⁰ and the adoption of allergen reference doses by regulatory agencies that would reduce manufacturer uncertainty.^{10,47,48}

Data Availability. Shotgun proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE²⁴ partner repository with the data set identifier PXD007688 (<http://dx.doi.org/10.6019/PXD007688>). Code and examples for matrix-dependent interference correction are available on Github: <https://github.com/dcroote/madic>.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.9b01388.

Experimental methods, supporting figures, and references (PDF)

(Table S1) Peptide targets for allergen detection with associated detection limits and calibration curve linear regression parameters; (Table S2) Information for each unique sample, including relevant ingredients, precautionary labeling statements, and precautionary phrasing type; NIST = National Institute of Standards and Technology, GF = gluten-free; “A” sample name prefix indicates the allergen was used as a calibrator; for bread samples 37–48, both the crumb (A suffix) and crust (B suffix) were analyzed; not listed: 9, 65, and 66, which are different packages of products 0, 33, and 59, respectively; (Table S3) MRM transitions with associated instrument parameters; and (Table S4)

Liquid chromatography gradient using 0.1% formic acid in water (mobile phase A) and 0.1% formic acid in acetonitrile (mobile phase B); two linear segments were used for peptide elution: the first from 7% B to 21.5% B contained the majority of peptides, while the second from 21.5% B to 30% B contained fewer peptides and was consequently designed with a steeper slope; following peptide elution, the column was washed with 95% B and re-equilibrated (XLSX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: quake@stanford.edu.

ORCID

Derek Croote: 0000-0003-4907-1865

Ido Braslavsky: 0000-0001-8985-8211

Author Contributions

D.C. and I.B. planned and executed experiments and analyzed the data. D.C. developed the software. S.R.Q. directed the research. All authors wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to acknowledge the staff at the Vincent Coates Foundation Mass Spectrometry Laboratory at Stanford University (SUMS): Karolina Krasinska, Allis Chien, and Theresa McLaughlin for helpful discussions related to quantitative method development and instrument performance, as well as Christopher Adams and Ryan Leib for helpful discussions related to shotgun proteomics and data analysis. This research was supported by the Simons Foundation (SFLIFE #288992 to S.R.Q.), a SUMS seed grant, and the Chan Zuckerberg Biohub. D.C. is supported by an NSF Graduate Research Fellowship and the Kou-I Yeh Stanford Graduate Fellowship. I.B. acknowledges support from Stanford University and from The Hebrew University of Jerusalem.

REFERENCES

- (1) Sicherer, S. H.; Sampson, H. A. *J. Allergy Clin. Immunol.* **2014**, *133*, 291–307 quiz 308.
- (2) Allen, K. J.; Turner, P. J.; Pawankar, R.; Taylor, S.; Sicherer, S.; Lack, G.; Rosario, N.; Ebisawa, M.; Wong, G.; Mills, E. N. C.; Beyer, K.; Fiocchi, A.; Sampson, H. A. *World Allergy Organ. J.* **2014**, *7*, 10.
- (3) Gendel, S. M. *Regul. Toxicol. Pharmacol.* **2012**, *63*, 279–285.
- (4) Remington, B. C.; Taylor, S. L.; Marx, D. B.; Petersen, B. J.; Baumert, J. L. *Food Chem. Toxicol.* **2013**, *62*, 485–491.
- (5) Gendel, S. M.; Khan, N.; Yajnik, M. J. *Food Prot.* **2013**, *76*, 302–306.
- (6) Jackson, L. S.; Al-Taher, F. M.; Moorman, M.; DeVries, J. W.; Tippet, R.; Swanson, K. M. J.; Fu, T.-J.; Salter, R.; Dunaif, G.; Estes, S.; Albillos, S.; Gendel, S. M. *J. Food Prot.* **2008**, *71*, 445–458.
- (7) Zurzolo, G. A.; Koplin, J. J.; Mathai, M. L.; Tang, M. K. L.; Allen, K. J. *Med. J. Aust.* **2013**, *198*, 621–623.
- (8) Ben-Shoshan, M.; Sheth, S.; Harrington, D.; Soller, L.; Fragapane, J.; Joseph, L.; St Pierre, Y.; La Vieille, S.; Elliott, S.; Wasserman, S.; Alizadehfar, R.; Harada, L.; Allen, M.; Allen, M. H.; Clarke, A. E. *J. Allergy Clin. Immunol.* **2012**, *129*, 1401–1404.
- (9) Noimark, L.; Gardner, J.; Warner, J. O. *Pediatr Allergy Immunol* **2009**, *20*, 500–504.
- (10) Allen, K. J.; Taylor, S. L. *J. Allergy Clin Immunol Pract* **2018**, *6*, 400–407.
- (11) Marchisotto, M. J.; Harada, L.; Blumenstock, J. A.; Bilaver, L. A.; Wasserman, S.; Sicherer, S.; Boloh, Y.; Regent, L.; Said, M.;

