bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Light contamination in stable isotope-labelled internal peptide standards is frequent and a potential source of false discovery and quantitation error in proteomics

Mogjiborahman Salek[1,2]†*, Jonas D. Förster[1,2,3]†, Wolf-Dieter Lehmann[4], Angelika B. Riemer[1,2]*

[1] Immunotherapy & Immunoprevention, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2] Molecular Vaccine Design, German Center for Infection Research (DZIF), partner site Heidelberg, Heidelberg, Germany
[3] Faculty of Biosciences, Heidelberg University, Heidelberg, Germany
[4] Molecular Structure Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany

† equally contributed

* Corresponding authors. Street address: Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.
Email addresses: m.salek@dkfz.de, a.riemer@dkfz.de

ORCID-IDs

M. Salek: 0000-0002-6664-4752
J. D. Förster: 0000-0003-3437-6738
A.B. Riemer: 0000-0002-5865-0714

Statements and Declarations

Competing Interests: The authors have no competing interests to declare that are relevant to the content of this article.

Funding: This work was funded through the German Cancer Research Center (DKFZ), the National Center for Tumor Diseases (NCT) Heidelberg, and the German Center for Infectious Diseases (DZIF).

Acknowledgements: The authors gratefully acknowledge Rebecca Köhler for excellent technical assistance and Dr. Jonas Becker for careful proofreading and editing of the manuscript.

Author contributions: Conceptualization, methodology, data acquisition: MS and JDF. Formal bioinformatics analysis: JDF. Data interpretation: MS, JDF, WDL. Writing – original draft preparation: MS. Writing – review and editing: MS, JDF, WDL, ABR. Funding acquisition and resources: ABR. Supervision: MS and ABR.

ABSTRACT

In mass spectrometry-based proteomics, heavy internal standards are used to validate target peptide detections and to calibrate peptide quantification. Here we report light contamination present in heavy labelled synthetic peptides of high isotopic enrichment. Application of such peptides as assay-internal standards potentially compromises the detection and quantification especially of low abundant cellular peptides. Therefore, it is important to adopt guidelines to prevent false-positive identifications of endogenous light peptides as well as errors in their quantification from biological samples.

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

INTRODUCTION

Stable isotope labelling (SIL) of amino acids in general starts by growing microorganisms (e.g. green algae, cyanobacteria) in the presence of highly ^{13}C-enriched CO_2 and/or highly ^{15}N-enriched ammonia (each at >99.5 atom% enrichment) as sole source of carbon and nitrogen, respectively [1,2]. Labelled amino acids obtained in this way are then incorporated into peptides either by chemical synthesis or metabolic labelling [3]. Peptides containing at least one amino acid highly enriched in ^{13}C and/or ^{15}N are often named "heavy" to set them off against their cognates with natural isotopic abundances, named "light" in this context.

In targeted mass spectrometry (MS)-based proteomics and in particular peptidomics studies, heavy internal standards are regarded as the current gold-standard to validate peptide detection and quantitation. For example, with peptides presented by human leukocyte antigen (HLA) class I molecules, this identification is directly linked to subsequent immunotherapy and anti-cancer vaccine design [4-6]. The challenge in achieving accurate quantification is to effectively eliminate systematic and random errors due to various instrumental parameters or matrix effects. Isotope dilution MS (IDMS) based on heavy internal standards fits this purpose and provides proof of detection and absolute quantification of the highest precision and accuracy fit for clinical validation [7-9]. More recently, internal standards were used to dynamically control signal acquisition for endogenous light peptides via live instrument control [10]. With these broad applications in the field of mass spectrometry and in particular those assays aiming at clinical translation, high isotopic purity of heavy peptides used as internal reference is of key importance.

However, we here report that heavy peptides generated by chemical synthesis frequently contain sufficient levels of light cognates to bias interpretation of MS data, thus wrongly validating the detection of cellular peptides.

