De-Novo-Sequencing using Spectrum-Graphs, enabling Open Searches

19. Juni 2023

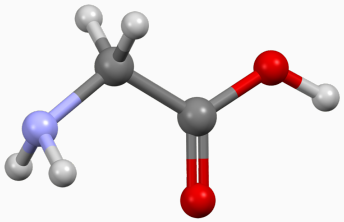
Dominik Habermann

Ruhr Universität Bochum

Gliederung

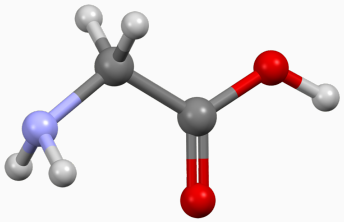
Hintergrund

Hintergrund



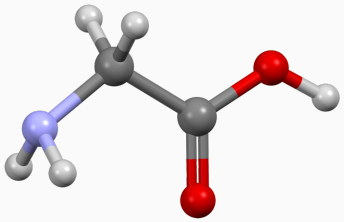
- Peptide: Kette von Aminosäuren (AS)

Hintergrund



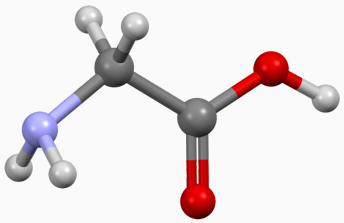
- Peptide: Kette von Aminosäuren (AS)
- 20 relevante AS

Hintergrund



- Peptide: Kette von Aminosäuren (AS)
- 20 relevante AS
- Reihenfolge von AS ist weitestgehend beliebig

Hintergrund



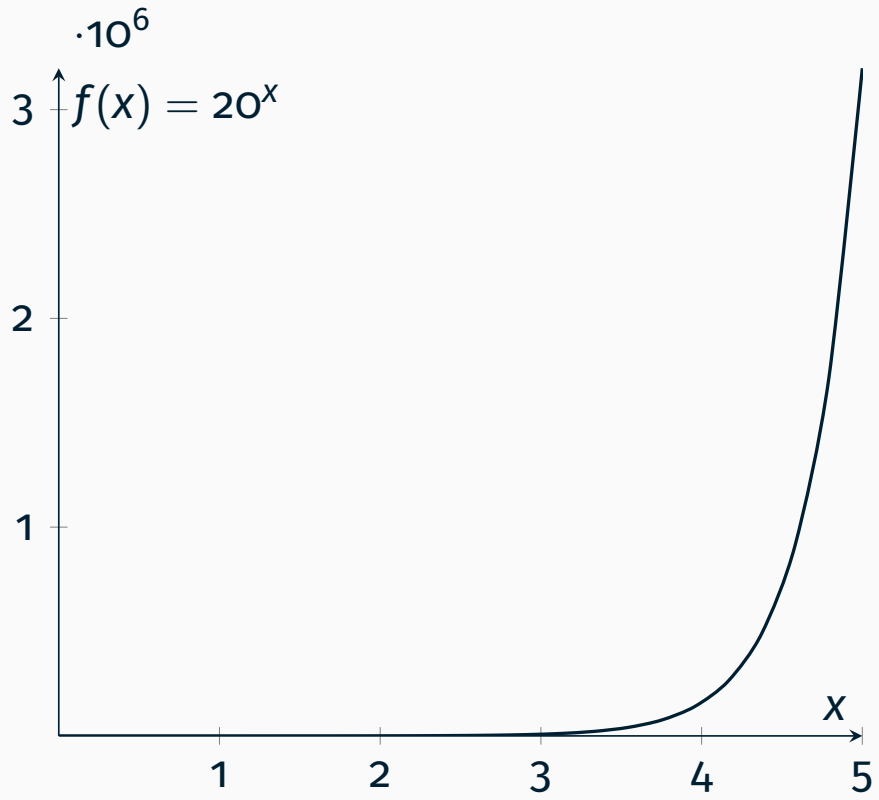
- Peptide: Kette von Aminosäuren (AS)
- 20 relevante AS
- Reihenfolge von AS ist weitestgehend beliebig
- $f(x) = 20^x \quad x : \textit{Anzahl an AS}$

Anzahl an Kombinationen

- Bereits bei wenigen AS: Kaum händelbare Anzahl an Kombinationen

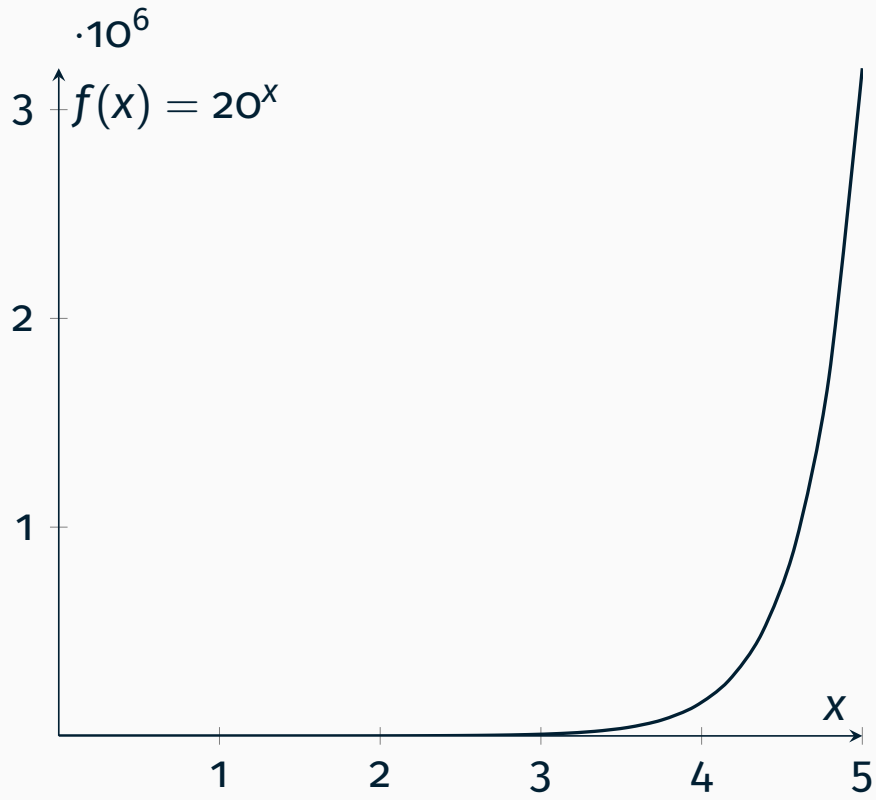
Anzahl an Kombinationen

- Bereits bei wenigen AS: Kaum händelbare Anzahl an Kombinationen



Anzahl an Kombinationen

- Bereits bei wenigen AS: Kaum händelbare Anzahl an Kombinationen



- Zum Vergleich: Proteine bis zu mehreren zehntausend AS

Biomedizinische Fragestellung

- Zuverlässige Bestimmung *kurzkettiger* Peptide möglich?

Biomedizinische Fragestellung

- Zuverlässige Bestimmung *kurzkettiger* Peptide möglich?
- Biomedizinisch relevant:

Biomedizinische Fragestellung

- Zuverlässige Bestimmung *kurzkettiger* Peptide möglich?
- Biomedizinisch relevant:
  - Katalogisierung von Proteinen

Biomedizinische Fragestellung

- Zuverlässige Bestimmung *kurzkettiger* Peptide möglich?
- Biomedizinisch relevant:
  - Katalogisierung von Proteinen
  - Wechselwirkungen von Proteinen

Biomedizinische Fragestellung

- Zuverlässige Bestimmung *kurzkettiger* Peptide möglich?
- Biomedizinisch relevant:
  - Katalogisierung von Proteinen
  - Wechselwirkungen von Proteinen
  - Analyse von Enzymen

AS Sequenzierung

AS Sequenzierung

- AS Sequenzierung: Bestimmung der AS-Sequenz

AS Sequenzierung

- AS Sequenzierung: Bestimmung der AS-Sequenz
- Hilfsmittel: Massenspektrometrie (MS)

AS Sequenzierung

- AS Sequenzierung: Bestimmung der AS-Sequenz
- Hilfsmittel: Massenspektrometrie (MS)
- MS kann chemische Strukturen bestimmen

