Automated Antibody Sequencing Software

Wilfred H. Tang, Marshall W. Bern, Chris Becker, K. Ilker Sen – Protein Metrics Inc., San Carlos, CA

Contact: wtang@proteinmetrics.com

www.proteinmetrics.com

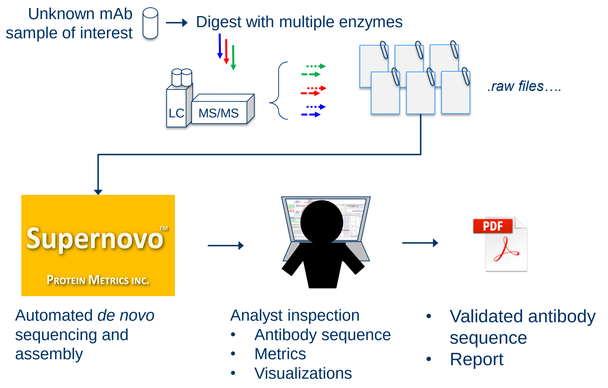
Introduction

Sequencing monoclonal antibodies (mAbs) is a labor-intensive process. It is also critical to the biopharmaceutical industry; *de novo* sequencing by tandem mass spectrometry provides an alternative approach in the discovery of new antibody drug candidates and tool reagents for research. We present here Supernovo™ software that automates this analysis. The result of the analysis is a complete antibody sequence, and, just as importantly, metrics and visualizations for validating the sequence. This poster presents the results of several "stress tests" that show the strength of the underlying algorithms and the importance of the accompanying inspection tools.

Supernovo Overview

- Starting antibody scaffold is obtained:
  - Constructed based on databases of known antibodies, or
  - Supplied by user
- The sequence assembly algorithm then completes the CDR sequences and framework mutations. Although the analysis is similar in some ways to DNA assembly, mass spectrometric identifications are often incomplete and of short read lengths (even with high accuracy, high quality data), making the automatic assembly challenging yet possible.

Workflow



Automated *de novo* sequencing and assembly

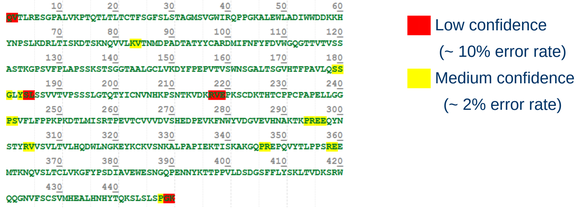
Analyst inspection
- Antibody sequence
- Metrics
- Visualizations

- Validated antibody sequence
- Report

Metrics/Visualizations for Validation

High level overview:

Highlight amino acid residues that should be inspected (in this instance, most questions are in the constant region (unlikely to be wrong)

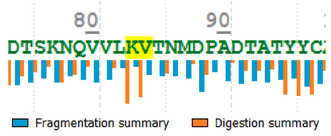


Low confidence (~ 10% error rate)

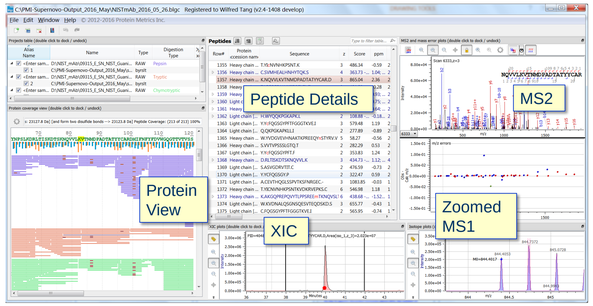
Medium confidence (~ 2% error rate)

Medium level view: Fragmentation and digestion summary

Summary of the accumulated evidence for each cleavage between amino acid residues – both MS^1 (digestion) and MS^2 (fragmentation)



Detailed view: Dashboard for in-depth inspection of peptides



Intact mass

Supernovo does not use the intact mass in its algorithms. Intact mass is a separate analytical measurement and can be used to provide an orthogonal check on the sequence composition.

Stress Testing Supernovo™

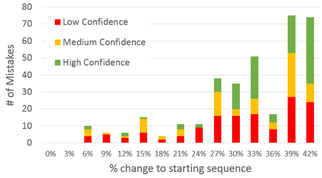
- Use NISTmAb, a mAb with known sequence
- Digest using trypsin, chymotrypsin, and pepsin
- Analyze on Orbitrap Elite → Total of 12,460 HCD MS/MS spectra

Supernovo's answer is completely correct (except for I vs. L)

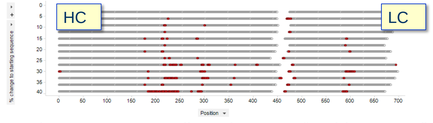
Stress test #1: Discard spectra (randomly) from the data set. Evaluate how Supernovo performs on sparser and sparser data sets.

- For up to ~ 80 – 90% spectra discarded, Supernovo's answer is completely correct or almost completely correct
- As the % spectra discarded increases above 90%, the number of mistakes in Supernovo's answer increases
- Most of the mistakes occur on amino acid residues that Supernovo marks as low confidence

Stress test #2: Supply Supernovo with a starting antibody scaffold that deviates (randomly) from the known answer. Evaluate how Supernovo performs as the initial scaffold deviates more and more from the known answer.



- For up to ~ 20 – 30% deviation over the entire protein from the answer, Supernovo's mistakes are relatively low (and manageable)
- At high levels of scaffold inaccuracy, some mistakes marked as high confidence are assembly errors – correct sequence placed in the wrong location



- Mistakes tend to occur repeatedly at the same "problem spots" (which correspond to residues with less mass spec evidence)

How Long Does It Take?

Supernovo run (an hour to a day of computer time) followed by 20 minutes of human time will **triage** the project:

1. **Complete or almost complete.** Only exact-mass substitutions (residue order, GA / Q, GG / N, NG vs DG) left. Human expert can correct, validate, and prepare a report in 1 – 2 hours.
2. **Possible.** Intact mass mismatch, but human expert might be able to finish with some effort.
3. **Impossible.** If Supernovo does not get constant part > 95% correct, go collect more spectra! Sequence coverage is too low.

Applications

Sequence commercially available mAbs with unknown sequence

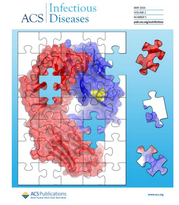
We used Supernovo to analyze a batch of 6 unknown mAbs from a pharmaceutical partner

- 5 mAbs were **category 1 (complete or nearly complete)**
- 1 mAb was borderline **category 1 or 2**. We were able to figure out the **complete sequence with modest effort**

Resurrect a mAb with lost hybridoma (Ref 1)

PL-2 anti-astrovirus mAb

- Neutralizing mAb made in 1994
- Unknown sequence
- Hybridoma long gone!
- Sequencing started using X-ray crystallography
- Sequencing finished using mass spectrometry



And more… See poster WP 033

Conclusions

Supernovo makes routine *de novo* sequencing of mAbs possible by providing:

- "Hands free" operation
- Robust results – as demonstrated by stress tests
- Metrics and visualizations to help the scientist validate the results

References and Acknowledgments

1. W. A. Bogdanoff et al., "De Novo Sequencing and Resurrection of a Human Astrovirus-Neutralizing Antibody," *ACS Infect. Dis.*, **2016**, *2*, pp 313–321.

Protein Metrics gratefully acknowledges support in part from the NIH, grant GM103362.