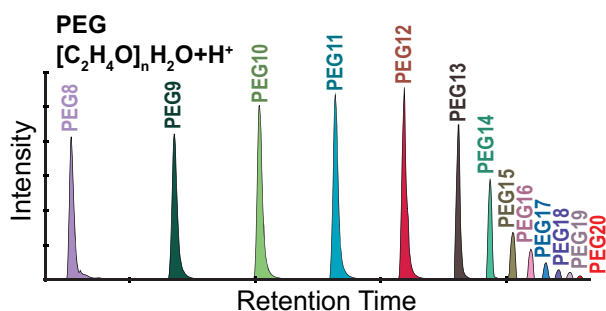CrossMark

APPLICATION NOTE

Rapid Assessment of Contaminants and Interferences in Mass Spectrometry Data Using Skyline

Matthew J. Rardin

Discovery Attribute Sciences, Amgen, South San Francisco, CA 94080, USA

Abstract. Proper sample preparation in proteomic workflows is essential to the success of modern mass spectrometry experiments. Complex workflows often require reagents which are incompatible with MS analysis (e.g., detergents) necessitating a variety of sample cleanup procedures. Efforts to understand and mitigate sample contamination are a continual source of disruption with respect to both time and resources. To improve the ability to rapidly assess sample contamination from a diverse array of sources, I developed a molecular library in Skyline for rapid extraction of contaminant precursor signals using MS1 filtering. This contaminant template library is easily managed and can be modified for a diverse array of mass spectrometry sample preparation workflows. Utilization of this template allows rapid assessment of sample integrity and indicates potential sources of contamination.

Keywords: Contaminants, Interferences, Skyline, MS1 filtering, Proteomics

Introduction

The analysis of peptides, proteins, and metabolites by liquid chromatography-mass spectrometry (LC-MS) is susceptible to a wide variety of contaminants that can compromise downstream analysis. The introduction of these contaminants may lead to intensive examination of workflows and reagents to identify the source, costing both time and money. In addition to the source of the sample, the complexity of sample preparation and workflows can lead to the introduction of new reagents and materials which may present unknown interferences to downstream analysis (Fig. 1a). One of the most common contaminants observed in proteomic workflows are surfactants like polyethylene glycol (PEG) that are introduced during sample preparation [1, 2]. These contaminants are particularly disruptive as they lead to ion suppression and often interfere with the target ion(s) of interest [3, 4]. Other sources of common contaminants include plasticizers such as phthalate esters

[5] and slip agents such as erucamide [6]. In addition to contamination of the liquid phase, there is also the potential for gas phase contamination from the laboratory air environment. Polydimethylcyclosiloxanes are common additives to skin care and cosmetic products and are ubiquitous in the laboratory air environment leading to high background signals in nanoflow LC-MS [7]. Fortunately, this type of contamination can be partially mitigated using active background ion reduction devices. For an extensive review on the sources and types of contamination in LC-MS, see the following review [6].

There are currently a variety of peptide standards and tools for assessing the performance of a mass spectrometer as well as quality control metrics [8–10]. However, despite the multitude of contamination entry points to proteomic workflows and their prevalence in samples, containers, and reagents, the modern protein chemist does not have the ability to rapidly assess MS data for the presence and levels of known contaminants beyond the manual interrogation of raw data. Here, I present an approach for rapidly assessing sample contamination using full-scan MS1 filtering in Skyline with a customizable transition list that provides a starting point for the rapid identification of common contaminants in proteomic workflows. Skyline is an open-source label-free quantitation application originally developed for multiple reaction monitoring experiments [11] and