AS Sequenzierung

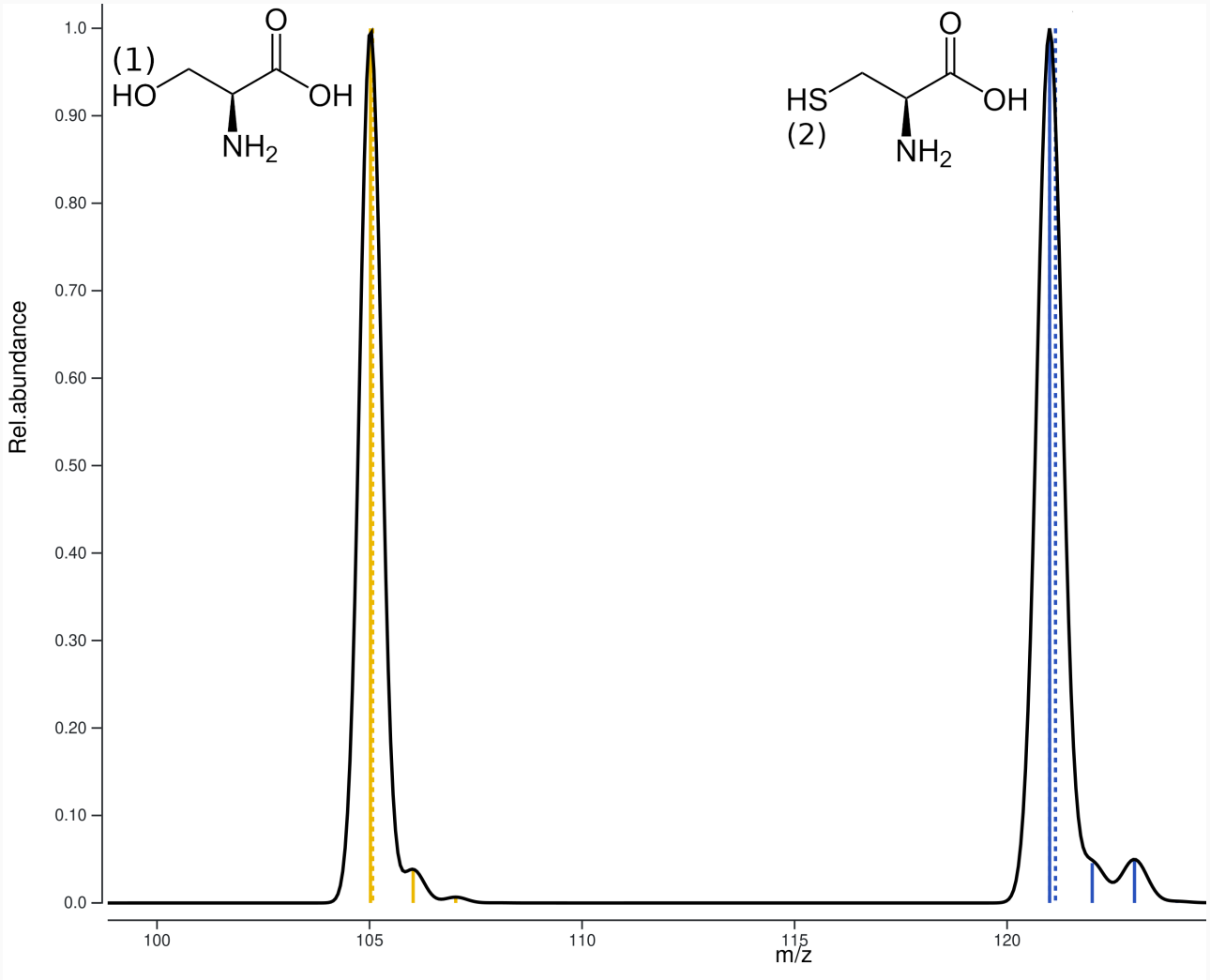
- AS Sequenzierung: Bestimmung der AS-Sequenz
- Hilfsmittel: Massenspektrometrie (MS)
- MS kann chemische Strukturen bestimmen
- Rückschluss auf die AS-Sequenz

Beispiel Spektrogramm

- Ergebnisse einer MS: Spektrogramm

Beispiel Spektrogramm

- Ergebnisse einer MS: Spektrogramm



De-Novo-Sequenzierung

Sequenzierung und De-Novo-Sequenzierung

- Beide Verfahren bestimmen die AS-Sequenz

Sequenzierung und De-Novo-Sequenzierung

- Beide Verfahren bestimmen die AS-Sequenz

AS Sequenzierung	De-Novo-Sequenzierung
Datenbanken als Hilfsmittel	

Sequenzierung und De-Novo-Sequenzierung

- Beide Verfahren bestimmen die AS-Sequenz

AS Sequenzierung	De-Novo-Sequenzierung
Datenbanken als Hilfsmittel	
Identifizierung von *bekannten* Sequenzen	

Sequenzierung und De-Novo-Sequenzierung

- Beide Verfahren bestimmen die AS-Sequenz

AS Sequenzierung	De-Novo-Sequenzierung
Datenbanken als Hilfsmittel	Ohne weitere Hilfsmittel
Identifizierung von *bekannten* Sequenzen	

Sequenzierung und De-Novo-Sequenzierung

- Beide Verfahren bestimmen die AS-Sequenz

AS Sequenzierung	De-Novo-Sequenzierung
Datenbanken als Hilfsmittel	Ohne weitere Hilfsmittel
Identifizierung von *bekannten* Sequenzen	Bestimmung *unbekannter* Sequenzen

Sequenzierung und De-Novo-Sequenzierung

- Beide Verfahren bestimmen die AS-Sequenz

AS Sequenzierung	De-Novo-Sequenzierung
Datenbanken als Hilfsmittel	Ohne weitere Hilfsmittel
Identifizierung von *bekannten* Sequenzen	Bestimmung *unbekannter* Sequenzen

- De novo: lat. „Von neuem“

De-Novo-Sequenzierung – Übersicht

- Zusätzliche Informationen notwendig

De-Novo-Sequenzierung – Übersicht

- Zusätzliche Informationen notwendig
- Verwendung einer 2. MS

De-Novo-Sequenzierung – Übersicht

- Zusätzliche Informationen notwendig
- Verwendung einer 2. MS
- Verfahren: Tandem-Massenspektrometrie MS2

De-Novo-Sequenzierung – 1. MS

- Ionen aus m/z Bereich auswählbar machen

De-Novo-Sequenzierung – 1. MS

- Ionen aus m/z Bereich auswählbar machen
- Quasi eine Filterung

De-Novo-Sequenzierung – 1. MS

- Ionen aus m/z Bereich auswählbar machen
- Quasi eine Filterung
- Ausgewählte Ionen werden für 2. MS verwendet

De-Novo-Sequenzierung – 2. MS

- Ionen aus 1. MS „fragmentieren“:

De-Novo-Sequenzierung – 2. MS

- Ionen aus 1. MS „fragmentieren“:
  - Energiezuführung

De-Novo-Sequenzierung – 2. MS

- Ionen aus 1. MS „fragmentieren“:
  - Energiezuführung
  - Ionen zerfallen

De-Novo-Sequenzierung – 2. MS

- Ionen aus 1. MS „fragmentieren“:
  - Energiezuführung
  - Ionen zerfallen
  - Ergebnis: „Fragment-Ionen“

De-Novo-Sequenzierung – 2. MS

- Ionen aus 1. MS „fragmentieren“:
  - Energiezuführung
  - Ionen zerfallen
  - Ergebnis: „Fragment-Ionen“
- Verschiedene Fragmentierungsmethoden mit spezifischen Fragmenten

De-Novo-Sequenzierung – 2. MS

- Ionen aus 1. MS „fragmentieren“:
  - Energiezuführung
  - Ionen zerfallen
  - Ergebnis: „Fragment-Ionen“
- Verschiedene Fragmentierungsmethoden mit spezifischen Fragmenten
- 2. MS wird auf Fragment-Ionen angewendet

De-Novo-Sequenzierung – MS2

- Höhere Genauigkeit durch Filterung nach 1. MS

De-Novo-Sequenzierung – MS2

- Höhere Genauigkeit durch Filterung nach 1. MS
- Bessere Selektivität beim 2. MS

De-Novo-Sequenzierung – MS2

- Höhere Genauigkeit durch Filterung nach 1. MS
- Bessere Selektivität beim 2. MS
- Ionen zerfallen spezifisch

De-Novo-Sequenzierung – MS2

- Höhere Genauigkeit durch Filterung nach 1. MS
- Bessere Selektivität beim 2. MS
- Ionen zerfallen spezifisch
- $\rightarrow$ Rekonstruktion der ursprünglichen Ionen möglich

De-Novo-Sequenzierung – MS2

- Höhere Genauigkeit durch Filterung nach 1. MS
- Bessere Selektivität beim 2. MS
- Ionen zerfallen spezifisch
- $\rightarrow$ Rekonstruktion der ursprünglichen Ionen möglich

- MS2 Ergebnisse haben eine höhere Qualität

pNovo+ Algorithmus

pNovo+ Algorithmus – Übersicht

- Algorithmus für die De-Novo-Sequenzierung

pNovo+ Algorithmus – Übersicht

- Algorithmus für die De-Novo-Sequenzierung
- Auswertung von MS2 Spektrogrammen

pNovo+ Algorithmus – Übersicht

- Algorithmus für die De-Novo-Sequenzierung
- Auswertung von MS2 Spektrogrammen
- Rekonstruktion der AS-Sequenz

pNovo+ Algorithmus – Übersicht

- Algorithmus für die De-Novo-Sequenzierung
- Auswertung von MS2 Spektrogrammen
- Rekonstruktion der AS-Sequenz
- Hilfsmittel: Spektrums-Graph

pNovo+ Algorithmus – Hauptansatz

- Zwei MS2 Spektren verwenden

pNovo+ Algorithmus – Hauptansatz

- Zwei MS2 Spektren verwenden
- Unterschiedliche Fragmentierungsmethoden pro Spektrum

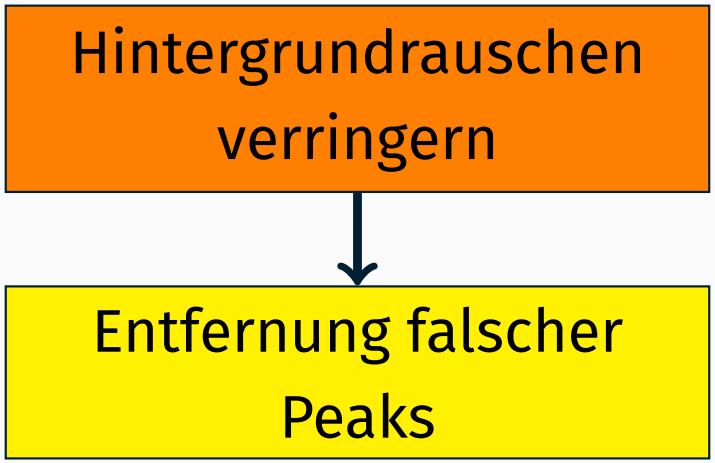
pNovo+ Algorithmus – Hauptansatz

- Zwei MS2 Spektren verwenden
- Unterschiedliche Fragmentierungsmethoden pro Spektrum
- Ziel: bessere Sequenzierungsergebnisse

