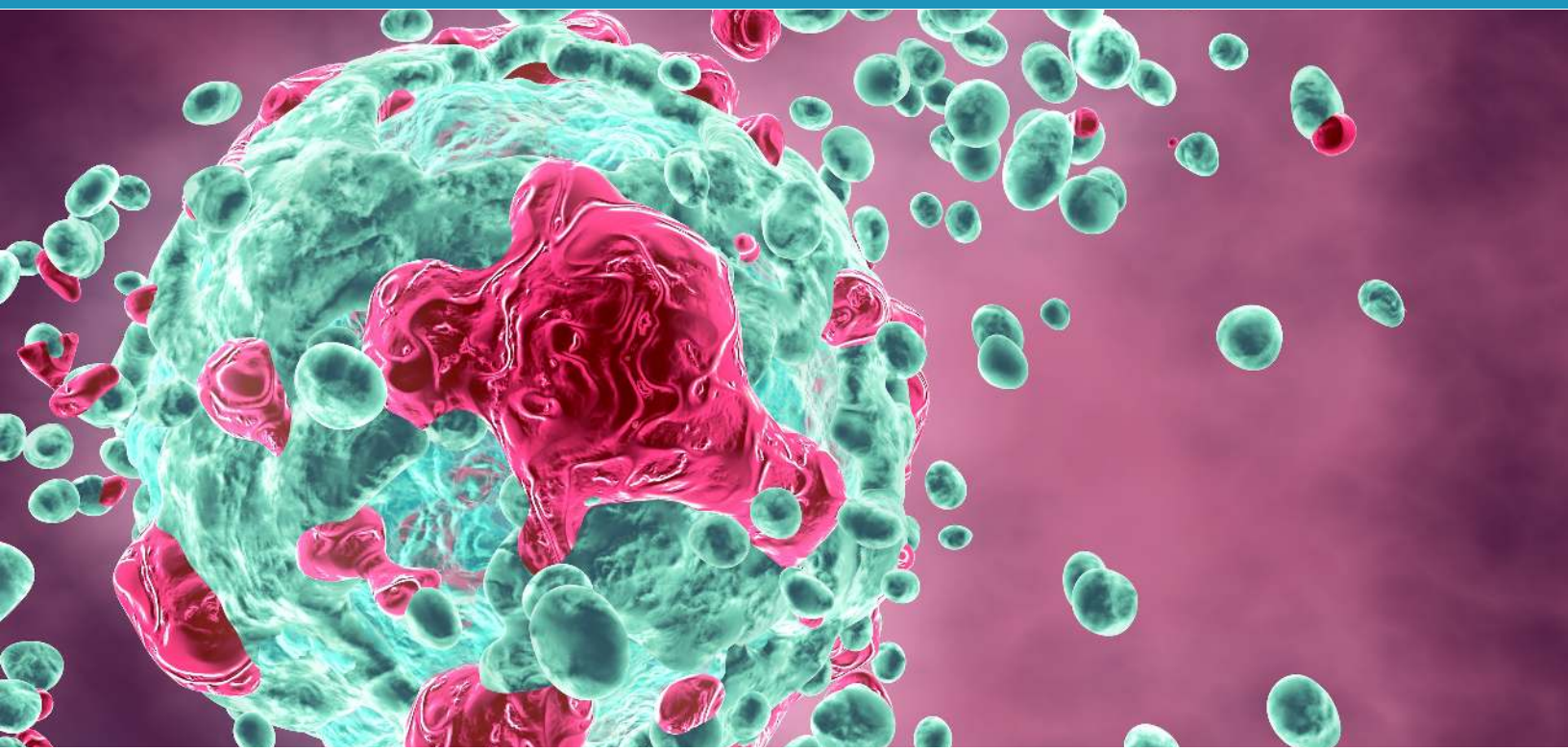


Spectral Library Construction and Search in PEAKS Online

Chaochao Chen, PhD, Application Scientist
Kyle Hoffman, PhD, Applications Manager

Bioinformatics Solutions Inc., Waterloo, Canada

Aim:

DDA files and DIA files were used in this study to examine the performance of PEAKS Online in spectral library construction and searching.

