



Predikce struktury RNA

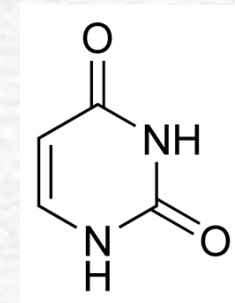
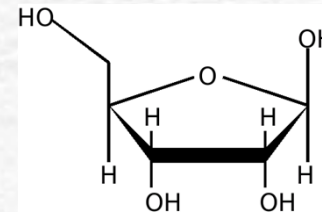


Obsah

- RNA
- Struktura RNA
 - Struktura tRNA
- Reprezentace sekundární struktury RNA
- Predikce sekundární struktury RNA
 - Nusinův algoritmus
 - Zuckerův algoritmus
 - Analýza kovariance sekvencí

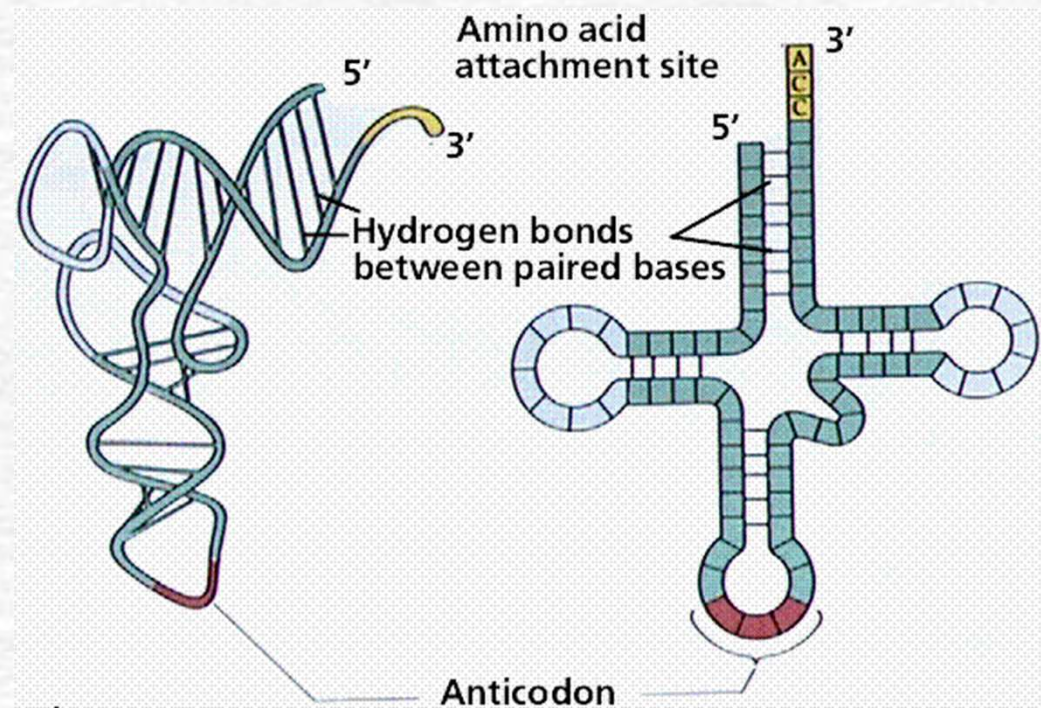
RNA

- ✓ Složení: dusíkatá báze (A, U, C, G), ribóza, fosfát
- ✓ Polynukleotidový řetězec
- ✓ Makromolekuly RNA jsou lineární, většinou jednovláknové, mnohem menší než molekuly DNA
- ✓ Molekuly RNA jsou více flexibilní – různá struktura
- ✓ 3 hlavní typy RNA:
 - ribozomální rRNA (80 - 85%)
 - mediátorová mRNA (1 – 4%)
 - transferová tRNA (15%)
- ✓ Sekvence nukleotidů ovlivňuje sekundární i terciální strukturu a funkci RNA
- ✓ Small nuclear RNA, Small Nucleolar RNA, MicroRNA, Short Interfering RNA



tRNA

- Přenáší jednotlivé aminokyseliny
- Nezbytná při proteosyntéze
- 20 aminokyselin ~ nejméně 20 různých tRNA
- O tom, která aminokyselina se váže na tRNA rozhoduje antikodon



RNA

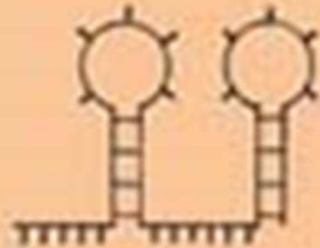
- Desítky až několik tisíc nukleotidů
- mRNA – žádná specifická struktura (regulační funkce)
- tRNA, rRNA – struktura ovlivňuje funkci
- Sekundární struktura:
 - vzory, podle kterých se části vláken ohýbají a tvoří lokální dvouvláknové struktury
 - může být reprezentována 2D diagramem
 - různé motivy
- Vlákná spojeny pomocí vodíkových vazeb
- Párování bází:
 - A-U, G-C
 - G-U**
- tRNA – tvar čtyřlístku



