Poster #23B

A novel Data-Independent-Acquisition (DIA) approach to *de novo* antibody Sequencing

Jonathan R Krieger, Jun Ma, Baozhen Shan
Bioinformatics Solutions Inc., Waterloo, Ontario, Canada

PEAKS AB Protein *de novo* Sequencing Service

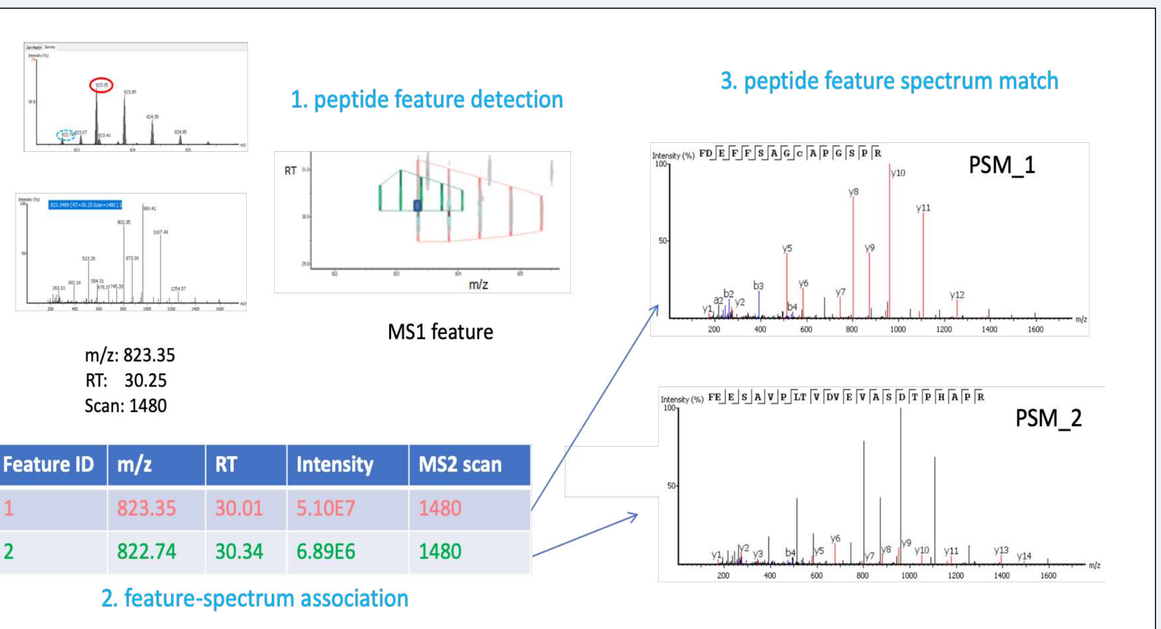
Abstract

The discovery of biological therapeutics, specifically antibodies, in treating a multitude of diseases has led to an explosion in the research and development of these pharmaceuticals. This rapid expansion has also fueled the demand for *de novo* antibody sequencing. The need to extensively characterize monoclonal antibodies at the molecular level presents a unique challenge to drug developers and manufactures. We have developed proprietary *de novo* antibody protein sequence technology and software to solve this problem. Our PEAKS AB software and service are able to **perform end to end sequencing on a monoclonal antibody of any isotype without any genetic or sequence information.**

While traditional antibody *de novo* sequencing is performed on mass spectrometry data generated from spectra collected in data-dependent (DDA) mode, we have recently shown that acquiring data in data-independent mode (DIA) can greatly assist in the accuracy and confidence of antibody sequences, and provides a better overall **complete antibody characterization**. The use of DIA also sheds the inherent abundance bias that plagues DDA based analysis, and thus, the presence of contaminating proteins and stabilizers does not have as great an effect on the ability to sequence the antibody, and thus also allows for accurate host cell protein analysis from the same sample. Furthermore, the DIA analysis of antibodies allows for the unbiased accurate identification and quantification of sequence variants.

Finally, and perhaps the largest potential with our new method, is the ability to sequence antibody mixtures. Using our DIA approach, all potential peptides are captured and fragmented by the mass spectrometer. Our software algorithm is able to accurately and efficiently assemble *de novo* peptides into complete antibody sequences. **Herein, we demonstrate our novel approach to *de novo* antibody sequencing and complete antibody and host cell protein characterization using standard antibodies**.

