Project Metrics Application Note

Release 1.0

Monika Bhatt

Feb 11, 2018

CONTENTS

SEQUENCE VARIANT ANALYSIS (SVA) PART I: BYONIC™ SEARCHES

Marshall Bern, Ph.D., K. Ilker Sen, Ph.D., Eric Carlson, Ph.D., and Chris Becker, Ph.D. Protein Metrics Inc. www.proteinmetrics.com June 2016

1.1 What's New in this version

1.1.1 Use of isotropically labeled acid as alkalyting agents

This version illustrates a new method of searching for sequence variants. Refer to the *Discussion regarding the use of isotopically labeled iodoacetic acid or iodoacetimide as alkylating agents* section for details.

1.2 Benefits summary

- Best practices for running a sequence variant search

- How-to for using Byonic™ software to effectively search for sequence variants, for subsequent inspection and quantification by Byologic® software.

1.3 Method

Almost all therapeutic protein data sets include mass spectra from peptides that differ from the corresponding database peptide by one amino acid substitution at some concentration. Sequence variants may arise from mutations or mistranslations in the host cell line or problems with the bioreactor feed. Searching for sequence variants is especially challenging, because there are a large number of possible amino acid substitutions, and many of the mass deltas are either exactly or approximately the same as the mass deltas from posttranslational modifications and sample preparation artifacts. In this application note, we show how to search for sequence variants in biotherapeutics. Sequence variant analysis aims to identify and quantify all the variant peptides down to very small percentages of their "wild type" (native or expected sequence) counterparts. Because the protein database is very small for a therapeutic protein, typically only the product plus trypsin or other digestive enzyme and decoys, Byonic running time is not a concern. It is feasible to search for all possible one-AA-substitution peptides in a few minutes. The challenge is distinguishing the relatively rare true sequence variants from false positives. This Application Note focuses on the first step, the Byonic search for SVA.

See also:

My internal notes (not to be shown to customer, but only to internal team). The above sentence describes issue x and y, but we haven't covered z yet. We plan on doing this with JIRA ticket ABC-1234.

Table 1.1: Summary of best practice steps for sequence variant Byonic searches

Step	Detail
Choice of alkylating agent	Many amino acid substitutions have the exact same mass difference as the common alkylating agents. We recommend choosing one based on the desired analysis: • Ideally, use C13 labeled Iodoacetic acid. • Iodoacetic acid tends to generate fewer over-alkylated peptides than iodoacetamide, and has fewer delta mass overlaps with all possible amino acid substitutions list. • If, however, only single base substitutions that lead to sequence variants will be considered, rather than the entire list, carbamidomethyl may be the desired agent as there are no conflicting masses within this shorter list.
Use a high resolution mass spectrometer	Use high resolution MS1. Ideally, also use high resolution MS2. Make sure the instrument is well calibrated and operating near peak performance.
MSMS Search: Modification rules settings.	Identify artifacts first with Protein Metrics' Preview™ or Byonic Wildcard Search™. • Do not include deamidation in the list of variants (Asn->Asp). Instead include them as named rare1 modifications. Include all the types of non-variant modifications known to be present in your sample, including all the sample preparation artifacts particular to your laboratory. Use rare1 for all modifications and sequence variants except common1 for pyro-glu modifications. • Use common max 1 and rare max 1. • See example in Table 2 for a typical modification list. Have the sequence variants in the search as rare1 modifications (see Figure 1 below for the full list, and an exemplary short list in Table 3). • A list of modifications for sequence variants is available from Protein Metrics. This list can be copied from a text editor and pasted into the Modification pop-up box in Byonic. • In addition, for mAbs with N-glycosylation, include the table of 50 N-glycans on the Glycan tab, and check Show all N-glycopeptides on the Advanced tab. Use common1 setting.

