AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

22BIO211: Intelligence of Biological Systems - 2

LEADERBOARD CYCLOPEPTIDE SEQUENCING.

Dr. Manjusha Nair M
Amrita School of Computing, Amritapuri

Email : manjushanair@am.amrita.edu
Contact No: 9447745519

Leaderboard Cyclopeptide Sequencing

- In our Branch and Bound algorithm with scoring function,
  - *To limit the number of candidate linear peptides under consideration, we will replace Peptides with* ***Leaderboard*** *for further extension.*
  - *Leaderboard - N highest scoring candidate peptides*
- At each step, expand all candidate peptides found in Leaderboard
- then eliminate those peptides whose newly calculated scores are not high enough to keep them on the Leaderboard.
- Leaderboard should be trimmed down to the “N highest-scoring linear peptides including ties”.

Leaderboard Cyclopeptide Sequencing

- Given a list of peptides *Leaderboard,* a spectrum *Spectrum*, and an integer *N*,
  - *define Trim(Leaderboard, Spectrum, N)*
  - *as the collection of the top N highest-scoring linear peptides in Leaderboard (including ties) with respect to Spectrum.*
- LeaderboardCyclopeptideSequencing is a heuristic, not guaranteed to correctly solve the Cyclopeptide Sequencing Problem

Leaderboard Cyclopeptide Sequencing

```
LeaderboardCyclopeptideSequencing(Spectrum, N)
  Leaderboard ← set containing only the empty peptide
  LeaderPeptide ← empty peptide
  while Leaderboard is non-empty
    Leaderboard ← Expand(Leaderboard)
    for each Peptide in Leaderboard
      if Mass(Peptide) = ParentMass(Spectrum)
        if Score(Peptide, Spectrum) > Score(LeaderPeptide, Spectrum)
          LeaderPeptide ← Peptide
      else if Mass(Peptide) > ParentMass(Spectrum)
        remove Peptide from Leaderboard
    Leaderboard ← Trim(Leaderboard, Spectrum, N)
  output LeaderPeptide
```

Leaderboard Cyclopeptide Sequencing

- Consider Spectrum_{10} of Tyrocidine B1 with 10% missing or false masses
  - *blue masses are not actually in the spectrum, but we show them so that it is clear which masses are missing.*

0	97	99	**113**	114	**128**	128	147	147	163	186	227
241	242	244	260	261	262	283	291	333	340	357	**385**
388	389	390	390	405	430	430	447	485	487	503	504
518	543	544	552	575	577	584	**631**	632	650	651	671
672	690	691	738	745	747	770	778	779	804	818	819
820	835	837	875	892	**892**	917	932	932	933	934	965
982	989	**1030**	**1031**	1039	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322	

- Applying LeaderboardCyclopeptideSequencing to this spectrum (with N = 1000) results in the correct cyclic peptide VKLFPWFNQY, which has a score of 86

Leaderboard Cyclopeptide Sequencing

- As the number of errors in the spectrum increases, so does the likelihood that this algorithm will return an incorrect peptide.
- Consider $Spectrum_{25}$ of Tyrocidine B1 with 25% missing or false masses.

0	97	99	113	114	**115**	128	128	147	147	163	186
227	241	242	**244**	244	**256**	260	261	262	283	291	**309**
330	333	340	**347**	**357**	**385**	388	389	390	390	405	**430**
430	**435**	447	485	487	503	504	518	**543**	544	552	575
577	584	**599**	**608**	631	632	650	651	**653**	**671**	672	690
691	**717**	738	745	**747**	770	**778**	779	804	818	819	**827**
835	837	875	892	892	917	932	932	933	934	965	982
989	**1031**	1039	1060	**1061**	1062	1078	1080	1081	1095	1136	1159
1175	1175	1194	1194	1208	1209	1223	**1225**	1322			

- Applying LeaderboardCyclopeptideSequencing to this spectrum (with N = 1000) identifies
  - *VKLFP**AD**FNQY (score: 83) as a highest-scoring cyclic peptide*
  - *instead of the correct peptide VKLFP**W**FNQY (score: 82).*
  - *These two peptides are similar, owing to the fact that the combined mass of A (71) and D (115) is equal to the mass of W (186).*

Leaderboard Cyclopeptide Sequencing

- Although the correct and incorrect peptides are similar, their amino acid compositions differ.
- If we could figure out the amino acid composition of Tyrocidine B1 from its spectrum alone and run LeaderboardCyclopeptideSequencing on this smaller alphabet (rather than on the alphabet of all amino acids), then we could eliminate the incorrect peptide VKLFPADFNQY from consideration.
- Pros:
  - ○ *Works with imperfect spectra*
- Cons:
  - ○ *Still very slow.*
  - ○ *Doesn't always accurately reproduce the correct peptide.*

Summary

- Leaderboard Cyclopeptide sequencing
  - *Leaderboard*
  - *Algorithm*
  - *Results and Scores for Tyrocidene B1*