PEAKS uses Feature Based Peptide Identification



Feature ID	m/z	RT	Intensity	MS2 scan
1	823.35	30.01	5.10E7	1480
2	822.74	30.34	6.89E6	1480

Figure 1 - PEAKS identification method into its unique *de novo* assisted identification workflow.

- MS1 scans are extracted and pre-processed, from which the peptide features and their elution profiles are detected.
- According to the isolation window and the retention time, the relevant features are associated with the MS2 spectra.
- Therefore, each MS2 can be associated with multiple features and they are all *de novo* sequenced, allowing for deeper protein/antibody coverage, the ability to detect low level variants, and an increase in PSMs

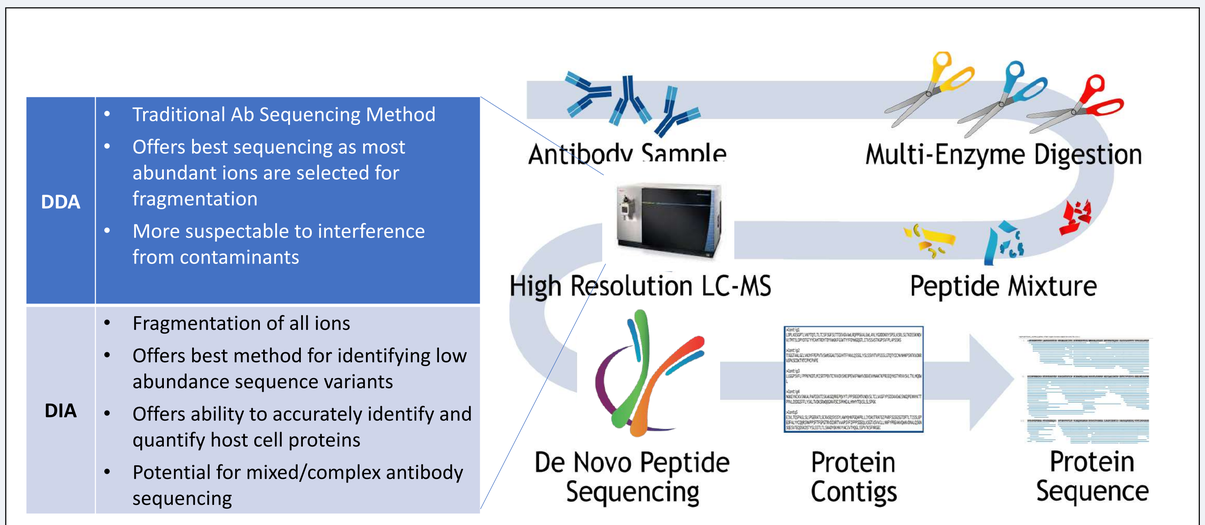


Figure 2 – Overview of our antibody sequencing workflow and characterization. Typically we take a multi-enzyme digest and perform high resolution LC-MS/MS in DDA and DIA mode to generate spectra for downstream *de novo* sequencing.

