

# Kako sekvenciramo antibiotike?

Algoritmi grube sile

*Bioinformatics Algorithms:  
an Active Learning Approach*

~Poglavlje 4~

# Pregled

- **Otkriće antibiotika**

- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grube sile za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do preko 100 aminokiselina
- Spektralna konvolucija
- Spektri u realnosti

# Šta su antibiotici?

**Antibiotik** → „supstanca koja ubija bakterije“

- Postoje u prirodi
- Proizvode ih organizmi iz porodice gljiva (npr. buđi) i bakterija



# Antibiotici na molekularnom nivou

Posmatraćemo *tirocidin B1*, antibiotik koji proizvodi bakterija *Bacillus Brevis*

Tirocidin B1 je **peptid** (mali protein):  
kratka niska *aminokiselina*

Valine	Leucine	Proline	Phenylalanine	Glutamine					
<b>Val</b>	<b>Lys</b>	<b>Leu</b>	<b>Phe</b>	<b>Pro</b>	<b>Trp</b>	<b>Phe</b>	<b>Asn</b>	<b>Gln</b>	<b>Tyr</b>
<b>V</b>	<b>K</b>	<b>L</b>	<b>F</b>	<b>P</b>	<b>W</b>	<b>F</b>	<b>N</b>	<b>Q</b>	<b>Y</b>
	Lysine	Phenylalanine	Tryptophan	Asparagine	Tyrosine				

# Pitanja

Zbog čega su antibiotici posebni peptidi?

Kako nastaju antibiotici?

Kako sekvencioniramo antibiotike?

???-???-???-???-???-???-???-???-???-???

# Pitanja

Zbog čega su antibiotici posebni peptidi?

Kako nastaju antibiotici?

Kako sekvencioniramo antibiotike?

**Val-Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr**

# Pregled

- Otkriće antibiotika
- **Kako bakterije prave antibiotike?**
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grubom silom za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do više od 100 aminokiselina
- Spektralna konvolucija
- Spektri u realnosti

# Kako se prave proteini?

DNK se prepisuje u RNK

DNA

5' GTGAAACTTTTTTCCTTGGTTTAATCAATAT 3'  
3' CACTTTGAAAAAGGAACCAAATTAGTTATA 5'

# Kako se prave proteini?

DNK se prepisuje u RNK

Transcribed RNA

GUGAAACUUUUUCCUUGGUUUAAUCAUAU

DNA

5' GTGAAACTTTTTTCCTTGGTTTAATCAATAT 3'

3' CACTTTGAAAAAGGAACCAAATTAGTTATA 5'

Transcribed RNA

CACUUUGAAAAAGGAACCAAUUAGUUAUA

# Kako se prave proteini?

DNK se prepisuje u RNK

Transcribed RNA

GUGAAACUUUUUCCUUGGUUUUAAUCAUAU

DNA

5' GTGAAACTTTTTCCTTGGTTTAATCAATAT 3'

3' CACTTTGAAAAAGGAACCAAATTAGTTATA 5'

Transcribed RNA

CACUUUGAAAAAGGAACCAAUUAGUUAUA

T (timin) se menja sa U (uracil)

# Kako se prave proteini?

RNK se **prevodi** u proteine

# Kako se prave proteini?

RNK se **prevodi** u proteine

**A**denin

**C**itosin

4 nukleotida

**G**uanin

**U**racil

# Kako se prave proteini?

RNK se **prevodi** u proteine

**A**denin

**C**itosin

**G**uanin

**U**racil

4 nukleotida  20 aminokiselina

Amino acid	3-letter code
Alanine	Al a
Cysteine	Cys
Aspartic acid	Asp
Glutamic acid	Gl u
Phenylalanine	Phe
Glycine	Gl y
Histidine	Hi s
Isoleucine	I l e
Lysine	Lys
Leucine	Leu
Methionine	Met
Asparagine	Asn
Proline	Pr o
Glutamine	Gl n
Arginine	Ar g
Serine	Ser
Threonine	Thr
Valine	Val
Tryptophan	Trp
Tyrosine	Tyr

# Kako se prave proteini?

Možemo li da prevodimo po dva nukleotida u jednu aminokiselinu?

AA GA  
AC GC  
AG GG  
AU GU  
CA UA  
CC UC  
CG UG  
CU UU

→  
16 bigrama  
20 aminokiselina

Amino acid	3-letter code
Alanine	Al a
Cysteine	Cys
Aspartic acid	Asp
Glutamic acid	Gl u
Phenylalanine	Phe
Glycine	Gl y
Histidine	Hi s
Isoleucine	I l e
Lysine	Lys
Leucine	Leu
Methionine	Met
Asparagine	Asn
Proline	Pr o
Glutamine	Gl n
Arginine	Ar g
Serine	Ser
Threonine	Thr
Valine	Val
Tryptophan	Trp
Tyrosine	Tyr

# Kako se prave proteini?

Možemo li da prevodimo po tri nukleotida u jednu aminokiselinu?

AAA CAA GAA UAA  
AAC CAC GAC UAC  
AAG CAG GAG UAG  
AAU CAU GAU UAU  
ACA CCA GCA UCA  
ACC CCC GCC UCC  
ACG CCG GCG UCG  
ACU CCU GCU UCU  
AGA CGA GGA UGA  
AGC CGC GGC UGC  
AGG CGG GGG UGG  
AGU CGU GGU UGU  
AUA CUA GUA UUA  
AUC CUC GUC UUC  
AUG CUG GUG UUG  
AUU CUU GUU UUU

→  
64 trigrama  
20 aminokiselina

Amino acid	3-letter code
Alanine	Al a
Cysteine	Cys
Aspartic acid	Asp
Glutamic acid	Gl u
Phenylalanine	Phe
Glycine	Gl y
Histidine	Hi s
Isoleucine	Il e
Lysine	Lys
Leucine	Leu
Methionine	Met
Asparagine	Asn
Proline	Pr o
Glutamine	Gl n
Arginine	Ar g
Serine	Ser
Threonine	Thr
Valine	Val
Tryptophan	Trp
Tyrosine	Tyr

# Kako se prave proteini?

**Kodon:** Triplet (trigram) nukleotida

**Genetski kod:** preslikavanje skupa kodona u skup aminokiselina

# Kako se prave proteini?

**Kodon:** Triplet (trigram) nukleotida

**Genetski kod:** preslikavanje skupa kodona u skup aminokiselina

UGG



Trp  
(W)

# Kako se prave proteini?

**Kodon:** Triplet (trigram) nukleotida

**Genetski kod:** preslikavanje skupa kodona u skup aminokiselina

CUA  
CUC  
CUG  
CUU  
UUA  
UUG



Lys  
(L)

# Kako se prave proteini?

**Kodon:** Triplet (trigram) nukleotida

**Genetski kod:** preslikavanje skupa kodona u skup aminokiselina

## Stop Codons

UAA

UAG

UGA



# Centralna dogma molekularne biologije

**DNK**

# Centralna dogma molekularne biologije

DNK  $\xrightarrow{\text{Transkripcija}}$  RNK

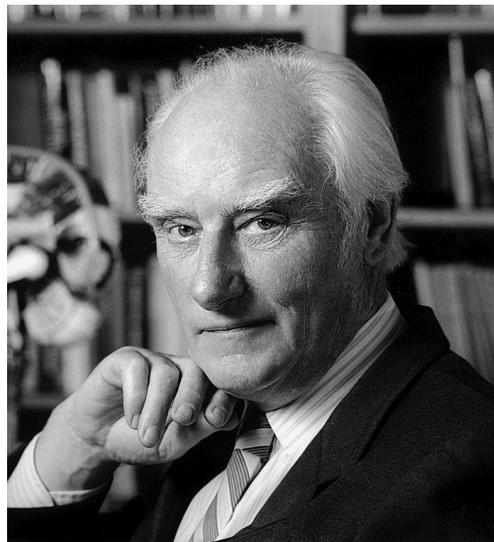
# Centralna dogma molekularne biologije



# Centralna dogma molekularne biologije

DNK  $\xrightarrow{\text{Transkripcija}}$  RNK  $\xrightarrow{\text{Translacija}}$  Protein

- [animacija](#)



Francis Crick

# Potruga za tirocidinom B1



**Cilj:** Naći 30-gram u genomu *Bacillus brevis* koji se prepisuje i prevodi u tirocidin B1 (peptid dužine 10).

# Potruga za tirocidinom B1

*Hiljade* različitih 30-grama  
se mogu prevesti u tirocidin B1.

# Potruga za tirocidinom B1

*Hiljade* različitih 30-grama  
se mogu prevesti u tirocidin B1.

**GTTAAATTATTTCTTGGTTTAATCAATAT**  
**ValLysLeuPheProTrpPheAsnGlnTyr**

# Potruga za tirocidinom B1

*Hiljade* različitih 30-grama  
se mogu prevesti u tirocidin B1.

GTTAAATTATTTCTTGGTTTAATCAATAT  
**ValLysLeuPheProTrpPheAsnGlnTyr**

**GTCAAGCTTTTCCCCTGGTTCAACCAGTAC**  
**ValLysLeuPheProTrpPheAsnGlnTyr**

# Potruga za tirocidinom B1

*Hiljade* različitih 30-grama  
se mogu prevesti u tirocidin B1.

GTTAAATTATTTTCCTTGGTTTAATCAATAT  
**ValLysLeuPheProTrpPheAsnGlnTyr**

GTCAAGCTTTTCCCCTGGTTCAACCAGTAC  
**ValLysLeuPheProTrpPheAsnGlnTyr**

**GTAAACTATTTCCGTGGTTCAATCAATAT**  
**ValLysLeuPheProTrpPheAsnGlnTyr**

# Potruga za tirocidinom B1

*Hiljade* različitih 30-grama  
se mogu prevesti u tirocidin B1.

**GTTAAATTATTTCTTGGTTTAATCAATAT**

**GTCAAGCTTTTCCCCTGGTTCAACCAGTAC**

**GTAAAAC TATTTCCGTGGTTCAATCAATAT**

# Potruga za tirocidinom B1

*Hiljade* različitih 30-grama  
se mogu prevesti u tirocidin B1.

GT**TAAATTAATTCCTTGGTTTAATCAATAT**

GT**CAAGCTTTCCCTGGTTCAACCAGTAC**

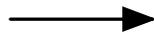
GT**AAAACATAATTCCTGGTTCAATCAATAT**

Nisu previše slični.

