The Buyers' Guide to Protein Sequencing

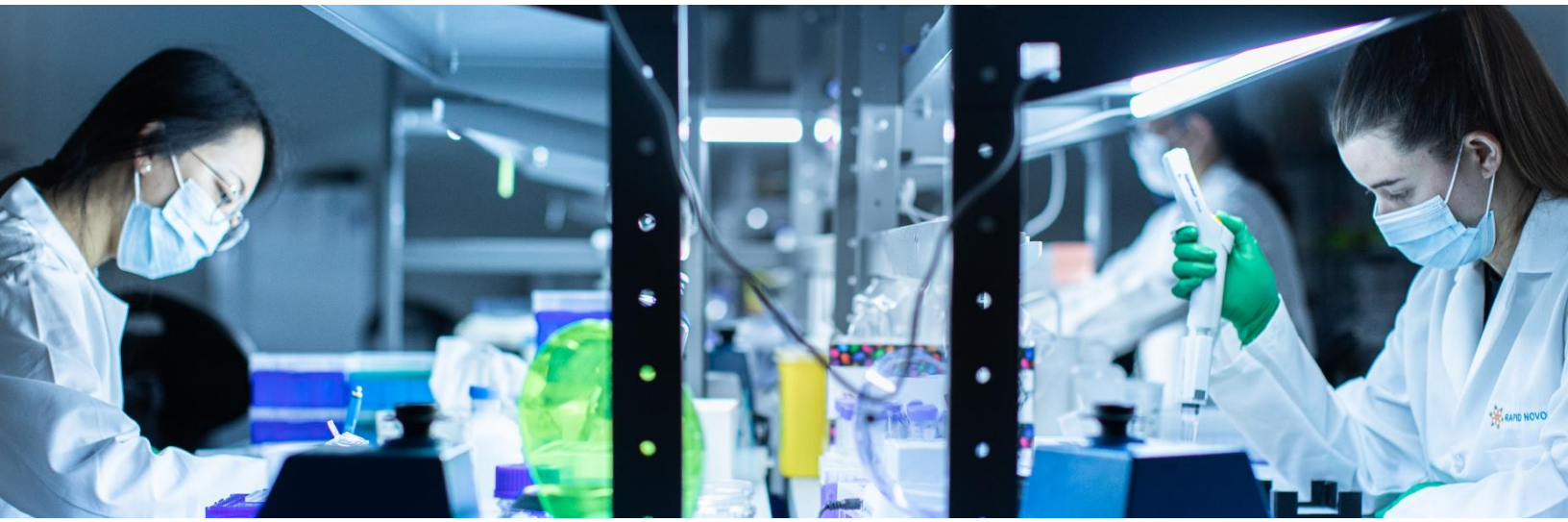
Rapid Novor | 2022

rapid novor

Introduction

This guide is intended to help prospective users determine if protein sequencing is a fit for their needs, what to expect, and what to consider when choosing a service provider.

- Understand protein sequencing and related/complementary methods
- Determine what methods best suit your needs
- Considerations for different types of proteins and samples
- How to evaluate a lab or vendor's capabilities
- What to expect in terms of pricing, timelines and results reports



About Rapid Novor

Rapid Novor helps biotech and pharmaceutical companies extract the full potential of antibodies in ways never before possible. We provide services and develop new technology that enables timely and reliable discovery and development of novel reagents, diagnostics, and therapeutics. Thanks to our Next Generation Protein Sequencing and antibody discovery services, researchers have furthered thousands of projects, patented antibody therapeutics, and developed the first recombinant polyclonal antibody diagnostics.

Our mission: To empower life science breakthroughs with next generation protein sequencing.

