Úvod do dynamického programovania, proteomika

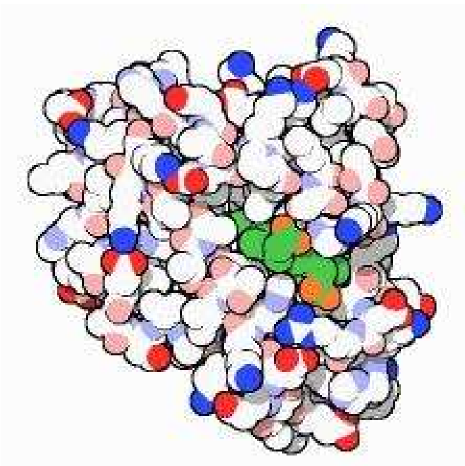
Askar Gafurov

7.10.2021

Proteomika

Proteín: sekvencia pozostáva z 20 rôznych aminokyselín

```
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASE
DLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKH
PGDFGADAQGAMNKALELFRKDMASNYKELGFQG
```



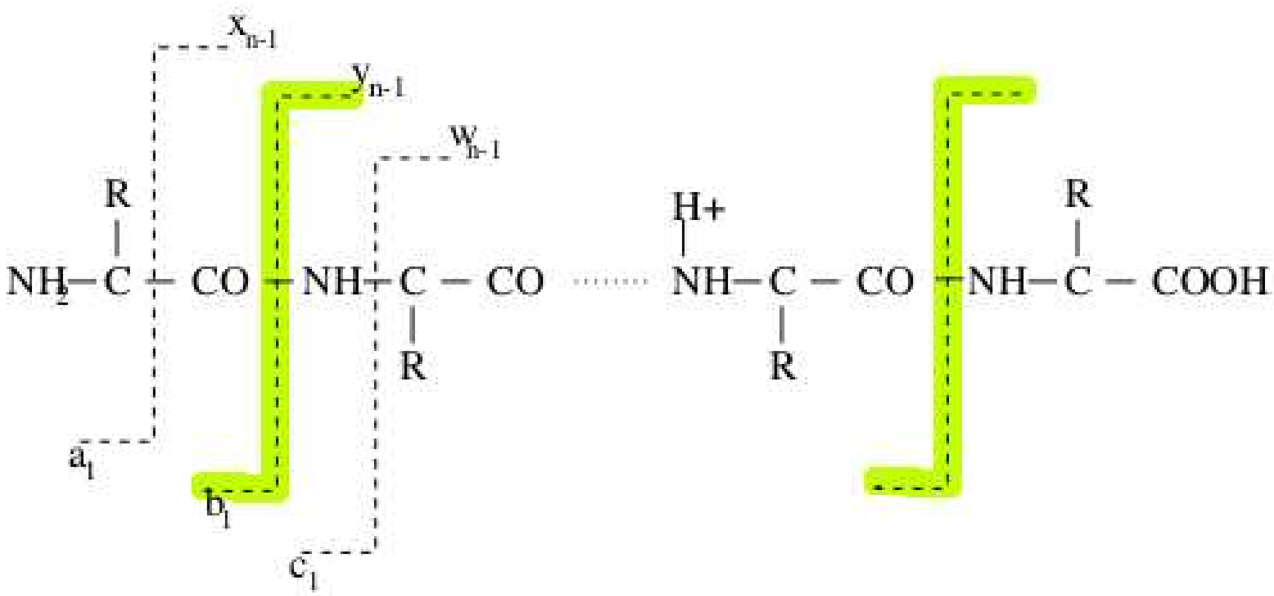
Z bunky sme izolovali určitý proteín, chceme zistiť jeho sekvenciu.