MATERIALS AND METHODS

Synthetic peptides and sample preparation

Peptide Retention Time Calibration (PRTC) Mixture (88321, Pierce™) was purchased from Thermo Fisher Scientific and is prepared at the same quality standards as "AQUA Ultimate" grade peptides of high chemical purity (>99%). Custom-synthetized heavy peptides were obtained from JPT Peptide Technologies (Berlin, Germany) and Synpeptide Co., Ltd (Shanghai, China) for low chemical purity (>70%) The above vendors are indicated in Table S1 as vendor 1, 2 and 3, respectively. For all vendors, the isotope purity is reported >99 atom% ^{13}C and ^{15}N, and incorporated heavy amino acids are V ($^{13}C_5$, $^{15}N_1$), L/I ($^{13}C_6$, $^{15}N_1$), K ($^{13}C_6$, $^{15}N_2$) and R ($^{13}C_6$, $^{15}N_4$). Upon arrival, peptides were dissolved in DMSO at 0.5nmol/µl, brought to a final concentration of 5 pmol/µl, dried, and stored at –80°C until LC-MS analysis.

Immunoprecipitation of HLA-displayed peptides

HLA immunoprecipitation (IP) was performed according to previously published protocols [11,12]. Briefly, $1x10^8$ C33A cells, a HPV-negative cervical cancer cell line, were lysed with a lysis buffer containing 1% N-octyl-β-D glucopyranoside, 0.25% Na-Deoxycholate, protease inhibitor cocktail (Sigma-Aldrich, Mannheim, Germany) /PMSF (Carl Roth, Karlsruhe, Germany) in PBS. After centrifugation at 40,000xg, 4°C, for 30 min, HLA-peptide complexes were immuno-isolated by

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

incubation with BB7.2 mouse anti-human HLA-A2 monoclonal antibody crosslinked to protein G Sepharose beads (Cytiva, Marlborough, MA, USA) for 4h at 4°C under constant mixing on a rotating wheel. Supernatant was discarded after centrifugation at 3200xg, for 3 min at RT. Pelleted HLA-peptide complexes bound to antibody-beads were washed 3 times in each of the following steps: first with ice-cold 20mM Tris-HCl pH 8 containing 150mM NaCl, followed by the same buffer but with 400 NaCl, and finally with 20mM Tris-HCl alone. Peptides were eluted from HLA bound to antibody-beads by 0.3% TFA. Resulting peptides were desalted by reverse-phase purification using a SepPak 96-well plate and dried by vacuum centrifugation (Concentrator plus, Eppendorf, Hamburg, Germany).

LC-MS

Samples were dissolved in 5% ACN, 0.1% TFA. Synthetic peptides were injected up to 280 fmol per peptide on column. All samples were analysed by liquid chromatography (U-3000, Thermo Fisher Scientific) coupled to Q Exactive HFX or Orbitrap Exploris 480 (Thermo Fisher Scientific). The LC gradient consisted of a first segment going from 3% to 7.5% B (19.9% H_2O, 0.1% FA, 80% ACN) and 92,5% A (0.1% FA in H_2O) over 5 min, followed by a second segment reaching 35.6% B in 75 min. Finally, the % B was increased in 2 steps to 95% in 4 min, followed by 5 min wash before 11 min equilibration at 2% B. For quantifying light contamination, Orbitrap Exploris was operated with multiplexed tSIM scans for simultaneous mass measurement of multiple peptide precursor ions using 240k resolution at 200 m/z, 100 ms injection time (IT), 100% AGC target and 2-3 separate quadrupole isolation windows, focused on the heavy and the light precursors. Protonated polycyclodimethylsiloxane (PCM-6, a background ion originating from ambient air) at 445.12 m/z served as a lock mass. MS2 data was acquired with PRM scans using 60k resolution at 200 m/z with a narrow <1 m/z isolation window tuned per precursor. Heavy precursor settings were 30ms IT, 50% AGC target. Light precursor masses were measured with 1000ms IT, 1000% AGC target. These settings are identical to those we use for the analysis of endogenous peptides in a biological sample.

To test the matrix effect, either 10 or 40 ng HeLa Protein Digest Standard (Pierce™, Thermo Scientific™) served as a test matrix which were spiked with 150 fmol (Fig 2 a), or 100 fmol (Fig 2 b c) of heavy target peptides. As a control, an IP sample obtained as described above was spiked with 50 fmol heavy target peptides (Fig 2 d). MS2 data was acquired with PRM scans using 60k resolution at 200 m/z with a narrow < 1 m/z isolation window tuned per precursor. To mirror the conditions of analysis of a real biological sample, the IP sample from C33A cells spiked with heavy peptides was analysed with parameters for heavy and light precursor mass measurements at 150 ms IT and 100% AGC target for the former, 1000 ms and 1000% for the latter. To explore the impact of the matrix effect on the sensitivity of peptide detection, MassPREP™ E.coli digestion standard (MPDS E.coli, Water, Milford, MA, USA) was used in the specified amounts to test the detection of the peptides from vendors 1 and 2 (Fig 2 e-g, Table S1). This data was acquired on a Q Exactive HF-X with an MS scan followed by 8 PRM scans, each with 0.7 m/z isolation window for a maximum of 350 ms IT and 5e5 ions AGC target with 23% normalized collision energy (NCE).