Correspondence to: Matthew Rardin; *e-mail:* mrardin@amgen.com

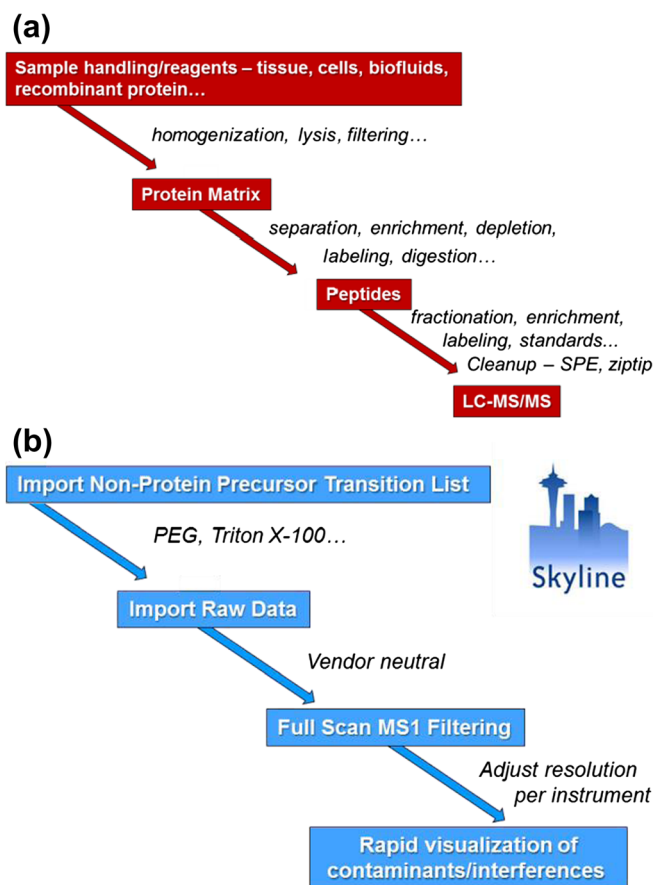**Figure 1.** (**a**) Schematic representation of potential sources of contamination during proteomic workflows. (**b**) Workflow for utilizing Skyline's molecular extraction tools to generate a transition list of non-proteinaceous molecules that can be monitored at the MS1 scan level. SPE, solid phase extraction; LC-MS/MS, liquid chromatography-tandem mass spectrometry; PEG, polyethylene glycol

later expanded to full-scan MS1 data [12–14]. Skyline features tools for viewing graphical displays of extracted ion chromatograms and is capable of processing data from most major vendors [15], making the approach described here to monitor common contaminants widely accessible.

Experimental

Non-Proteinaceous Transition List in Skyline

The list of molecular contaminants used in the current version (Supplemental Table S1) was compiled from a collection of reviews and reports on interferences and contaminants in mass spectrometry [1, 6, 7, 16]. Inserting a non-protein transition list into Skyline requires several pieces of information: molecule list name, precursor name, molecular formula, adduct ion (e.g., H^+, Na^+, NH_4^+), precursor mass-to-charge, and charge state. All molecules were listed as singly charged based on previous reports [6, 17]. For polymers such as PEG, the molecular list name remains constant while the precursor name varies with polymer length. Total PEG contamination, as with other

polymers, is then viewed by highlighting only the molecular list name in the Skyline transition tree.

MS1 Filtering in Skyline

Skyline is an open-source software application that is freely available for download [11, 15]. For additional details and tutorials, visit the Skyline website (http://proteome.gs. washington.edu/software/skyline). Full-scan (MS1) features were set to Orbitrap for precursor mass analyzer, resolving power of 120,000 at 400 *m/z* and one isotopic peak. Instrument scan range was set to 350–1500 *m/z*. Raw data files were imported directly into Skyline (v4.1.0.11714) and ion intensity chromatograms are displayed for a single isotopic peak. The Skyline contamination template file can be viewed and downloaded via the Panorama Public data repository: https://panoramaweb.org/labkey/contaminants.url.

Instrumentation

Data was acquired utilizing a Waters nanoACQUITY M-class system (Waters, Milford, MA) in-line with an Orbitrap Fusion tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a Digital PicoView nanospray source (DPV550, New Objective, Woburn, MA). Samples were separated on a 150 mm × 75 μm C18 charged surface hybrid column with 1.7-μm particle size (Waters, Milford, MA) at a flow rate of 300 nL/min. Data was acquired in positive ion mode using a top speed method at an MS1 resolution of 120,000.