Abstract:

A published dataset [1] was used to illustrate the performance of PEAKS Online in spectral library construction with a large DDA dataset and subsequent DIA data search against the generated spectral library. With the ability to run on multi-CPU/GPU cluster and cloud server, PEAKS Online offers accelerated data analysis, which is beneficial to large-scale projects.

Introduction:

Data-dependent acquisition (DDA) and data-independent acquisition (DIA) methods are two data collection modes in tandem mass spectrometry (MS2) of proteomics. In DDA mode, the mass spectrometer selects a fixed number of the most intense precursor ions from the MS1 scan and fragments those precursor ions for MS2 analysis. On the contrary, in DIA, during MS2 spectrum acquisition, all precursor ions within a selected mass-to-charge (m/z) window are fragmented. Although DDA has been the most used methodology for qualitative and quantitative proteomics, DIA has gained popularity due to its improved detection and reproducibility. However, the direct relationship between a precursor ion and its fragment ions is lost since the fragment ions in DIA MS2 spectra could result from multiple precursor ions.

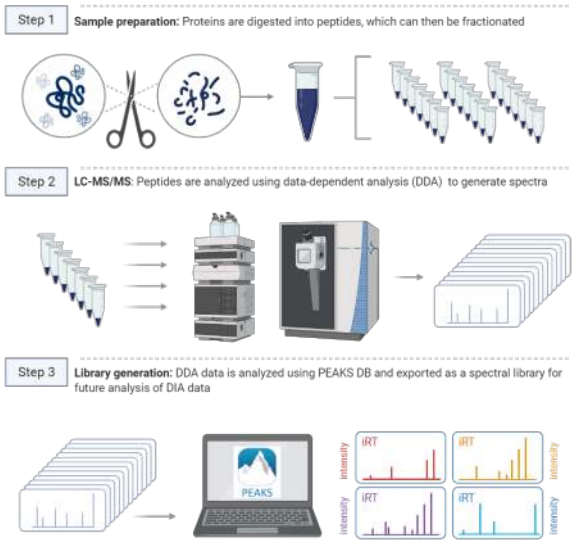


Fig 1. How to Generate a Spectral Library

Because of the complexity of MS2 spectra generated in DIA analysis, a high quality and high coverage spectral library is preferred to aid the interpretation of DIA data. A spectral library is typically acquired from multiple fractionated DDA analysis of the same or same type of sample on the same instrument and searched against a protein sequence database for identification. After computational deconvolution of the multiplexed DIA MS2 spectra, individual fragment ions are assigned to their precursor ions based on the eluting profiles of the fragment ions and precursor ions. Peptides are identified by matching the fragment ion peaks in MS2 spectra to the peptide fragment ion pattern in the spectral library together with other features such as errors in m/z and retention time (RT) [2].

Methods:

A published dataset was downloaded and used to illustrate the construction of a spectral library from 156 DDA files and DIA data search against the created spectral library in PEAKS Online. Details of the dataset are presented below.

Experimental Design:

For the construction of the spectral library, a total of 4 cell lines and 6 pooled tumor tissues were lysed, digested, and fractionated by high pH reversed-phase (HpRP) chromatography using StageTip or HPLC column, and phosphopeptides were enriched by iron-based immobilized metal affinity chromatography (Fe-IMAC) by StageTip and analyzed in DDA mode (Fig. 2a). Synthetic phosphopeptides associated with lung cancer-related driver genes were also included to build the spectral library for enhanced coverage. Single-shot DIA analyses of 2 cell lines were acquired and used to evaluate the spectral library.