Data Analysis

MS1 data analysis was performed with R (v. 4.1) [13] with the "tidyverse" suite of packages (v. 1.3.0) [14]. Theoretical isotopic envelopes were calculated with the IsoSpec R package (v. 2.1.3) [15]. For theoretical heavy envelopes, several isotopic purity settings for heavy isotopes ^{13}C and ^{15}N were

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

probed and 97% isotopic purity was selected for best fits. Theoretical isotopic envelopes and tSIM spectra were centroided and matched with 4 ppm mass tolerance. The matched envelope was then fitted with a linear model, minimizing the sum of squares error.

MS2 data was analysed with the Skyline software (v. 20.2) [16]. Top 4 intensity product ions were extracted with 10 ppm mass tolerance. Detected peaks were manually curated. Light peaks were discarded when peak retention times or shapes did not match with the heavy reference, when the normalized spectral contrast angle (dotp) [17] was low, or when too few transitions were detected. For the complex peptide matrix (CPM) analysis, the "nls.multstart" (v. 1.2.0) R package [18] was used to fit the response of the peptide MS2 intensity (I) to the CPM load with the function $\mathrm{I} = a\ ([\mathrm{CPM}] - b)^{10\wedge c}$.

RESULTS AND DISCUSSION

The sequences and extent of light contamination in the synthetic heavy peptides we tested are given in full in supplementary Table S1 and Figure S1. These different sets of peptides were in use at the time in our laboratory and reflect different suppliers and purities. They include 9-15 amino acids long human leukocyte antigen (HLA) binders or tryptic peptides containing 1 or 2 labelled residues.

Representative for the peptides of highest available quality (99.5% isotope purity, >97% chemical purity, concentration precision ±10%, absolute quantification (AQUA) grade), peptide #43 GISNEGQNASI**K** shows precursor isotopic envelopes at both light (monoisotopic 0) and heavy (monoisotopic +8) m/z (Figure 1a). The strong similarity of the fragmentation patterns between the light and heavy monoisotopic precursor ions in comparison to an external library, as captured by high spectral contrast angle (dotp) values (0.96, 0.98), confirms the sequence identity of the light and heavy peptides (Figure 1b). The correct peptide identification is further supported by the alignment of their retention times, indicating virtually perfect chromatographic co-elution. The light contamination of peptide #43 GISNEGQNASI**K** is quantified as 219 ppm according to the MS2 peak area of the shown transitions (Figure 1c and Figure S1, for other peptides from this set: #40 (4507 ppm) to #54 (25 ppm)).

Although sources of such contamination can be multiple, the absence of a gradual increase in the intensity of intermediate signals between the light (0) and heavy (+8) monoisotopic peaks, excludes the possibility of incomplete incorporation of ^{13}C and ^{15}N isotopes during the initial metabolic labelling of the amino acids. More likely, the source of contamination would therefore be the introduction of completely unlabelled, *i.e.* "light", amino acids from an external source during amino acid labelling, extraction or peptide synthesis.

In around 50% of all heavy synthetic peptides we analysed as mixtures within separate batches from different vendors, the level of light contamination was higher than 100 ppm (0.01%, Figure 1c). The highest level of contamination measured was 4.5% for peptide #40 TASEFDSAIAQD**K** belonging to the AQUA grade peptide mixture (Figure 1c and Table S1). Initial measurements indicated exceedingly high contamination of >900 ppm for a total of 6 peptides. When confirming these peptides individually, the contamination assessment could be corrected down by factors of 100 to 1500 for those peptides, where the source of contamination was tracked down to another peptide of related sequence in the same mixture (Table S1, peptides #92, #95, #99, #100, #113). For instance, the peptide #92 RALYVDSLFF initially showed contamination of 32812 ppm but could be

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

corrected down to 84 ppm when measured separately, whereas the source of major contamination was determined to be the peptide #84 RALYVDSLFF**L** which introduced RALYVDSLFF as incomplete synthesis product. Thus, spike-in of such related sequences in a biological sample would be problematic. Nevertheless, it is important to note that the light contamination reported here cannot be attributed solely to cross-contamination due to mixing of peptides of low chemical purity. The separate analysis of single peptides still shows light contamination (Table S1, #1 K**L**GEIVTTI). In addition, when analysing three independent batches of the AQUA grade peptides (chemical purity >97%), the contamination remained high (#40: 4.5%, 1.1% or 0.04%, data not shown).