- Schnadt, S.; Allen, K. J.; Muraro, A.; Taylor, S. L.; Gupta, R. S. *Allergy* **2016**, *71*, 1081–1085.
- (12) Montowska, M.; Fornal, E. *LWT - Food Science and Technology* **2018**, *87*, 310–317.
- (13) Gu, S.; Chen, N.; Zhou, Y.; Zhao, C.; Zhan, L.; Qu, L.; Cao, C.; Han, L.; Deng, X.; Ding, T.; Song, C.; Ding, Y. *Food Control* **2018**, *84*, 89–96.
- (14) Pilolli, R.; De Angelis, E.; Monaci, L. *Anal. Bioanal. Chem.* **2018**, *410*, S653–S662.
- (15) Parker, C. H.; Khuda, S. E.; Pereira, M.; Ross, M. M.; Fu, T.-J.; Fan, X.; Wu, Y.; Williams, K. M.; DeVries, J.; Pulvermacher, B.; Bedford, B.; Zhang, X.; Jackson, L. S. *J. Agric. Food Chem.* **2015**, *63*, 10669–10680.
- (16) Planque, M.; Arnould, T.; Dieu, M.; Delahaut, P.; Renard, P.; Gillard, N. *J. Chromatogr. A* **2016**, *1464*, 115–123.
- (17) Planque, M.; Arnould, T.; Dieu, M.; Delahaut, P.; Renard, P.; Gillard, N. *J. Chromatogr. A* **2017**, *1530*, 138–151.
- (18) Goma, A.; Boye, J. *Food Chem.* **2015**, *175*, S85–S92.
- (19) Gillette, M. A.; Carr, S. A. *Nat. Methods* **2013**, *10*, 28–34.
- (20) Gallien, S.; Duriez, E.; Demeure, K.; Domon, B. *J. Proteomics* **2013**, *81*, 148–158.
- (21) Carr, S. A.; Abbatiello, S. E.; Ackermann, B. L.; Borchers, C.; Domon, B.; Deutsch, E. W.; Grant, R. P.; Hoofnagle, A. N.; Hüttenhain, R.; Koomen, J. M.; Liebler, D. C.; Liu, T.; MacLean, B.; Mani, D. R.; Mansfield, E.; Neubert, H.; Paulovich, A. G.; Reiter, L.; Vitek, O.; Aebersold, R.; Anderson, L.; Bethem, R.; Blonder, J.; Boja, E.; Botelho, J.; Boyne, M.; Bradshaw, R. A.; Burlingame, A. L.; Chan, D.; Keshishian, H.; Kuhn, E.; Kinsinger, C.; Lee, J. S. H.; Lee, S.-W.; Moritz, R.; Oses-Prieto, J.; Rifai, N.; Ritchie, J.; Rodriguez, H.; Srinivas, P. R.; Townsend, R. R.; Van Eyk, J.; Whiteley, G.; Wiita, A.; Weintraub, S. *Mol. Cell. Proteomics* **2014**, *13*, 907–917.
- (22) Radauer, C.; Nandy, A.; Ferreira, F.; Goodman, R. E.; Larsen, J. N.; Lidholm, J.; Pomés, A.; Raulf-Heimsoth, M.; Rozynek, P.; Thomas, W. R.; Breiteneder, H. *Allergy* **2014**, *69*, 413–419.
- (23) Croote, D.; Quake, S. R. *Npj Syst. Biol. Appl.* **2016**, *2*, 16022.
- (24) Vizcaino, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; Xu, Q.-W.; Wang, R.; Hermjakob, H. *Nucleic Acids Res.* **2016**, *44*, D447–S6.
- (25) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. *Bioinformatics* **2010**, *26*, 966–968.
- (26) Planque, M.; Arnould, T.; Delahaut, P.; Renard, P.; Dieu, M.; Gillard, N. *Food Chem.* **2019**, *274*, 35–45.
- (27) Abbatiello, S. E.; Schilling, B.; Mani, D. R.; Zimmerman, L. J.; Hall, S. C.; MacLean, B.; Albertolle, M.; Allen, S.; Burgess, M.; Cusack, M. P.; Gosh, M.; Hedrick, V.; Held, J. M.; Inerowicz, H. D.; Jackson, A.; Keshishian, H.; Kinsinger, C. R.; Lyssand, J.; Makowski, L.; Mesri, M.; Rodriguez, H.; Rudnick, P.; Sadowski, P.; Sedransk, N.; Shaddox, K.; Skates, S. J.; Kuhn, E.; Smith, D.; Whiteaker, J. R.; Whitwell, C.; Zhang, S.; Borchers, C. H.; Fisher, S. J.; Gibson, B. W.; Liebler, D. C.; MacCoss, M. J.; Neubert, T. A.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A. *Mol. Cell. Proteomics* **2015**, *14*, 2357–2374.
- (28) Bao, Y.; Waldemarson, S.; Zhang, G.; Wahlander, A.; Ueberheide, B.; Myung, S.; Reed, B.; Molloy, K.; Padovan, J. C.; Eriksson, J.; Neubert, T. A.; Chait, B. T.; Fenyö, D. *Methods* **2013**, *61*, 299–303.
- (29) Abbatiello, S. E.; Mani, D. R.; Keshishian, H.; Carr, S. A. *Clin. Chem.* **2010**, *56*, 291–305.
- (30) Campbell, J. L.; Le Blanc, J. C. Y.; Kibbey, R. G. *Bioanalysis* **2015**, *7*, 853–856.
- (31) Korte, R.; Brockmeyer, J. *Anal. Bioanal. Chem.* **2016**, *408*, 7845–7855.
- (32) Domon, B.; Aebersold, R. *Nat. Biotechnol.* **2010**, *28*, 710–721.
- (33) Ford, L. S.; Taylor, S. L.; Pacenza, R.; Niemann, L. M.; Lambrecht, D. M.; Sicherer, S. H. *J. Allergy Clin. Immunol.* **2010**, *126*, 384–385.