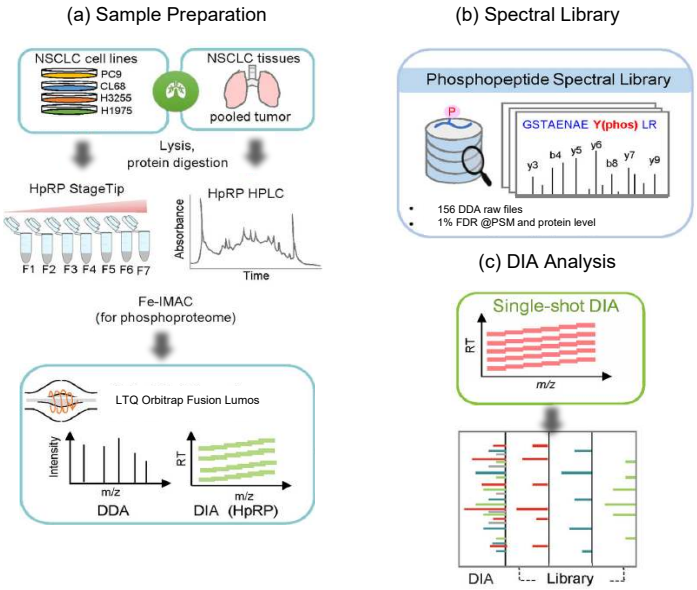


Fig 2. DDA-MS vs. DIA-MS comparison

Data Analysis:

A hundred and fifty-six DDA raw files from fractionated cell line and cancer tissue samples were used to build the spectral library (Fig. 2b). A precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.05 Da were applied. Carbamidomethylation (+57.022 Da) of cysteine residues was set as a fixed modification. Oxidation of methionine (+15.995 Da) residues and acetylation on protein N terminus (+42.016 Da), as well as phosphorylation (+79.966 Da) on Ser, Thr, and Tyr residues were added as variable modifications. Semi-tryptic peptides and a missed cleavage of 2 were included, and a maximum of 3 variable modifications were used. A Database search was performed in PEAKS Online against a Uniprot Human database (Homo sapiens = 26590 entries). A cut-off of 1% FDR was applied at PSM and protein levels. Single-shot DIA data acquired on the same instrument was searched against the spectral library using the same parameters for the library generation (Fig. 2c).

Results and Discussion

Sample Name	# MS Run	# MS1	# MS2	#PSM	#Peptides	# Sequences	# Proteins	# Protein Groups	# PSM / # MS2
All (156 samples)	156	3272073	9298706	3956090	232761	114596	16386	8252	43%
Sample 1	1	21003	46789	16670	8627	6691	2076	1029	36%
Sample 2	1	21146	46632	16286	8353	6524	2056	1023	35%
Sample 3	1	21950	45623	15373	7359	5752	1831	891	34%
Sample 4	1	21811	46134	15887	7191	5610	1771	874	34%
Sample 5	1	22568	44980	15724	7262	5763	1807	883	35%

Fig 3. PEAKS DB Search results on 156 DDA files

Spectral Library Construction Using DDA Dataset

Using lung cancer cell lines and patient-derived tissues, a phosphoproteome spectral library of 184383 phosphopeptides on 8252 protein groups (unique peptide was set to 2) was constructed within 48 hours from 156 DDA files (Fig. 3). In the generated spectral library, each peptide entry contains useful information including its dominant charge states, whether it is unique to a protein group, its m/z, intensity, and RT, and m/z and relative intensity of matched fragment ions.