Results and Discussion

Characterization of proteins and peptides by mass spectrometry utilizes a wide variety of sample preparation methods from intact protein analysis to diverse procedures requiring isolation and homogenization of tissues for generating a protein matrix (Fig. 1a). Protein mixtures can then undergo a number of procedures such as enrichment or depletion followed by proteolytic digestion. The resulting peptide mixtures can then be further processed by fractionation or labeling prior to a desalting step before analysis by LC-MS/MS. Each stage or reagent in the workflow is a potential source of contamination and mitigation of potential interfering compounds is a time-consuming and difficult process. To rapidly assess mass spectrometry data for known sources of contamination, a molecular library was developed using previously compiled databases [6, 16] and the open-source application Skyline [11] (Fig. 1b). The molecular transition list consists of 64 parent molecules and 800 molecular species (Supplemental Table S1). This transition list contains commonly observed contaminants in proteomic-based workflows including surfactants like PEG and Triton X-100, plasticizers such as diisoocytl phthalate, slip agents like erucamide, polysiloxanes commonly found in beauty products, and bittering agents like denatonium from low-purity solvents (Table 1). In addition to the protonated form of the molecule,

Table 1. Abbreviated list of common types of contaminants routinely observed in proteomic workflows that can be monitored using Skyline

Molecule list name	Precursor ion formula	Precursor m/z[a]	Category
Polyethylene glycol	$C_{24}H_{48}O_{12}H_2OH$	547.3324	Surfactant
Polyethylene glycol, ammonium adduct	$C_{24}H_{48}O_{12}H_2ONH_4$	564.3590	Surfactant
Triton X-100, X-114, X-405, or X-45	$C_{14}H_{22}OC_{18}H_{36}O_9H$	603.4103	Detergent
Triton X-100, X-114, X-405, or X-45, sodium adduct	$C_{14}H_{22}OC_{18}H_{36}O_9HNa$	625.3392	Detergent
Polysiloxane	$C_{10}H_{30}Si_5O_5H$	371.1012	Cosmetics/deodorants
Denatonium benzoate	$C_{21}H_{28}N_2OH$	325.2274	Bittering agent
Diisooctyl phthalate	$C_{24}H_{36}O_4H$	391.2843	Plasticizer
Dimethyl sulfoxide	$C_2H_6OSCH_3CNH$	120.0478	Solvent
Erucamide	$C_{22}H_{43}NOH$	338.3417	Slip Agent

[a]Represents singly charged species

ammoniated or sodiated forms are also included in some cases. Using this template in Skyline allows one to rapidly assess their samples for known contaminants that may interfere with downstream analysis.