- (34) Zagon, J.; Dittmer, J.; Elegbede, C. F.; Papadopoulos, A.; Braeuning, A.; Crépet, A.; Lampen, A. *J. Food Compos. Anal.* **2015**, *44*, 196–204.
- (35) Surojanametukul, V.; Khaiprapai, P.; Jithan, P.; Varayanond, W.; Shoji, M.; Ito, T.; Tamura, H. *Food Control* **2012**, *23*, 1–6.
- (36) Baumgartner, S.; Fürtler-Leitzenberger, I.; Drs, E.; Molinelli, A.; Krska, R.; Immer, U.; Schmitt, K.; Bremer, M.; Haasnoot, W.; Danks, C.; Romkies, V.; Reece, P.; Wilson, P.; Kiening, M.; Weller, M.; Niessner, R.; Corsini, E.; Mendonça, S. *Food Contaminants: Mycotoxins and Food Allergens*; Siantar, D. P., Trucksess, M. W., Scott, P. M., Herman, E. M., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2008; Vol. 1001, pp 370–381.
- (37) Hefle, S. L.; Furlong, T. J.; Niemann, L.; Lemon-Mule, H.; Sicherer, S.; Taylor, S. L. *J. Allergy Clin. Immunol.* **2007**, *120*, 171–176.
- (38) Pele, M.; Brohée, M.; Anklam, E.; Van Hengel, A. J. *Food Addit. Contam.* **2007**, *24*, 1334–1344.
- (39) Bush, R. K.; Taylor, S. L.; Nordlee, J. A.; Busse, W. W. *J. Allergy Clin. Immunol.* **1985**, *76*, 242–245.
- (40) Awazu, H.; Kawai, B.; Matsui, K.; Komiyama, Y. *Clin. Exp. Allergy* **1998**, *28*, 1559–1564.
- (41) Paschke, A.; Zunker, K.; Wigotzki, M.; Steinhart, H. *J. Chromatogr., Biomed. Appl.* **2001**, *756*, 249–254.
- (42) Porras, O.; Carlsson, B.; Fällström, S. P.; Hanson, L. A. *Int. Arch. Allergy Immunol.* **2004**, *78*, 30–32.
- (43) Müller, U.; Weber, W.; Hoffmann, A.; Franke, S.; Lange, R.; Vieths, S. *Zeitschrift für Lebensmitteluntersuchung und -forschung A* **1998**, *207*, 341–351.
- (44) Taylor, S. L.; Remington, B. C.; Panda, R.; Goodman, R. E.; Baumert, J. L. In *Handbook of food allergen detection and control*; Elsevier, 2015; pp 341–366.
- (45) Committee on Food Allergies: Global Burden, Causes, Treatment, Prevention, and Public Policy. *Finding a path to safety in food allergy: assessment of the global burden, causes, prevention, management, and public policy*; Oria, M. P.; Stallings, V. A., Eds.; National Academies Press (US): Washington, DC, 2016.
- (46) Taylor, S. L.; Baumert, J. L. Allergenicity of Soybean Lecithin: Expert Opinion Statement, *Food Allergy Research & Resource Program*; University of Nebraska https://farrp.unl.edu/documents/OpinionsSummaries/2017_0928_Soy%20Lecithin.pdf (accessed Feb 17, 2019).
- (47) Allen, K. J.; Remington, B. C.; Baumert, J. L.; Crevel, R. W. R.; Houben, G. F.; Brooke-Taylor, S.; Kruizinga, A. G.; Taylor, S. L. *J. Allergy Clin. Immunol.* **2014**, *133*, 156–164.
- (48) Taylor, S. L.; Baumert, J. L.; Kruizinga, A. G.; Remington, B. C.; Crevel, R. W. R.; Brooke-Taylor, S.; Allen, K. J.; Allergen Bureau of Australia & New Zealand; Houben, G. *Food Chem. Toxicol.* **2014**, *63*, 9–17.
- (49) Zhu, J.; Pouillot, R.; Kwagyr-Afful, E. K.; Luccioli, S.; Gendel, S. M. *Food Chem. Toxicol.* **2015**, *80*, 92–100.
- (50) Crotty, M. P.; Taylor, S. L. *J. Allergy Clin. Immunol.* **2010**, *125*, 935–937.
- (51) U.S. Food and Drug Administration. A Survey of Milk in Dark Chocolate Products. <https://www.fda.gov/Food/IngredientsPackagingLabeling/FoodAllergens/ucm446646.htm> (accessed Sep 28, 2017).
- (52) Hefle, S. L.; Lambrecht, D. M. *J. Food Prot.* **2004**, *67*, 1933–1938.
- (53) Khuda, S. E.; Sharma, G. M.; Gaines, D.; Do, A. B.; Pereira, M.; Chang, M.; Ferguson, M.; Williams, K. M. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* **2016**, *33*, 1274–1282.
- (54) Koerner, T. B.; Cléroux, C.; Poirier, C.; Cantin, I.; Alimkulov, A.; Elamparo, H. *Food Addit. Contam., Part A* **2011**, *28*, 705–710.
- (55) Do, A. B.; Khuda, S. E.; Sharma, G. M. *J. AOAC Int.* **2018**, *101*, 23–35.
- (56) Gélinas, P.; McKinnon, C. M.; Mena, M. C.; Méndez, E. *Int. J. Food Sci. Technol.* **2008**, *43*, 1245–1252.