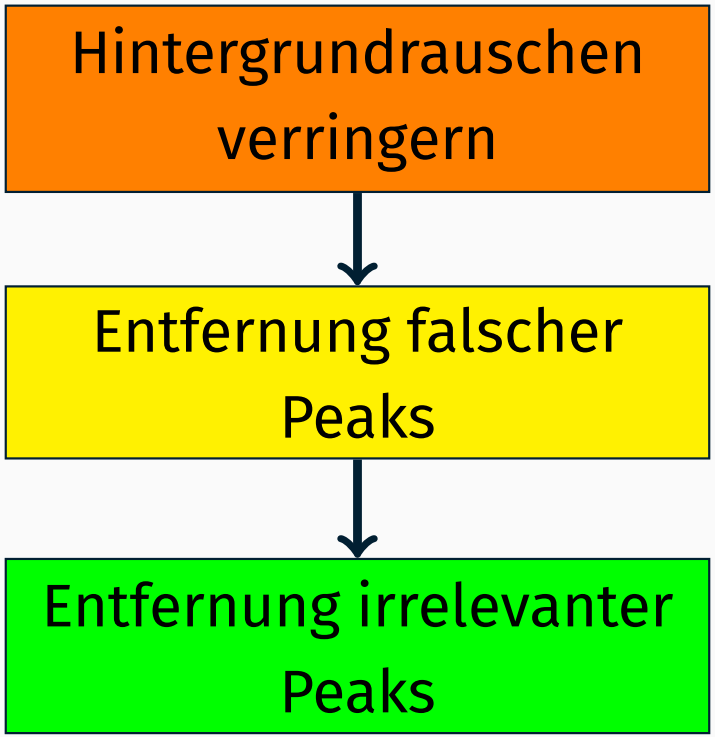
pNovo+ Algorithmus – Datenaufbereitung

Hintergrundrauschen verringern

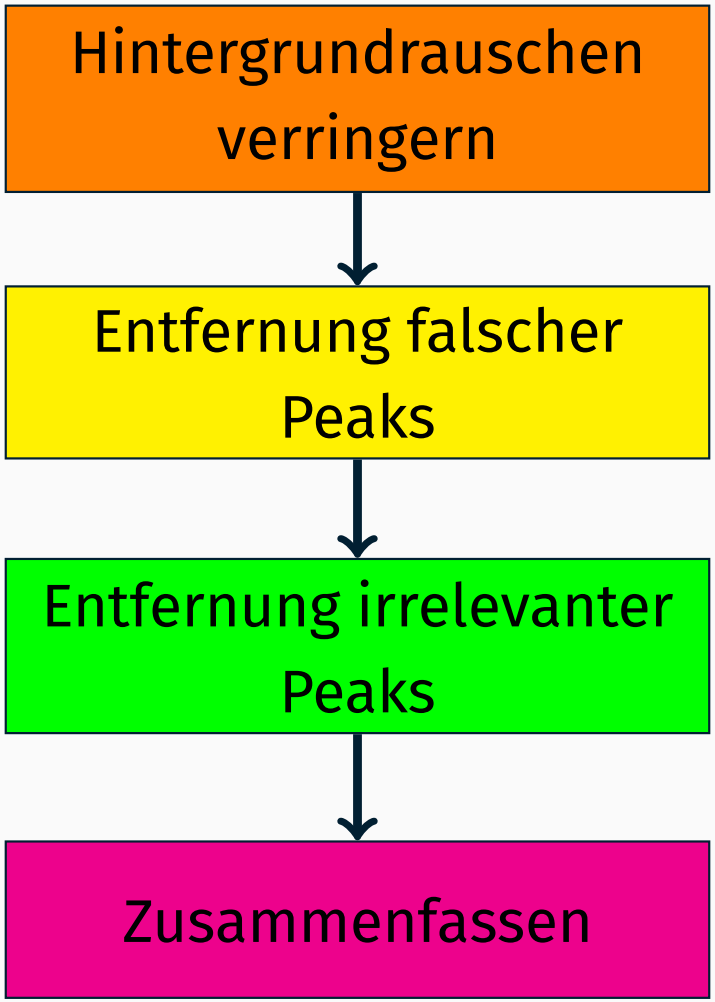
pNovo+ Algorithmus – Datenaufbereitung



pNovo+ Algorithmus – Datenaufbereitung

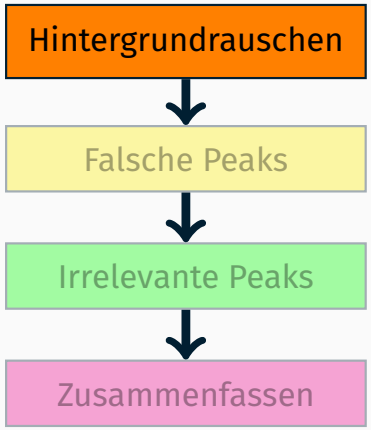


pNovo+ Algorithmus – Datenaufbereitung



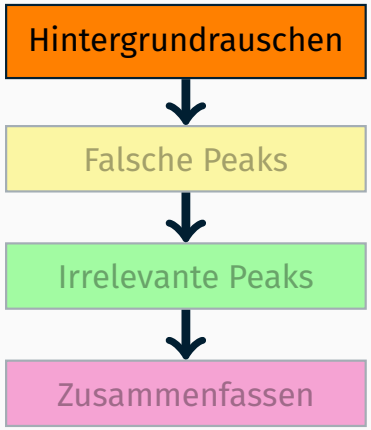
pNovo+ Algorithmus – Datenaufbereitung

- Überpriorisierung fehlerhafter Daten vermeiden



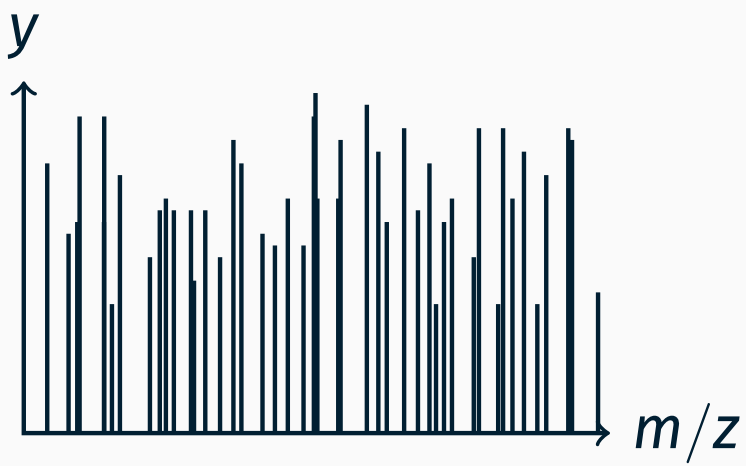
pNovo+ Algorithmus – Datenaufbereitung

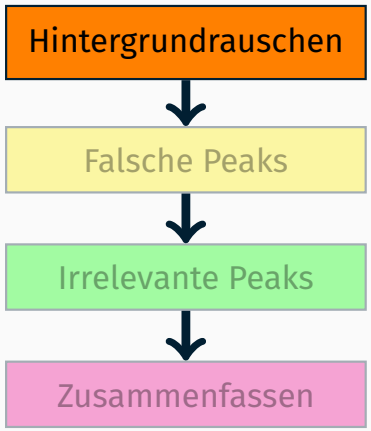
- Überpriorisierung fehlerhafter Daten vermeiden
- Tool: *ln*()



pNovo+ Algorithmus – Datenaufbereitung

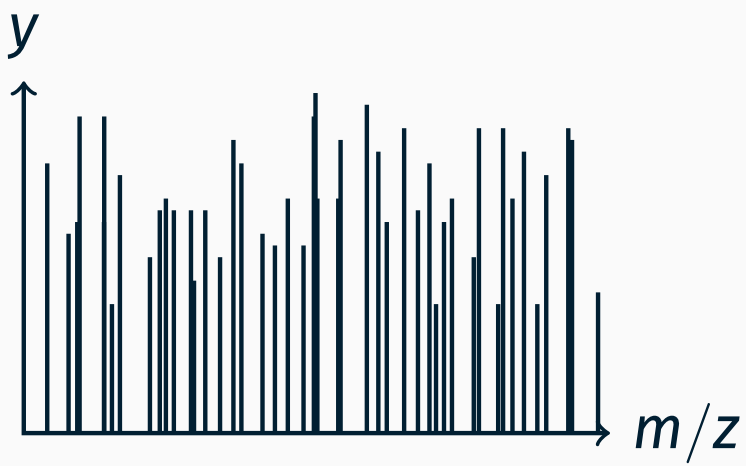
- Überpriorisierung fehlerhafter Daten vermeiden
- Tool: *ln*()

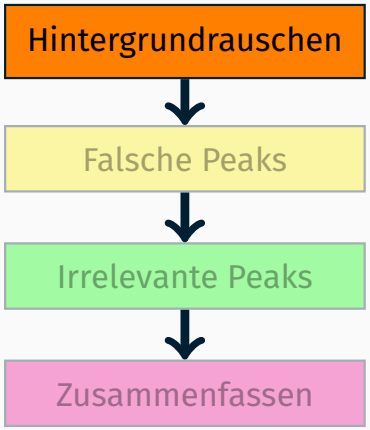




pNovo+ Algorithmus – Datenaufbereitung

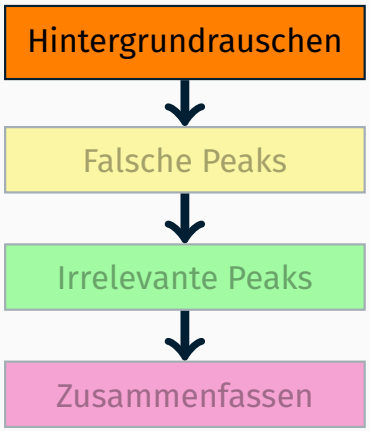
- Überpriorisierung fehlerhafter Daten vermeiden
- Tool: *ln*()