1.3.1 Discussion regarding the use of isotopically labeled iodoacetic acid or iodoacetimide as alkylating agents

We suggest using iodoacetic acid for alkylation of Cysteines because this typically produces fewer sample preparation artifacts compared to iodoacetimide. In either case however, there are a significant number of confusable masses of sequence variants with alkylation artifacts. Table 4 below lists confusable masses for unlabeled and C13-labeled alkylations and sequence variants. Examination of the table shows that for commercially available labeled iodoacetic acid and iodoacetimide, that the single-C13-labeled iodoacetic acid will give the best separation and hence the fewest (or none) confusion with sequence variants. Note that if a three-C13-labeled iodoacetic acid is available, the +61 modification has no confusable masses with sequence variants.

% single-letter substitutions for A	% single-letter substitutions for I	% single-letter substitutions for R
Ala->Cys / 31.97207+58.005479 @ A | rare1	Ile->Ala / -42.04695 @ I | rare1	Arg->Ala / -85.06400 @ R | rare1
Ala->Asp / 43.98983 @ A | rare1	Ile->Cys / -10.07488+58.005479 @ I | rare1	Arg->Cys / -53.09193+58.005479 @ R | rare1
Ala->Glu / 58.00548 @ A | rare1	Ile->Asp / 1.94288 @ I | rare1	Arg->Asp / -41.07417 @ R | rare1
Ala->Phe / 76.03130 @ A | rare1	Ile->Glu / 15.95853 @ I | rare1	Arg->Glu / -27.05852 @ R | rare1
Ala->Gly / -14.01565 @ A | rare1	Ile->Phe / 33.98435 @ I | rare1	Arg->Phe / -9.03270 @ R | rare1
Ala->His / 66.02180 @ A | rare1	Ile->Gly / -56.06260 @ I | rare1	Arg->Gly / -99.07965 @ R | rare1
Ala->Lys / 57.05785 @ A | rare1	Ile->His / 23.97485 @ I | rare1	Arg->His / -19.04220 @ R | rare1
Ala->Leu / 42.04695 @ A | rare1	Ile->Lys / 15.01090 @ I | rare1	Arg->Lys / -28.00615 @ R | rare1
Ala->Met / 60.00337 @ A | rare1	Ile->Met / 17.95642 @ I | rare1	Arg->Leu / -43.01705 @ R | rare1
Ala->Asn / 43.00582 @ A | rare1	Ile->Asn / 0.95887 @ I | rare1	Arg->Met / -25.06063 @ R | rare1
Ala->Pro / 26.01565 @ A | rare1	Ile->Pro / -16.03130 @ I | rare1	Arg->Asn / -42.05818 @ R | rare1
Ala->Gln / 57.02147 @ A | rare1	Ile->Gln / 14.97452 @ I | rare1	Arg->Pro / -59.04835 @ R | rare1
Ala->Arg / 85.06400 @ A | rare1	Ile->Arg / 43.01705 @ I | rare1	Arg->Gln / -28.04253 @ R | rare1
Ala->Ser / 15.99492 @ A | rare1	Ile->Ser / -26.05203 @ I | rare1	Arg->Ser / -69.06908 @ R | rare1
Ala->Thr / 30.01057 @ A | rare1	Ile->Thr / -12.03638 @ I | rare1	Arg->Thr / -55.05343 @ R | rare1
Ala->Val / 28.03130 @ A | rare1	Ile->Val / -14.01565 @ I | rare1	Arg->Val / -57.03270 @ R | rare1
Ala->Trp / 115.04220 @ A | rare1	Ile->Trp / 72.99525 @ I | rare1	Arg->Trp / 29.97820 @ R | rare1
Ala->Tyr / 92.02622 @ A | rare1	Ile->Tyr / 49.97927 @ I | rare1	Arg->Tyr / 6.96222 @ R | rare1

