De Novo Peptide Sequencing

Tomas Hruz, Simon Rösch, Thomas Tschager, and Peter Widmayer
Institute of Theoretical Computer Science, ETH Zürich

1 Tandem Mass Spectrometry

Tandem Mass Spectrometry

m/z
m/z
h
m/z
m/z
MS1
MS2
Tandem Mass Spectrometry
ADSFDGF
ADS|FDGF
A|DSFDGF
DAS|DGDF
ADSFDGF
DASDGDF
1
2
3
1
2
3
h
h
h
DASDGDF
DASDG|DF

2 Problem Definition

For a sequence \( s \) over an alphabet \( \Sigma \), the theoretical spectrum \( TS(s) \) of \( s \) is the set of all prefix and suffix masses of \( s \).

Single Peptide De Novo Sequencing

Input: peptide mass \( M \) (measured in MS1), set of fragment masses \( X \) (measured in MS2),
Maximum Intersection: find a sequence \( s \) with mass \( M \) that maximizes \( |X \cap TS(s)| \).
Symmetric Difference: find a sequence \( s \) with mass \( M \) that minimizes \( |X \setminus TS(s)| + |TS(s) \setminus X| \).

Multiple Peptides De Novo Sequencing

Input: \( k \) peptide masses \( M_1, \ldots, M_k \) (MS1), set of fragment masses \( X \) (MS2)
find sequences \( s_1, \ldots, s_k \), such that
(i) sequence \( s_i \) has mass \( M_i \) for \( i = 1, \ldots, k \)
(ii) \( |X \cap \bigcup_{i=1}^{k} TS(s_i)| \) is maximized.

3 Single Peptide De Novo Sequencing

Spectrum graph \( SG = (V, E) \) (directed acyclic graph)
\( V = \{m, M - m \mid m \in X \cup \{0, M\}\} \)
\( E = \{(v, w) \mid v, w \in V, \exists s : s \text{ has mass } w - v\} \)
Maximum Intersection: \( O(|V|\cdot|E|) \)-algorithm (Chen et al., 2001)
Symmetric Difference: \( O(max(|V|, |E|, p \cdot |E|^2)) \)-algorithm, where \( p \) is maximal edge label length (Master Thesis of Simon Rösch)

4 Multiple Peptides De Novo Sequencing

Assumption: \( \Gamma = \max_{i,j} |M_i - M_j| < \min \text{ mass}(s) \)
Outline (for \( k = 2 \)): Compute directed acyclic graph \( G = (V, E) \):
- Vertices \((a, b, c, d)\) with \( a, b \in V(SG_1) \) and \( c, d \in V(SG_2) \)
- Edge from \((a, b, c, d)\) to \((a, b, e, d)\) if \((c, e) \in E(SG_2) \) and \( c = \min(a, M_1 - b, c, M_2 - d) \)
- \( k \) longest paths in \( O(|V| + |E| + k) \) (Eppstein, 1998).

Partners: Dr. Ludovic Gillet and Prof. Ruedi Aebersold, Institute of Molecular Systems Biology, ETH Zurich, Hannes Roest, Department of Genetics, Stanford University.