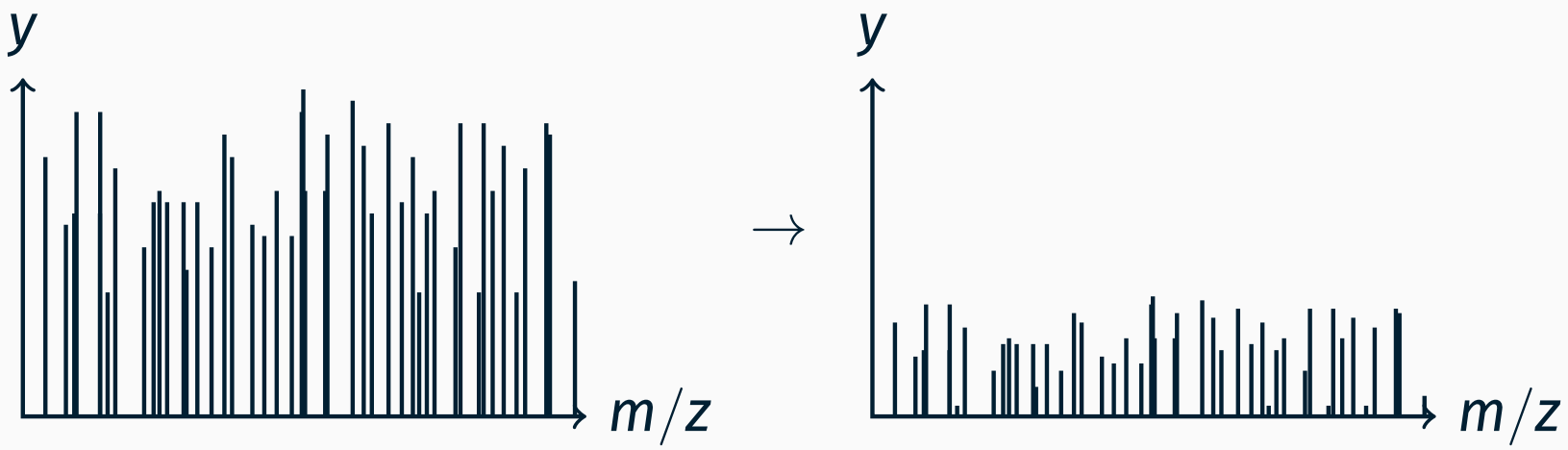




pNovo+ Algorithmus – Datenaufbereitung

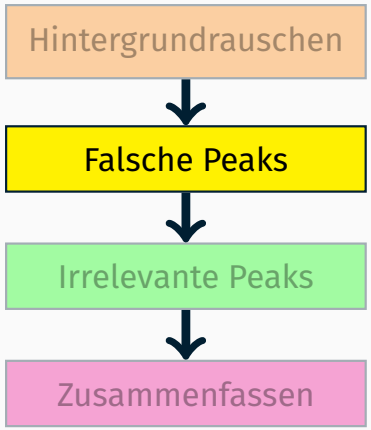
- Überpriorisierung fehlerhafter Daten vermeiden
- Tool: *ln*()





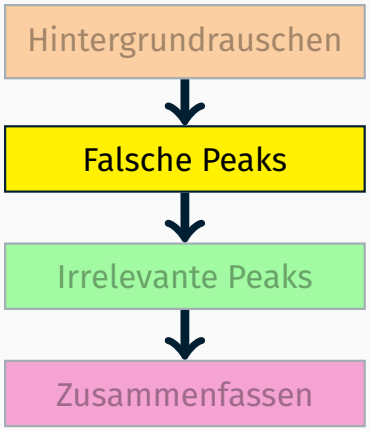
pNovo+ Algorithmus – Datenaufbereitung

- Hintergrundrauschen könnte auch AS mit Isotop sein



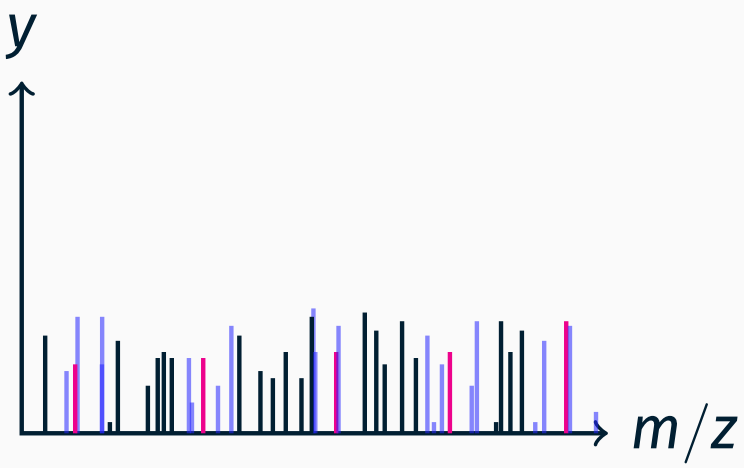
pNovo+ Algorithmus – Datenaufbereitung

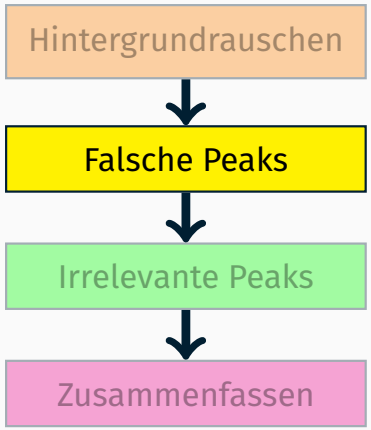
- Hintergrundrauschen könnte auch AS mit Isotop sein
- m/z wählen, die garantiert von AS stammen



pNovo+ Algorithmus – Datenaufbereitung

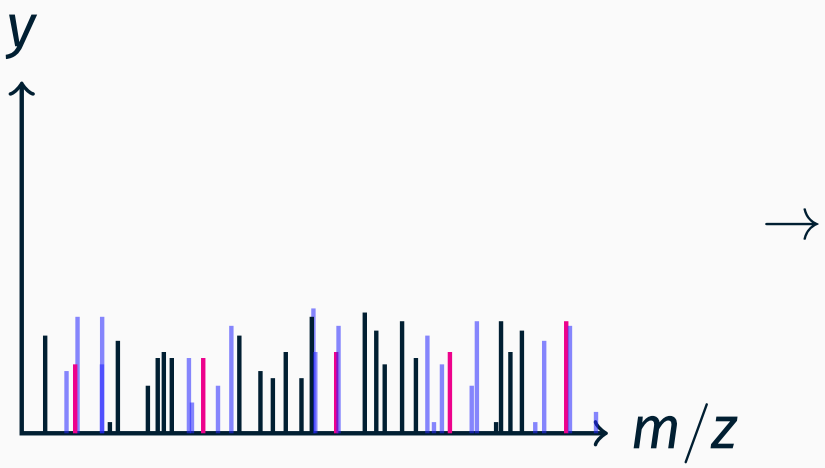
- Hintergrundrauschen könnte auch AS mit Isotop sein
- m/z wählen, die garantiert von AS stammen
- Peaks mit definierten Abstand auswählen

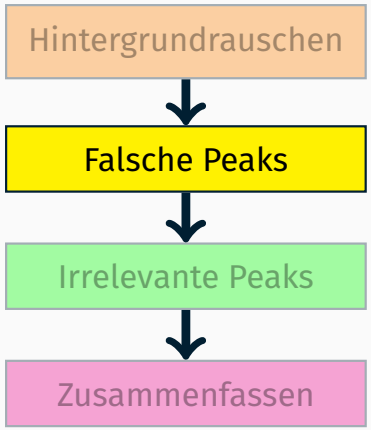




pNovo+ Algorithmus – Datenaufbereitung

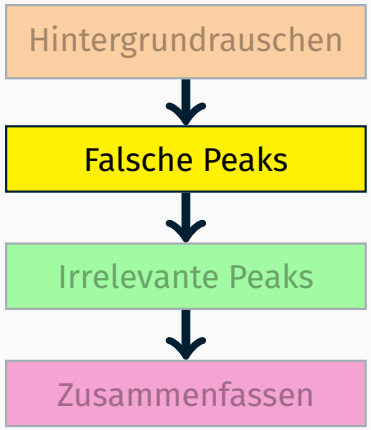
- Hintergrundrauschen könnte auch AS mit Isotop sein
- m/z wählen, die garantiert von AS stammen
- Peaks mit definierten Abstand auswählen

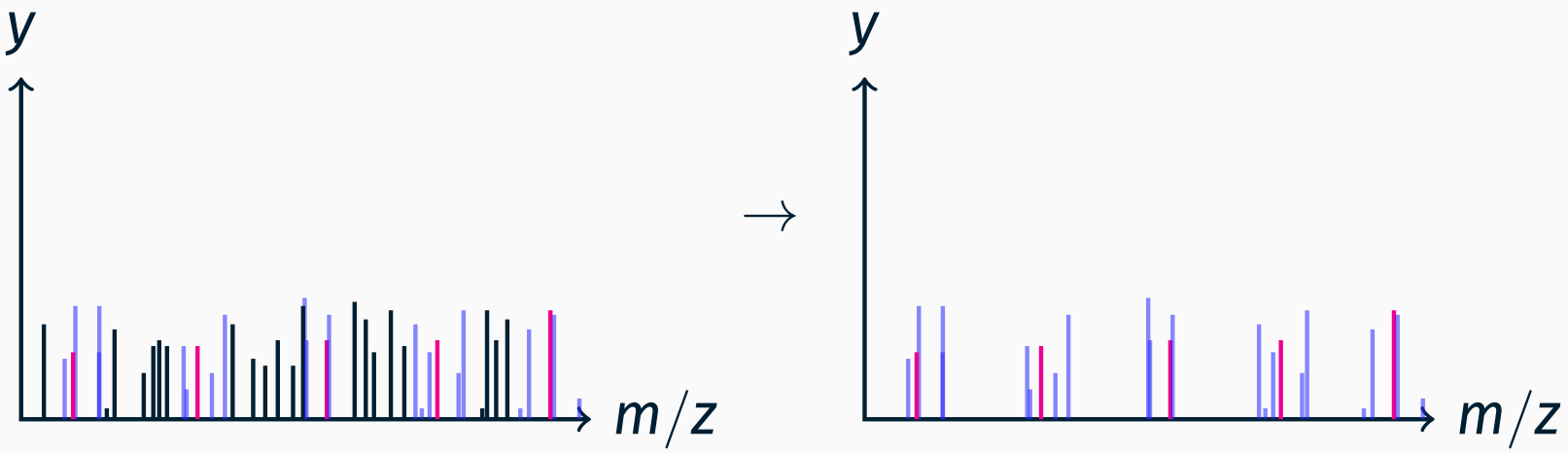




pNovo+ Algorithmus – Datenaufbereitung

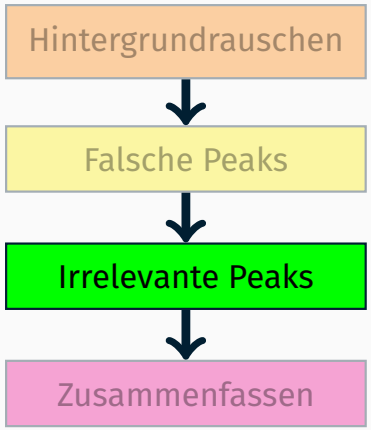
- Hintergrundrauschen könnte auch AS mit Isotop sein
- m/z wählen, die garantiert von AS stammen
- Peaks mit definierten Abstand auswählen