% single-letter substitutions for C	% single-letter substitutions for K	% single-letter substitutions for S
Cys->Ala / -31.97207-58.005479 @ C | rare1	Lys->Ala / -57.05785 @ K | rare1	Ser->Ala / -15.99492 @ S | rare1
Cys->Asp / 12.01776-58.005479 @ C | rare1	Lys->Cys / -25.08578+58.005479 @ K | rare1	Ser->Cys / 15.97715+58.005479 @ S | rare1
Cys->Glu / 26.03341-58.005479 @ C | rare1	Lys->Asp / -13.06802 @ K | rare1	Ser->Asp / 27.99491 @ S | rare1
Cys->Phe / 44.05923-58.005479 @ C | rare1	Lys->Glu / 0.94763 @ K | rare1	Ser->Glu / 42.01056 @ S | rare1
Cys->Gly / -45.98772-58.005479 @ C | rare1	Lys->Phe / 18.97345 @ K | rare1	Ser->Phe / 60.03638 @ S | rare1
Cys->His / 34.04973-58.005479 @ C | rare1	Lys->Gly / -71.07350 @ K | rare1	Ser->Gly / -30.01057 @ S | rare1
Cys->Lys / 25.08578-58.005479 @ C | rare1	Lys->His / 8.96395 @ K | rare1	Ser->His / 50.02688 @ S | rare1
Cys->Leu / 10.07488-58.005479 @ C | rare1	Lys->Leu / -15.01090 @ K | rare1	Ser->Lys / 41.06293 @ S | rare1
Cys->Met / 28.03130-58.005479 @ C | rare1	Lys->Met / 2.94552 @ K | rare1	Ser->Leu / 26.05203 @ S | rare1
Cys->Asn / 11.03375-58.005479 @ C | rare1	Lys->Asn / -14.05203 @ K | rare1	Ser->Met / 44.00845 @ S | rare1
Cys->Pro / -5.95642 -58.005479@ C | rare1	Lys->Pro / -31.04220 @ K | rare1	Ser->Asn / 27.01090 @ S | rare1
Cys->Gln / 25.04940 -58.005479@ C | rare1	Lys->Gln / -0.03638 @ K | rare1	Ser->Pro / 10.02073 @ S | rare1
Cys->Arg / 53.09193 -58.005479@ C | rare1	Lys->Arg / 28.00615 @ K | rare1	Ser->Gln / 41.02655 @ S | rare1
Cys->Ser / -15.97715-58.005479 @ C | rare1	Lys->Ser / -41.06293 @ K | rare1	Ser->Arg / 69.06908 @ S | rare1
Cys->Thr / -1.96150-58.005479 @ C | rare1	Lys->Thr / -27.04728 @ K | rare1	Ser->Thr / 14.01565 @ S | rare1
Cys->Val / -3.94077 -58.005479@ C | rare1	Lys->Val / -29.02655 @ K | rare1	Ser->Val / 12.03638 @ S | rare1
Cys->Trp / 83.07013-58.005479 @ C | rare1	Lys->Trp / 57.98435 @ K | rare1	Ser->Trp / 99.04728 @ S | rare1
Cys->Tyr / 60.05415-58.005479 @ C | rare1	Lys->Tyr / 34.96837 @ K | rare1	Ser->Tyr / 76.03130 @ S | rare1