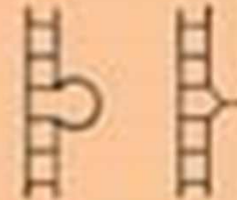
dvoušroubovice



vlásenka
se smyčkou



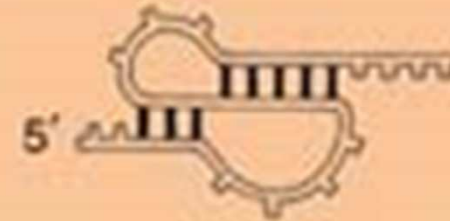
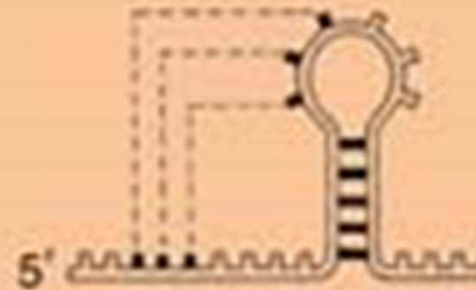
jednořetězcová
oblast



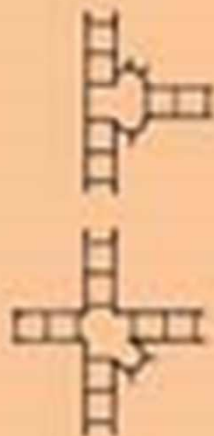
vyboulená smyčka



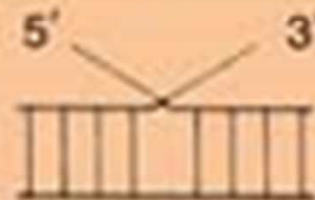
symetrická a asymetrická
vnitřní smyčka



pseudoknot a mechanismus
