Structural Bioinformatics Lab 4

Ain Shams University
Computer and Information Sciences
Bioinformatics 3rd Year
TA. Esraa Hamdi – TA Walaa El Kady – TA Islam Sharshar

Agenda

- Cyclopeptide Sequencing Problem

- Sequencing Antibiotics by Shattering them into Pieces:
  - Branch and Bound Approach

Cyclopeptide Sequencing Problem

- What about reversing it?

- What about going from the spectrum to the peptide?

Cyclopeptide Sequencing Problem

- The mass spectrometer generates a Spectrum And then we want to determine a peptide that came from it.

- That's going to be a harder problem, and we call this the **Cyclopeptide Sequencing Problem**,

  To reconstruct a cyclic peptide from its theoretical spectrum.

Cyclopeptide Sequencing Problem

Spectrum of TMDH				Spectrum of NQEL	
" "	0			" "	0
T	101			L	113
D	115			N	114
M	131			Q	128
H	137	Their spectra completely		E	129
TM	232	disagree!		LN	227
HT	238			NQ	242
MD	246			EL	242
DH	252	How can we use this?		QE	257
TMD	347			LNQ	355
DHT	353			ELN	356
HTM	369			QEL	370
MDH	383			NQE	371
TMDH	484			NQEL	484

5

Branch and Bound

Let's run CYCLOPEPTIDESEQUENCING on the following *Spectrum*:

0	97	97	99	101	103	196	198	198	200	202
295	297	299	299	301	394	396	398	400	400	497

Download This Spectrum

CYCLOPEPTIDESEQUENCING first expands *List* into the set of all 1-mers consistent with *Spectrum*:

97	99	101	103
P	V	T	C

The algorithm next appends each of the 18 amino acid masses to each of the 1-mers above. The resulting *List* containing $4 \cdot 18 = 72$ peptides of length 2 is then trimmed to keep only the 10 peptides that are consistent with *Spectrum*:

97-99	97-101	97-103	99-97	99-101
PV	PT	PC	VP	VT
99-103	101-97	101-99	103-97	103-99
VC	TP	TV	CP	CV

Branch and Bound

After expansion and trimming in the next iteration, *List* contains 15 consistent 3-mers:

97-99-103	**97-99-101**	97-101-97	**97-101-99**	97-103-99
PVC	**PVT**	PTP	**PTV**	PCV
99-97-103	99-97-101	**99-101-97**	99-103-97	101-97-99
VPC	VPT	**VTP**	VCP	TPV
101-97-103	**101-99-97**	103-97-101	**103-97-99**	103-99-97
TPC	**TVP**	CPT	**CPV**	CVP

Branch and Bound

With one more iteration, *List* contains 10 consistent 4-mers. Observe that the six 3-mers highlighted in red above failed to expand into any 4-mers below, and so we now know that CYCLOPEPTIDESEQUENCING may generate some incorrect *k*-mers.

97-99-103-97	97-101-97-99	97-101-97-103	97-103-99-97	99-97-101-97
PVCP	PTPV	PTPC	PCVP	VPTP
99-103-97-101	101-97-99-103	101-97-103-99	103-97-101-97	103-99-97-101
VCPT	TPVC	TPCV	CPTP	CVPT

Branch and Bound

In the final iteration, we generate 10 consistent 5-mers:

97-99-103-97-101	97-101-97-99-103	97-101-97-103-99	97-103-99-97-101	99-97-101-97-103
PVCPT	PTPVC	PTPCV	PCVPT	VPTPC
99-103-97-101-97	101-97-99-103-97	101-97-103-99-97	103-97-101-97-99	103-99-97-101-97
VCPTP	TPVCP	TPCVP	CPTPV	CVPTP

Thank you

Any Questions ?