# Potruga za tirocidinom B1

Translacija može početi na bilo kojoj poziciji u genomu;

imamo 6 različitih čitajućih okvira



Translated peptides

GluThrPheSerLeuValXXXSerIle  
XXXAsnPhePheLeuGlyLeuIleAsn  
ValLysLeuPheProTrpPheAsnGlnTyr

Transcribed RNA

GUGAAACUUUUCCUUGGUUUAUCAUAU

DNA

5' GTGAAACTTTTTCTTGGTTTAATCAATAT 3'  
3' CACTTTGAAAAAGGAACCAAATTAGTTATA 5'

Transcribed RNA

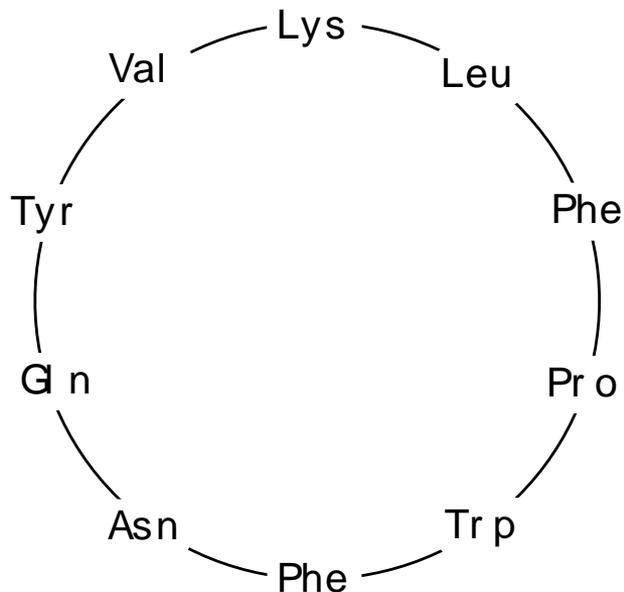
CACUUUGAAAAAGGAACCAAUUAGUUUAU

Translated peptides

HisPheLysLysArgProLysIleLeuIle  
SerValLysGluLysThrXXXAspIle  
PheSerLysGlyGlnAsnLeuXXXTyr



# Tirocidin B1 je cikličan



*Deset različitih linearnih reprezentacija:*

**Val-Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr**

Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr-Val

...

Tyr-Val-Lys-Leu-Phe-Pro-Trp-Phe-Asn-Gln

# Potruga za tirocidinom B1

Koliko 30-grama u genomu *Bacillus brevis* kodira jednu linearnu reprezentaciju tirocidina B1?

# Potruga za tirocidinom B1

Koliko 30-grama u genomu *Bacillus brevis* kodira jednu linearnu reprezentaciju tirocidina B1?

...obrada...

...obrada...

...obrada...

...obrada...

...obrada...

Nijedan?! :o

# Dogma ne važi uvek



1963: američki genetičar Edvard Tatum je inhibirao ribozom bakterije *Bacillus brevis*.

Nastavljena je proizvodnja nekih peptida, uključujući i tirocidine!



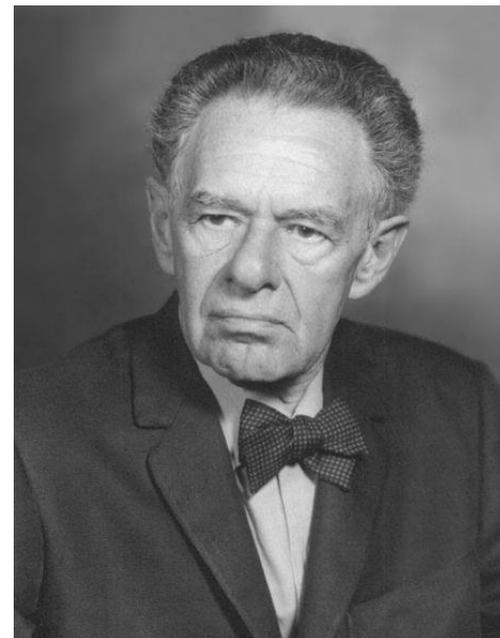
Edvard Tatum

# Dogma ne važi uvek



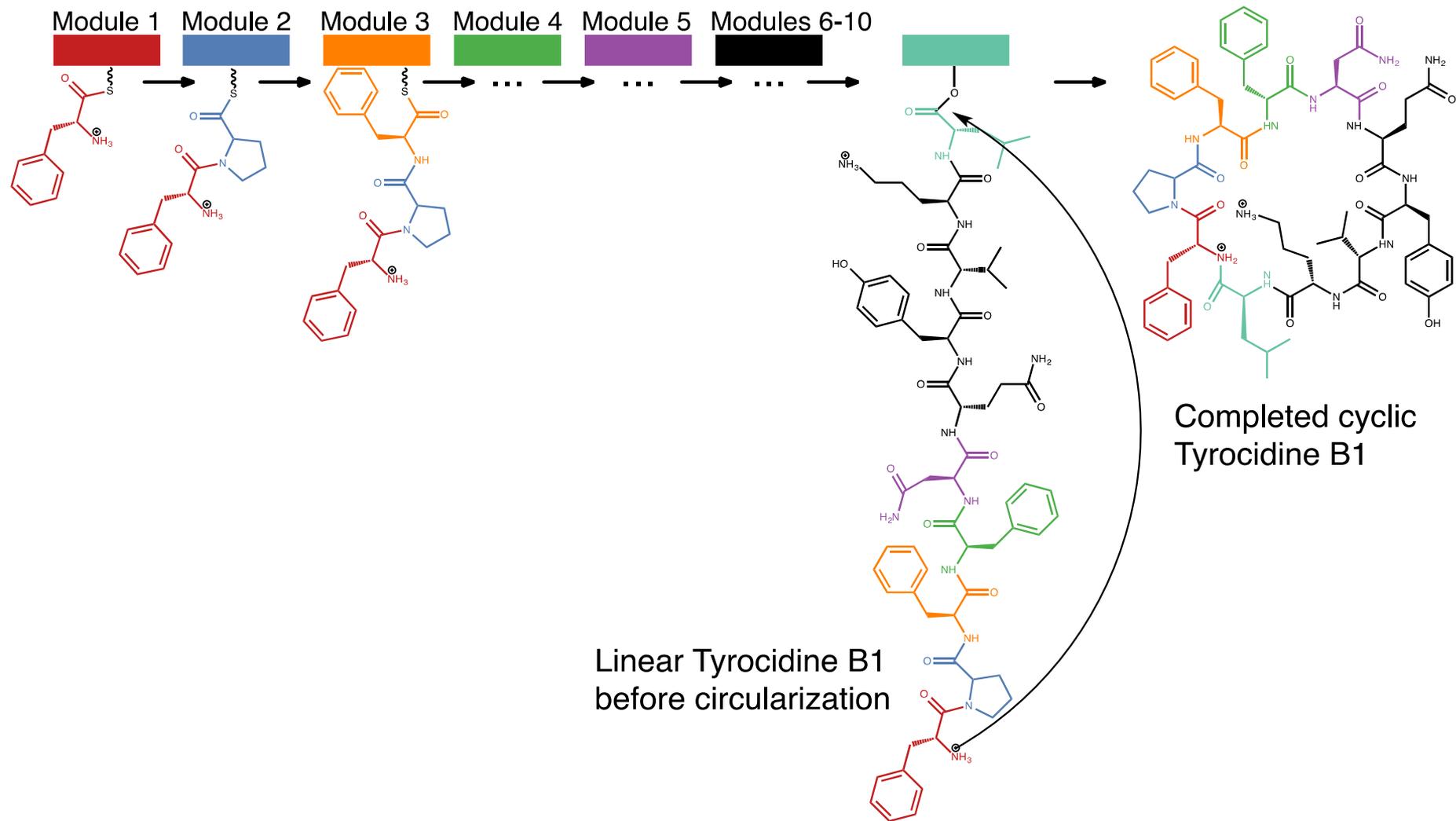
1969: američko-nemački biohemičar Fric Lipman je pokazao da su tirocidini **ne-ribozomalni peptidi (NRP-ovi)**.

Za sintezu NRP-ova nisu odgovorni ribozomi već enzimi **NRP sintetaze**.



Fric Lipman

# Sinteza tirocidina B1 pomoću NRP sintetaze



# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- **Sekvencioniranje antibiotika razbijanjem na komade**
- Algoritam grube sile za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do više od 100 aminokiselina
- Spektralna konvolucija
- Spektri u realnosti

# Sekvencioniranje tirocidina

- Sekvencioniranjem genoma bakterije *Bacillus brevis* nećemo moći da utvrdimo sekvencu tirocidina B1.
- Tirocidin B1 se ne dobija sintezom unutar ribozoma bakterije *Bacillus brevis* kao mnogi drugi proteini. Zbog toga metode za sekvencioniranje DNK nisu od pomoći u ovom slučaju, već moramo direktno sekvencionirati peptid.

# Maseni spektrometar

- Sekvencioniranjem genoma bakterije *Bacillus brevis* nećemo moći da utvrdimo sekvencu tirocidina B1.

**Maseni spektrometar**  
„skupa molekularna vaga“



# Kako merimo težinu molekula?

**1 Dalton (Da)  $\approx$  masa protona/neutrona**

Masa molekula  $\approx$  suma masa protona/neutrona

# Kako merimo težinu molekula?

**1 Dalton (Da)  $\approx$  masa protona/neutrons**

Masa molekula  $\approx$  suma masa protona/neutrons

Masa glicina ( $\text{C}_2\text{H}_3\text{ON}$ )  $\approx$  **12**  $\cdot$  **2**

# Kako merimo težinu molekula?

**1 Dalton (Da)**  $\approx$  masa protona/neutrona

Masa molekula  $\approx$  suma masa protona/neutrona

Masa glicina ( $\text{C}_2\text{H}_3\text{ON}$ )  $\approx 12 \cdot 2 + 1 \cdot 3$

# Kako merimo težinu molekula?

**1 Dalton (Da)  $\approx$  masa protona/neutrona**

Masa molekula  $\approx$  suma masa protona/neutrona

Masa glicina ( $\text{C}_2\text{H}_3\text{ON}$ )  $\approx 12 \cdot 2 + 1 \cdot 3 + 16$

# Kako merimo težinu molekula?

**1 Dalton (Da)**  $\approx$  masa protona/neutrona

Masa molekula  $\approx$  suma masa protona/neutrona

Masa glicina ( $\text{C}_2\text{H}_3\text{ON}$ )  $\approx 12 \cdot 2 + 1 \cdot 3 + 16 + 14$

# Kako merimo težinu molekula?

**1 Dalton (Da)**  $\approx$  masa protona/neutrona

Masa molekula  $\approx$  suma masa protona/neutrona

Masa glicina ( $\text{C}_2\text{H}_3\text{ON}$ )  $\approx 12 \cdot 2 + 1 \cdot 3 + 16 + 14$   
 $\approx 57 \text{ Da}$

# Kako merimo težinu molekula?

**1 Dalton (Da)  $\approx$  masa protona/neutrona**

Masa molekula  $\approx$  suma masa protona/neutrona

Masa glicina ( $\text{C}_2\text{H}_3\text{ON}$ )  $\approx 12 \cdot 2 + 1 \cdot 3 + 16 + 14$   
 $\approx 57 \text{ Da}$

Stvarna masa: 57.02 Da

**Celobrojna masa: 57**

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VKLFPWFNQY)

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VKLFPWFNQY)

V  
99

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	<b>K</b>	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	<b>128</b>	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (V**K**LFPWFNQY)

$$V \quad \mathbf{K}$$
$$99 + \mathbf{128}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VK**L**FPWFNQY)

$$V \quad K \quad L$$
$$99+128+113$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	<b>F</b>	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	<b>147</b>	156	163	186

Koja je masa of tirocidina B1? (VKL**F**PWFNQY)

$$V \quad K \quad L \quad \mathbf{F}$$
$$99+128+113+\mathbf{147}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	<b>P</b>	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	<b>97</b>	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VKLFPWFNQY)

$$V \quad K \quad L \quad F \quad \mathbf{P}$$
$$99+128+113+147+\mathbf{97}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	<b>W</b>
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	<b>186</b>

Koja je masa of tirocidina B1? (VKLFP**W**FNQY)

$$V \quad K \quad L \quad F \quad P \quad \mathbf{W}$$
$$99+128+113+147+97+\mathbf{186}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	<b>F</b>	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	<b>147</b>	156	163	186