Terminology and Related Techniques

🕮 *de novo* protein sequencing

/ dē ˈnōvō/ /ˈprō˛tēn/ /ˈsēkwənsing/

definition: The method by which the amino acid sequence of a protein is determined by tandem mass spectrometry without the assistance of a database

synonyms: *de novo* sequencing, de novo antibody sequencing, antibody sequencing, amino acid sequencing, [see also: next generation protein sequencing]

distinct from: peptide mapping, hybridoma sequencing, B cell sequencing, single cell sequencing, edman sequencing, N-terminal sequencing

Related definitions

Coverage: This refers to the amount of the amino acid sequence that contains peptide evidence for the sequence. Rapid Novor's minimum standard is 100% coverage across the entire antibody, with minimum 30x overlap by individual peptides over each amino acid.

Monoclonal: Antibodies in the sample are of the same amino acid sequence.

Polyclonal: Antibodies in the sample have high diversity in their sequences. These mixtures contain a very large number of different antibodies.

Oligoclonal: The sample contains a limited number of antibodies with different sequences, typically two to five.

Comparing types of protein sequencing

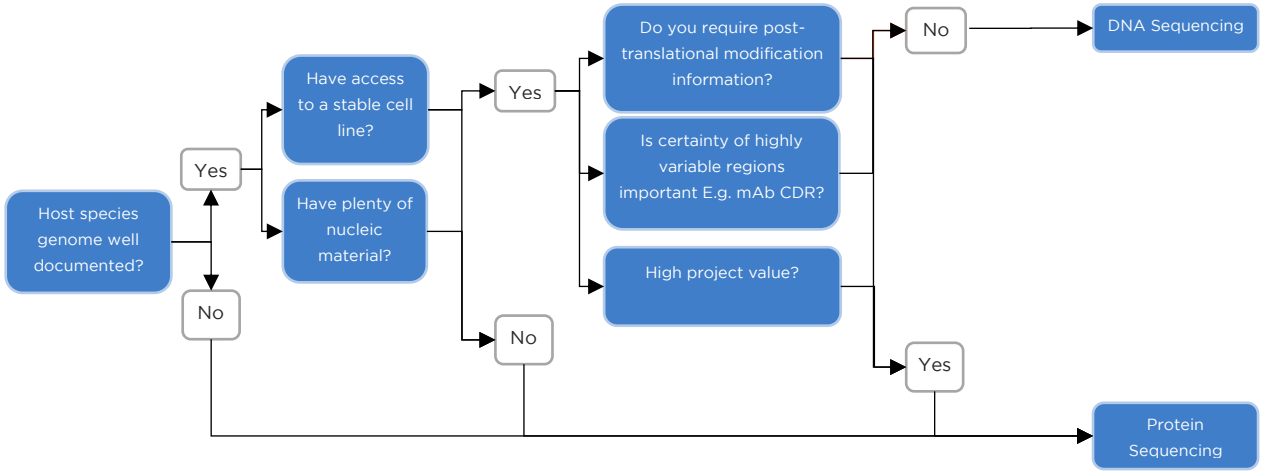
Edman Degradation	Peptide Mapping	*de novo* Protein Sequencing
N-terminal sequencing analysis identifies the amino-terminal residues through a cyclical reaction and chromatography. In practice, cannot fully sequence anything longer than 30 amino acids. Will not work if the N-terminus has been chemically modified, and sequencing will stop if a non-alpha-amino acid is encountered.	A commonly used strategy in protein identification, sequence confirmation, or characterization. Peptides are analyzed through mass spectrometry and are compared to a database to obtain sequence information. Only peptide masses are measured and contaminations can interfere. Only proteins in the database are identified.	Obtain the full amino acid sequence from any protein, without any prior knowledge of the sample. The protein is digested using a protease cocktail and analyzed via LC-MS/MS. Novel proteins are sequenced with extremely high accuracy at Rapid Novor. This includes determination of ambiguous residues such as Ile/Leu, and PTM identification. Additional sample information such as impurities or additional chains may also be identified.

Next Generation Protein Sequencing (NGPS)

This term has been used by several groups who are trying to make protein sequencing faster, cheaper, and more accessible. This term could be related to any of the above techniques, but overall, it represents a significant advancement in the state of the art. Rapid Novor's NGPS technology builds on 20+ years of innovation in de novo protein sequencing.