% single-letter substitutions for D	% single-letter substitutions for L	% single-letter substitutions for T
Asp->Ala / -43.98983 @ D | rare1	Leu->Ala / -42.04695 @ L | rare1	Thr->Ala / -30.01057 @ T | rare1
Asp->Cys / -12.01776 +58.005479@ D | rare1	Leu->Cys / -10.07488+58.005479 @ L | rare1	Thr->Cys / 1.96150+58.005479 @ T | rare1
Asp->Glu / 14.01565 @ D | rare1	Leu->Asp / 1.94288 @ L | rare1	Thr->Asp / 13.97926 @ T | rare1
Asp->Phe / 32.04147 @ D | rare1	Leu->Glu / 15.95853 @ L | rare1	Thr->Glu / 27.99491 @ T | rare1
Asp->Gly / -58.00548 @ D | rare1	Leu->Phe / 33.98435 @ L | rare1	Thr->Phe / 46.02073 @ T | rare1
Asp->His / 22.03197 @ D | rare1	Leu->Gly / -56.06260 @ L | rare1	Thr->Gly / -44.02622 @ T | rare1
Asp->Lys / 13.06802 @ D | rare1	Leu->His / 23.97485 @ L | rare1	Thr->His / 36.01123 @ T | rare1
Asp->Leu / -1.94288 @ D | rare1	Leu->Lys / 15.01090 @ L | rare1	Thr->Lys / 27.04728 @ T | rare1
Asp->Met / 16.01354 @ D | rare1	Leu->Met / 17.95642 @ L | rare1	Thr->Leu / 12.03638 @ T | rare1
Asp->Asn / -0.98401 @ D | rare1	Leu->Asn / 0.95887 @ L | rare1	Thr->Met / 29.99280 @ T | rare1
Asp->Pro / -17.97418 @ D | rare1	Leu->Pro / -16.03130 @ L | rare1	Thr->Asn / 12.99525 @ T | rare1
Asp->Gln / 13.03164 @ D | rare1	Leu->Gln / 14.97452 @ L | rare1	Thr->Pro / -3.99492 @ T | rare1
Asp->Arg / 41.07417 @ D | rare1	Leu->Arg / 43.01705 @ L | rare1	Thr->Gln / 27.01090 @ T | rare1
Asp->Ser / -27.99491 @ D | rare1	Leu->Ser / -26.05203 @ L | rare1	Thr->Arg / 55.05343 @ T | rare1
Asp->Thr / -13.97926 @ D | rare1	Leu->Thr / -12.03638 @ L | rare1	Thr->Ser / -14.01565 @ T | rare1
Asp->Val / -15.95853 @ D | rare1	Leu->Val / -14.01565 @ L | rare1	Thr->Val / -1.97927 @ T | rare1
Asp->Trp / 71.05237 @ D | rare1	Leu->Trp / 72.99525 @ L | rare1	Thr->Trp / 85.03163 @ T | rare1
Asp->Tyr / 48.03639 @ D | rare1	Leu->Tyr / 49.97927 @ L | rare1	Thr->Tyr / 62.01565 @ T | rare1