Koja je masa of tirocidina B1? (VKLFPW**F**NQY)

$$V \quad K \quad L \quad F \quad P \quad W \quad \mathbf{F}$$
$$99+128+113+147+97+186+\mathbf{147}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	<b>N</b>	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	<b>114</b>	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VKLFPWF**N**QY)

$$V \quad K \quad L \quad F \quad P \quad W \quad F \quad \mathbf{N}$$
$$99+128+113+147+97+186+147+\mathbf{114}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VKLFPWFN**Q**Y)

$$V \quad K \quad L \quad F \quad P \quad W \quad F \quad N \quad Q$$
$$99+128+113+147+97+186+147+114+128$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	<b>Y</b>	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	<b>163</b>	186

Koja je masa of tirocidina B1? (VKLFPWFNQ**Y**)

$$V \quad K \quad L \quad F \quad P \quad W \quad F \quad N \quad Q \quad \mathbf{Y}$$
$$99+128+113+147+97+186+147+114+128+\mathbf{163}$$

# Tabela celobrojnih masa

Celobrojne mase svih 20 aminokiselina

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

Koja je masa of tirocidina B1? (VKLFPWFNQY)

$$\begin{array}{cccccccccc} V & K & L & F & P & W & F & N & Q & Y \\ 99+128+113+147+97+186+147+114+128+163 & = & \mathbf{1322} \end{array}$$

# Tabela celobrojnih masa

Primetimo da neke aminokiseline imaju jednake mase:

G	A	S	P	V	T	C	I	L	N	D	K	Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	113	114	115	128	128	129	131	137	147	156	163	186

# Tabela celobrojnih masa

Primetimo da neke aminokiseline imaju jednake mase:

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

# Tabela celobrojnih masa

Primetimo da neke aminokiseline imaju jednake mase:

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

20 aminokiselina → 18 celobrojnih masa

# Kako radi maseni spektrometar

**NQEL**

**NQEL**

**NQEL**

**NQEL**

**NQEL**

**NQEL**

**NQEL**

**NQEL**

**NQEL**

# Kako radi maseni spektrometar

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL



# Kako radi maseni spektrometar

podpeptid

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL

NQEL



L

N

Q

E

LN

NQ

EL

QE

LNQ

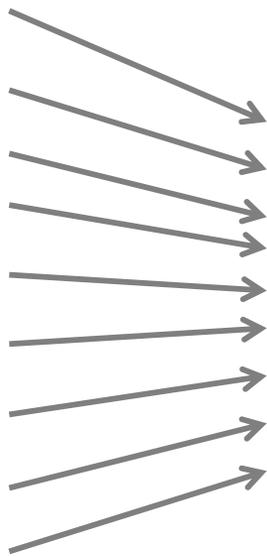
ELN

QEL

NQE

# Kako radi maseni spektrometar

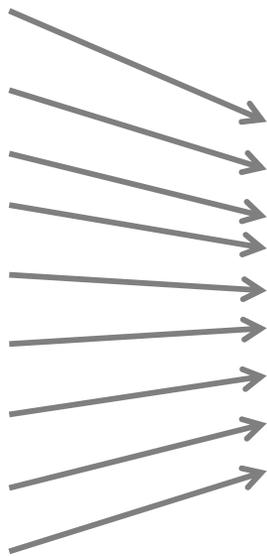
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL



podpeptid	masa
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371

# Kako radi maseni spektrometar

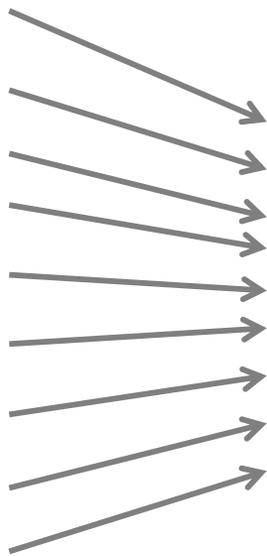
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL



podpeptid	masa
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371

# Kako radi maseni spektrometar

NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL  
NQEL



podpeptid	masa
L	113
N	114
Q	128
E	129
LN	227
NQ	<b>242</b>
EL	<b>242</b>
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371

# Formulacija računarskog problema

**Teorijski spektar peptida:** mase svih mogućih podpeptida, uključujući  $\emptyset$  i masu celog peptida.

podpeptid	masa
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371
NQEL	484

# Formulacija računarskog problema

podpeptid      masa

**Teorijski spektar peptida:** mase svih mogućih podpeptida, uključujući  $\emptyset$  i masu celog peptida.

L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371
NQEL	484



# Formulacija računarskog problema

podpeptid	masa
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371
NQEL	484

**Teorijski spektar peptida:** mase svih mogućih podpeptida, uključujući  $\emptyset$  i masu celog peptida.



# Formulacija računarskog problema

masa

**Teorijski spektar peptida:** mase svih mogućih podpeptida, uključujući  $\emptyset$  i masu celog peptida.

113

114

128

129

227

242

242

257

355

356

370

371

484



**Problem sekvenciranja ciklopeptida:**  
*Rekonstruisati ciklični peptid na osnovu njegovog teorijskog spektra.*

# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- **Algoritam grube sile za sekvencioniranje ciklopeptida**
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do više od 100 aminokiselina
- Spektralna konvolucija
- Spektri u realnosti

# Sekvencioniranje ciklopeptida grubom silom

Masa celog peptida je obično poznata.

## Algoritam:

1. Generisati sve **peptide** sa datom masom (1322).
2. Za svakog formirati teorijski spektar.
3. Uporediti sve dobijene spektre sa datim **spektrom**

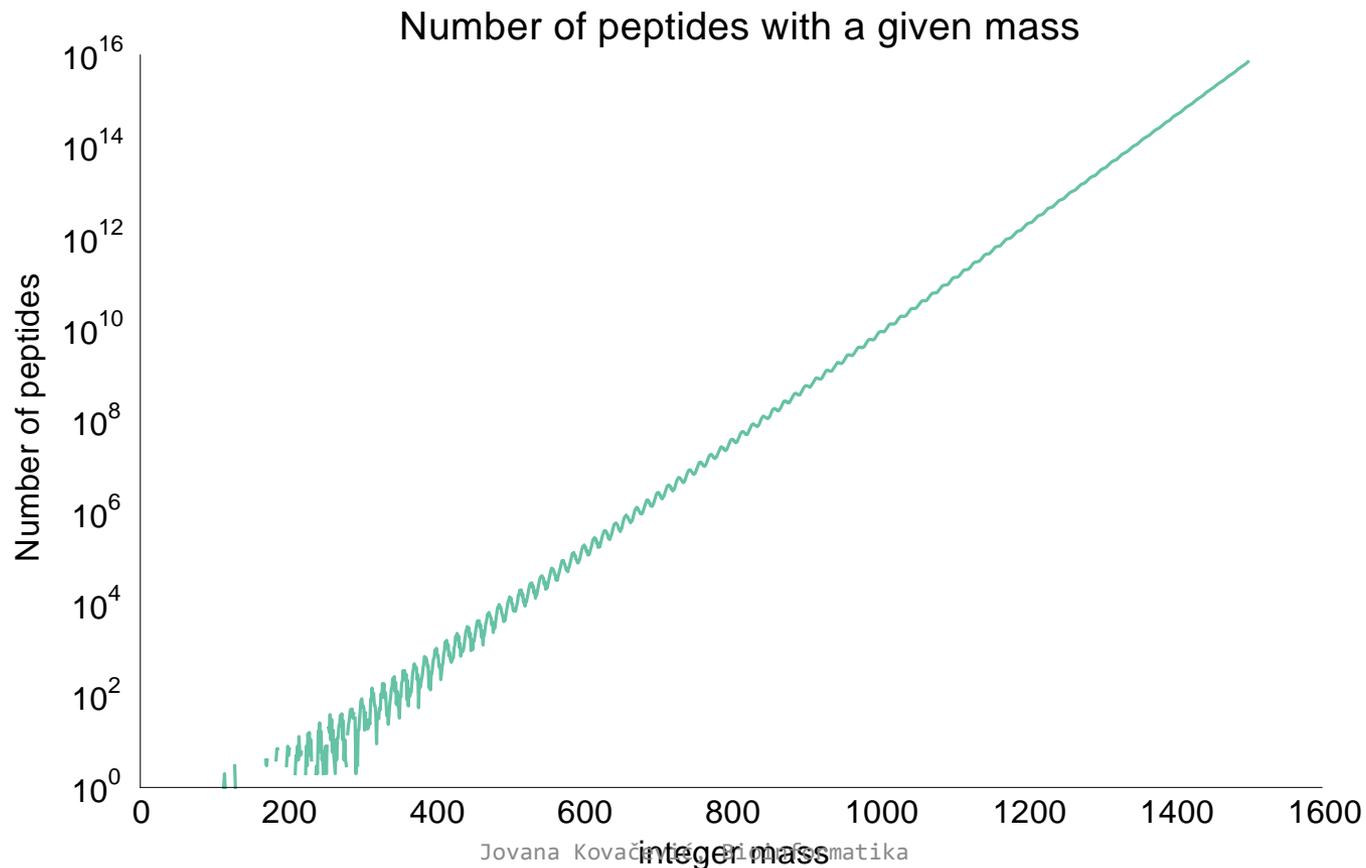
**Algoritam grube sile:** Ispitati sve kandidate.

# Sekvencioniranje ciklopeptida grubom silom

Koliko peptida ima celobrojnu masu = 1322?

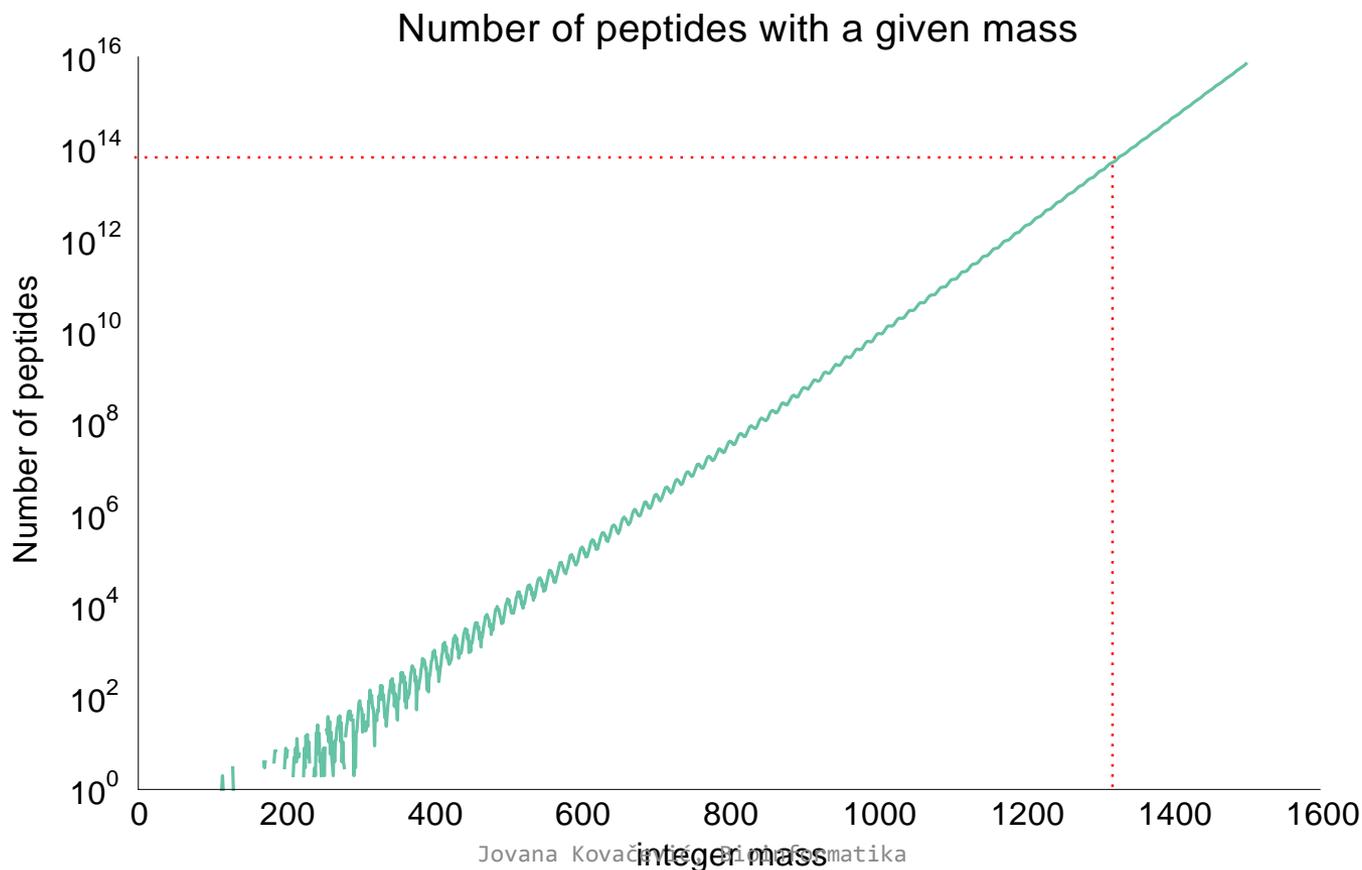
# Sekvencioniranje ciklopeptida grubom silom