- (57) Farage, P.; de Medeiros Nóbrega, Y. K.; Pratesi, R.; Gandolfi, L.; Assunção, P.; Zandonadi, R. P. *Public Health Nutr* **2017**, *20*, 413–416.
- (58) Hattersley, S.; Ward, R.; Baka, A.; Crevel, R. W. R. *Food Chem. Toxicol.* **2014**, *67*, 255–261.
- (59) Turner, P. J.; Kemp, A. S.; Campbell, D. E. *BMJ*. **2011**, *343*, No. d6180.
- (60) Gupta, R. S.; Taylor, S. L.; Baumert, J. L.; Kao, L. M.; Schuster, E.; Smith, B. M. *J. Food Prot.* **2017**, *80*, 1719–1725.

Supporting Information for

Addressing Complex Matrix Interference Improves Multiplex Food Allergen Detection by Targeted LC-MS/MS

Derek Croote¹, Ido Braslavsky^{1,2}, Stephen R. Quake^{1,3,4*}

¹Department of Bioengineering, Stanford University, Stanford, California 94305, United States

²Robert H. Smith Faculty of Agriculture, Food, and Environment, The Hebrew University of Jerusalem, Rehovot 7610001, Israel

³Department of Applied Physics, Stanford University, Stanford, California 94305, United States

⁴Chan Zuckerberg Biohub, San Francisco, California 94158, United States

*Corresponding author. E-mail: quake@stanford.edu

Table of contents

Experimental methods, references	2-4
Supplementary figures	
Figure S1	5
Figure S2	6
Figure S3	6
Figure S4	7
Supplementary tables	See separate excel file for all four tables
Table S1	Tab 1
Table S2	Tab 2
Table S3	Tab 3
Table S4	Tab 4

Experimental Methods

Shotgun proteomics

The following pure allergens were prepared as described in the main Materials and Methods section and subjected to shotgun proteomics: raw almonds, raw brazil nuts, raw cashews, celery seeds, National Institute of Standards and Technology (NIST) RM 8445 (spray-dried whole egg for allergen detection), raw hazelnuts, lupin flour, NIST SRM 1549a (whole milk powder), mustard seeds, peanut flour (partially defatted 12% fat light roast, kindly provided by Golden Peanut and Tree Nuts), raw pecans, raw pistachios, sesame seeds, soy flour, raw walnuts, whole wheat flour, barley flour, buckwheat flour, wheat gluten flour, oat flour, and rye flour. The inclusion of barley, oat, buckwheat, and rye was based on our initial desire to build a method that targeted gluten peptides in addition to peptides derived from allergenic proteins, but we found few sufficiently sensitive tryptic peptides, likely due to the absence of alcohol in our extraction solvent that preferentially solubilizes glutes.¹⁻³ Shotgun proteomics data were acquired on an Orbitrap Elite using data dependent acquisition. The top 15 most intense multiply charged precursor ions were initially scanned in the orbitrap at a resolution of 60,000 or 120,000 with a target ion population of 750,000 and a maximum injection time of 120 ms. Precursor ions above a signal threshold of 500 were fragmented with a normalized collision energy of 35% and subsequently dynamically excluded for a duration of 60s. The automatic gain control target in the ion trap was 10^4 with a max inject time 25 ms. Chromatographic separation was achieved using either a ThermoFisher EasyNano LC1000 or a Waters NanoAcquity with 20 cm columns pulled and packed in house with Dr. Maisch ReproSil Pur 2.4 μ m C18 matrix. The gradient was linear from 4% to 40% mobile phase B over either 80 min or 120 min at a flow rate of either 300 nL/min (EasyNano LC1000) or 450 nL/min (Waters nanoAcquity). Peptide identification was performed using the Byonic (Protein Metrics) bottom-up search engine run with a 1% protein FDR cutoff, maximum of two missed cleavages, and carbamidomethyl cysteine (from iodoacetamide alkylation) as the only fixed modification. The fasta files used as database search input were composed of common contaminants as well as all available proteins of the given species from the NCBI protein database and totaled in size from 86 to tens of thousands of proteins. The small number of available proteins for some species would normally pose a problem for discovery experiments aiming to identify large number of proteins due to a high false discovery rate; however, our use case of screening specific allergenic proteins for desirable peptide targets is less affected. For large-scale identification, the addition of proteins belonging to phylogenetically similar organisms is one potential solution.