To demonstrate the utility of this approach, a raw data file with regularly spaced peaks in the chromatogram was examined (Fig. 2a). The extracted MS1 scan (350–1500 m/z) from this region of the gradient displays two ion series separated by repeating units of 44.026 (Fig. 2b). This ion series is a hallmark of polymer contamination and both ion series from the chromatogram correspond to the protonated and ammoniated form of PEG ($[C_2H_4O]_nH_2O+H^+$ and $[C_2H_4O]_nH_2O+NH_4^+$, respectively). The raw data file was then imported into Skyline containing the molecular contaminant transition list
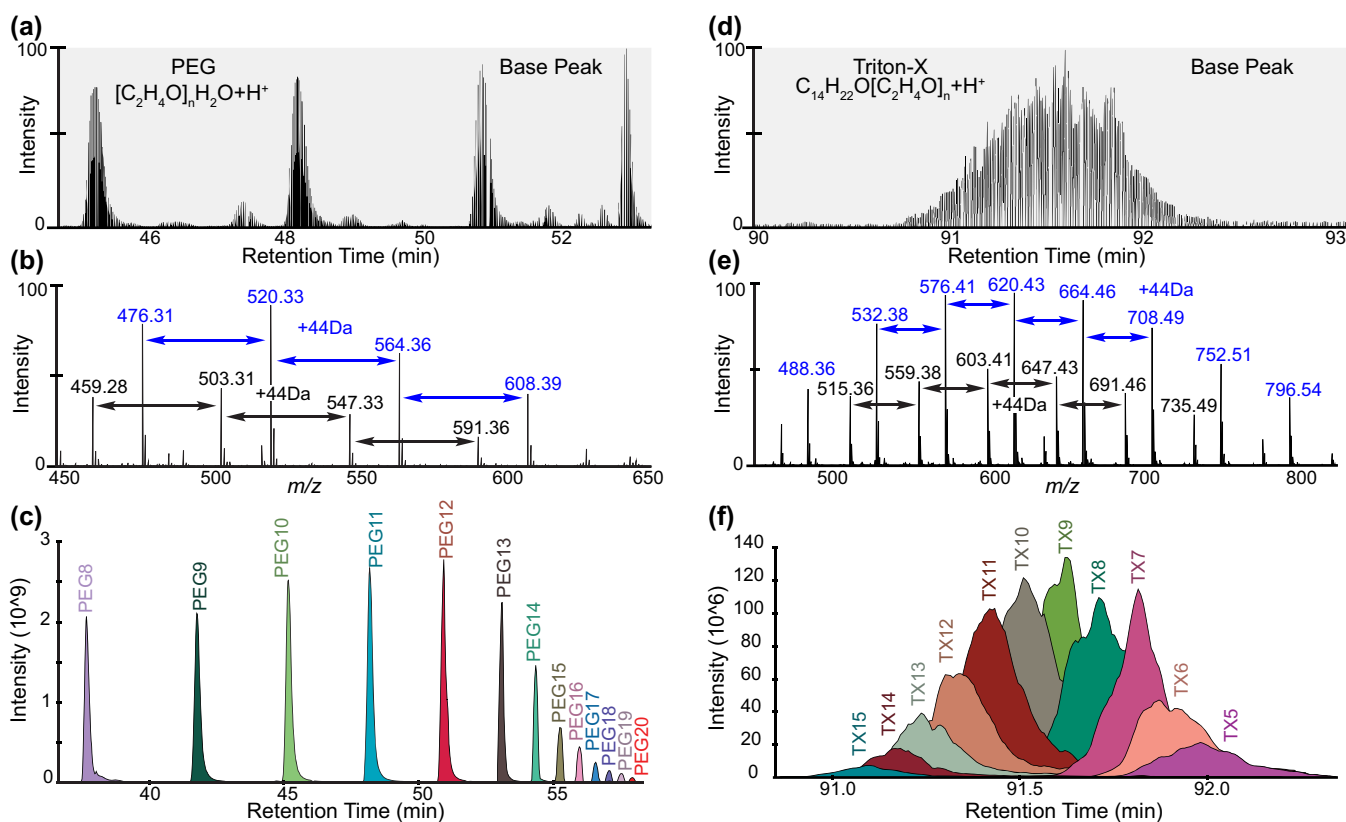

Figure 2. Monitoring contaminants in Skyline. (**a**) Base peak of evenly spaced peaks observed during an LC-MS/MS acquisition containing PEG. (**b**) Extracted MS1 scan of chromatogram shown in (**a**) above, with two ion series separated by repeating units of ethylene oxide at 44.026 Da. The protonated form of PEG; ($[C_2H_4O]_nH_2O+H^+$) is separated by black arrows and the blue arrows represent the ammoniated forms of PEG ($[C_2H_4O]_nH_2O+NH_4^+$). (**c**) Chromatograms, peak intensities, and retention times for PEG molecules extracted and visualized in Skyline. Peak labels represent the number of ethylene oxide units. (**d**) Triton X (may contain Triton X-100, X-114, X-405, or X-45) base peak observed during an LC-MS/MS acquisition. (**e**) Extracted MS1 scan of chromatogram shown in (**d**) above with two ion series separated by repeating units of ethylene glycol 44.026 Da. The protonated form of Triton X-100 ($C_{14}H_{22}O[C_2H_4O]_n+H^+$) is separated by black arrows and the blue arrow represents the ammoniated form of Triton X-100 ($C_{14}H_{22}O[C_2H_4O]_n+NH_4^+$). (**f**) Chromatograms, peak intensities, and retention times for Triton X molecules extracted and visualized in Skyline. Peak labels TX5–18 represent the number of ethylene oxide units