Koliko peptida ima celobrojnu masu = 1322?



# Sekvencioniranje ciklopeptida grubom silom

Koliko peptida ima celobrojnu masu = 1322?



# Nedostaci algoritma grube sile

**114-128-129-113**

N Q E L

**101-131-115-137**

T M D H

# Nedostaci algoritma grube sile

**114-128-129-113**

N Q E L

**101-131-115-137**

T M D H

ukupna masa: **484**

# Nedostaci algoritma grube sile

**114-128-129-113**

N Q E L

ukupna masa: **484**

**101-131-115-137**

T M D H

ukupna masa: **484**

# Nedostaci algoritma grube sile

**114-128-129-113**

N Q E L

ukupna masa: **484**

**101-131-115-137**

T M D H

ukupna masa: **484**

Ovi peptidi su *potpuno različiti*.

# Nedostaci algoritma grube sile

114-128-129-113

N Q E L

ukupna masa: 484

101-131-115-137

T M D H

ukupna masa: 484

Ovi peptidi su *potpuno različiti*.

Kako možemo isključiti **pogrešan** peptid?

# Nedostaci algoritma grube sile

114-128-129-113

N Q E L

ukupna masa: 484

101-131-115-137

T M D H

ukupna masa: 484

Ovi peptidi su *potpuno različiti*.

Kako možemo isključiti **pogrešan** peptid?

# Nedostaci algoritma grube sile

spektar za **TMDH**

""	0
T	101
D	115
M	131
H	137
TM	232
HT	238
MD	246
DH	252
TMD	347
DHT	353
HTM	369
MDH	383
<b>TMDH</b>	<b>484</b>

spektar za **NQEL**

""	0
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371
<b>NQEL</b>	<b>484</b>

# Nedostaci algoritma grube sile

spektar za **TMDH**

""	0
T	101
D	115
M	131
H	137
TM	232
HT	238
MD	246
DH	252
TMD	347
DHT	353
HTM	369
MDH	383
<b>TMDH</b>	<b>484</b>

Njihovi spektri  
su potpuno različiti!

spektar za **NQEL**

""	0
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371
<b>NQEL</b>	<b>484</b>

# Nedostaci algoritma grube sile

spektar za **TMDH**

""	0
T	101
D	115
M	131
H	137
TM	232
HT	238
MD	246
DH	252
TMD	347
DHT	353
HTM	369
MDH	383
<b>TMDH</b>	484

spektar za **NQEL**

""	0
L	113
N	114
Q	128
E	129
LN	227
NQ	242
EL	242
QE	257
LNQ	355
ELN	356
QEL	370
NQE	371
<b>NQEL</b>	484

Njihovi spektri  
su potpuno različiti!

*Kako možemo to  
da iskoristimo?*

# Novi pristup

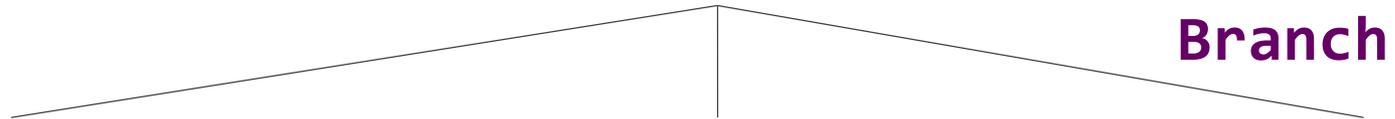
**Postepeno** konstruišemo kandidate za rešenja od manjih *linearnih* peptida.

Na taj način smanjićemo ukupni broj linearnih peptida koje posmatramo.

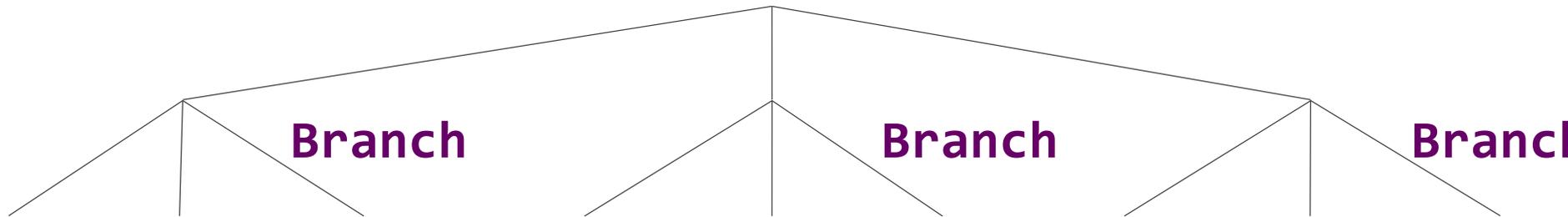
# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grube sile za sekvencioniranje ciklopeptida
- **Branch-and-Bound algoritam za sekvencioniranje ciklopeptida**
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do više od 100 aminokiselina
- Spektralna konvolucija
- Spektri u realnosti