Targeted LC-MS/MS

Data were generated on an AB Sciex QTRAP 6500+ in positive ion mode operated with the following optimized source parameters: CUR: 25, GS1: 50, GS2: 55, IS: 5500, EP: 10, CXP: 20, and temperature: 400°C. Q1 and Q3 resolutions were set to unit, MRM target cycle time was set to 1.2 s and the MRM detection window was set to 60 s (encompassing 30 s before and after the expected peptide retention time). Values for transition declustering potential and collision energy were provided by Skyline (Table S3). Reverse phase liquid chromatographic separation of 5 μ L sample injections was accomplished with a Waters Acquity UPLC and Acquity UPLC Peptide CSH C18 column (130 Å pore size, 1.7 μ m particles, 2.1 mm ID, 150 mm length) operated at 35°C and a flow rate of 0.250 mL/min. An Acquity UPLC Peptide CSH C18 VanGuard pre-column (130 Å pore size, 1.7 μ m particles, 2.1 mm ID, 5 mm length) was used to extend the lifespan of the column and was exchanged when column pressures began to rise. Water with 0.1% formic acid and acetonitrile with 0.1% formic acid were used as mobile phases for the 30 min gradient consisting of two linear segments spanning 7-30% acetonitrile as well as wash and re-equilibration steps (Table S4). Samples were injected in duplicate with at least two water blanks in between samples.

Method development

Although we attempted to exclude peptides sharing specificity to other food species based on aggregated sequence repository data,⁴ this informatic approach cannot guarantee specificity due to the inherent incompleteness of such databases and the homology of some allergens across multiple species. For example, the selected almond peptides are nonspecific with regard to peach, however the latter is unlikely to be a contaminating ingredient in most processed foods and any positive results can be assessed within this context of nonspecificity. Furthermore, after completing method development we found two nonspecific peptides: sesame *Ses i 6* peptide SPLAGYTSVIR was found in pecan, while pistachio *Pis v 2* peptide VTSINALNLPILR was found in cashew (Figure S4). Unsurprisingly, both pecan and cashew are poorly represented in protein databases such as UniProt, indicating that allergen detection method specificity validation using a variety of foods and allergens will likely remain a necessity for the near future. Note that it was later observed that the latter nonspecificity had been listed in a supplementary table of a tree nut investigation⁵, evidence that the field could benefit from the practice of mass spectrometry data submission to public repositories. Table S1 contains the final set of peptides included in the method.

Pecan protein *Car i 1* peptides QCCQQLSQMEEQCQCEGLR and GEEMEEMVQCASDLPK were included in the scheduled MRM method, but ultimately excluded from analysis due to the degradation of the SIL peptide internal standard and low sensitivity. These peptides were originally chosen because they were two of the few peptides that are specific to pecan *Car i 1*;⁴ alternatively in the future it may be worthwhile assessing peptides specific to multiple tree nuts e.g. SGGYDEDNQR (shares specificity with walnut) in order to find more sensitive pecan peptide targets. One of the three egg *Gal d 4* peptides, NTDGSTDYQILQINSR, was excluded for similar sensitivity reasons. Also included in our method was one peptide, WLQLSAER, that we considered could be useful as a multiple tree nut marker because it is specific to hazelnut, pecan, pistachio, and walnut, but we found it also shares specificity with potato and was therefore excluded from analysis. Hydrophobic milk peptides YLGYLEQLLR and FFVAPFPEVFGK exhibited carryover into water blanks and subsequent samples following a sample containing high amounts of milk and in such instances were excluded from analysis. Having redundancy through the additional milk *Bos d 9* peptide HQGLPQEVLNENLLR enabled the quantification of milk even when ignoring these peptides in the cases of carryover.

Figure S1B shows how the gradient length and retention time window affect the number of concurrent method transitions, where higher values corresponds to lower transition dwell times according to the formula $d_t = \frac{c - pn_t}{n_t}$, where c is the cycle time, p is the pause time between transitions, and n_t is the number of transitions at time t . We chose a conventional 5 ms pause time and a cycle time of 1.2 s given a 14.4 s median peak width so that at least 10 points per peak were sampled. Compared to an 11 min elution (wash and reequilibration excluded) with 120 s scheduled retention time windows, a 24 min elution with 60 s scheduled retention time windows achieves an eight-fold greater median transition dwell time (Figure S1C) and was consequently chosen for the final method (Figure S1D).