	Peptide	-10LgP	Mass	Length	ppm	m/z	RT	Delta RT	Scan	#Spec	PTM
1	VLHEAEGHIVTC(+57.02)ETNTGEVYR	51.30	2413.1333	21	-1.0	604.2867	57.38	0.50	44637:PhosphoCL6...	12	C
2	LKS(+79.97)EDELRPEVDEHTQK	51.30	2131.9788	17	2.3	534.0002	44.34	0.66	34226:PhosphoCL6...	12	P
3	VGGGGTAGGDRWEGEDEDEDVKDNWDDDDDEK	51.30	3481.3574	32	0.4	871.3422	71.37	-1.04	55833:PhosphoCL6...	6	
4	SSSDPGIPGGPQAIPATNS(+79.97)PDHSDHTLSVSSDSGHST...	51.30	4063.7781	41	-0.5	813.7580	76.39	-4.24	59856:PhosphoCL6...	6	P
5	NVQQDNSEAGTQPQVQTDAQQTSQSPPS(+79.97)PELTSEEN...	51.30	4716.0947	43	1.1	944.2222	86.77	-0.77	68182:PhosphoCL6...	6	P
6	RVGDPPQPLPEEPM(+15.99)EVQGAERAS(+79.97)PEPQR	51.30	3191.4707	28	-0.8	798.8699	73.47	-2.32	57559:PhosphoCL6...	6	O P
7	RTSDPS(+79.97)AAVNHVSSTTSLGENYEDDDLVNSDEVM(+1...	51.30	3877.6472	35	0.2	970.4140	91.36	-1.92	71856:PhosphoCL6...	6	O P
8	TFLESKEELSHS(+79.97)PEPC(+57.02)TK	51.30	2197.9602	18	0.1	733.6567	59.73	0.27	46486:PhosphoCL6...	6	C P
9	RRS(+79.97)LEPAENVHGAGGGAFPASQTPSK	51.30	2700.2769	26	-0.5	676.0725	52.10	0.29	40411:PhosphoCL6...	7	P
10	LEDTAGDTGHSSLEAPRS(+79.97)PDTLAPVASER	51.30	3058.3879	29	-0.3	765.5999	80.09	-2.31	62809:PhosphoCL6...	6	P

Fig 4. Spectral library search peptide view results

Spectral Library Search with DIA Data

The spectral library search result includes a list of peptides accurately matched from the DIA MS2 spectra to the library spectra as evidence of the identifications, as shown in Fig. 4.

For each identified peptide, the extracted ion chromatographs from the precursor ions and top fragment ions can be visualized (Fig. 5). In addition, the alignment of fragment ion peaks in DIA MS2 spectra to the peptide fragment ion pattern in the spectral library and ion match table (Fig. 6) are provided for further assessment.

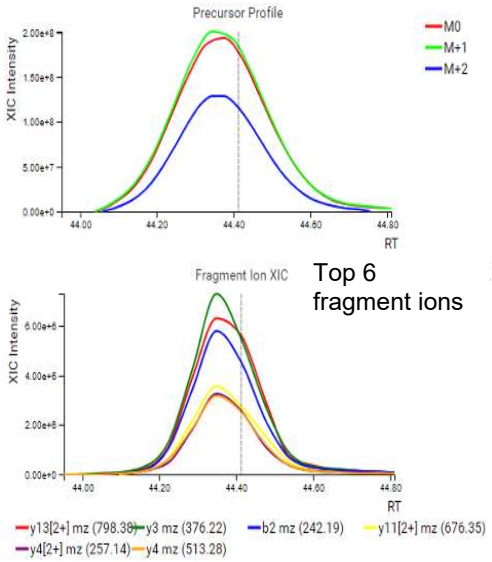
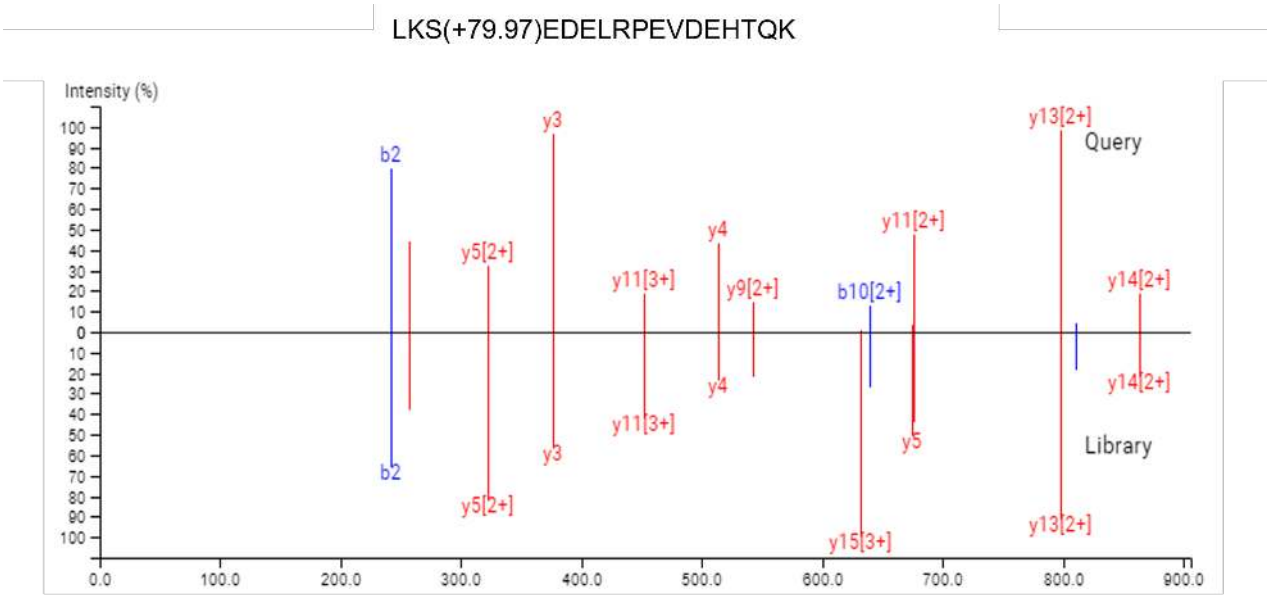