Data acquisition with instrument parameters tailored for high sensitivity inevitably increases the chances of detecting light contaminations as low as 50 ppm (Figure 2a, Table S1, #8 RTLED**L**LMGT). Similarly, in quantitative MS the well-documented matrix effect [19-21], often attributed to the changes in the efficiency of electrospray ionization, can enhance or suppress a target peptide signal. Indeed, we observed that the analysis of heavy peptides in the presence of a complex peptide matrix (CPM, here: HeLa tryptic digests), enhances both the signal of the heavy peptide and of its light contamination, which might have remained undetected in the absence of the matrix (Figure 2b). Increasing the amount of CPM showed similar effects as increasing instrument sensitivity of detection (Figure 2c). It has been suggested that consistent quantification requires analysis in a fixed amount of matrix [22], implying that in biological experiments negative controls including heavy internal standards need to be performed in the presence of cellular or closely related matrices. In the absence of such a control, peptide detection in a biological sample would lack sufficient level of objective evidence to distinguish the light contamination introduced by the heavy internal standard, from a *bona fide* endogenous peptide (Figure 2d).

To further investigate the observed CPM effect, we analysed 54 (Table S1, vendors 1 and 2) heavy peptides covering the entire chromatographic retention time range in the presence of increasing amounts of *E. coli* tryptic digests. The summed intensities of the top 6 fragment ions for three representative peptides with early, intermediate and late retention times indicate that hydrophilic peptides, represented by peptide #49 SAAGAFGPELS**R**, need much less CPM to reach their maximum intensity than hydrophobic peptides, represented by peptide #48 LSSEAPALFQFDL**K**. Collectively, the response of the intensity (I) for each of the 54 peptides to the amount of CPM is best captured by $I = a\ ([\mathrm{CPM}] - b)^{10\wedge c}$, where c is representing the curvature as negative, null or positive, a and b are scaling and offset factors, respectively. c delineates three responsiveness behaviours to the amount of CPM, as fast, linear and slow (Figure 2f). Indeed, this peptide-specific responsiveness correlates linearly with the chromatographic retention time of the peptides (Figure 2g), suggesting that the MS signal enhancement by CPM in our data is mainly caused by preventing peptide losses via hydrophobic interactions to vessel surfaces.

The consequences of light contamination present in heavy internal standards are even more important for interpretation of LC-MS data in biological contexts, as modern MS instrumentations are sensitive by design. Newly developed sophisticated data acquisition approaches [10,23-25] further increase the basic instrumental sensitivity and thereby the importance of taking into account the unappreciated source of error introduced by light contamination of heavy peptides. For instance, recently introduced intelligent instrument control improves IDMS by dynamic adjustment of detection parameters based on heavy internal standards. A low resolution survey of heavy standards added to the sample at a higher concentration will trigger real time measurement of light

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

endogenous peptides of low abundance at high resolution and sensitivity [10]. Here, light contamination, introduced by the addition of heavy labelled peptides of seemingly high isotopic purity, can be easily mistaken as endogenous light peptides leading to false positive identifications and flawed quantifications.

We therefore recommend the adoption of a few simple and rational guidelines to prevent false-positive detection of endogenous peptides. First, the impact of light contamination in batch analyses with multiple peptides can be controlled by consistent design of the peptide labelling. For instance in the case of KSVLTAFLM**L**W and KS**V**LTAFLM, the first peptide contains an unlabelled segment identical to the second, which should be prevented by carefully selecting the site of labelling, and if necessary introducing a second labelling within the peptide sequence. Second, it is paramount to determine systematically that the employed LC-MS method does not produce any light target signal when performed with a suitable negative control matrix spiked with the heavy internal standard peptide of interest. Additionally, in targeted MS workflows, it is still common to use light synthetic peptides, physically indistinguishable from the endogenous targets, for the optimisation of instrument parameters. We strongly recommend that the LC-MS system is not exposed to light synthetic target peptides under any circumstances, as carry-over could again lead to false-positive detection.