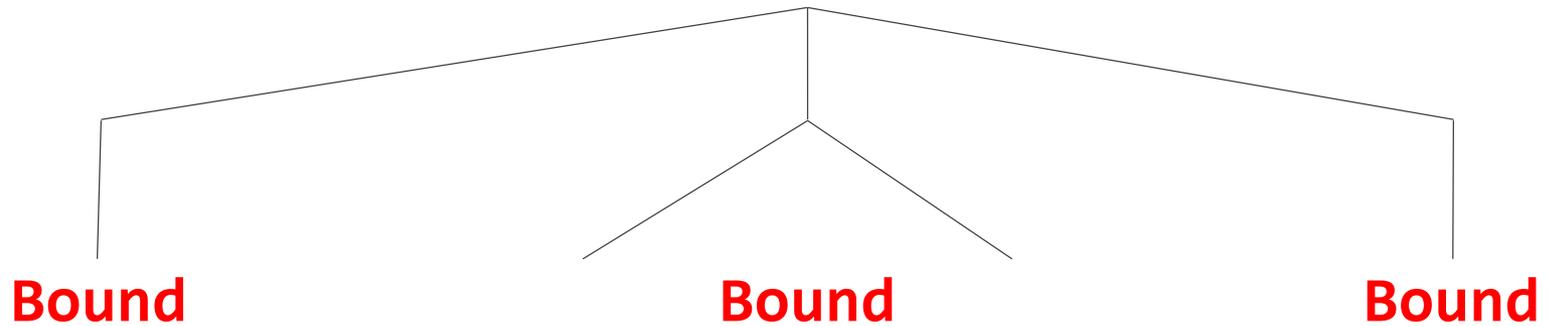
# Branch-and-Bound Algoritmi



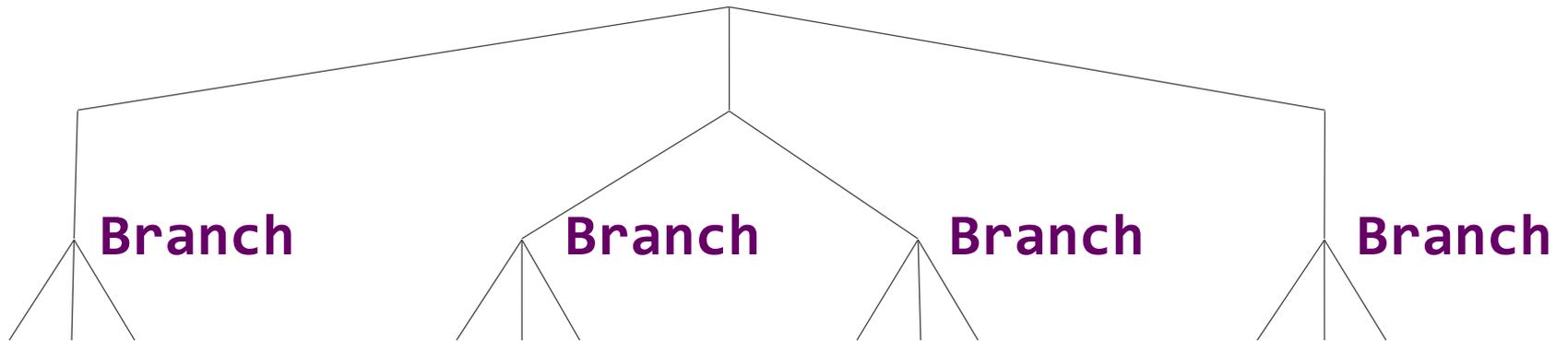
# Branch-and-Bound Algoritmi



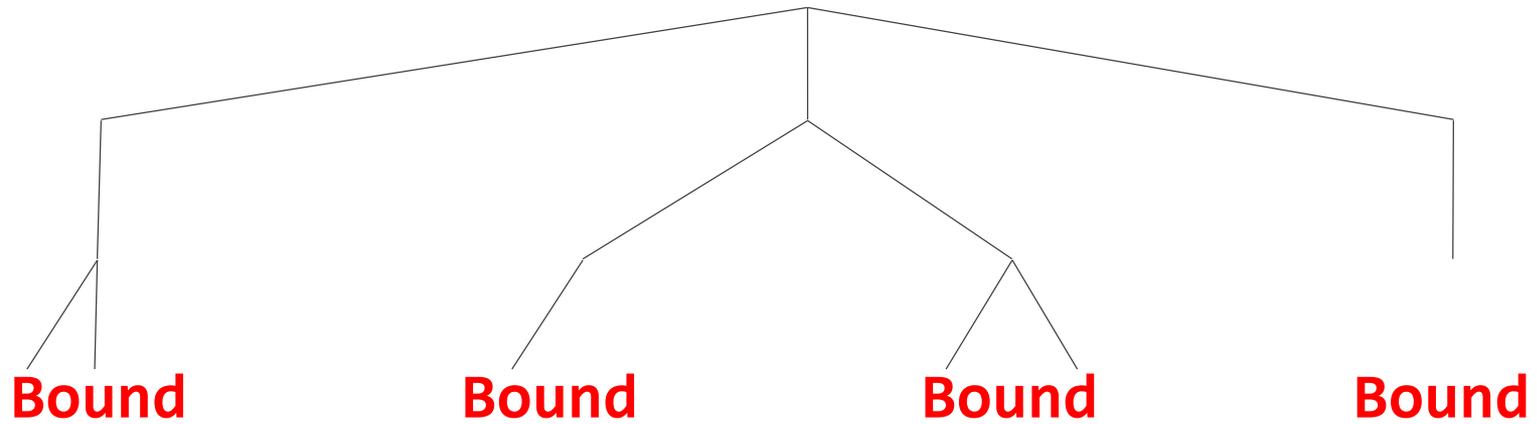
# Branch-and-Bound Algoritmi



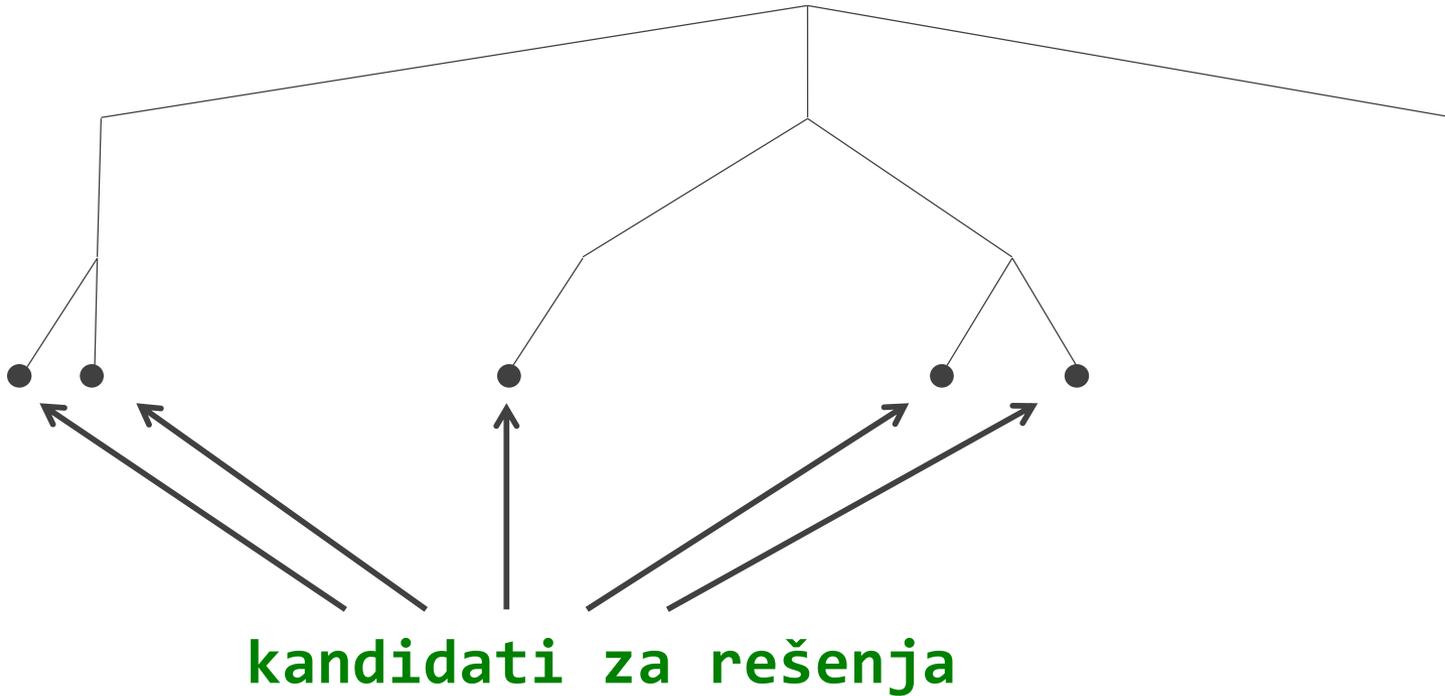
# Branch-and-Bound Algoritmi



# Branch-and-Bound Algoritmi



# Branch-and-Bound Algoritmi



# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

Da li se u spektru nalaze  
mase nekih aminokiselina?

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

Da li se u spektru nalaze  
mase nekih aminokiselina?

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

Da li se u spektru nalaze mase nekih aminokiselina?

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

Da li se u spektru nalaze mase nekih aminokiselina?

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

Počecemo sa 4 unigrama -  
peptida dužine 1:  
P, V, T, C

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

**Proširimo** unigrame u sve moguće bigrame:

<b>P</b> A	<b>V</b> A	<b>T</b> A	<b>C</b> A
<b>P</b> C	<b>V</b> C	<b>T</b> C	<b>C</b> C
<b>P</b> D	<b>V</b> D	<b>T</b> D	<b>C</b> D
<b>P</b> E	<b>V</b> E	<b>T</b> E	<b>C</b> E
...	...	...	...
<b>P</b> W	<b>V</b> W	<b>T</b> W	<b>C</b> W
<b>P</b> Y	<b>V</b> Y	<b>T</b> Y	<b>C</b> Y

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

**Proširimo** unigrame u sve moguće bigrame:

PA	VA	TA	CA
PC	VC	TC	CC
PD	VD	TD	CD
PE	VE	TE	CE
...	...	...	...
PW	VW	TW	CW
PY	VY	TY	CY

Kako možemo da **skratimo** ovu listu?

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

PV je konzistentno sa spektrom:

$$\text{Mass(P)} = 97$$

$$\text{Mass(V)} = 99$$

$$\text{Mass(PV)} = 196$$

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

PV je konzistentno sa spektrom:

$$\text{Mass}(P) = \mathbf{97}$$

$$\text{Mass}(V) = \mathbf{99}$$

$$\text{Mass}(PV) = \mathbf{196}$$

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

PV je **konzistentno** sa spektrom:

$$\text{Mass(P)} = 97 \quad \text{Mass(V)} = 99 \quad \text{Mass(PV)} = 196$$

CD je **nekonzistentno** sa spektrom:

$$\text{Mass(C)} = 103 \quad \text{Mass(D)} = 115 \quad \text{Mass(CD)} = 218$$

## B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

PV je konzistentno sa spektrom:

$$\text{Mass(P)} = 97$$

$$\text{Mass(V)} = 99$$

$$\text{Mass(PV)} = 196$$

CD je nekonzistentno sa spektrom:

$$\text{Mass(C)} = 103$$

$$\text{Mass(D)} = 115$$

$$\text{Mass(CD)} = 218$$

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

## Lista konzistentnih bigrama:

PV	PT	PC	VP	VT
VC	TP	TV	CP	CV

**proširi**, pa **skrati**...

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

## Lista konzistentnih trigramama:

PVC	PVT	PTP	PTV	PCV
VPC	VPT	VTP	VCP	TPV
TPC	TVP	CPT	CPV	CVP

**proširi**, pa **skrati**...

# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

## Lista konzistentnih tetragrama:

PVCP	PTPV	PTPC	PCVP	VPTP
VCPT	TPVC	TPCV	CPTP	CVPT

**proširi**, pa **skrati**...

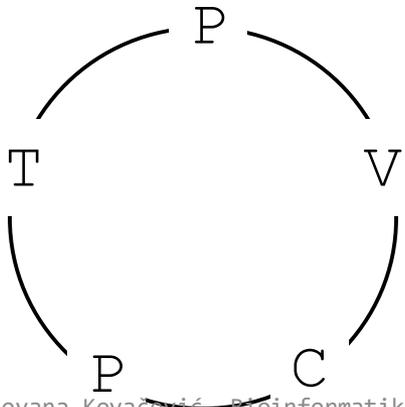
# B-&-B za sekvencioniranje ciklopeptida

*Spektar*

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

## Lista konzistentnih pentagrama:

- |       |       |       |       |       |
|-------|-------|-------|-------|-------|
| PVCPT | PTPVC | PTPCV | PCVPT | VPTPC |
| VCPTP | TPVCP | TPCVP | CPTPV | CVPTP |



**Jedan ciklični peptid!**

## CyclopeptideSequencing(*Spectrum*)

*CandidatePeptides* ← a set containing only the empty peptide

*FinalPeptides* ← empty list of strings

**while** *CandidatePeptides* is nonempty

*// prosirujemo sve kandidate za jednu aa*

*CandidatePeptides* ← **Expand**(*CandidatePeptides*)

for each peptide *Peptide* in *CandidatePeptides*

**if** *Mass*(*Peptide*) = *ParentMass*(*Spectrum*) *// kandidat za FinalPeptides*

**if** *Cyclospectrum*(*Peptide*) = *Spectrum* and *// ako se spektri poklapaju*

*Peptide* is not in *FinalPeptides* *// i Peptide nije vec dodat*

*append Peptide to FinalPeptides // dodajemo ga*

*remove Peptide from CandidatePeptides // svakako ga izbacujemo iz*

*// liste kandidata*

**else** *// nije kandidat za FinalPeptides, da li ga izbacujemo*

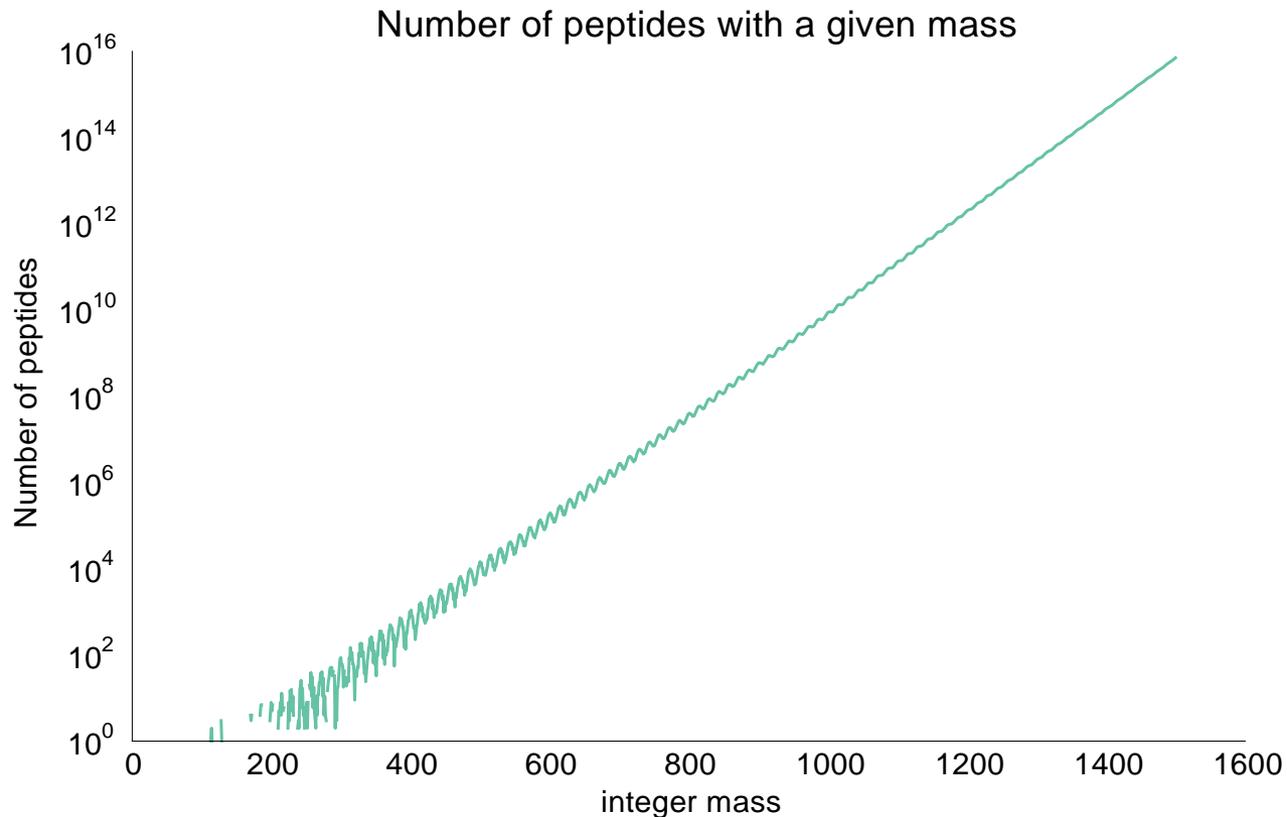
**if** *Peptide* is not consistent with *Spectrum* *// da, ako nije konzi-*

*remove Peptide from CandidatePeptides // stentan sa spektrom*

**return** *FinalPeptides*

# Da li je ovaj B-&-B pristup efikasan?

Algoritam grube sile za sekvencioniranje ciklopeptida je eksponencijalan.