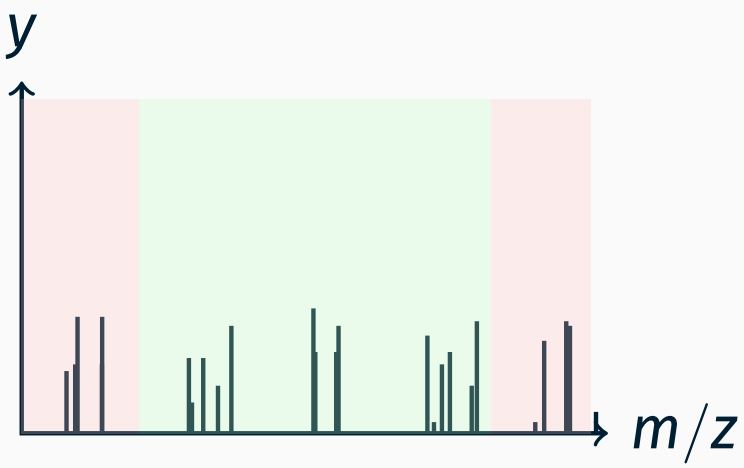
pNovo+ Algorithmus – Datenaufbereitung

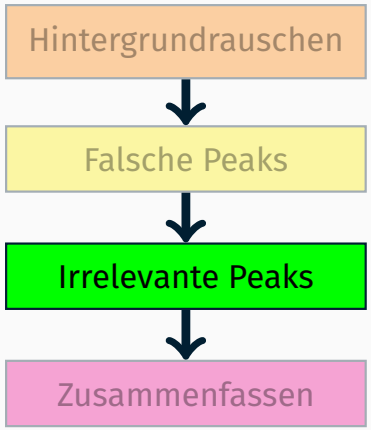
- Peaks aus irrelevantem Intervall entfernen



pNovo+ Algorithmus – Datenaufbereitung

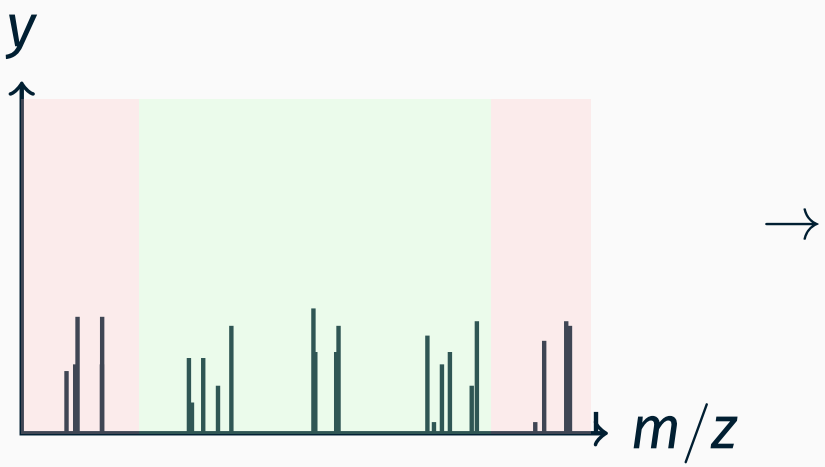
- Peaks aus irrelevantem Intervall entfernen

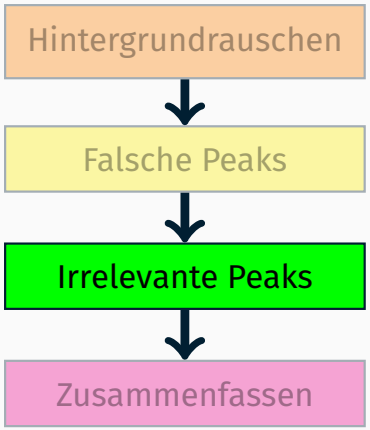




pNovo+ Algorithmus – Datenaufbereitung

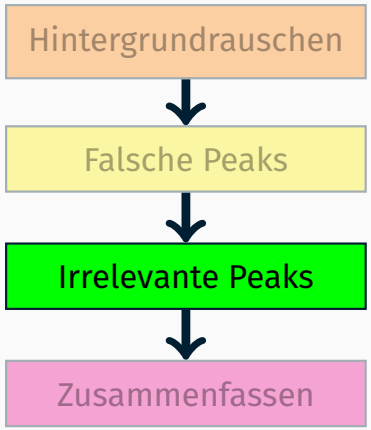
- Peaks aus irrelevantem Intervall entfernen

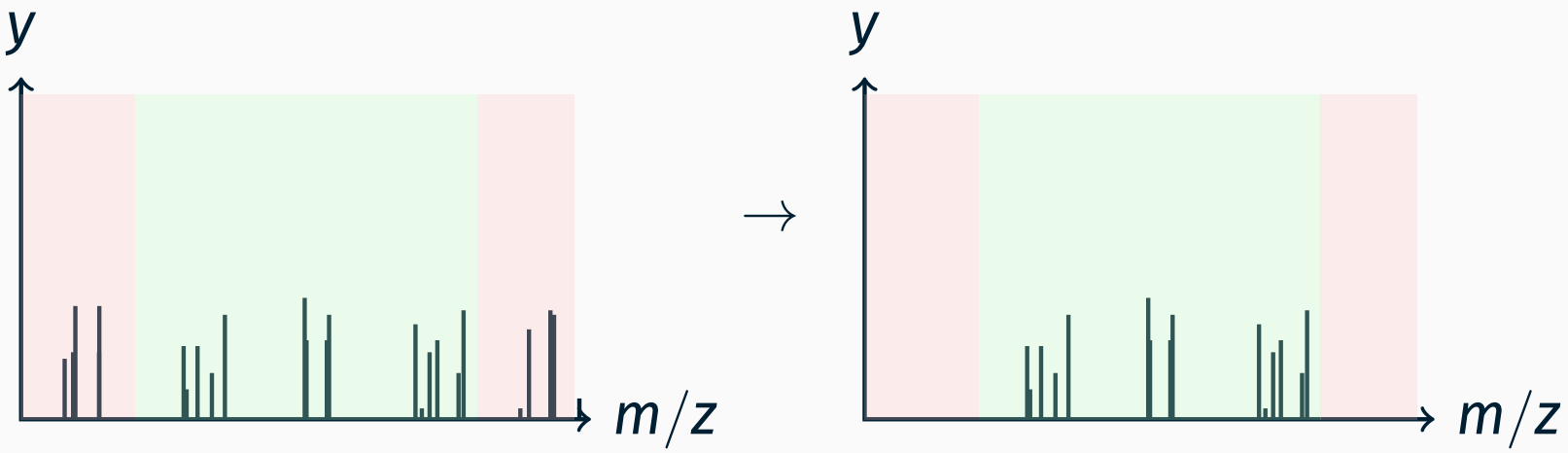




pNovo+ Algorithmus – Datenaufbereitung

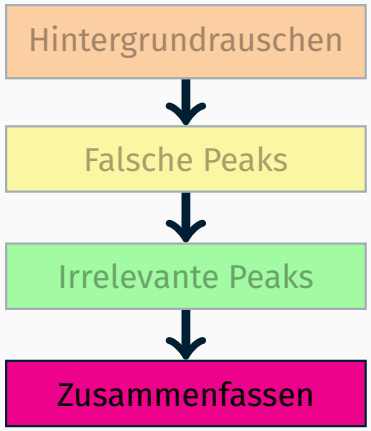
- Peaks aus irrelevantem Intervall entfernen