Hmotnostná spektrometria (mass spectometry)

- Meria pomer hmostnosť/náboj molekúl vo vzorke
- Používa sa na identifikáciu proteínov
- Proteín nasekáme enzýmom trypsín (seká na `[KR]{P}`) na peptidy
- Meriame hmostnosť kúskov, porovnáme s databázou proteínov.
- Tandemová hmotnostná spektrometria (MS/MS) ďalej fragmentuje každý kúsok a dosiahne podrobnejšie spektrum, ktoré obsahuje viac informácie
- V niektorých prípadoch tak vieme sekvenciu proteínu určiť priamo z MS/MS, bez databázy proteínov

Tandemová hmotnostná spektrometria MS/MS

Štiepenie peptidu na prefixy a sufixy



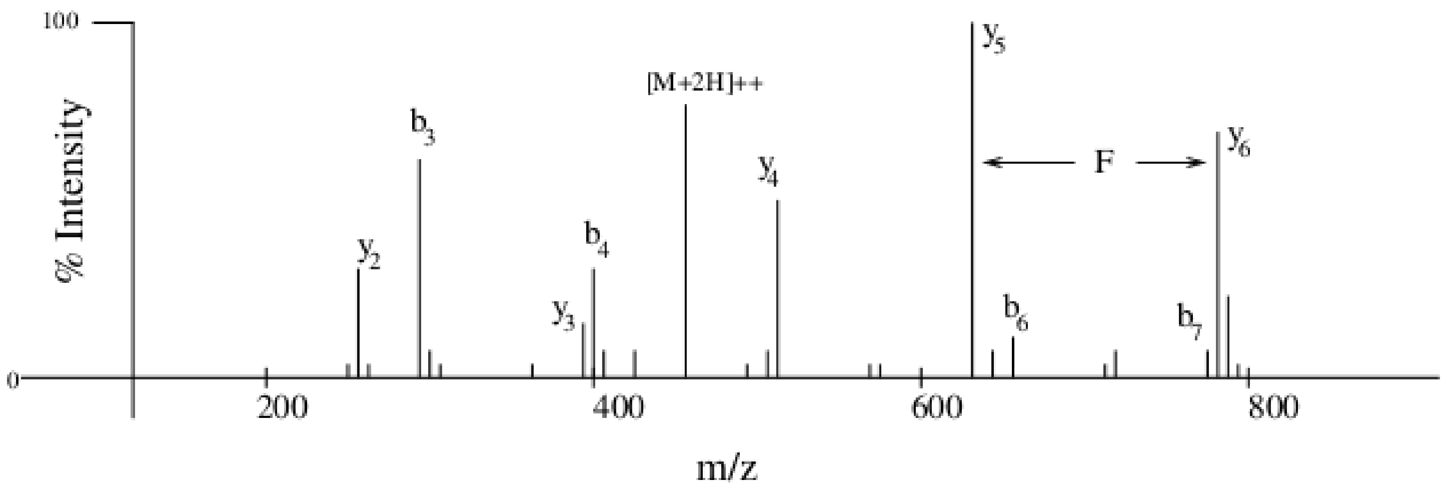
zdroj: Bafna and Reinert

b-ióny: prefixy

y-ióny: sufixy

Tandemová hmotnostná spektrometria MS/MS

88	145	292	405	534	663	778	924	b-ions
S	G	F	L	E	E	D	K	
924	837	780	633	520	391	262	141	y-ions



zdroj: Bafna and Reinert

Sekvenovanie peptidov pomocou MS/MS

Vstup: celková hmotnosť peptidu M,
hmotnosti aminokyselín $a[1], \ldots, a[20]$ (celé čísla),
spektrum ako tabuľka $f[0], \ldots, f[M]$, ktorá hmotnosti určí skóre podľa signálu v okolí príslušného bodu grafu

Označenie:
Nech $x = x_1 \ldots x_k$ je postupnosť aminokyselín
Nech $m(x) = \sum_{j=1}^{k} a[x_j]$ je hmotnosť x
Nech $\mathcal{M}_P(x) = \{m(x_1 \ldots x_j) \mid j = 1, \ldots, k\}$ sú hmotnosti prefixov x
Nech $\mathcal{M}_S(x) = \{m(x_j \ldots x_k) \mid j = 1, \ldots, k\}$ sú hmotnosti sufixov x

Problém 1: uvažujeme iba b-ióny (prefixy)

Výstup: postupnosť aminokyselín x taká, že $m(x) = M$ a $\sum_{m \in \mathcal{M}_P(x)} f[m]$ je maximálna možná