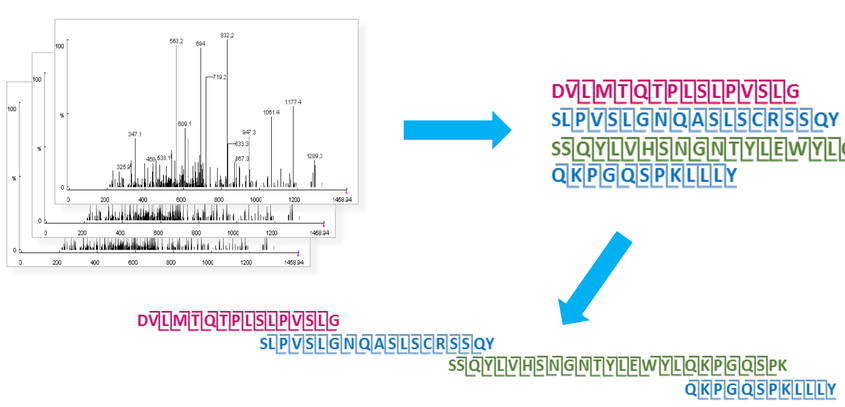


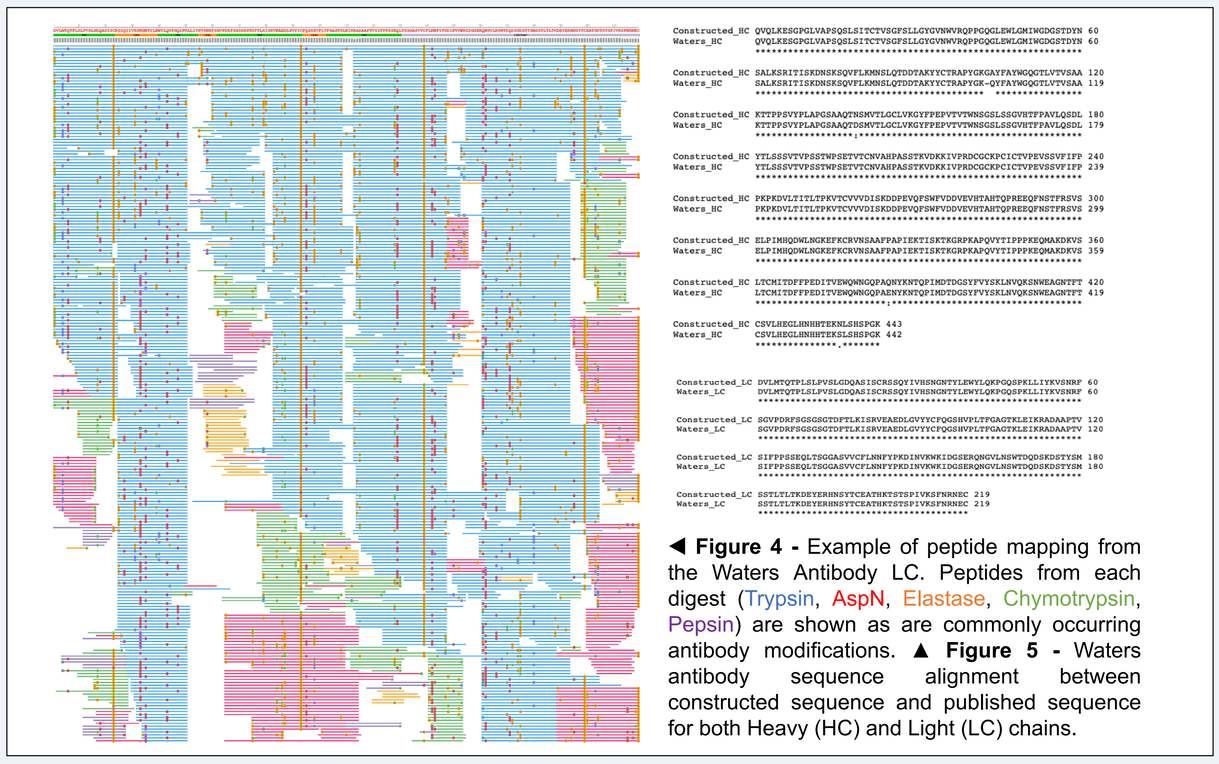
Figure 3 - Our PEAKS AB software uses an algorithm to assemble *de novo* sequencing tags into full sequences of heavy and light chains of antibodies. These are de novo sequencing results obtained from MS2 spectra. The overlaps are detected between de novo sequencing tags with high confidence, and based on the overlapping information, de novo sequencing tags are assembled to a full sequence[1].

NIST and Waters Standard Antibody Data

Table 1: Percentage of sequence identified using DIA alone or DIA with 2 DDA runs relative to known sequence. Manual verification is standard with any antibody sequencing project, regardless of data acquisition.