Data analysis, quantification, and statistics

Data were acquired with Analyst 1.6.3 and processed with Skyline v4.2. Peak boundaries were chosen by Skyline and manually adjusted if necessary to align with SIL peptide transition boundaries. Reports and chromatograms exported from Skyline were processed and analyzed in Python using MADIC and the numpy, scipy, pandas, matplotlib, jupyter, and seaborn software packages. A Savitsky-Golay filter was applied to smooth raw chromatograms.

To quantify allergen contamination in commercial foods, a peptide calibration curve was constructed using an allergen extract mixture and wheat flour matrix. Wheat was chosen as a background because of its ubiquity in food and its moderate, representative protein content. Note the choice of wheat as a background therefore restricts the method to detecting instead of also quantifying wheat in an unknown sample. An initial mixture of 19 allergen extracts was generated by combining equal extract masses of the 14 target allergens plus barley flour, oat flour, rye flour, raw pecans and celery seeds (Table S2). An 8 point calibration corresponding to 1, 5, 10, 25, 50, 100, 250, and 500 ppm was then constructed by serial dilution of this mixture into the wheat matrix, following which a constant amount of SIL peptide mix was added. Detection limits for each peptide were established based on the four stringent peptide filtering criteria previously described: signal-to-noise, replicate agreement, expected relative transition ratios, and target transition retention time agreement with SIL peptide retention time. The ratio of native to SIL peptide transition areas for the top two transitions formed the basis for quantification of each peptide within Skyline. For points above each peptide's detection limit, a linear regression was constructed in Skyline using a 1/x weighting to prioritize accurate quantification of low ppm allergen contamination. For an unknown sample, ppm values were interpolated from this calibration curve and then multiplied by the ratio of measured protein extract concentration from the sample (Table S2) to measured protein extract concentration from the wheat calibration matrix, which was necessary to correct for the normalized total peptide injection mass of 20 µg used for all samples and to report ppm on a food, rather than protein, basis. As each sample was injected in duplicate, the final ppm value is given as the mean of these two values. For the large majority of allergic individuals, the limits of detection we have demonstrated are sufficiently sensitive to detect reactive allergen quantities. However, depending on the food serving size and allergen, an estimated 1-5% of allergic individuals who can react to 0.5 mg or less of allergenic protein⁶ may fall below these limits of detection.

Soy lecithin samples

We examined the soy protein content of three soy lecithin powders (sample 73-75) by adding 10% soy lecithin powder by mass, a much higher percentage normally found in foods, to a wheat flour carrier. Soy was found in only one sample (sample 74), which was the only one to contain PAL. After adjusting for the 10% mass fraction of lecithin, the calculated soy concentration in this sample was over 2000 ppm.

References

- (1) Colgrave, M. L.; Byrne, K.; Blundell, M.; Howitt, C. A. *J Proteomics* **2016**, *147*, 169–176.
- (2) Colgrave, M. L.; Byrne, K.; Blundell, M.; Heidelberger, S.; Lane, C. S.; Tanner, G. J.; Howitt, C. A. *Anal Chem* **2016**, *88*, 9127–9135.
- (3) Lock, S. *Foods* **2013**, *3*, 13–29.
- (4) Croote, D.; Quake, S. R. *npj Syst. Biol. Appl.* **2016**, *2*, 16022.
- (5) Korte, R.; Lepski, S.; Brockmeyer, J. *Anal Bioanal Chem* **2016**, *408*, 3059–3069.
- (6) Taylor, S. L.; Baumert, J. L.; Kruizinga, A. G.; Remington, B. C.; Crevel, R. W. R.; Brooke-Taylor, S.; Allen, K. J.; Allergen Bureau of Australia & New Zealand; Houben, G. *Food Chem Toxicol* **2014**, *63*, 9–17.