(Supplemental Table S1) and the MS1 peak area was extracted for each molecular species corresponding to PEG_{1-20} (Fig. 2c). The graphical display in Skyline demonstrates the sample is heavily contaminated with PEG polymers ranging from PEG8 ($[C_2H_4O]_8H_2O+H^+$ - 371.2276^+) to PEG20 ($[C_2H_4O]_{20}H_2O+H^+$ - 899.5421^+) with individual peaks spread across several minutes of the gradient. Another common contaminant observed in proteomic workflows is the detergent Triton X-100 often used for solubilization of biological samples. In contrast to PEG contamination which tends to elute with regularly spaced peaks spread across the gradient (Fig. 2a), polymers of Triton X-100 elute as one broad peak (Fig. 2d). Similar to PEG, Triton X-100 also displays a molecular ion series separated by 44.026 Da and in this case both the protonated and ammoniated forms are also present, $C_{14}H_{22}O[C_2H_4O]_n+H^+$ and $C_{14}H_{22}O[C_2H_4O]_n+NH_4^+$, respectively (Fig. 2e). Extraction of the MS1 scan in Skyline reveals a series of overlapping peaks that co-elute within a few minutes of each other (Fig. 2f). These two examples demonstrate the feasibility of using Skyline for assessing sample integrity during proteomic-based workflows for non-protein-based contaminants. In addition, once a species is added to the list of molecules to monitor, one no longer needs to undertake the tedious task of matching up ions manually from published databases.

Conclusion

Although several tools and approaches have been developed to assess instrument performance metrics such as reproducibility and sensitivity, little effort has been done to help researchers rapidly interrogate the integrity of their samples for molecular interferences. The current work provides an approach for rapidly assessing contamination of mass spectrometry data by non-proteinaceous molecules saving both time and valuable resources. The current molecular transition list is not meant to be comprehensive, but rather a starting point for which one can easily modify and adapt to various analytical needs. Although this approach does not identify unknown species, I have found that utilizing mass to formula calculators [18] can readily serve this purpose. Finally, by adapting an open-source vendor-neutral software platform like Skyline, this approach is easily adaptable to most proteomic workflows and mass spectrometry platforms.

References

1. Tong, H., Bell, D., Tabei, K., Siegel, M.M.: Automated data massaging, interpretation, and E-mailing modules for high throughput open access mass spectrometry. J. Am. Soc. Mass Spectrom. **10**, 1174–1187 (1999)

2. Weaver, R., Riley, R.J.: Identification and reduction of ion suppression effects on pharmacokinetic parameters by polyethylene glycol 400. Rapid Commun. Mass Spectrom. **20**, 2559–2564 (2006)

3. Annesley, T.M.: Ion suppression in mass spectrometry. Clin. Chem. **49**, 1041–1044 (2003)

4. Furey, A., Moriarty, M., Bane, V., Kinsella, B., Lehane, M.: Ion suppression; a critical review on causes, evaluation, prevention and applications. Talanta. **115**, 104–122 (2013)

5. Verge, K.M., Agnes, G.R.: Plasticizer contamination from vacuum system O-rings in a quadrupole ion trap mass spectrometer. J. Am. Soc. Mass Spectrom. **13**, 901–905 (2002)

6. Keller, B.O., Sui, J., Young, A.B., Whittal, R.M.: Interferences and contaminants encountered in modern mass spectrometry. Anal. Chim. Acta. **627**, 71–81 (2008)

7. Schlosser, A., Volkmer-Engert, R.: Volatile polydimethylcyclosiloxanes in the ambient laboratory air identified as source of extreme background signals in nanoelectrospray mass spectrometry. J. Mass Spectrom. **38**, 523–525 (2003)