		Waters Reference Ab	NIST Reference Ab
DIA (5 enzymes)	HC Fully Automated	90.87%	90.77%
	HC Automated + Manual Verification	99%	99%
	LC Fully Automated	100%	100%
	LC Automated + Manual Verification	100%	100%
DIA (5 enzymes) + DDA (2 enzymes)	HC Fully Automated	98.19%	97.56%
	HC Automated + Manual Verification	100%	100%
	LC Fully Automated	100%	100%
	LC Automated + Manual Verification	100%	100%

- Light chain sequences of both antibodies can be completely sequenced using our software
- Including 2 digests in DDA acquisition method improves fully automated HC sequencing by approximately 7%
- In all cases, 100% of antibody sequence could be obtained
- Sequencing was performed in a blinded manner such that standard antibody was revealed after sequencing was complete



◀ **Figure 4 -** Example of peptide mapping from the Waters Antibody LC. Peptides from each digest (Trypsin, AspN, Elastase, Chymotrypsin, Pepsin) are shown as are commonly occurring antibody modifications. ▲ **Figure 5 -** Waters antibody sequence alignment between constructed sequence and published sequence for both Heavy (HC) and Light (LC) chains.

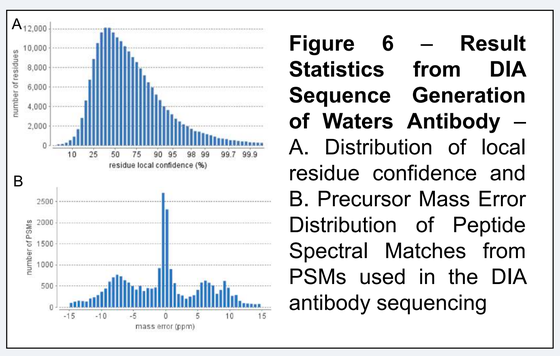


Figure 6 – Result Statistics from DIA Sequence Generation of Waters Antibody – A. Distribution of local residue confidence and B. Precursor Mass Error Distribution of Peptide Spectral Matches from PSMs used in the DIA antibody sequencing

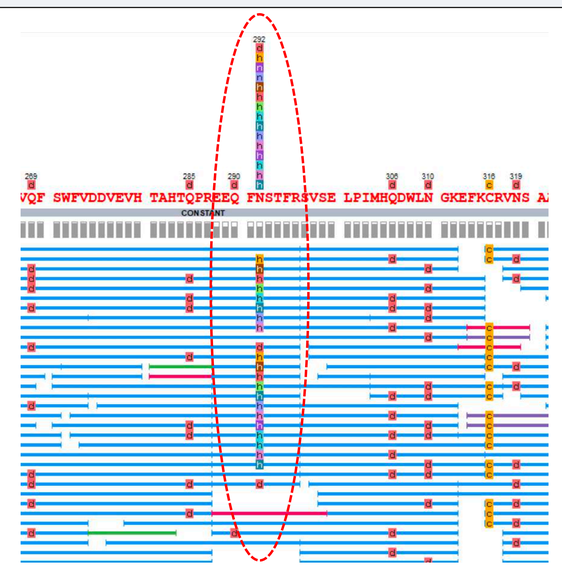
Sample	# MS2		
	All	ALC (%)	
		> 30	> 50
~~Sample~~ 1	33622	14936	4601
~~Sample~~ 2	38068	19318	5008
~~Sample~~ 3 Chymotrypsin	40488	18814	4590
~~Sample~~ 4 Pepsin	24558	8744	2124
~~Sample~~ 5 Trypsin	27624	19787	12318
~~Sample~~ 6 Trypsin – 1hr*	22250	8634	5919
~~Sample~~ 7 Trypsin – ON*	22597	9111	6431
Total	**209207**	**99344**	**40991**

Sample	# Peptides/PSMs	
	Contaminants	Target Proteins
AspN	75/76	1746/2818
Elastase	34/38	903/1246
Chymotrypsin	33/33	1267/1702
Pepsin	11/11	481/730
Trypsin	77/83	1461/2536
Trypsin – 1hr*	7/10	1485/6667
Trypsin – ON*	62/195	1254/7500
Total	**299/446**	**8597/23199**

Figure 7 – Distribution of Average Local Confidence (ALC) scores from each individual digest as well as database search results for contaminant/enzyme proteins and target proteins in the dataset.