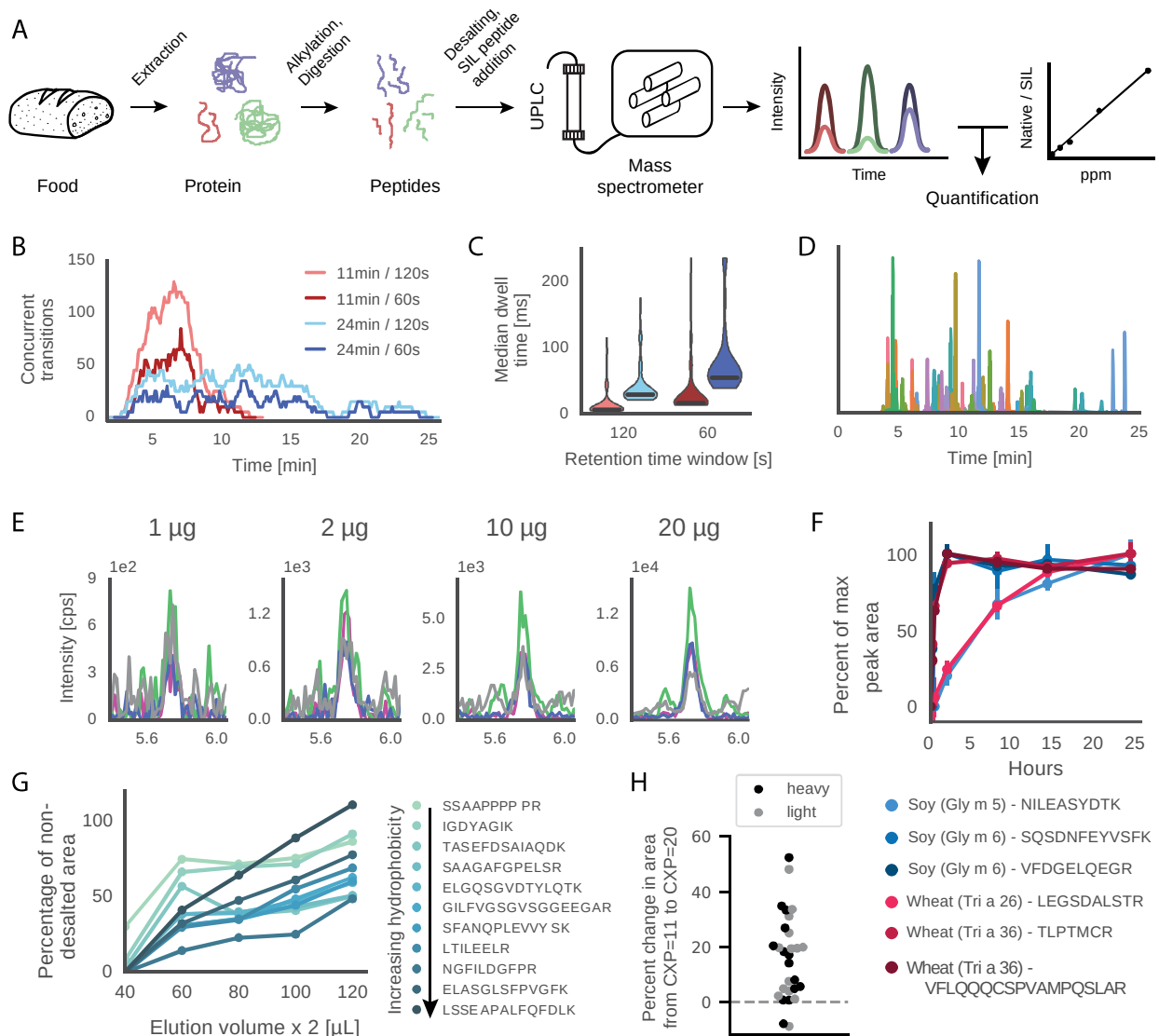


Figure S1. Assay design and optimization. (A) Assay overview. (B) Number of concurrent transitions for 11 min and 24 min elutions with 60 s or 120 s scheduled retention time windows. (C) Violin plots of median dwell times for all 265 transitions partitioned by method configuration in (B). (D) Final 30 min method with a 24 min elution and 60 s scheduled retention time windows for 53 peptides. (E) Increasing total peptide mass injected via higher reconstitution concentration increases transition intensity as well as signal to noise. (F) Tryptic digestion time course of a cookie (sample 25) containing wheat and soy. At each time point, an aliquot from the ongoing digest was acidified and stored at -20°C. All aliquots were subsequently desalted in parallel using a C18 spin plate. (G) Evaluation of peptide recovery as a function of C18 desalting elution volume spanning the manufacturer's recommended range using Pierce Retention Time Calibration peptides at a final concentration of 50 fmol / µL in a carrier of bovine serum albumin tryptic peptides. To maximize recovery, listed volumes correspond to two rounds of elution with that volume. Recovery was improved with larger elution volumes in a hydrophobicity-dependent manner with the exception of LSSEAPALFQFDLK. (H) Increasing the cell exit potential (CXP) from 11 to 20 increased peptide area for the native (light) and SiL (heavy) versions of 13 / 14 randomly selected peptides.