jeho vzniku



uzel

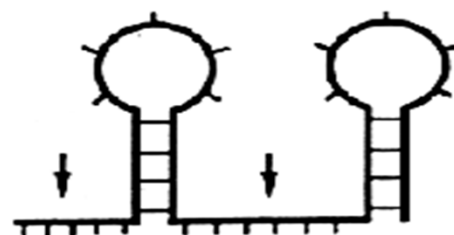


triplex

a. DUPLEXES



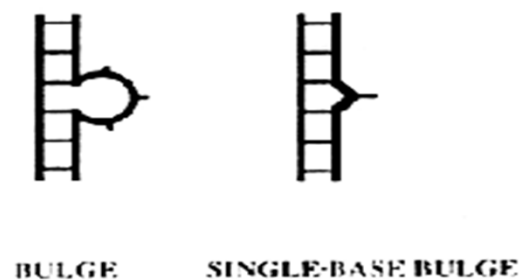
b. SINGLE STRANDED REGIONS



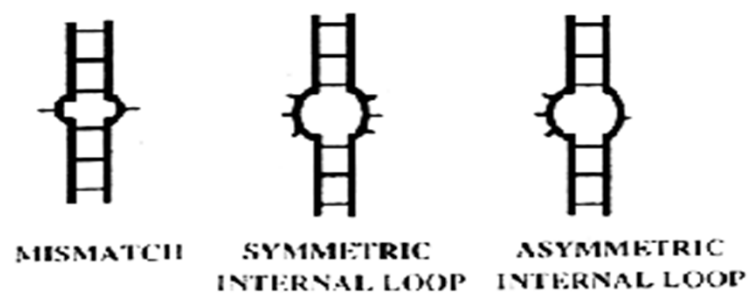
c. HAIRPINS



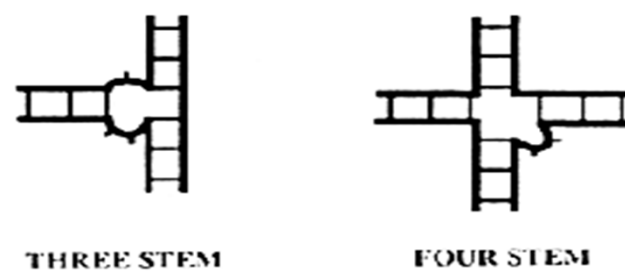
d. BULGES

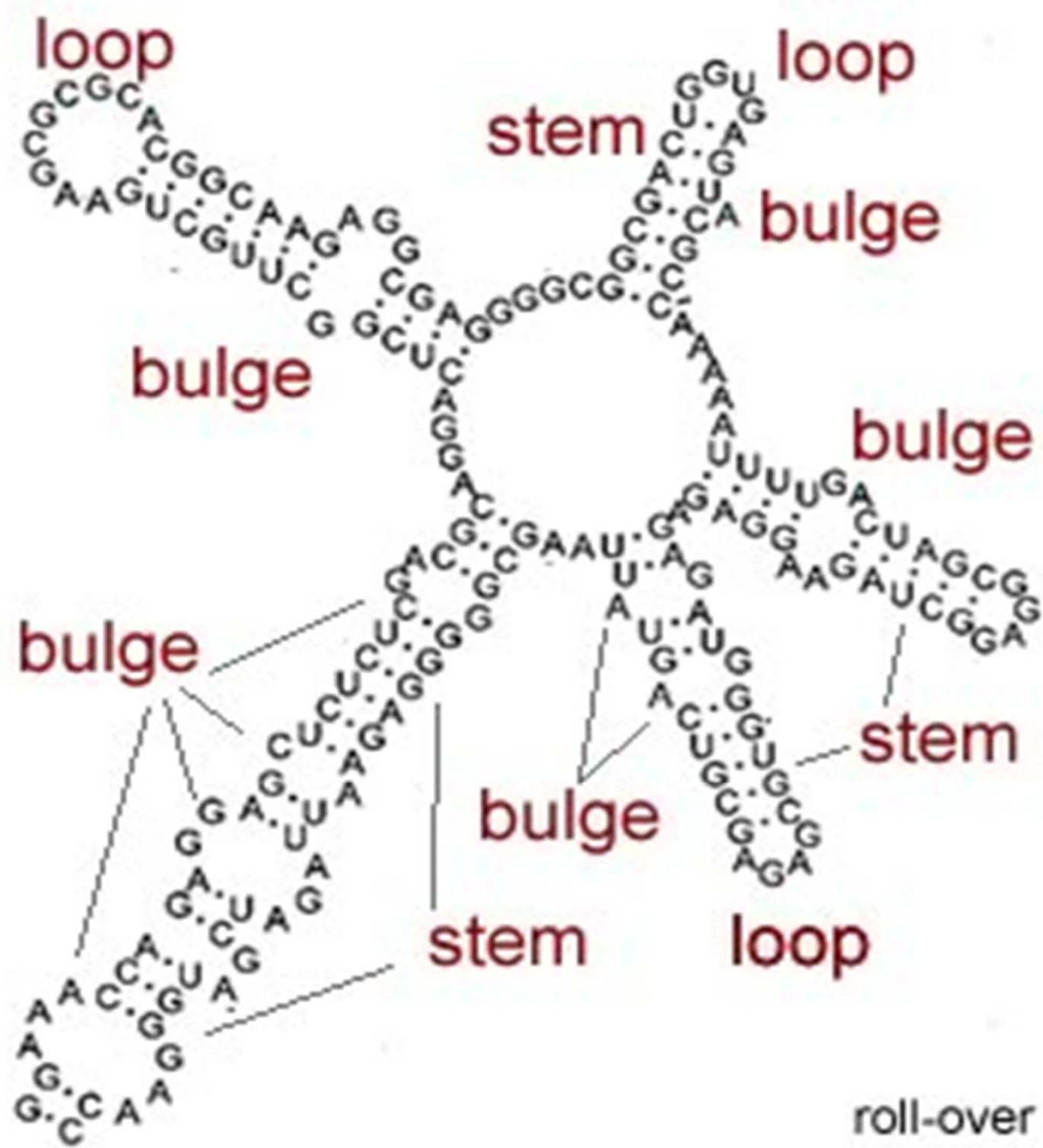


e. INTERNAL LOOPS



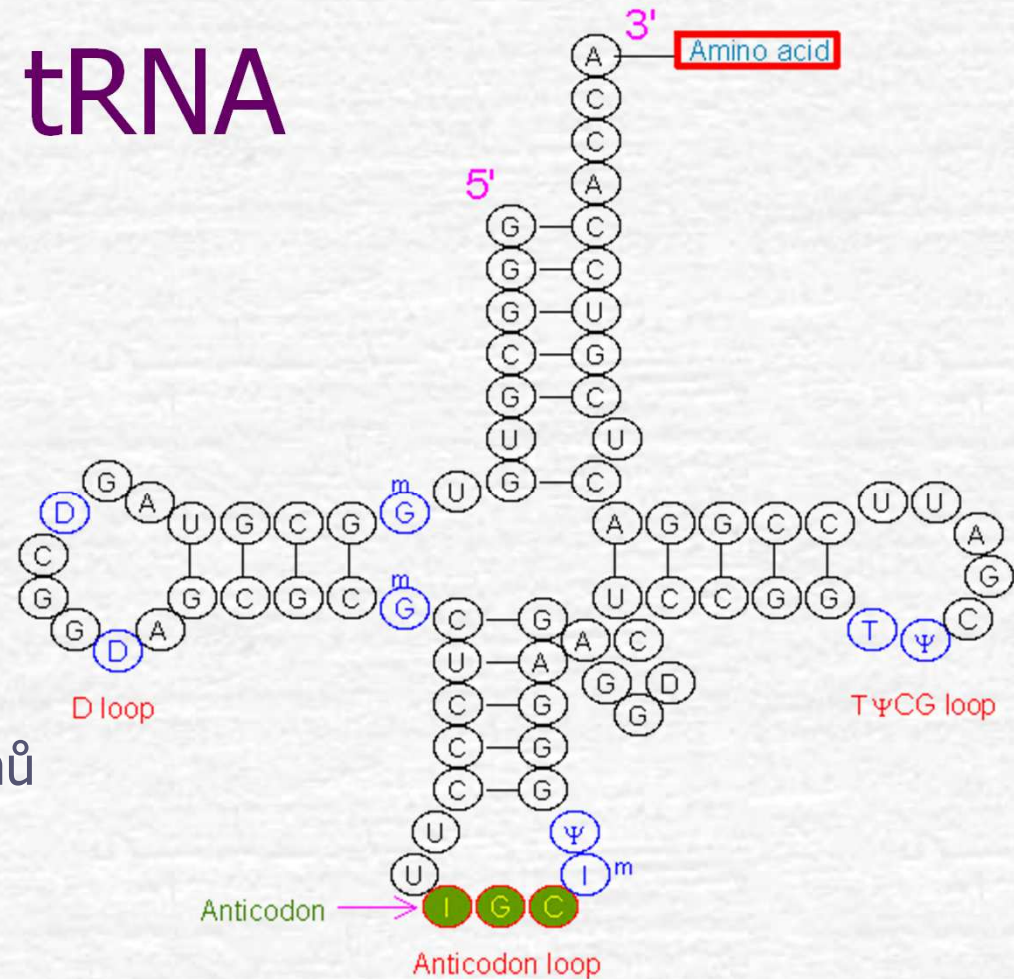