Da li je ovaj B-&-B pristup efikasan?

B&B algoritam za sekvencioniranje ciklopeptida *može biti* eksponencijalan za neke peptide...



Da li je ovaj B-&-B pristup efikasan?

B&B algoritam za sekvencioniranje ciklopeptida *može biti* eksponencijalan za neke peptide...



...ali je u praksi veoma brz!

# Can We Go Home Now?



Can We Go Home Now?

**NO!**

# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grube sile za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- **Prilagođavanje sekvenciranja za spektre sa greškama**
- Od 20 do više od 100 aminokiselina
- Spektralna konvolucija
- Istina o spektrima

Od teorijskog spektra  
do spektra sa šumom

**Eksperimentalni spektri**  
često sadrže greške.

# Od teorijskog spektra do spektra sa šumom

**Eksperimentalni spektri  
često sadrže greške.**

Posmatrajmo sledeće spektre za NQEL:

<b>teorijski:</b>	0		113	114	128	129	227	242	242	257		355	356	370	371	484
<b>eksperimentalni:</b>	0	99	113	114	128		227			257	299	355	356	370	371	484

# Od teorijskog spektra do spektra sa šumom

**Eksperimentalni spektri  
često sadrže greške.**

Posmatrajmo sledeće spektre za NQEL:

teorijski:	0	113	114	128	129	227	242	242	257	355	356	370	371	484	
eksperimentalni:	0	99	113	114	128	227			257	299	355	356	370	371	484

**Lažne mase:** prisutne u eksperimentalnom  
spektru, nema ih u teorijskom spektru

# Od teorijskog spektra do spektra sa šumom

**Eksperimentalni spektri  
često sadrže greške.**

Posmatrajmo sledeće spektre za NQEL:

teorijski:	0		113	114	128	129	227	242	242	257		355	356	370	371	484
eksperimentalni:	0	99	113	114	128		227			257	299	355	356	370	371	484

**Lažne mase:** prisutne u eksperimentalnom spektru, nema ih u teorijskom spektru

**Nedostajuće mase:** prisutne u teorijskom spektru, nema ih u eksperimentalnom spektru

# Neophodan je novi algoritam

Do sada: teorijski spektar peptida je morao da se **tačno** poklapa sa eksperimentalnim spektrom.

teorijski:	0		113	114	128	129	227	242	242	257		355	356	370	371	484
eksperimentalni:	0	99	113	114	128		227			257	299	355	356	370	371	484

# Neophodan je novi algoritam

Do sada: teorijski spektar peptida je morao da se **tačno** poklapa sa eksperimentalnim spektrom.

teorijski:	0		113	114	128	129	227	242	242	257		355	356	370	371	484
eksperimentalni:	0	99	113	114	128		227			257	299	355	356	370	371	484

Umesto toga: **skor** peptida koji pokazuje koliko masa njegov spektar **deli** sa eksperimentalnim spektrom.

$Score(\text{NQEL}, \text{ExperimentalSpectrum}) = 11$

# Odsecanje u golf turniru

**Odsecanje:** uzima u obzir samo igrače koji su konkurentni

Golfer	Score
Cabrera	-6
Woods	-4
Watson	-1
McDowell	-1
Scott	+1
Daly	+14

# Odsecanje u golf turniru

**Odsecanje:** uzima u obzir samo igrače koji su konkurentni

Golfer	Score
Cabrera	-6
Woods	-4
Watson	-1
McDowell	-1
Scott	+1
Daly	+14

Zadrži tri  
najbolja igrača

# Odsecanje u golf turniru

**Odsecanje:** uzima u obzir samo igrače koji su konkurentni

Golfer	Score
Cabrera	-6
Woods	-4
<b>Watson</b>	<b>-1</b>
<b>McDowell</b>	<b>-1</b>
Scott	+1
Daly	+14

Zadrži tri  
najbolja igrača

# Odsecanje u golf turniru

**Odsecanje:** uzima u obzir samo igrače koji su konkurentni

Golfer	Score
Cabrera	-6
Woods	-4
<b>Watson</b>	<b>-1</b>
<b>McDowell</b>	<b>-1</b>
Scott	+1
Daly	+14

Zadrži tri  
najbolja igrača  
*„with ties“*

# Odsecanje u golf turniru

**Odsecanje:** uzima u obzir samo igrače koji su konkurentni

Golfer	Score
Cabrera	-6
Woods	-4
Watson	-1
McDowell	-1

Zadrži tri  
najbolja igrača  
*„with ties“*

## LeaderboardCyclopeptideSequencing(*Spectrum*, *N*)

*Leaderboard* ← set containing only the empty peptide

*LeaderPeptide* ← empty peptide

**while** *Leaderboard* is non-empty

*// prosirujemo sve peptide iz Leaderboard za jednu aa*

*Leaderboard* ← **Expand**(*Leaderboard*)

**for** each *Peptide* in *Leaderboard*

**if** *Mass*(*Peptide*) = *ParentMass*(*Spectrum*) *// kandidat za resenje*

**if** *Score*(*Peptide*, *Spectrum*) > *Score*(*LeaderPeptide*, *Spectrum*)

*LeaderPeptide* ← *Peptide* *// ako mu je skor veci od trenutnog*  
*// nutnog maksimalnog, maksimalni se azurira*

**else** *// ako nije kandidat za resenje, da li da ga zadrzimo?*

**if** *Mass*(*Peptide*) > *ParentMass*(*Spectrum*)

*remove Peptide from Leaderboard* *// ako mu je masa*

*// veca od roditeljske mase spectra, nece se poboljsati*

*// izbacujemo ga*

*Leaderboard* ← **Trim**(*Leaderboard*, *Spectrum*, *N*) *// skracujemo listu*

*// na N with ties*

**output** *LeaderPeptide*

# Testiranje na spektru tirocidina B1

*Spectrum*<sub>10</sub>: 10% lažnih/nedostajućih masa

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	1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

# Testiranje na spektru tirocidina B1

*Spectrum*<sub>10</sub>: 10% **lažnih**/nedostajućih masa

0	97	99	113	114	128	128	147	147	163	186	227	241	242
244	260	261	262	283	291	333	340	357	<b>385</b>	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	<b>820</b>	835	837	875	892	892	917	932	932	933
934	965	982	989	<b>1030</b>	1031	1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

# Testiranje na spektru tirocidina B1

*Spectrum*<sub>10</sub>: 10% **lažnih**/**nedostajućih** masa

0	97	99	<b>113</b>	114	<b>128</b>	128	147	147	163	186	227	241	242
244	260	261	262	283	291	333	340	357	<b>385</b>	<b>388</b>	389	390	390
405	430	430	447	485	487	503	504	518	543	544	552	575	577
584	<b>631</b>	632	650	651	671	672	690	691	738	745	747	770	778
779	804	818	819	<b>820</b>	835	837	875	892	<b>892</b>	917	932	932	933
934	965	982	989	<b>1030</b>	<b>1031</b>	1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

# Testiranje na spektru tirocidina B1

*Spectrum*<sub>10</sub>: 10% **lažnih**/**nedostajućih** masa

0	97	99		114		128	147	147	163	186	227	241	242
244	260	261	262	283	291	333	340	357	<b>385</b>		389	390	390
405	430	430	447	485	487	503	504	518	543	544	552	575	577
584		632	650	651	671	672	690	691	738	745	747	770	778
779	804	818	819	<b>820</b>	835	837	875	892		917	932	932	933
934	965	982	989	<b>1030</b>		1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

# Testiranje na spektru tirocidina B1

*Spectrum*<sub>10</sub>: 10% **lažnih/**nedostajućih masa

0	97	99		114		128	147	147	163	186	227	241	242
244	260	261	262	283	291	333	340	357	<b>385</b>		389	390	390
405	430	430	447	485	487	503	504	518	543	544	552	575	577
584		632	650	651	671	672	690	691	738	745	747	770	778
779	804	818	819	<b>820</b>	835	837	875	892		917	932	932	933
934	965	982	989	<b>1030</b>		1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

Peptid sa najvišim skorom: VKLFPWFNQY 😊

# Šumovitiji spektar tirocidina B1

*Spectrum*<sub>25</sub>: 25% lažnih/nedostajućih masa

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							

# Šumovitiji spektar tirocidina B1

*Spectrum*<sub>25</sub>: 25% **lažnih**/nedostajućih masa

0	97	99	113	114	<b>115</b>	128	128	147	147	163	186	227	241
242	<b>244</b>	244	<b>256</b>	260	261	262	283	291	<b>309</b>	<b>330</b>	333	340	<b>347</b>
357	<b>385</b>	388	389	390	390	405	430	430	<b>435</b>	447	485	487	503
504	518	543	544	552	575	577	584	<b>599</b>	<b>608</b>	631	632	650	651
<b>653</b>	671	672	690	691	<b>717</b>	738	745	747	770	778	779	804	818
819	<b>827</b>	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							

# Šumovitiји spektar tirocidina B1

*Spectrum*<sub>25</sub>: 25% **lažnih**/**nedostajućih** masa

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

# Šumovitiji spektar tirocidina B1

*Spectrum*<sub>25</sub>: 25% **lažnih**/**nedostajućih** masa

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

# Šumovitiji spektar tirocidina B1

*Spectrum*<sub>25</sub>: 25% **lažnih**/**nedostajućih** masa

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

Tačan peptid:

VKLFP WFNQY

Peptid sa najvišim skorom:

VKLFP**AD**FNQY ☹️

Tačan peptid: VKLFP WFNQY  
Peptid sa najvišim skorom: VKLFP**AD**FNQY ☹️

G	<b>A</b>	S	P	V	T	C	I/L	N	<b>D</b>	K/Q	E	M	H	F	R	Y	<b>W</b>
57	<b>71</b>	87	97	99	101	103	<b>113</b>	114	<b>115</b>	<b>128</b>	129	131	137	147	156	163	<b>186</b>

- Da li bismo mogli da na osnovu spektra detektujemo od kojih se aminokiselina peptid sastoji i da primenimo algoritam samo za te aminokiseline?
- Zbog nedostajućih masa, to se ne vidi direktno iz spektra

# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grube sile za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- **Od 20 do više od 100 aminokiselina**
- Spektralna konvolucija
- Istina o spektrima

# Od 18 do preko 100 aminokiselina

S obzirom da ne nastaju prema pravilima Centralne dogme, NRP-ovi mogu da sadrže **nestandardne** aminokiseline

## Tirocidin B

Val-**Orn**-Leu-Phe-Pro-Trp-Phe-Asn-Gln-Tyr

**Ornitin**: nestandardna aminokiselina

Mase nestandardnih aminokiselina nisu uključene u tabelu celobrojnih masa. Pretpostavimo zato da bilo koji ceo broj između 57 i 200 može biti masa neke aminokiseline.