CONCLUSION

We here report that heavy synthetic peptides produced with the highest chemical and isotope purity can still contain substantial amounts of light cognates. The data was initially presented at the HUPO 2019 conference and its HUPO-HIPP satellite meeting in Adelaide [26]. To the best of our knowledge, this under-appreciated source of error is only described in one other publication, as one of several pitfalls that can obscure the analysis of HLA ligandome data [27]. We here provide an in-depth analysis of light contamination by i) testing heavy peptides from multiple providers, ii) measuring the overall extent of light contamination, and iii) describing the matrix effect that enhances the intensity of light contamination that otherwise may remain undetected in quality controls. Given the high sensitivity of modern mass spectrometers, light contamination of internal standards, even as low as 50 ppm, will inevitably lead to a high level of false discovery. In contrast to protein detection relying on multiple peptides, the impact of the neglected source of error reported here is expected to be large on the analysis of single peptides such as low abundance post-translational modifications in cellular signalling [28] and tumor-derived HLA-displayed peptides [5], often identified at the limit of detection. To prevent the generation of false-positive detections by such studies in the future, we recommend easy-to-implement protective measures.

REFERENCES

1. Bernstein RB (1957) Simple laboratory method for producing enriched Carbon-13. Science 126 (3264):119-120. doi:10.1126/science.126.3264.119
2. Cohen JS, Horsley W, Sternlicht H (1970) The isolation of carbon-13 enriched amino acids. Biochim Biophys Acta 222 (2):521-523. doi:10.1016/0304-4165(70)90144-3

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1 (5):376-386. doi:10.1074/mcp.m200025-mcp200
4. Chong C, Muller M, Pak H, Harnett D, Huber F, Grun D, Leleu M, Auger A, Arnaud M, Stevenson BJ, Michaux J, Bilic I, Hirsekorn A, Calviello L, Simo-Riudalbas L, Planet E, Lubinski J, Bryskiewicz M, Wiznerowicz M, Xenarios I, Zhang L, Trono D, Harari A, Ohler U, Coukos G, Bassani-Sternberg M (2020) Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat Commun 11 (1):1293. doi:10.1038/s41467-020-14968-9
5. Ebrahimi-Nik H, Michaux J, Corwin WL, Keller GL, Shcheglova T, Pak H, Coukos G, Baker BM, Mandoiu, II, Bassani-Sternberg M, Srivastava PK (2019) Mass spectrometry driven exploration reveals nuances of neoepitope-driven tumor rejection. JCI Insight 5. doi:10.1172/jci.insight.129152
6. Douglass J, Hsiue EH, Mog BJ, Hwang MS, DiNapoli SR, Pearlman AH, Miller MS, Wright KM, Azurmendi PA, Wang Q, Paul S, Schaefer A, Skora AD, Molin MD, Konig MF, Liu Q, Watson E, Li Y, Murphy MB, Pardoll DM, Bettegowda C, Papadopoulos N, Gabelli SB, Kinzler KW, Vogelstein B, Zhou S (2021) Bispecific antibodies targeting mutant RAS neoantigens. Sci Immunol 6 (57). doi:10.1126/sciimmunol.abd5515
7. Villanueva J, Carrascal M, Abian J (2014) Isotope dilution mass spectrometry for absolute quantification in proteomics: concepts and strategies. J Proteomics 96:184-199. doi:10.1016/j.jprot.2013.11.004
8. Whiteaker JR, Lin C, Kennedy J, Hou L, Trute M, Sokal I, Yan P, Schoenherr RM, Zhao L, Voytovich UJ, Kelly-Spratt KS, Krasnoselsky A, Gafken PR, Hogan JM, Jones LA, Wang P, Amon L, Chodosh LA, Nelson PS, McIntosh MW, Kemp CJ, Paulovich AG (2011) A targeted proteomics-based pipeline for verification of biomarkers in plasma. Nat Biotechnol 29 (7):625-634. doi:10.1038/nbt.1900
9. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 100 (12):6940-6945. doi:10.1073/pnas.0832254100
10. Gallien S, Kim SY, Domon B (2015) Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). Mol Cell Proteomics 14 (6):1630-1644. doi:10.1074/mcp.O114.043968
11. Bassani-Sternberg M, Braunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, Straub M, Weber J, Slotta-Huspenina J, Specht K, Martignoni ME, Werner A, Hein R, D HB, Peschel C, Rad R, Cox J, Mann M, Krackhardt AM (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. Nat Commun 7:13404. doi:10.1038/ncomms13404
12. Chong C, Marino F, Pak H, Racle J, Daniel RT, Muller M, Gfeller D, Coukos G, Bassani-Sternberg M (2018) High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferongamma-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. Mol Cell Proteomics 17 (3):533-548. doi:10.1074/mcp.TIR117.000383
13. Team RC (2021) R: A Language and Environment for Statistical Computing.
14. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. Journal of Open Source Software 4 (43). doi:10.21105/joss.01686
15. Lacki MK, Valkenborg D, Startek MP (2020) IsoSpec2: Ultrafast Fine Structure Calculator. Anal Chem 92 (14):9472-9475. doi:10.1021/acs.analchem.0c00959
16. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