Fig 5. Extracted ion chromatographs from precursor ions and fragment ions



#	Label	Library M/Z	Library Intensity(%)	Query M/Z	Query Intensity
1	b2	242.1868	66.2	242.1854	4.56e+6(81.0%)
2	y4[2+]	257.1390	38.3	257.1416	2.57e+6(45.7%)
3	y5[2+]	321.6603	82.0	321.6626	1.89e+6(33.7%)
4	y3	376.2191	56.7	376.2174	5.52e+6(98.0%)
5	y11[3+]	451.2321	42.3	451.2344	1.12e+6(19.9%)
6	y4	513.2780	23.5	513.2762	2.52e+6(44.7%)
7	y9[2+]	541.7556	21.7	541.7571	9.27e+5(16.5%)
8	y15[3+]	631.2690	100.0	631.2713	1.31e+5(2.3%)
9	b10[2+]	639.2889	27.0	639.2906	8.29e+5(14.7%)
10	y5	673.9673	50.9	673.9698	2.67e+5(4.7%)
11	y11[2+]	676.3482	43.9	676.3500	2.74e+6(48.8%)
12	y13[2+]	798.3829	91.3	798.3839	5.63e+6(100.0%)
13	b13[2+]	810.8578	18.5	810.8586	3.31e+5(5.9%)
14	y14[2+]	862.9042	21.9	862.9055	1.14e+6(20.3%)

Fig 6. Spectral library mirror and ion match table

Sample Name	# MS Run	# MS1	# MS2	#PSM	#Peptides	# Sequences	# PSM / # MS2
All (6 samples)	6	14823	741150	422603	87972	49380	57%
Sample 1	1	2469	123450	69863	62463	35302	57%
Sample 2	1	2473	123650	70128	62780	35586	57%
Sample 3	1	2471	123550	69936	62618	35420	57%
Sample 4	1	2471	123550	70928	63779	36458	57%
Sample 5	1	2470	123500	71032	63812	36570	58%
Sample 6	1	2469	123450	70716	63675	36268	57%

Fig 7. PEAKS LIB Search results on 6 DIA files

Conclusion:

PEAKS Online allows efficient handling of large-scale proteomics projects. By searching against the spectral library generated from a large DDA dataset in PEAKS Online, the interpretation of complex DIA data can be achieved (Fig. 7).

References:

1. Kitata, R.B., Choong, W.K., Tsai, C.F., Lin, P.Y., Chen, B.S., Chang, Y.C., Nesvizhskii, A.I., Sung, T.Y. and Chen, Y.J., 2021. A data-independent acquisition-based global phosphoproteomics system enables deep profiling. Nature communications, 12(1), pp.1-14.
2. Li, K.W., Gonzalez-Lozano, M.A., Koopmans, F. and Smit, A.B., 2020. Recent developments in data independent acquisition (DIA) mass spectrometry: application of quantitative analysis of the brain proteome. Frontiers in molecular neuroscience, p.248.