When to use Protein vs DNA Sequencing

Although often more costly than DNA sequencing, protein sequencing is the key to understanding the functional molecule by looking directly at the amino acids present. Protein sequencing remains a staple of antibody development, particularly because of its ability to access post-translational modifications (PTMs) and additional chains, in contrast to nucleotide sequencing methods.



Reliability and validity of *de novo* protein sequencing

Recent Advances

Mass spectrometry based protein sequencing had traditionally been known as slow, inaccurate and incomplete. Software and chemistry techniques have been developed in the last 5 years to overcome these challenges. Specifically, Rapid Novor has developed proprietary algorithms and software to sequence peptides and assemble hundreds of overlapping peptide sequences into full length protein sequences. Additionally, Rapid Novor uses patent-pending chemistry to enhance the analysis and deliver sequences with 100% accuracy and coverage.

Literature review

Paper: In-Depth Characterization of Monoclonal Antibodies with a Single Experiment and Fully Automated Data Analysis

Webinar: Fully Harnessing the Power of Immunotherapy through Protein Sequencing

Case Study: Prevalence of Secondary Light Chains - Rapid Novor's technology will also detect and sequence unexpected assets, such as additional light and heavy chains.

Overcoming Common Ambiguities in Mass Spectrometry

It's important to use an approach that can distinguish between Isoleucine and Leucine because these amino acids are isobaric, meaning they have the same mass. Common mass spectrometry techniques cannot distinguish them. If you cannot distinguish between these two, you'll end up with a sequence that has ambiguities, which means you'll have costly trial and error downstream. The best approach involves examining w-ions via electron-transfer/higher-energy collision dissociation (EThcD) mass spectrometric analysis, as in Rapid Novor's WILD® technology.

Common practice

De novo protein sequencing is quickly becoming a mainstream technique, particularly in the study of antibodies. Rapid Novor alone has sequenced over 5000 proteins in the last few years. There are dozens of publications and patents citing *de novo* sequencing services and software. Rapid Novor has created a routine, robust, and high-throughput workflow for the de novo sequencing of antibodies and other proteins.

Types of Proteins and Mixtures

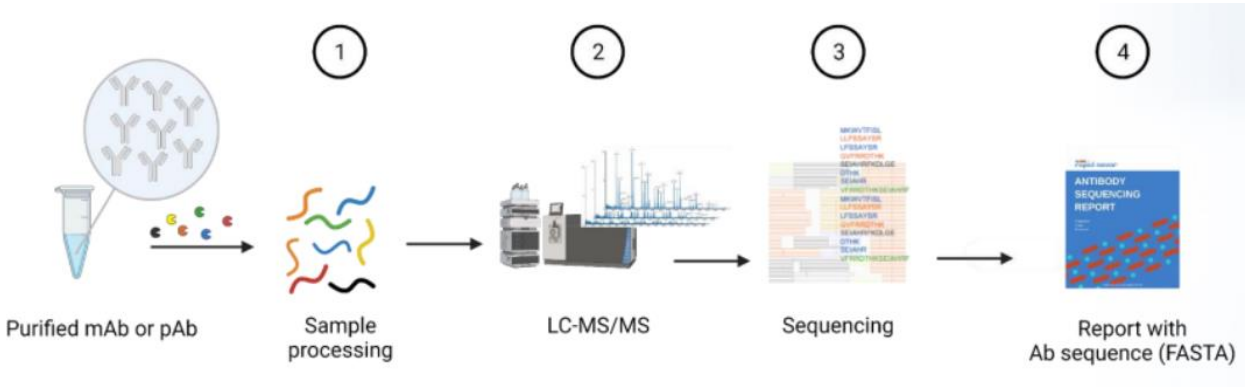
Sequencing Non-Antibody Proteins

Although Rapid Novor specializes in antibody protein sequencing, at its core the process is still protein sequencing, and as such, any protein can be sequenced using the workflow. The complexity of sequencing proteins depends in part on the size and structure of the protein. Conformation and hydrophobic/hydrophilic regions may impact the enzymatic digest and separation prior to mass spectrometry and lead to areas of low/no coverage. Near-real-time data analysis is required in order to identify these regions and make adjustments to ensure full coverage.