1460.53	Neutral Loss->(HexNAc)4(Hex)4	
1647.61	Neutral Loss->(HexNAc)5(Hex)3(Fuc)1	
1444.53	(HexNAc)4(Hex)3(Fuc)1	
1606.59	(HexNAc)4(Hex)4(Fuc)1	
1768.64	(HexNAc)4(Hex)5(Fuc)1	
1298.48	(HexNAc)4(Hex)3	
1241.45	(HexNAc)3(Hex)3(Fuc)1	
2075.73	(HexNAc)4(Hex)5(Fuc)1(307)1	
1930.69	(HexNAc)4(Hex)6(Fuc)1	
1460.53	(HexNAc)4(Hex)4	
2092.74	(HexNAc)4(Hex)7(Fuc)1	
1403.51	(HexNAc)3(Hex)4(Fuc)1	
1095.40	(HexNAc)3(Hex)3	
1913.68	(HexNAc)4(Hex)4(Fuc)1(307)1	

Figure 8 – Glycan identification on consensus glycosylation site of Waters Antibody. Using the DIA approach, we observe an increase in the number of glycan sites identified, and a 3 fold increase in the reported glycosylation of the antibody. All glycan identification came from the trypsin digestion fraction.



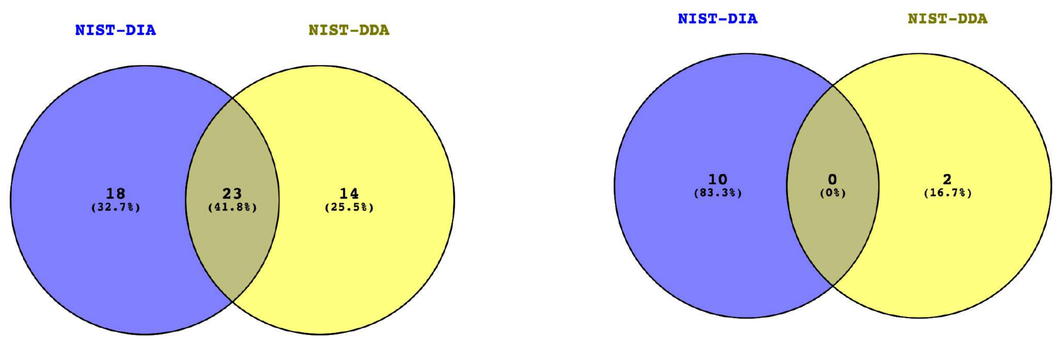


Figure 9 – Venn Diagrams showing identified mouse host cell proteins from NIST antibody. A. Identified proteins by DIA or DDA Direct Database Search in PEAKS X+. IgG proteins have been removed. **B.** Identified proteins after IgG proteins and Contaminants have been removed. DIA data is able to identify more host cell contaminant proteins due to the nature of the data acquisition.

Contact Us

Bioinformatics Solutions Inc.
http://www.bioinfor.com
peaks@bioinfor.com
1-855-885-8288

References

[1]Complete *De Novo* Assembly of Monoclonal Antibody Sequences. *Sci Rep.* 2016 Aug 26;6:31730

Tran NH, Qiao R, Xin L, Chen X, Liu C, Zhang X, Shan B, Ghodsi A, Li M. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. Nature Methods. 16(1), 63-66. 20/12/2018
Tran NH, Zhang X, Xin L, Shan B, Li M.

De novo peptide sequencing by deep learning. Proceedings of the National Academy of Sciences of the United States of America. 114(29). 18/7/2017
Tran NH, Rahman MZ, He L, Xin L, Shan B, Li M.

Complete De Novo Assembly of Monoclonal Antibody Sequences Scientific Reports. 6(31730). 26/08/2016.

Conclusions, Applications, & Future Directions

- Complete *de novo* antibody sequencing can be achieved using PEAKS X+ and PEAKS AB software
- A combination of DIA runs with 2 Trypsin DDA runs achieves fully automated 100% sequence coverage
- DIA allows for a better characterization of the glycans attached to the antibodies
- DIA analysis allows for the deeper characterization of host cell proteins and full antibody characterization
- This method will be a useful method to employ in our antibody sequencing service, and will be beneficial for assessing, characterizing, and sequencing antibody mixtures and polyclonal antibodies.

Let BSI handle all your antibody characterization needs. Please contact us or stop by our booth for a demonstration, example report, and more information!