8. Burkhart, J.M., Premsler, T., Sickmann, A.: Quality control of nano-LC-MS systems using stable isotope-coded peptides. Proteomics. **11**, 1049–1057 (2011)

9. Bereman, M.S.: Tools for monitoring system suitability in LC MS/MS centric proteomic experiments. Proteomics. **15**, 891–902 (2015)

10. Abbatiello, S.E., Mani, D.R., Schilling, B., Maclean, B., Zimmerman, L.J., Feng, X., Cusack, M.P., Sedransk, N., Hall, S.C., Addona, T., Allen, S., Dodder, N.G., Ghosh, M., Held, J.M., Hedrick, V., Inerowicz, H.D., Jackson, A., Keshishian, H., Kim, J.W., Lyssand, J.S., Riley, C.P., Rudnick, P., Sadowski, P., Shaddox, K., Smith, D., Tomazela, D., Wahlander, A., Waldemarson, S., Whitwell, C.A., You, J., Zhang, S., Kinsinger, C.R., Mesri, M., Rodriguez, H., Borchers, C.H., Buck, C., Fisher, S.J., Gibson, B.W., Liebler, D., Maccoss, M., Neubert, T.A., Paulovich, A., Regnier, F., Skates, S.J., Tempst, P., Wang, M., Carr, S.A.: Design, implementation and multisite evaluation of a system suitability protocol for the quantitative assessment of instrument performance in liquid chromatography-multiple reaction monitoring-MS (LC-MRM-MS). Mol. Cell. Proteomics. **12**, 2623–2639 (2013)

11. 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.: Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. **26**, 966–968 (2010)

12. Schilling, B., Rardin, M.J., MacLean, B.X., Zawadzka, A.M., Frewen, B.E., Cusack, M.P., Sorensen, D.J., Bereman, M.S., Jing, E., Wu, C.C., Verdin, E., Kahn, C.R., Maccoss, M.J., Gibson, B.W.: Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. Mol. Cell. Proteomics. **11**, 202–214 (2012)

13. Rardin, M.J., Newman, J.C., Held, J.M., Cusack, M.P., Sorensen, D.J., Li, B., Schilling, B., Mooney, S.D., Kahn, C.R., Verdin, E., Gibson, B.W.: Label-free quantitative proteomics of the lysine acetylome in mitochondria identifies substrates of SIRT3 in metabolic pathways. Proc. Natl. Acad. Sci. U. S. A. **110**, 6601–6606 (2013)

14. Rardin, M.J., Schilling, B., Cheng, L.Y., MacLean, B.X., Sorensen, D.J., Sahu, A.K., MacCoss, M.J., Vitek, O., Gibson, B.W.: MS1 peptide ion intensity chromatograms in MS2 (SWATH) data independent acquisitions. Improving post acquisition analysis of proteomic experiments. Mol. Cell. Proteomics. **14**, 2405–2419 (2015)

15. Pino, L.K., Searle, B.C., Bollinger, J.G., Nunn, B., MacLean, B., MacCoss, M.J.: The skyline ecosystem: informatics for quantitative mass spectrometry proteomics. Mass Spectrom. Rev. (2017). https://doi.org/10.1002/mas.21540

16. Weber, R.J., Li, E., Bruty, J., He, S., Viant, M.R.: MaConDa: a publicly accessible mass spectrometry contaminants database. Bioinformatics. **28**, 2856–2857 (2012)

17. Bachor, R., Kluczyk, A., Stefanowicz, P., Szewczuk, Z.: Facile synthesis of deuterium-labeled denatonium cation and its application in the quantitative analysis of Bitrex by liquid chromatography-mass spectrometry. Anal. Bioanal. Chem. **407**, 6557–6561 (2015)

18. Strohalm, M., Hassman, M., Kosata, B., Kodicek, M.: mMass data miner: an open source alternative for mass spectrometric data analysis. Rapid Commun. Mass Spectrom. **22**, 905–908 (2008)