Príklad

Uvažujme len 3 aminokyseliny X,Y,Z

$M = 23$, $a[X] = 4$, $a[Y] = 6$, $a[Z] = 7$

m	4	6	7	11	12	17	18	19
$f[m]$	1	1	1	1	1	1	1	1

Hmotnosti prefixov $\mathcal{M}_P(XZYY) =$
$\{m(), m(X), m(XZ), m(XZYY), m(XZYY)\} = \{0, 4, 11, 17, 23\}$

Hmotnosti sufixov $\mathcal{M}_S(XZYY) =$
$\{m(), m(Y), m(YY), m(ZYY), m(XZYY)\} = \{0, 6, 12, 19, 23\}$

Skóre XZYY: $\sum_{m \in \mathcal{M}_P(ZYXX)} f[m] = 0 + 1 + 1 + 1 + 0 = 3$

Skóre XZXXX: $\sum_{m \in \mathcal{M}_P(ZYZZZ)} f[m] =$
$f[0] + f[4] + f[11] + f[15] + f[19] + f[23] = 0 + 1 + 1 + 0 + 1 + 0 = 3$

Sekvenovanie peptidov pomocou MS/MS

Problém 2: uvažujeme prefixy aj sufixy, sčítame ich skóre

Výstup: postupnosť aminokyselín x taká, že $m(x) = M$ a $\sum_{m \in \mathcal{M}_P(x)} f[m] + \sum_{m \in \mathcal{M}_S(x)} f[m]$ je maximálna možná

Problém 3: uvažujeme prefixy aj sufixy, sčítame ich skóre, ale každú hmotnosť započítame najviac raz

Výstup: postupnosť aminokyselín x taká, že $m(x) = M$ a $\sum_{m \in \mathcal{M}_P(x) \cup \mathcal{M}_S(x)} f[m]$ je maximálna možná

Príklad

$M = 23$, $a[X] = 4$, $a[Y] = 6$, $a[Z] = 7$

m	4	6	7	11	12	17	18	19
$f[m]$	1	1	1	1	1	1	1	1

$\mathcal{M}_P(XZYY) = \{0, 4, 11, 17, 23\}$ $\quad$ $\mathcal{M}_S(XZYY) = \{0, 6, 12, 19, 23\}$

$\mathcal{M}_P(XZXXX) = \{0, 4, 11, 15, 19, 23\}$
$\mathcal{M}_S(XZXXX) = \{0, 4, 8, 12, 19, 23\}$

Problém 2: $\sum_{m \in \mathcal{M}_P(x)} f[m] + \sum_{m \in \mathcal{M}_S(x)} f[m]$
Skóre XZYY: $0 + 1 + 1 + 1 + 0 + 0 + 1 + 1 + 1 + 0 = 6$
Skóre XZXXX: $0 + 1 + 1 + 0 + 1 + 0 + 0 + 1 + 0 + 1 + 1 + 0 = 6$

Problém 3: $\sum_{m \in \mathcal{M}_P(x) \cup \mathcal{M}_S(x)} f[m]$
XZYY: $\{0, 4, 6, 11, 12, 17, 19, 23\}$, $1 + 1 + 1 + 1 + 1 + 1 + 0 = 6$
XZXXX: $\{0, 4, 8, 11, 12, 15, 19, 23\}$, $1 + 0 + 1 + 1 + 0 + 1 + 0 = 4$

Ekvivalencia problémov

Problém 2: maximalizujeme $\sum_{m \in \mathcal{M}_P(x)} f[m] + \sum_{m \in \mathcal{M}_S(x)} f[m]$

Iná formulácia: maximalizujeme $\sum_{m \in \mathcal{M}_p(x)} g[m]$
kde $g[m] = f[m] + f[M - m]$

Ekvivalencia problémov

Problém 3: maximalizujeme $\sum_{m\in\mathcal{M}_P(x)\cup\mathcal{M}_S(x)} f[m]$

Iná formulácia: maximalizujeme $\sum_{m\in\mathcal{M}_p(x)\cup\mathcal{M}_S(x),m\leq M/2} h[m]$

kde $h[m] = \begin{cases} f[m] + f[M-m] & \text{ak } m < M/2 \\ f[m] & \text{ak } m = M/2 \end{cases}$