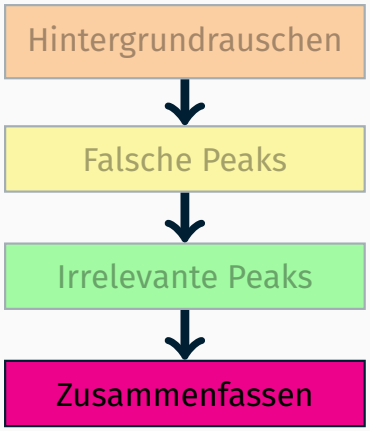
pNovo+ Algorithmus – Datenaufbereitung

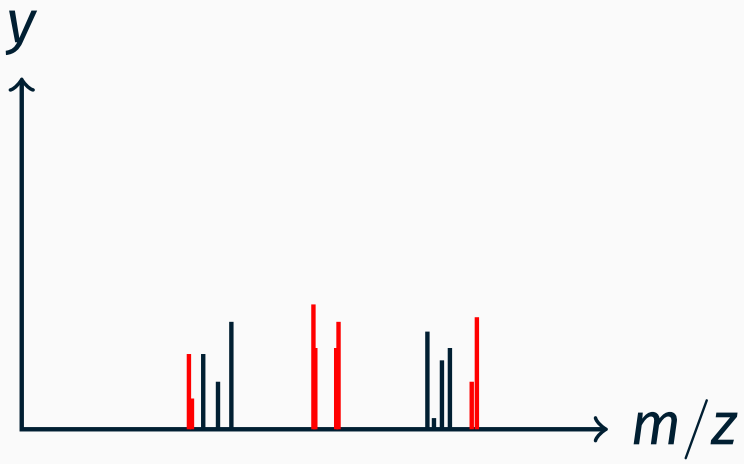
- Zusammenfassen von Peaks



pNovo+ Algorithmus – Datenaufbereitung

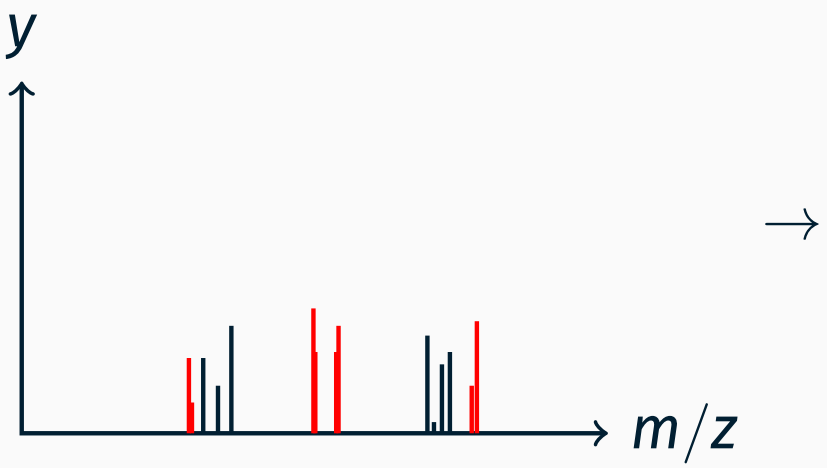
- Zusammenfassen von Peaks
- Abstand einen Schwellwert unterschreitet

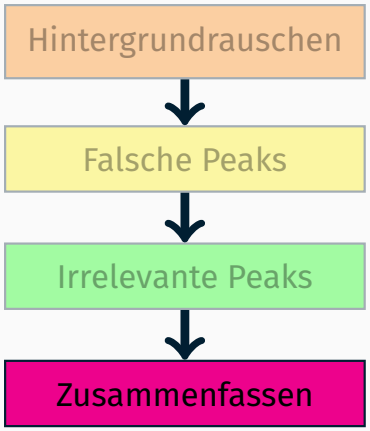




pNovo+ Algorithmus – Datenaufbereitung

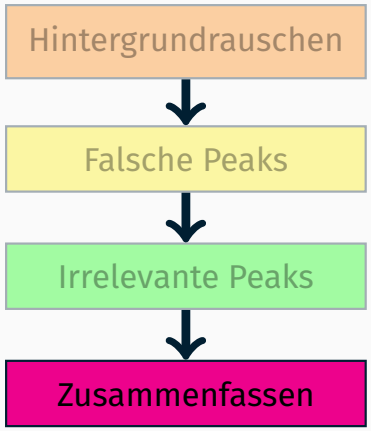
- Zusammenfassen von Peaks
- Abstand einen Schwellwert unterschreitet
- $y =$ Median aus zusammengefassten Peaks

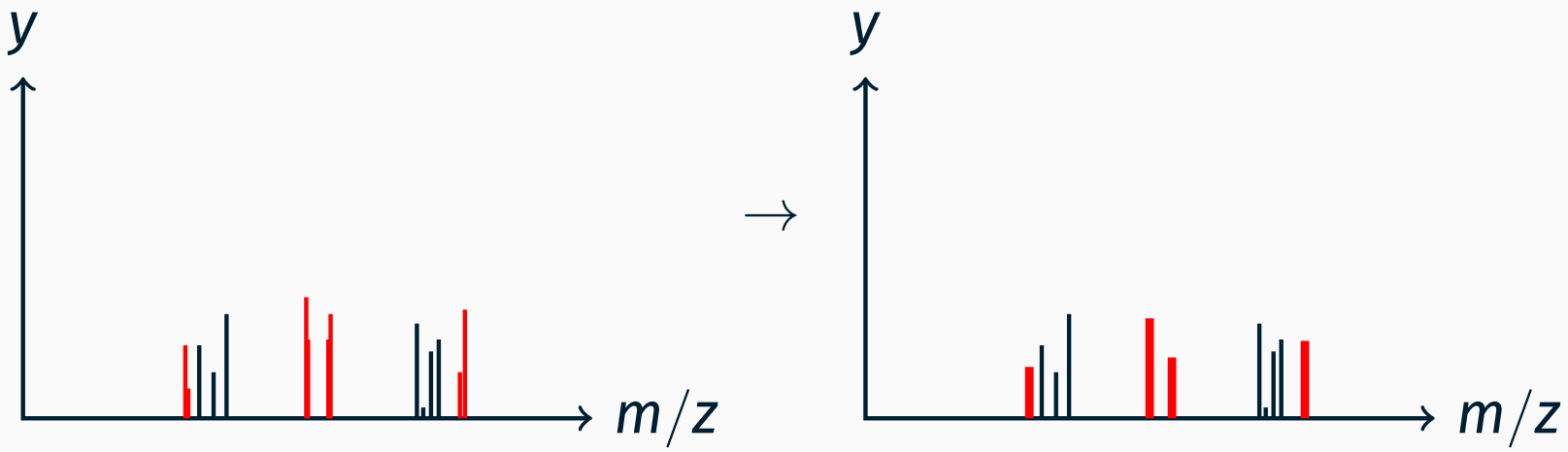




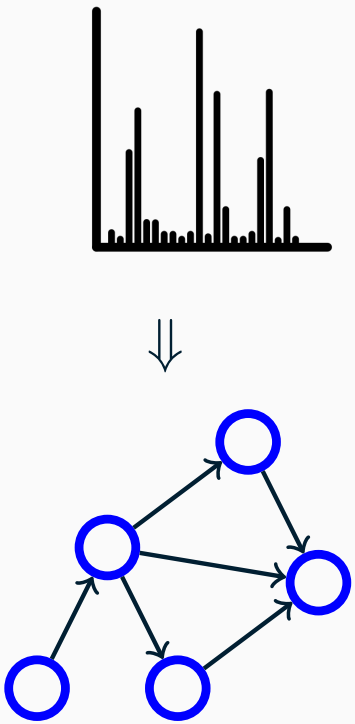
pNovo+ Algorithmus – Datenaufbereitung

- Zusammenfassen von Peaks
- Abstand einen Schwellwert unterschreitet
- $y =$ Median aus zusammengefassten Peaks



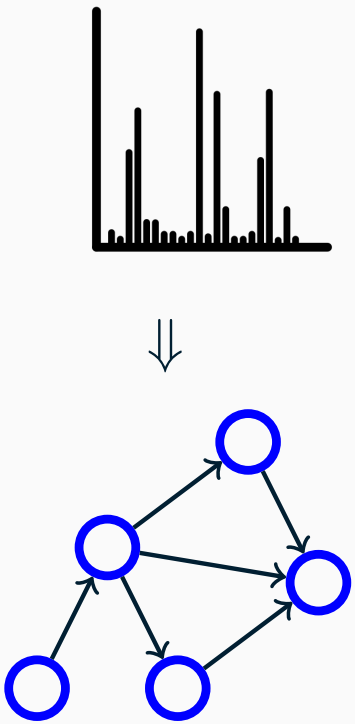


pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



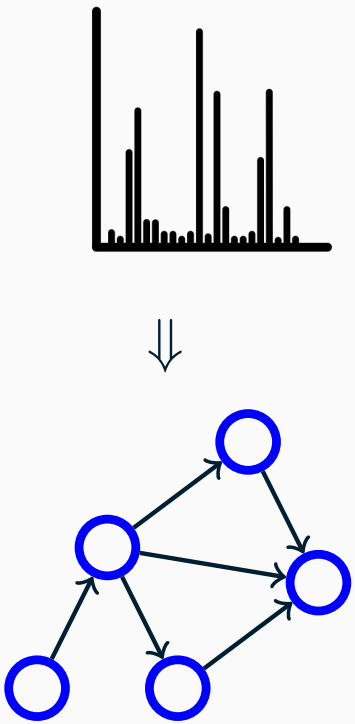
- Verwendung vorverarbeiteter MS2 Spektren

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



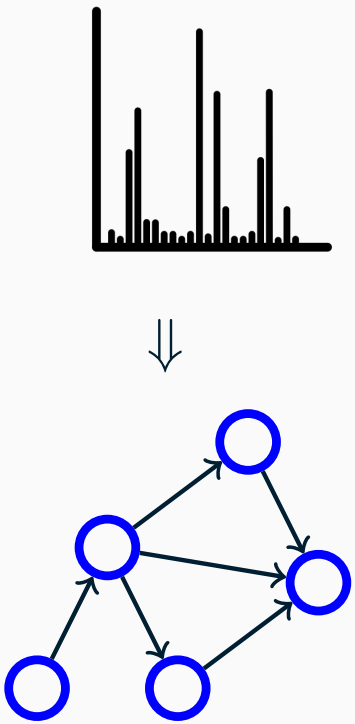
- Verwendung vorverarbeiteter MS2 Spektren
- Peaks $\widehat{=}$ Knoten

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



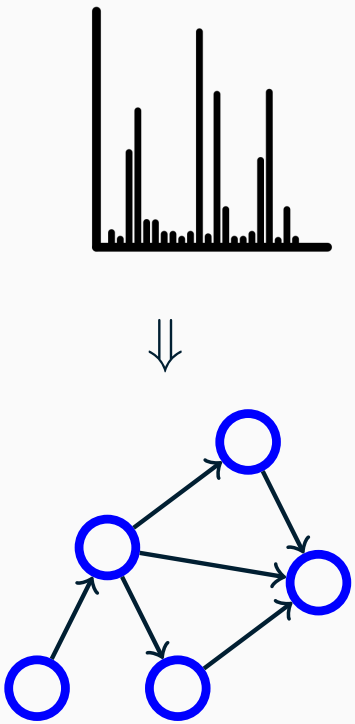
- Verwendung vorverarbeiteter MS2 Spektren
- Peaks $\widehat{=}$ Knoten
- Knoten bekommen eine „Masse“