analyzing targeted proteomics experiments. Bioinformatics 26 (7):966-968. doi:10.1093/bioinformatics/btq054

17. Toprak UH, Gillet LC, Maiolica A, Navarro P, Leitner A, Aebersold R (2014) Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. Mol Cell Proteomics 13 (8):2056-2071. doi:10.1074/mcp.O113.036475
18. Padfield D, Matheson G (2020) nls.multstart: Robust Non-Linear Regression using AIC Scores. 1.2.0 edn.,
19. IUPAC (2020) matrix effect. In: Compendium of Chemical Terminology. doi:10.1351/goldbook.M03759
20. Trufelli H, Palma P, Famiglini G, Cappiello A (2011) An overview of matrix effects in liquid chromatography-mass spectrometry. Mass Spectrom Rev 30 (3):491-509. doi:10.1002/mas.20298
21. Zhou W, Yang S, Wang PG (2017) Matrix effects and application of matrix effect factor. Bioanalysis 9 (23):1839-1844. doi:10.4155/bio-2017-0214
22. Heller DN (2007) Ruggedness testing of quantitative atmospheric pressure ionization mass spectrometry methods: the effect of co-injected matrix on matrix effects. Rapid Commun Mass Spectrom 21 (5):644-652. doi:10.1002/rcm.2882
23. Meier F, Geyer PE, Virreira Winter S, Cox J, Mann M (2018) BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. Nat Methods 15 (6):440-448. doi:10.1038/s41592-018-0003-5
24. Virreira Winter S, Meier F, Wichmann C, Cox J, Mann M, Meissner F (2018) EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. Nat Methods 15 (7):527-530. doi:10.1038/s41592-018-0037-8
25. Wichmann C, Meier F, Virreira Winter S, Brunner AD, Cox J, Mann M (2019) MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides. Mol Cell Proteomics 18 (5):982-994. doi:10.1074/mcp.TIR118.001131
26. Salek M, Förster J, Köhler R, Riemer AB Poster #820: Detection of ultra-low abundant epitopes by targeted mass spectrometry. In: HUPO 2019, Adelaide, Australia, 2019.
27. Fritsche J, Kowalewski DJ, Backert L, Gwinner F, Dorner S, Priemer M, Tsou CC, Hoffgaard F, Romer M, Schuster H, Schoor O, Weinschenk T (2021) Pitfalls in HLA Ligandomics-How to Catch a Li(e)gand. Mol Cell Proteomics 20:100110. doi:10.1016/j.mcpro.2021.100110
28. Stopfer LE, Flower CT, Gajadhar AS, Patel B, Gallien S, Lopez-Ferrer D, White FM (2021) High-Density, Targeted Monitoring of Tyrosine Phosphorylation Reveals Activated Signaling Networks in Human Tumors. Cancer Res 81 (9):2495-2509. doi:10.1158/0008-5472.CAN-20-3804

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

Figure 1. Presence of light contamination in heavy peptides.