Sequencing Monoclonal Antibodies

Given their nature, antibodies are highly variable and were notoriously difficult to sequence until recently. Rapid Novor has invented the technology necessary to do this with 100% accuracy, and at scale.

REmAb® is Rapid Novor's proprietary *de novo* sequencing platform, combining in-house mass spectrometry, proprietary chemistry and novel sequencing algorithms. REmAb® can deliver an accurate sequence for any species, any isotype with 100% accuracy and 100% coverage. The workflow is as follows:



A component of REmAb®, WILD® is the world's first commercially available solution to accurately distinguish between isoleucine and leucine using mass spectrometry. Our method employs proprietary protein chemistry & Orbitrap Fusion™ instruments to perform electron-transfer higher energy collision dissociation (EThcD).

Sequencing Polyclonal Antibodies

De novo protein sequencing is not only possible on monoclonal antibodies, but also polyclonal antibody (pAb) mixtures. Rapid Novor is the first and only company in the world to offer *de novo* polyclonal antibody sequencing. This has obvious implications for immortalizing and standardizing pAb reagents, but it also leads to an entirely new approach to antibody therapeutic discovery.

There are many advantages and disadvantages of using mAbs vs pAbs. Polyclonal Antibody sequencing unlocks a 'best of both worlds' option:

	pAbs	**mAbs**
Advantages	Easy to produce Many existing products on the market High affinity Broad biophysical diversity	Recombinant production possible Indefinite access with digital sequence Industry has standardized processes for up-scale production Easy to characterize
Disadvantages	Batch to batch variation Limited supply Challenging to up-scale production Binding not strictly defined Difficult to characterize	Harder to produce initially More challenges against targets
Best of both worlds	**Polyclonal antibody sequencing helps you sequence and recombinantly express the most abundant antibodies in your pAb, recapitulating the high performance of your pAb with a recombinant cocktail.**	

How to find and select a vendor for protein sequencing

Theoretically, anyone with a mass spectrometer can perform protein sequencing. However, *de novo* protein sequencing requires a very expensive specialized equipment investment, specialized software, and scientists to analyze the data to guarantee high accuracy. Manual sequencing or use of out-dated software causes inaccuracy and delay. Here's how to determine if you're working with experts:

Look for REmAb®

REmAb® is the world leading Next-Generation Protein Sequencing service. There are only a few technology partners and local service providers that are licensed to resell the REmAb® service on its own or as an integral part of their own service. If you see REmAb® technology, you know you're working with the best.

Key Questions to Ask

What Bioinformatics Software are they using?

Best in class: machine learning and decision tree based *de novo* peptide sequencing algorithms (e.g. Novor), automated sequence assembly

Previous standard: heuristics based *de novo* sequencing algorithms, manual sequence assembly

How are they doing Ile/Leu determination?

Best in class: w-ion analysis using EThcD fragmentation

Previous standard: estimate based on homologous sequences and expected frequency of how often these amino acids occur

What depth of coverage do they offer? How many enzymes do they use?

The REmAb® standard involves **7 enzymes** that cleave in different locations. Using more enzymes means more overlapping peptides are created, and more independent evidence is collected about each amino acid (more coverage). This also means more fractions of the sample to process, more mass spectrometry run time and more data analysis - so inexperienced or low cost providers will try to economize on the number of enzymes/fractions, putting accuracy at risk.

Rapid Novor's minimum requirement is 30x overlap by individual peptides over the whole amino acid sequence. This is our confidence level for 100% accuracy. If 30x is not achieved after the first round of digestions, Rapid Novor will perform additional experiments with additional enzymes (sometimes as many as 10) to achieve the coverage required across the entire protein sequence.

What's their team's focus and expertise?