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



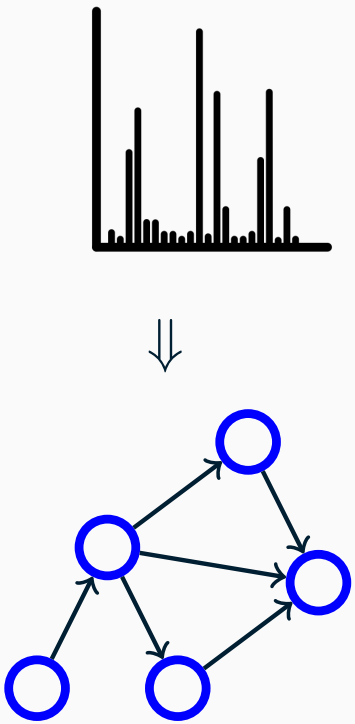
- Verwendung vorverarbeiteter MS2 Spektren
- Peaks $\widehat{=}$ Knoten
- Knoten bekommen eine „Masse“
- Masse $\widehat{=}$ m/z Wert

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



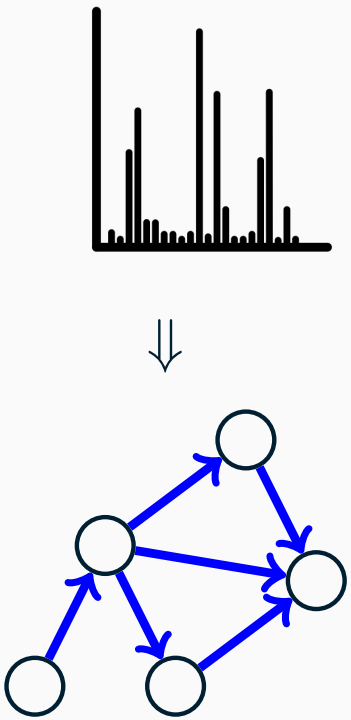
- Verwendung vorverarbeiteter MS2 Spektren
- Peaks $\widehat{=}$ Knoten
- Knoten bekommen eine „Masse“
- Masse $\widehat{=}$ m/z Wert
- Startknoten (Masse = 0)

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



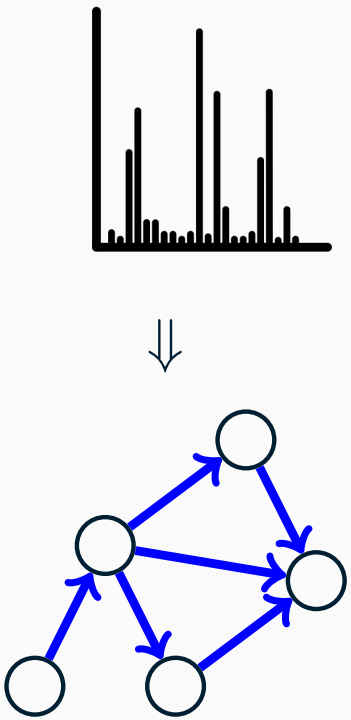
- Verwendung vorverarbeiteter MS2 Spektren
- Peaks $\widehat{=}$ Knoten
- Knoten bekommen eine „Masse“
- Masse $\widehat{=}$ m/z Wert
- Startknoten (Masse = 0)
- Endknoten (Masse = vorheriger Knoten - 18,02)

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



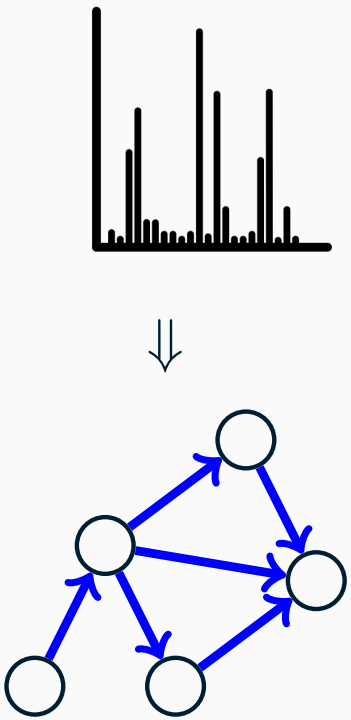
- Gerichtete Kanten zwischen Knotenpaar, wenn:

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



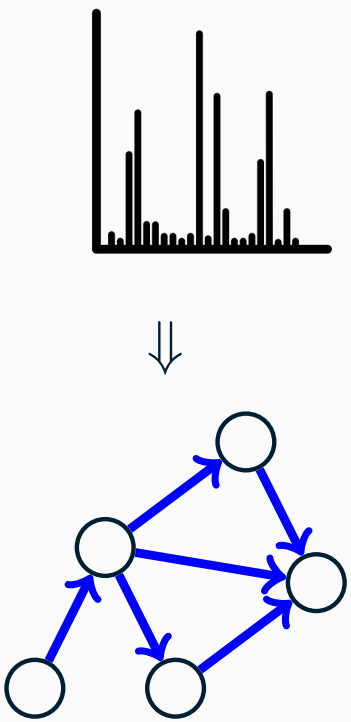
- Gerichtete Kanten zwischen Knotenpaar, wenn:
  - Massendifferenz genau Masse einer AS entspricht

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



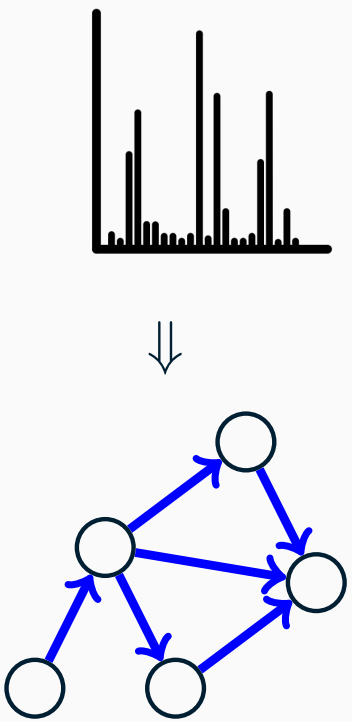
- Gerichtete Kanten zwischen Knotenpaar, wenn:
  - Massendifferenz genau Masse einer AS entspricht
  - Massendifferenz genau Masse zwei AS entsprechen

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



- Gerichtete Kanten zwischen Knotenpaar, wenn:
  - Massendifferenz genau Masse einer AS entspricht
  - Massendifferenz genau Masse zwei AS entsprechen
- $N + \binom{n+N-1}{N-1}$ Differenzen $n = 2$ $N = 20$

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen



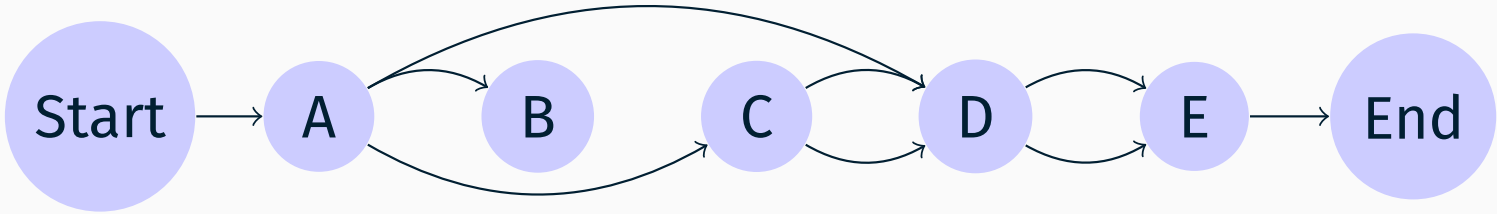
- Gerichtete Kanten zwischen Knotenpaar, wenn:
  - Massendifferenz genau Masse einer AS entspricht
  - Massendifferenz genau Masse zwei AS entsprechen
- $N + \binom{n+N-1}{N-1}$ Differenzen $n = 2$ $N = 20$
- 230 Differenzen

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen

- Ergebnis: Directed acyclic graph (DAG)

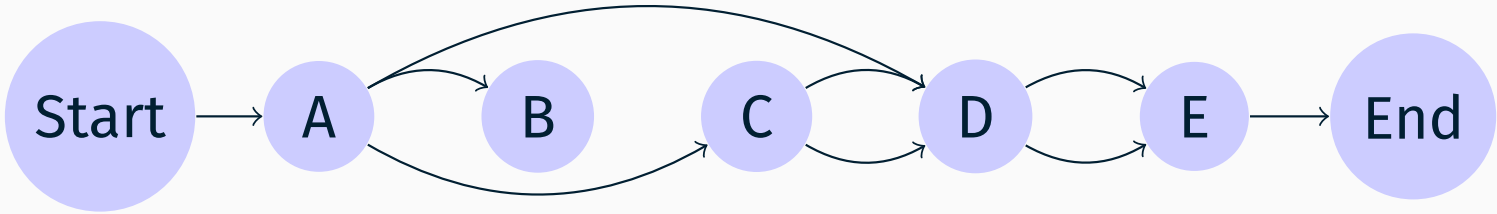
pNovo+ Algorithmus – Bildung eines Spektrumsgraphen

- Ergebnis: Directed acyclic graph (DAG)



pNovo+ Algorithmus – Bildung eines Spektrumsgraphen

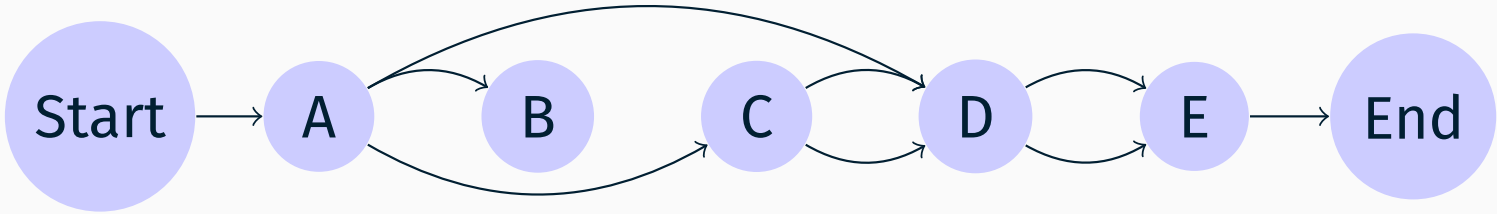
- Ergebnis: Directed acyclic graph (DAG)