# Primena *Leaderboard* algoritma na Spectrum10 sa proširenim skupom masa

*Spectrum*<sub>10</sub>: 10% **lažnih**/**nedostajućih** masa

0	97	99		114		128	147	147	163	186	227	241	242
244	260	261	262	283	291	333	340	357	<b>385</b>		389	390	390
405	430	430	447	485	487	503	504	518	543	544	552	575	577
584		632	650	651	671	672	690	691	738	745	747	770	778
779	804	818	819	<b>820</b>	835	837	875	892		917	932	932	933
934	965	982	989	<b>1030</b>		1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

# Primena *Leaderboard* algoritma na Spectrum10 sa proširenim skupom masa

*Spectrum*<sub>10</sub>: 10% **lažnih**/**nedostajućih** masa

0	97	99		114		128	147	147	163	186	227	241	242
244	260	261	262	283	291	333	340	357	<b>385</b>		389	390	390
405	430	430	447	485	487	503	504	518	543	544	552	575	577
584		632	650	651	671	672	690	691	738	745	747	770	778
779	804	818	819	<b>820</b>	835	837	875	892		917	932	932	933
934	965	982	989	<b>1030</b>		1309	1060	1061	1062	1078	1080	1081	1095
1136	1159	1175	1175	1194	1194	1208	1209	1223	1225	1322			

Peptid sa najvišim skorom: VKLFPWFNQ-**98-65** 😞

Tačan peptid:

VKLFPWENQY

Peptid sa najvišim skorom: VKLFPWENQ-98-65 ☹️

G	A	S	P	V	T	C	I/L	N	D	K/Q	E	M	H	F	R	Y	W
57	71	87	97	99	101	103	113	114	115	128	129	131	137	147	156	163	186

- *LeaderboardCyclopeptideSequencing* sa proširenim skupom masa ne radi ni za skup sa šumom od 10%
- Da li bismo mogli da na osnovu spektra detektujemo od kojih se aminokiselina peptid sastoji i da primenimo algoritam samo za te aminokiseline?

# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grube sile za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do više od 100 aminokiselina
- **Spektralna konvolucija**
- Istina o spektrima

# Restrikcija azbuke aminokiselina

- Težine aminokiselina mogu biti svi celi brojevi između 57 i 200
- To znači da razmatramo 144 aminokiseline (standardne i nestandardne)

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

# Restrikcija azbuke aminokiselina

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

Posmatrajmo sledeći spektar za NQEL:

eksperimentalni: 0 99 113 114 128 227 257 299 355 356 370 371 484

# Restrikcija azbuke aminokiselina

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

Posmatrajmo sledeći spektar za NQEL:

eksperimentalni: 0 99 113 114 128 227 257 299 355 356 370 371 484

Mass(**E**) = **129**, koja nedostaje, ali...

# Restrikcija azbuke aminokiselina

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

Posmatrajmo sledeći spektar za NQEL:

eksperimentalni: 0 99 113 114 128 227 257 299 355 356 370 371 484

Mass(**E**) = **129**, koja nedostaje, ali...

$$\text{Mass}(\text{Q}\mathbf{E}) - \text{Mass}(\text{Q}) = \mathbf{257} - \mathbf{128} = \mathbf{129}$$

# Restrikcija azbuke aminokiselina

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

Posmatrajmo sledeći spektar za NQEL:

eksperimentalni: 0 99 113 114 **128** 227 **257** 299 355 356 370 371 484

Mass(**E**) = **129**, koja nedostaje, ali...

$$\text{Mass}(\text{Q}\mathbf{E}) - \text{Mass}(\text{Q}) = \mathbf{257} - \mathbf{128} = \mathbf{129}$$

# Restrikcija azbuke aminokiselina

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

Posmatrajmo sledeći spektar za NQEL:

eksperimentalni: 0 99 113 114 128 **227** 257 299 355 **356** 370 371 484

Mass(**E**) = **129**, koja nedostaje, ali...

$$\text{Mass}(\mathbf{E}_{\text{LN}}) - \text{Mass}(\text{LN}) = \mathbf{356} - \mathbf{227} = \mathbf{129}$$

# Restrikcija azbuke aminokiselina

**Cilj:** smanjiti broj aminokiselina koje uzimamo u obzir.

Posmatrajmo sledeći spektar za NQEL:

eksperimentalni: 0 99 113 114 128 227 257 299 **355** 356 370 371 **484**

Mass(**E**) = **129**, koja nedostaje, ali...

$$\text{Mass}(\text{NQ}\mathbf{E}\text{L}) - \text{Mass}(\text{LNQ}) = \mathbf{484} - \mathbf{355} = \mathbf{129}$$

# Spektralna konvolucija

Apsolutna vrednost razlike  
između svake dve mase u spektru.

	" "	false	L	N	Q	LN	QE	false	LNQ	ELN	QEL	NQE
	0	99	113	114	128	227	257	299	355	356	370	371
0												
99	<b>99</b>											
113	<b>113</b>	14										
114	<b>114</b>	15	1									
128	<b>128</b>	29	15	14								
227	227	<b>128</b>	<b>114</b>	<b>113</b>	<b>99</b>							
257	257	158	144	143	<b>129</b>	30						
299	299	200	186	185	171	72	42					
355	355	256	242	241	227	<b>128</b>	98	56				
356	356	257	243	242	228	<b>129</b>	<b>99</b>	57	1			
370	370	271	257	256	242	143	<b>113</b>	71	15	14		
371	371	272	258	257	243	144	<b>114</b>	72	16	15	1	
484	484	385	371	370	356	257	227	185	<b>129</b>	<b>128</b>	<b>114</b>	<b>113</b>

# Spektralna konvolucija

Koje su najčešći elementi između 57 i 200?

	" "	false	L	N	Q	LN	QE	false	LNQ	ELN	QEL	NQE
0	0	99	113	114	128	227	257	299	355	356	370	371
99	99											
113	113	14										
114	114	15	1									
128	128	29	15	14								
227	227	128	114	113	99							
257	257	158	144	143	129	30						
299	299	200	186	185	171	72	42					
355	355	256	242	241	227	128	98	56				
356	356	257	243	242	228	129	99	57	1			
370	370	271	257	256	242	143	113	71	15	14		
371	371	272	258	257	243	144	114	72	16	15	1	
484	484	385	371	370	356	257	227	185	129	128	114	113

# Spektralna konvolucija

Koje su najčešći elementi između 57 i 200?

99

113

114

128

129

V

L

N

Q

E

# Spektralna konvolucija

Koje su najčešći elementi između 57 i 200?

99	113	114	128	129
V	L	N	Q	E

5 najčešćih elemenata u konvoluciji



4 aminokiseline peptida NQEL!

# *ConvolutionCyclopeptideSequencing*

1. Formirati spektralnu konvoluciju spektra.
2. Uzeti  $M$  najčešćih elemenata u konvoluciji (između 57 i 200).
3. Primeniti *LeaderboardCyclopeptideSequencing*, formirajući peptide samo na osnovu ovih  $M$  celih brojeva.

# Eksperiment na *Spectrum*<sub>10</sub>

1. Formirati konvoluciju spektra *Spectrum*<sub>10</sub>.

# Eksperiment na $Spectrum_{10}$

1. Formirati konvoluciju spektra $Spectrum_{10}$ .	147
	128
	97
2. Odabrati $M = 10$ najčešćih elemenata.	113
	114
	186
	57
	163
	99
	145

# Eksperiment na *Spectrum*<sub>10</sub>

1. Formirati konvoluciju spektra <i>Spectrum</i> <sub>10</sub> .	147	F
	128	K/Q
	97	P
2. Odabrati $M = 10$ najčešćih elemenata.	113	I/L
	114	N
	186	W
	57	G
	163	Y
	99	V
	145	

# Eksperiment na *Spectrum*<sub>10</sub>

1. Formirati konvoluciju spektra <i>Spectrum</i> <sub>10</sub> .	147	F
	128	K/Q
2. Odabrati $M = 10$ najčešćih elemenata.	97	P
	113	I/L
3. Pokrenuti algoritam...	114	N
	186	W
	57	G
	163	Y
	99	V
	145	

# Eksperiment na $Spectrum_{10}$

1. Formirati konvoluciju spektra $Spectrum_{10}$ .	147	F
	128	K/Q
2. Odabrati $M = 10$ najčešćih elemenata.	97	P
	113	I/L
3. Pokrenuti algoritam...	114	N
	186	W
	57	G
	163	Y
	99	V
	145	

Pronađeni peptid: **VKLF PWFNQY** 😊

# Eksperiment na *Spectrum*<sub>25</sub>

*ConvolutionCyclopeptideSequencing* čak uspeva da rekonstruiše tirocidim B1 iz šumovitijeg spektra *Spectrum*<sub>25</sub>.