% single-letter substitutions for A	% single-letter substitutions for I	% single-letter substitutions for R
Ala->Cys / 31.97207+58.005479 @ A | rare1	Ile->Ala / -42.04695 @ I | rare1	Arg->Ala / -85.06400 @ R | rare1
Ala->Asp / 43.98983 @ A | rare1	Ile->Cys / -10.07488+58.005479 @ I | rare1	Arg->Cys / -53.09193+58.005479 @ R | rare1
Ala->Glu / 58.00548 @ A | rare1	Ile->Asp / 1.94288 @ I | rare1	Arg->Asp / -41.07417 @ R | rare1
Ala->Phe / 76.03130 @ A | rare1	Ile->Glu / 15.95853 @ I | rare1	Arg->Glu / -27.05852 @ R | rare1
Ala->Gly / -14.01565 @ A | rare1	Ile->Phe / 33.98435 @ I | rare1	Arg->Phe / -9.03270 @ R | rare1
Ala->His / 66.02180 @ A | rare1	Ile->Gly / -56.06260 @ I | rare1	Arg->Gly / -99.07965 @ R | rare1
Ala->Lys / 57.05785 @ A | rare1	Ile->His / 23.97485 @ I | rare1	Arg->His / -19.04220 @ R | rare1
Ala->Leu / 42.04695 @ A | rare1	Ile->Lys / 15.01090 @ I | rare1	Arg->Lys / -28.00615 @ R | rare1
Ala->Met / 60.00337 @ A | rare1	Ile->Met / 17.95642 @ I | rare1	Arg->Leu / -43.01705 @ R | rare1
Ala->Asn / 43.00582 @ A | rare1	Ile->Asn / 0.95887 @ I | rare1	Arg->Met / -25.06063 @ R | rare1
Ala->Pro / 26.01565 @ A | rare1	Ile->Pro / -16.03130 @ I | rare1	Arg->Asn / -42.05818 @ R | rare1
Ala->Gln / 57.02147 @ A | rare1	Ile->Gln / 14.97452 @ I | rare1	Arg->Pro / -59.04835 @ R | rare1
Ala->Arg / 85.06400 @ A | rare1	Ile->Arg / 43.01705 @ I | rare1	Arg->Gln / -28.04253 @ R | rare1
Ala->Ser / 15.99492 @ A | rare1	Ile->Ser / -26.05203 @ I | rare1	Arg->Ser / -69.06908 @ R | rare1
Ala->Thr / 30.01057 @ A | rare1	Ile->Thr / -12.03638 @ I | rare1	Arg->Thr / -55.05343 @ R | rare1
Ala->Val / 28.03130 @ A | rare1	Ile->Val / -14.01565 @ I | rare1	Arg->Val / -57.03270 @ R | rare1
Ala->Trp / 115.04220 @ A | rare1	Ile->Trp / 72.99525 @ I | rare1	Arg->Trp / 29.97820 @ R | rare1
Ala->Tyr / 92.02622 @ A | rare1	Ile->Tyr / 49.97927 @ I | rare1	Arg->Tyr / 6.96222 @ R | rare1
% single-letter substitutions for C	% single-letter substitutions for K	% single-letter substitutions for S
Cys->Ala / -31.97207-58.005479 @ C | rare1	Lys->Ala / -57.05785 @ K | rare1	Ser->Ala / -15.99492 @ S | rare1
Cys->Asp / 12.01776-58.005479 @ C | rare1	Lys->Cys / -25.08578+58.005479 @ K | rare1	Ser->Cys / 15.97715+58.005479 @ S | rare1
Cys->Glu / 26.03341-58.005479 @ C | rare1	Lys->Asp / -13.06802 @ K | rare1	Ser->Asp / 27.99491 @ S | rare1
Cys->Phe / 44.05923-58.005479 @ C | rare1	Lys->Glu / 0.94763 @ K | rare1	Ser->Glu / 42.01056 @ S | rare1
Cys->Gly / -45.98772-58.005479 @ C | rare1	Lys->Phe / 18.97345 @ K | rare1	Ser->Phe / 60.03638 @ S | rare1
Cys->His / 34.04973-58.005479 @ C | rare1	Lys->Gly / -71.07350 @ K | rare1	Ser->Gly / -30.01057 @ S | rare1
Cys->Lys / 25.08578-58.005479 @ C | rare1	Lys->His / 8.96395 @ K | rare1	Ser->His / 50.02688 @ S | rare1
Cys->Leu / 10.07488-58.005479 @ C | rare1	Lys->Leu / -15.01090 @ K | rare1	Ser->Lys / 41.06293 @ S | rare1
Cys->Met / 28.03130-58.005479 @ C | rare1	Lys->Met / 2.94552 @ K | rare1	Ser->Leu / 26.05203 @ S | rare1
Cys->Asn / 11.03375-58.005479 @ C | rare1	Lys->Asn / -14.05203 @ K | rare1	Ser->Met / 44.00845 @ S | rare1