- Alle möglichen Pfade von Start nach End ermitteln

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen

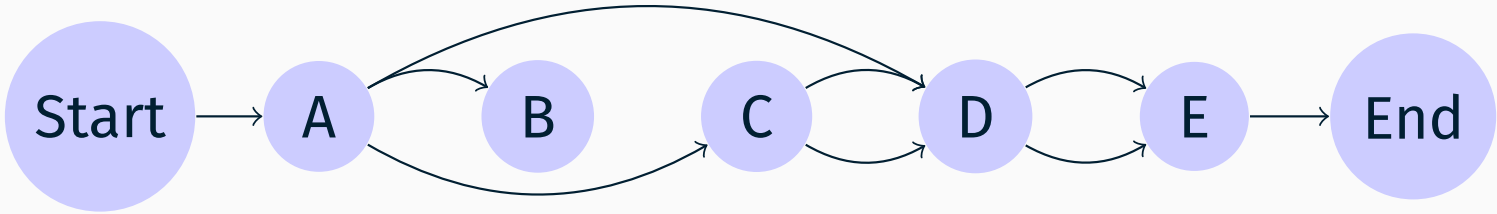
- Ergebnis: Directed acyclic graph (DAG)



- Alle möglichen Pfade von Start nach End ermitteln
- Scoring Funktion „bewertet“ jeden Pfad

pNovo+ Algorithmus – Bildung eines Spektrumsgraphen

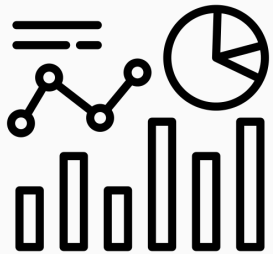
- Ergebnis: Directed acyclic graph (DAG)



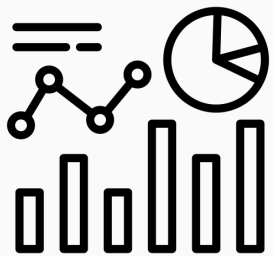
- Alle möglichen Pfade von Start nach End ermitteln
- Scoring Funktion „bewertet“ jeden Pfad
- Pfad mit dem höchsten Scoring Wert ist das Ergebnis

pNovo+ Algorithmus – Evaluierung

- 8677 Datensätze

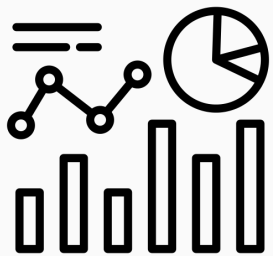


pNovo+ Algorithmus – Evaluierung



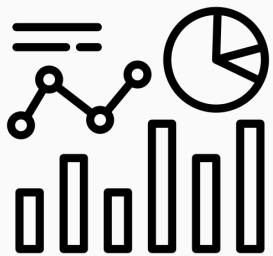
- 8677 Datensätze
- Erfolgreiche Sequenzierungen: $81,2\%$

pNovo+ Algorithmus – Evaluierung



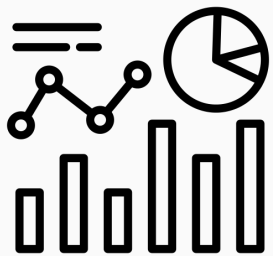
- 8677 Datensätze
- Erfolgreiche Sequenzierungen: 81,2%
- Konkurrenzalgorithmus: 71.8%

pNovo+ Algorithmus – Evaluierung



- 8677 Datensätze
- Erfolgreiche Sequenzierungen: 81,2%
- Konkurrenzalgorithmus: 71.8%
- pNovo+ besser als die Konkurrenz!

pNovo+ Algorithmus – Evaluierung



- 8677 Datensätze
- Erfolgreiche Sequenzierungen: 81,2%
- Konkurrenzalgorithmus: 71.8%
- pNovo+ besser als die Konkurrenz!
- Side Note: pNovo+ ist frei verfügbar

Open-pNovo Algorithmus

Open-pNovo Algorithmus – Hintergrund

- Peptide sind nicht zwingend stabil

Open-pNovo Algorithmus – Hintergrund

- Peptide sind nicht zwingend stabil
- Wechselwirkungen können die Sequenz abändern

Open-pNovo Algorithmus – Hintergrund

- Peptide sind nicht zwingend stabil
- Wechselwirkungen können die Sequenz abändern
- Posttranslationale Proteinmodifikationenen (PTM)

Open-pNovo Algorithmus – Hintergrund

- Peptide sind nicht zwingend stabil
- Wechselwirkungen können die Sequenz abändern
- Posttranslationale Proteinmodifikationenen (PTM)
- Mit De-Novo-Algorithmen an sich kein Problem

Open-pNovo Algorithmus – PTM

- Bildung von nicht proteinogenen AS möglich

Open-pNovo Algorithmus – PTM

- Bildung von nicht proteinogenen AS möglich
- AS, die normalerweise nicht in Peptiden vorkommen

Open-pNovo Algorithmus – PTM

- Bildung von nicht proteinogenen AS möglich
- AS, die normalerweise nicht in Peptiden vorkommen
- Spektrogramm zeigt solche AS

Open-pNovo Algorithmus – PTM

- Bildung von nicht proteinogenen AS möglich
- AS, die normalerweise nicht in Peptiden vorkommen
- Spektrogramm zeigt solche AS
- Änderungen können von pNovo+ nicht erkannt werden

Open-pNovo Algorithmus – PTM

- Bildung von nicht proteinogenen AS möglich
- AS, die normalerweise nicht in Peptiden vorkommen
- Spektrogramm zeigt solche AS
- Änderungen können von pNovo+ nicht erkannt werden

- pNovo+ erzeugt zwangsweise Fehler

Open-pNovo Algorithmus – RankBoost

- Neue Scoring Funktion: RankBoost

Open-pNovo Algorithmus – RankBoost

- Neue Scoring Funktion: RankBoost
- Machine Learning Algorithmus aus 2003

Open-pNovo Algorithmus – RankBoost

- Neue Scoring Funktion: RankBoost
- Machine Learning Algorithmus aus 2003
- Erweiterung des AdaBoost Algorithmus

Open-pNovo Algorithmus – RankBoost

- Neue Scoring Funktion: RankBoost
- Machine Learning Algorithmus aus 2003
- Erweiterung des AdaBoost Algorithmus
- Präferenzen in Datensätzen zu erkennen

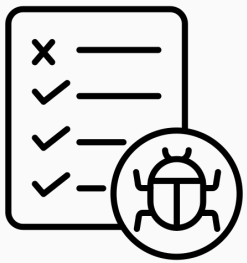
Open-pNovo Algorithmus – RankBoost

- Neue Scoring Funktion: RankBoost
- Machine Learning Algorithmus aus 2003
- Erweiterung des AdaBoost Algorithmus
- Präferenzen in Datensätzen zu erkennen

- Filterung der nicht gültigen AS

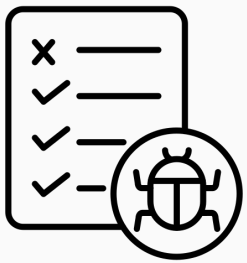
Open-pNovo Algorithmus – Evaluierung von Open-pNovo

- 20259 Datensätze

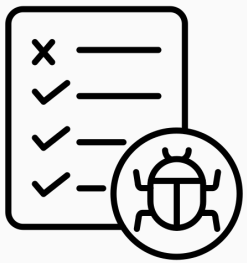


Open-pNovo Algorithmus – Evaluierung von Open-pNovo

- 20259 Datensätze
- Erfolgreiche Sequenzierungen: $76,3\%$

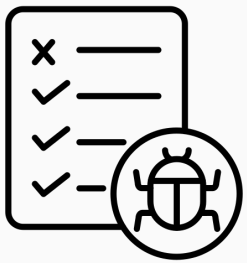


Open-pNovo Algorithmus – Evaluierung von Open-pNovo



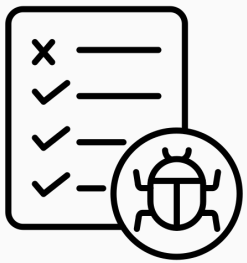
- 20259 Datensätze
- Erfolgreiche Sequenzierungen: $76,3\%$
- pNovo+: $68,5\%$

Open-pNovo Algorithmus – Evaluierung von Open-pNovo



- 20259 Datensätze
- Erfolgreiche Sequenzierungen: $76,3\%$
- pNovo+: $68,5\%$
- Zwei Konkurrenzalgorithmen: $65,8\%$ sowie $39,9\%$

Open-pNovo Algorithmus – Evaluierung von Open-pNovo



- 20259 Datensätze
- Erfolgreiche Sequenzierungen: $76,3\%$
- pNovo+: $68,5\%$
- Zwei Konkurrenzalgorithmen: $65,8\%$ sowie $39,9\%$
- Open-pNovo ebenfalls besser als die Konkurrenz!

Zusammenfassung

Zusammenfassung

- De-Novo-Sequenzierung leichter durchführbar

Zusammenfassung

- De-Novo-Sequenzierung leichter durchführbar
- Beide Algorithmen liefern erstklassige Ergebnisse

Zusammenfassung

- De-Novo-Sequenzierung leichter durchführbar
- Beide Algorithmen liefern erstklassige Ergebnisse
- pNovo+ Ansatz mit Spektrumsgraphen ist wirkungsvoll

Zusammenfassung

- De-Novo-Sequenzierung leichter durchführbar
- Beide Algorithmen liefern erstklassige Ergebnisse
- pNovo+ Ansatz mit Spektrumsgraphen ist wirkungsvoll
- Open-pNovo erkennt zuverlässig Proben mit PTMs