# Pregled

- Otkriće antibiotika
- Kako bakterije prave antibiotike?
- Sekvencioniranje antibiotika razbijanjem na komade
- Algoritam grube sile za sekvencioniranje ciklopeptida
- Branch-and-Bound algoritam za sekvencioniranje ciklopeptida
- Prilagođavanje sekvenciranja za spektre sa greškama
- Od 20 do više od 100 aminokiselina
- Spektralna konvolucija
- **Spektri u realnosti**

# Spektri u realnosti

*Spectrum<sub>25</sub> je mnogo manje šumovit nego spektri dobijeni u praksi.*

Takođe, maseni spektrometar ne meri jednostavno fragmente peptida



# Spektri u realnosti

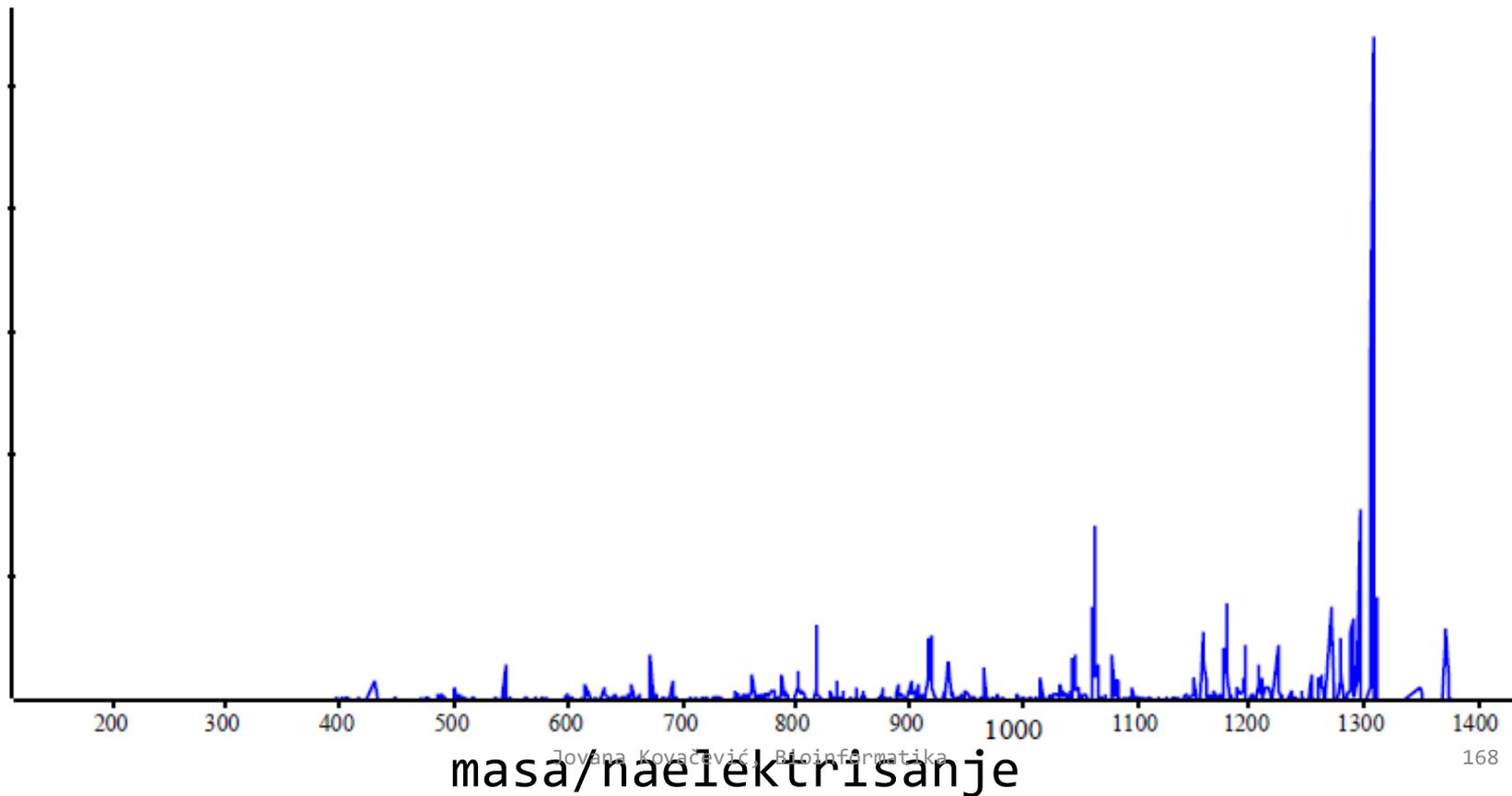
1. Jonizuje fragmente peptida.
2. Sortira fragmente koristeći elektromagnetno polje.
3. Meri odnos **masa/naelektrisanje** za svaki fragment.
4. Određuje **intenzitet** (kao broj jona) u svakom odnosu masa/naelektrisanje.



# Spektar tirocidina B1 u realnosti

Spektar: grafik intenziteta prema odnosu masa/naelektrisanje

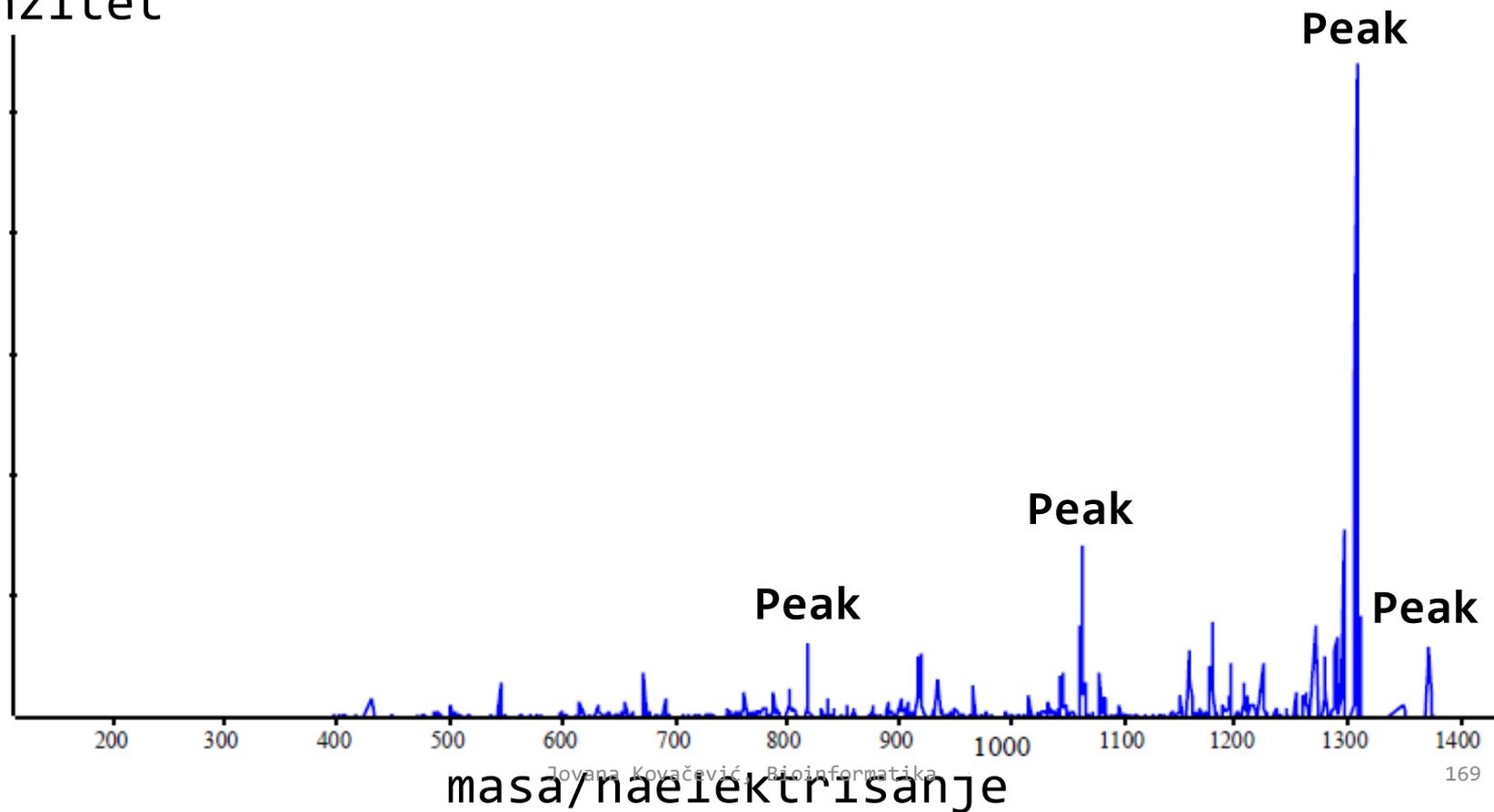
intenzitet



# Spektar tirocidina B1 u realnosti

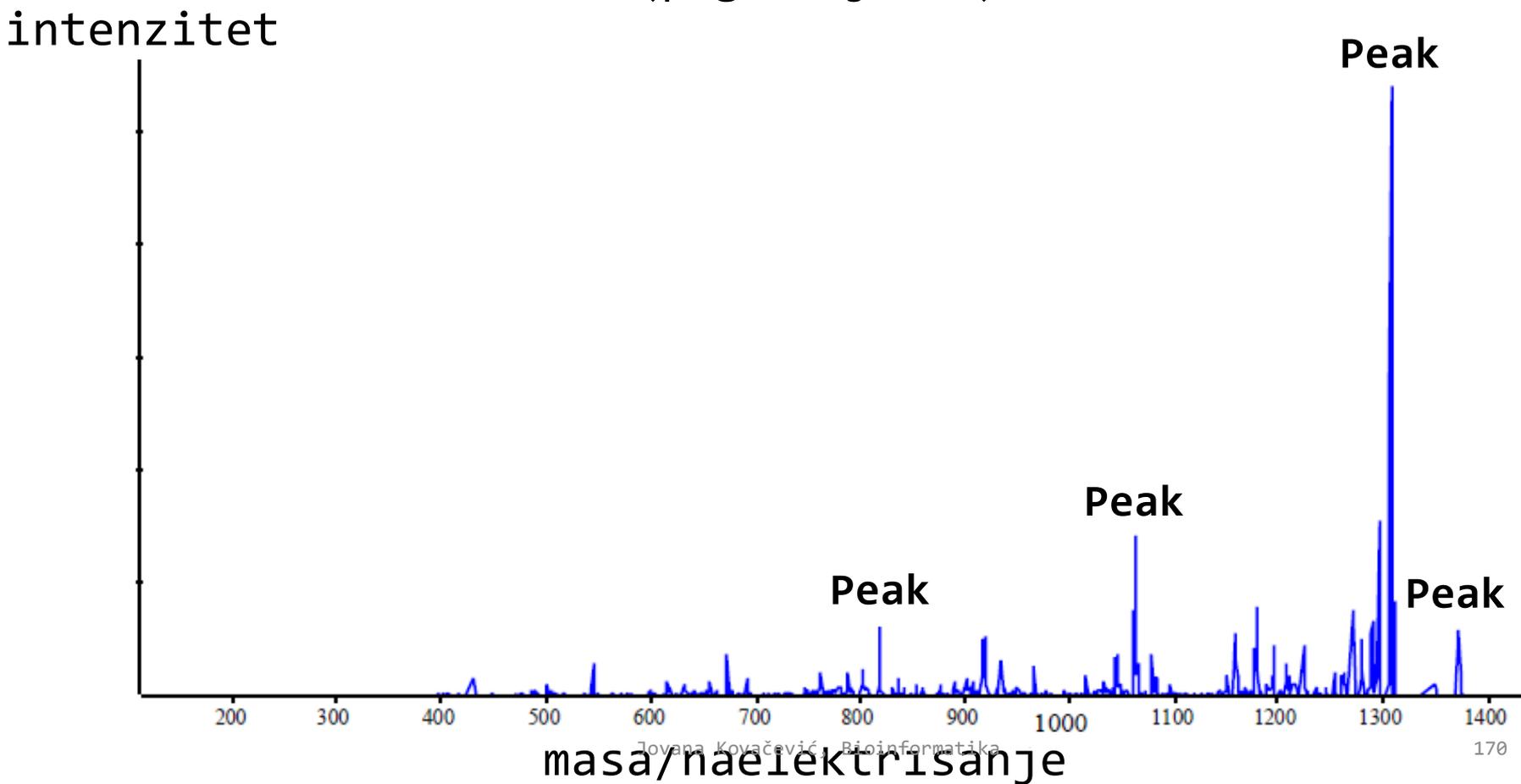
Spektar: grafik intenziteta prema odnosu masa/naelektrisanje

intenzitet



# Spektar tirocidina B1 u realnosti

Izazov: rekonstruisati peptid na osnovu realnog spektra  
(poglavlje 11)



- Slajdovi pokrivaju poglavlje 4 knjige *Bioinformatics Algorithms: an Active Learning Approach*
- Sadržaj slajdova je preuzet sa zvaničnih prezentacija autora i dodatno prilagođen