(a) Upper and lower panels show selected ion monitoring (SIM) data acquisition scans for the light and heavy forms of a representative peptide and its 4000x zoom in, respectively. The calculated (yellow) and experimentally measured (black) isotope patterns of the labelled peptide and its light contamination are displayed. The indicated number of heavy isotopes, "0" and "+8", corresponds to the light and heavy monoisotopic signals, respectively. Background peaks not matched to the isotope pattern are displayed in grey (b) Extracted fragment-ion chromatograms representing the light and heavy forms of the depicted peptide sequence (the labelled residue, lysine (K), is in bold, for which the sum of incorporated stable isotopes is 8 ($^{13}C_6$ and $^{15}N_2$), as indicated in (a)). The bars underneath the peptide sequence correspond to the detected fragments or transitions and are color-coded accordingly. dotp is the normalized spectral contrast angle that is scoring the similarity of the detected peptide´s fragmentation pattern of light (L) and heavy (H) precursors, compared to a reference library. The identity of the peptide is further confirmed by the co-elution of the light and heavy signals. (c) Extent of light contamination (L/H) for 113 tested heavy peptides, from three vendors (for colour scheme see Table S1). The light contamination for each peptide is shown in the left panel as the summed peak area for the top 4 fragments of the light (dots) and heavy (triangles) peptides. The right panel shows the extent of light contamination as parts per million (ppm). The arrow pinpoints the peptide #43 (219 ppm), for which the details of light contamination are shown in panels (a) and (b).

Figure 2. Injection time and matrix effects on the detection of light contamination.

(a-d) Extracted ion chromatograms representing the transitions (fragments) corresponding to the depicted heavy peptide sequence and its light contamination detected in different conditions. dotp comparison to reference library is annotated for light (L) and heavy (H) signals. (a) Increasing of injection time (sensitivity); (b) presence or absence of a CPM; (c) increasing amounts of CPM; (d) in the context of anti-HLA immunoprecipitation (IP) sample. Oxidized peptide is targeted in this assay. Heavy reference signal in the IP sample is trace-level carry-over from other experiments on the same LC system. This is common for heavy peptides injected at relatively high concentrations. (e-g) Correlation between the chemical nature of target peptides and the amount of CPM. Three technical replicates per condition. The response of peptide intensity to CPM amount is fitted with $\mathrm{I} = a\,([\mathrm{CPM}] - b)^{10\wedge c}$; (e) Three representative peptides (hydrophilic, intermediate, lipophilic) for different behaviour in the presence of CPM. (f) The effect of factor *c* and the modes that were commonly observed for peptides of different hydrophobicity (g) The factor *c* shows a linear correlation with peptide retention time (RT) (Pearson correlation coefficient of 0.8).

Figure S1. Full set of extracted ion chromatograms acquired for light contamination assessment.

Extracted ion chromatograms for heavy (bottom) and light (top) precursors. Where appropriate, panels showing peptide chromatograms from single peptide solutions are given next to chromatograms acquired from the full peptide mixture. dotp and ppm assessment of heavy vs light is calculated from the summed peak area of the top 4 transitions of the displayed chromatograms. Where light contamination signal was not selected as identified, ppm and dotp are given as "NA" (not available).

bioRxiv preprint doi: https://doi.org/10.1101/2021.11.08.467684; this version posted November 8, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table S1. Full list of peptides acquired in this study and the assessed ppm quantification for the light contamination.

Peptide numbering reflects the position in Figure 1c. Vendor information is given as specified in the experimental section. Light contamination signal is given as fraction of light precursor peak area vs heavy precursor peak area. Separate columns list measurements where peptide contaminations were confirmed individually as opposed to bulk measurement in mixture.