Cys->Pro / -5.95642 -58.005479@ C | rare1	Lys->Pro / -31.04220 @ K | rare1	Ser->Asn / 27.01090 @ S | rare1
Cys->Gln / 25.04940 -58.005479 @ C | rare1	Lys->Gln / -0.03638 @ K | rare1	Ser->Pro / 10.02073 @ S | rare1
Cys->Arg / 53.09193 -58.005479 @ C | rare1	Lys->Arg / 28.00615 @ K | rare1	Ser->Gln / 41.02655 @ S | rare1
Cys->Ser / -15.97715-58.005479 @ C | rare1	Lys->Ser / -41.06293 @ K | rare1	Ser->Arg / 69.06908 @ S | rare1
Cys->Thr / -1.96150-58.005479 @ C | rare1	Lys->Thr / -27.04728 @ K | rare1	Ser->Thr / 14.01565 @ S | rare1
Cys->Val / -3.94077 -58.005479 @ C | rare1	Lys->Val / -29.02655 @ K | rare1	Ser->Val / 12.03638 @ S | rare1
Cys->Trp / 83.07013-58.005479 @ C | rare1	Lys->Trp / 57.98435 @ K | rare1	Ser->Trp / 99.04728 @ S | rare1
Cys->Tyr / 60.05415-58.005479 @ C | rare1	Lys->Tyr / 34.96837 @ K | rare1	Ser->Tyr / 76.03130 @ S | rare1
% single-letter substitutions for D	% single-letter substitutions for L	% single-letter substitutions for T
Asp->Ala / -43.98983 @ D | rare1	Leu->Ala / -42.04695 @ L | rare1	Thr->Ala / -30.01057 @ T | rare1
Asp->Cys / -12.01776 +58.005479@ D | rare1	Leu->Cys / -10.07488+58.005479 @ L | rare1	Thr->Cys / 1.96150+58.005479 @ T | rare1
Asp->Glu / 14.01565 @ D | rare1	Leu->Asp / 1.94288 @ L | rare1	Thr->Asp / 13.97926 @ T | rare1
Asp->Phe / 32.04147 @ D | rare1	Leu->Glu / 15.95853 @ L | rare1	Thr->Glu / 27.99491 @ T | rare1
Asp->Gly / -58.00548 @ D | rare1	Leu->Phe / 33.98435 @ L | rare1	Thr->Phe / 46.02073 @ T | rare1
Asp->His / 22.03197 @ D | rare1	Leu->Gly / -56.06260 @ L | rare1	Thr->Gly / -44.02622 @ T | rare1
Asp->Lys / 13.06802 @ D | rare1	Leu->His / 23.97485 @ L | rare1	Thr->His / 36.01123 @ T | rare1
Asp->Leu / -1.94288 @ D | rare1	Leu->Lys / 15.01090 @ L | rare1	Thr->Lys / 27.04728 @ T | rare1
Asp->Met / 16.01354 @ D | rare1	Leu->Met / 17.95642 @ L | rare1	Thr->Leu / 12.03638 @ T | rare1
Asp->Asn / -0.98401 @ D | rare1	Leu->Asn / 0.95887 @ L | rare1	Thr->Met / 29.99280 @ T | rare1
Asp->Pro / -17.97418 @ D | rare1	Leu->Pro / -16.03130 @ L | rare1	Thr->Asn / 12.99525 @ T | rare1
Asp->Gln / 13.03164 @ D | rare1	Leu->Gln / 14.97452 @ L | rare1	Thr->Pro / -3.99492 @ T | rare1
Asp->Arg / 41.07417 @ D | rare1	Leu->Arg / 43.01705 @ L | rare1	Thr->Gln / 27.01090 @ T | rare1
Asp->Ser / -27.99491 @ D | rare1	Leu->Ser / -26.05203 @ L | rare1	Thr->Arg / 55.05343 @ T | rare1
Asp->Thr / -13.97926 @ D | rare1	Leu->Thr / -12.03638 @ L | rare1	Thr->Ser / -14.01565 @ T | rare1
Asp->Val / -15.95853 @ D | rare1	Leu->Val / -14.01565 @ L | rare1	Thr->Val / -1.97927 @ T | rare1
Asp->Trp / 71.05237 @ D | rare1	Leu->Trp / 72.99525 @ L | rare1	Thr->Trp / 85.03163 @ T | rare1
Asp->Tyr / 48.03639 @ D | rare1	Leu->Tyr / 49.97927 @ L | rare1	Thr->Tyr / 62.01565 @ T | rare1

Fig. 1.1: List of all possible one-letter sequence variants, for the case of cysteine alkylation by unlabeled iodoacetic acid. The % symbol signifies a comment line and instances corresponding to deamidation have been commented out (see yellow highlight on the next page). These variants are pasted into the modification pop-up box (as a single column). Continued on the next page.