Best in class: instruments and people dedicated to protein sequencing full time, lab and mass spectrometry and bioinformatics under one roof

Previous standard: occasional *de novo* sequencing work

Increased focus and expertise means less time setting up instruments and experiments. It means greater ability to handle difficult samples.

Quick Comparison

	REmAb®	**Other**
Throughput	50 to 100 per week	Typically one at a time
Accuracy	>99.9%	Often low
Speed	Typically 1-2 weeks.	Typically 2 months
Ability to handle the unexpected	Yes, additional chains, impurities, oligoclonal, etc.	Rarely
Software	Proprietary purpose built algorithms	Off the shelf software
Real-time sequencing for enhanced coverage	Yes	No
Accurate Ile/Leu determination	Yes, with WILD® technology	Rarely, often based on digestion patterns
Multi-enzyme digest	Minimum 5, up to 10 enzymes	Often 1 or 2
Data science + lab	Yes	No
Protein sequencing specialized	Yes	No
Additional QC checks	Yes	Rarely

The impact of sequence accuracy

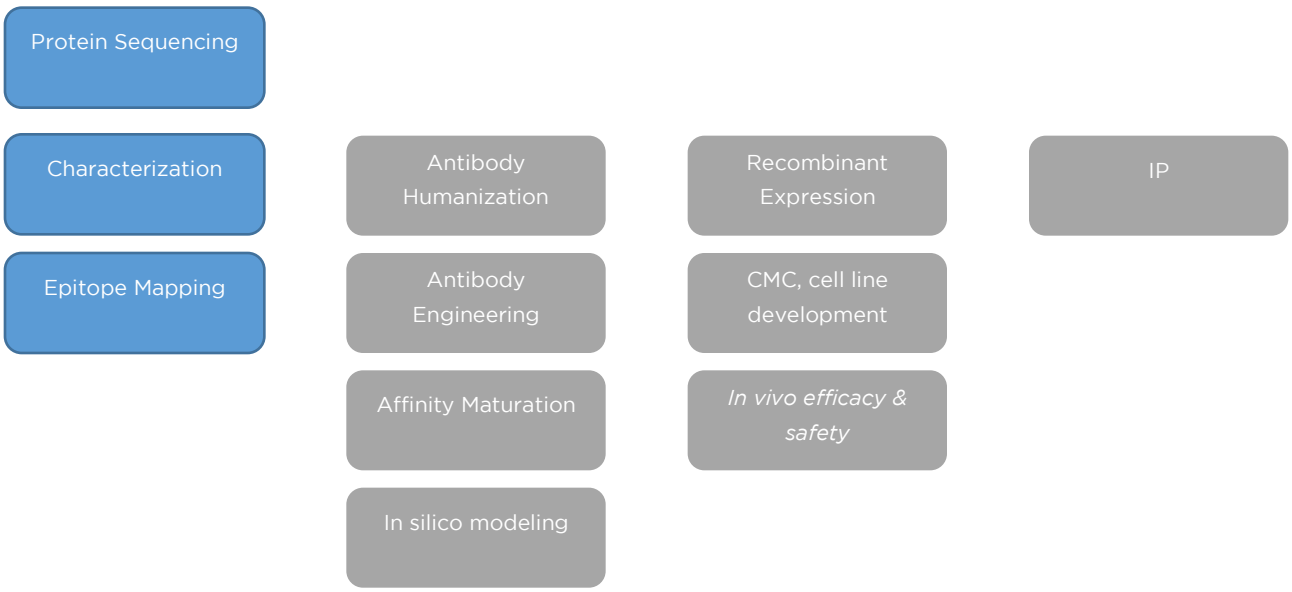
Next steps after protein sequencing often involve protein engineering and recombinant expression. Ambiguities in a sequence make downstream work exponentially more difficult. For instance if your sequence contains 1 ambiguous position with 2 possible amino acids, you must recombinantly express and test 2 variants. 2 ambiguities would require 4 variants to be tested. 3 would require 8 variants, and so on. Engineering efforts often result in multiple constructs to express and test, compounding the problem. Considering recombinant expression may cost $1000 for 1mg (at a minimum, $2000 is more likely), the difference between a 99.5% accurate sequence and a 99.9% accurate sequence (e.g. 3 amino acids in an Antibody) is measurable in dollars: $8000+

The difference between a 99.5% accurate sequence and a 99.9% accurate sequence is measurable in dollars: at least $8000

Rapid Novor offers a variety of aids to help with downstream work including free report walkthroughs, meetings with the bioinformatics team, and consulting from our antibody engineering expert.