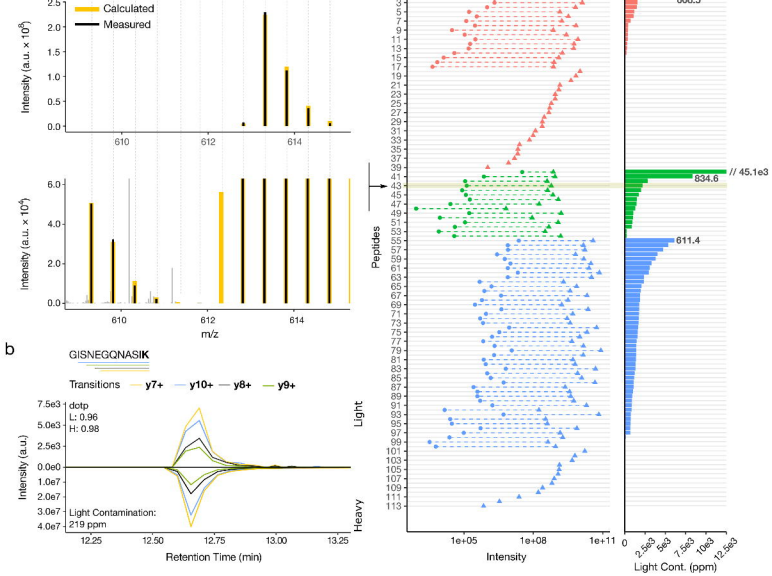
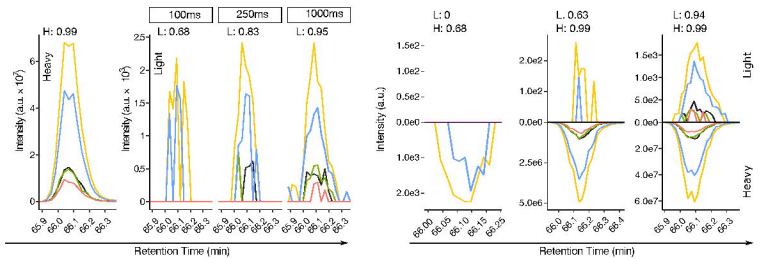
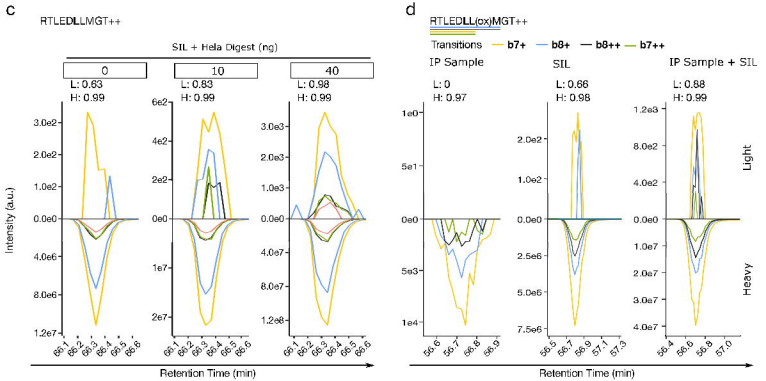

Calculated
Measured
Intensity (a.u. x 10^8)
m/z
b
GISNEGQNASIK
Transitions y7+ y10+ y8+ y9+
dotp
L: 0.96
H: 0.98
Light Contamination:
219 ppm
Intensity (a.u.)
Retention Time (min)
Light
Heavy
Peptides
Intensity
Light Cont. (ppm)
// 45.1e3
834.6
611.4

Figure 1


100ms
250ms
1000ms
H: 0.99
L: 0.68
L: 0.83
L: 0.95
Heavy
Light
Intensity (a.u. x 10^5)
Retention Time (min)
L: 0
H: 0.88
L: 0.63
H: 0.99
L: 0.94
H: 0.99
Intensity (a.u.)
Retention Time (min)


c
RTLEDLLMGT++
SIL + Hela Digest (ng)
0
10
40
L: 0.63
H: 0.99
L: 0.83
H: 0.99
L: 0.98
H: 0.99
Intensity (a.u.)
Retention Time (min)
d
RTLEDLL(ox)MGT++
Transitions
b7+
b8+
b8++
b7++
IP Sample
SIL
IP Sample + SIL
L: 0
H: 0.97
L: 0.66
H: 0.98
L: 0.88
H: 0.99
Light
Heavy
Retention Time (min)

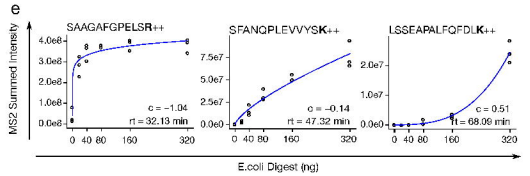
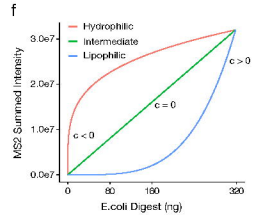
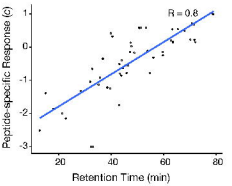

e
SAAGAFGPELSR++
SFANQPLEVVYSK++
LSSEAPALFQFDLK++
MS2 Summed Intensity
c = −1.04
rt = 32.13 min
c = −0.14
rt = 47.32 min
c = 0.51
rt = 68.09 min
E.coli Digest (ng)

Figure 2


f
Hydrophilic
Intermediate
Lipophilic
MS2 Summed Intensity
c > 0
c = 0
c < 0
E.coli Digest (ng)


g
R = 0.8
Peptide-specific Response (c)
Retention Time (min)