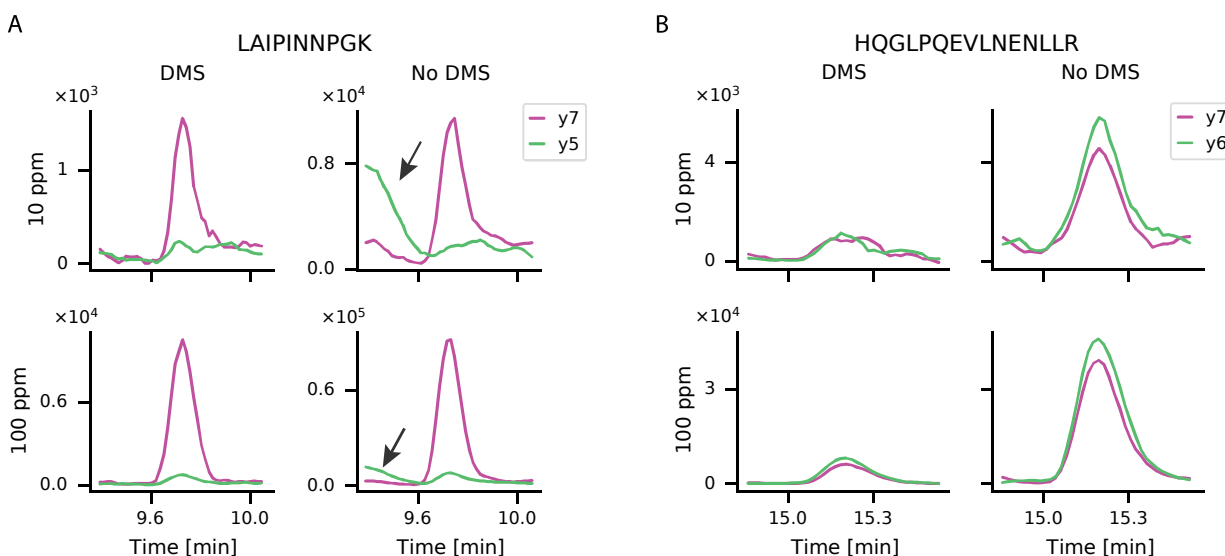


Figure S2. Tradeoff between signal intensity and interference reduction with the use of differential mobility spectrometry (DMS). DMS compensation voltages for each peptide were optimized using Skyline and three iterative injections of synthetic standards progressing from a 'rough tune' to 'fine tune.' No modifier solvent was used. (A) Lupin peptide LAIPINNPCK exhibits consistent off-peak wheat matrix interference in one transition (arrows) without DMS that is eliminated when using DMS at the cost of signal intensity. (B) Chromatograms for milk peptide HQGLPQEVLENENLLR featuring a shared y-axis scale between DMS and no-DMS conditions illustrate the loss in signal intensity from DMS.

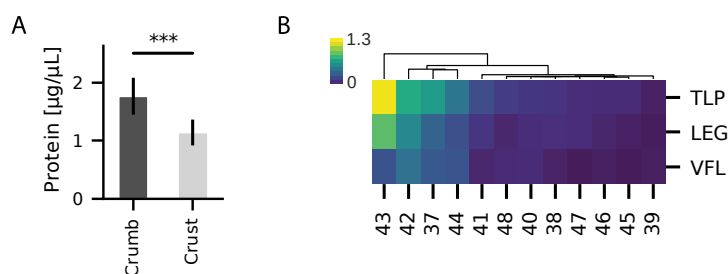


Figure S3. Effects of thermal processing on paired crumb and crust samples from breads. (A) Protein concentration in paired bread extracts, *** p < 0.001, paired sample t-test, n=12 breads, error bars = sd. (B) Heatmap illustrating fraction of SIL-normalized peptide area measured in the crust relative to the crumb for wheat *Tri a 26* peptide LEGSDALSTR (LEG) and *Tri a 36* peptides TLPTMCR (TLP) and VFLQQQCSPVAMPQSLAR (VFL).

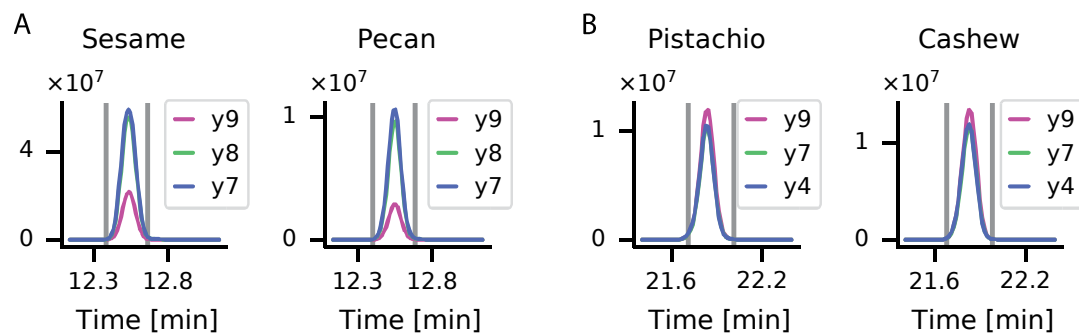


Figure S4. Nonspecific peptides. (A) Sesame *Ses i 6* peptide SPLAGYTSVIR is found in both sesame seeds (left; (sample A2)) and raw pecans (right; (sample A11)). (B) Pistachio *Pis v 2* peptide VTSINALNLPILR is found in raw pistachios (left; (sample A9)) and raw cashews (right; (sample A3)).