Downstream impacts

Sequence information is used to guide further studies (e.g. mechanism of action), engineering efforts, production efforts, patent filings/disputes and FDA filings. When absolute certainty of the sequence and biophysical characteristics is needed, the most accurate and comprehensive protein sequencing solution is required – and is an ideal complement for genome sequence information. Months of work and millions of dollars are at stake.



What to Expect

What can you expect to pay?

Monoclonal Antibody Sequencing

Market rates: typically in the range of $10,000 to $20,000 USD.

REmAb®: $13,000 ($5000 for first time clients)

Non-Antibody Protein Sequencing

Market rates: Larger proteins or more complex projects (involving additional characterization or multiple similar proteins) can be in the $50,000 to $100,000 range.

REmAb®: starting at $10,000

Polyclonal Antibody Sequencing

Market rates: n/a

REpAb®: please inquire

What are the sample requirements?

Monoclonal Antibody Sequencing

Our standard requirement for REmAb®: 100ug, 80% pure, any common buffer.

Non-Antibody Protein Sequencing

Sample requirements are the same for as for monoclonal antibodies

Polyclonal Antibody Sequencing

Sample requirements depend on the project goals, the antigen and additional information. If you have 1mg of affinity purified IgG there should be no problem. Please inquire.

Typical Timelines

In the past, *de novo* protein sequencing would take months, however software and advances in mass spectrometry have allowed for a more expedited timeline. REmAb® projects for both monoclonal antibodies and other proteins are typically completed in 2-3 weeks. Our record is 28 hours.

REpAb polyclonal sequencing and discovery projects take anywhere from 8 weeks to 6 months depending on the nature of the project.

What to Expect from your Results

When interpreting your protein sequencing report, it is important to look at the depth of coverage to get a sense of accuracy. Rapid Novor's minimum requirement is 30x overlap by individual peptides over each amino acid, to assume 100% accuracy.

A single amino acid switch could mean the difference between a successful antibody, or one that doesn't work as it should. Isoleucine/Leucine determination should be used in any *de novo* sequencing workflow, or else the results will be ambiguous. W-ion technology is the only method that can distinguish between these two amino acids with 100% accuracy. Other methods such as relying on the enzyme digestion rules, or referencing the germline sequence may be considered, but are not reliable.

Sample Report

A protein sequencing report should contain the amino acid sequence, notable observations (such as PTMs, presence of contaminants, etc), isoleucine/leucine determination, and the coverage view.

How to buy

If you are considering approaches

If you are considering how protein sequencing might solve a problem in your research or to improve your overall workflow, please contact us. We are scientists too and we like solving problems.

Book a meeting and we'll try to help. We can even refer you to other companies that our clients have had success with upstream or downstream of our work.

If You are Ready to Start (soon)

If you think Protein Sequencing could be a fit for your needs in the near future, please contact us to talk with one of our scientists.

1	A representative will contact you within minutes (during business hours)
2	Discuss details of your project and address any of your questions
3	We will generate a project proposal including a quote and summary of the service
4	Sample preparation and shipping instructions will be provided
5	Work will begin immediately upon receipt of your samples

For More Information

Visit www.rapidnovor.com for more information on our services.

These data represent typical results. Information, descriptions and specifications about Rapid Novor services listed in this publication are subject to change.

Please speak with a Rapid Novor representative to discuss your use case and requirements.

Unit 450, 137 Glasgow St.

Kitchener, Ontario, Canada N2G 4X8