f. JUNCTIONS





Struktura tRNA

- 70 – 90 nukleotidů
- acceptor stem
- D-loop
- T-loop
- Anticodon loop
- Variable loop (arm)
- Člověk – cca 600 tRNA genů



RNA

3D struktura – vyšší struktury:

- Pseudoknot
- Kissing hairpins
- Hairpin-bulge contacts



Pseudoknot

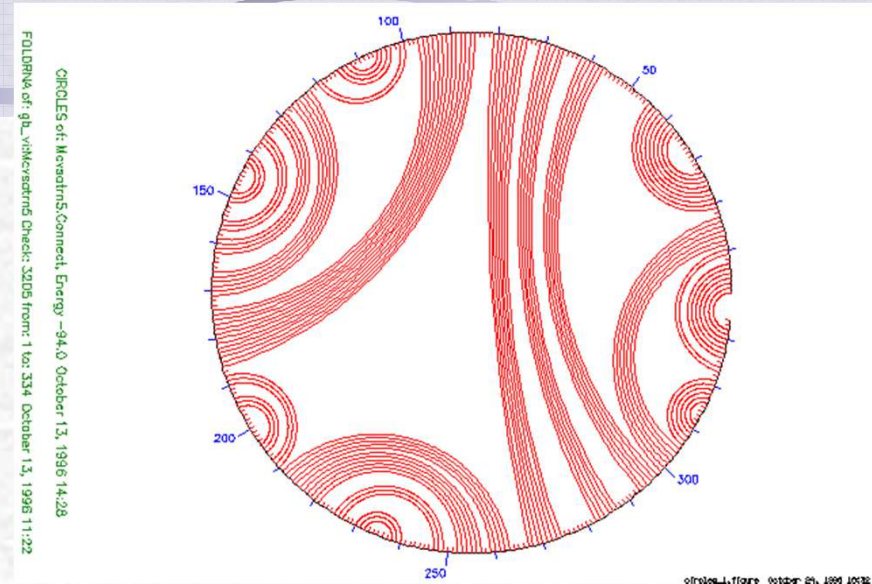
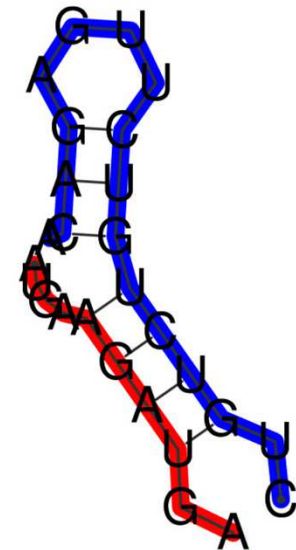
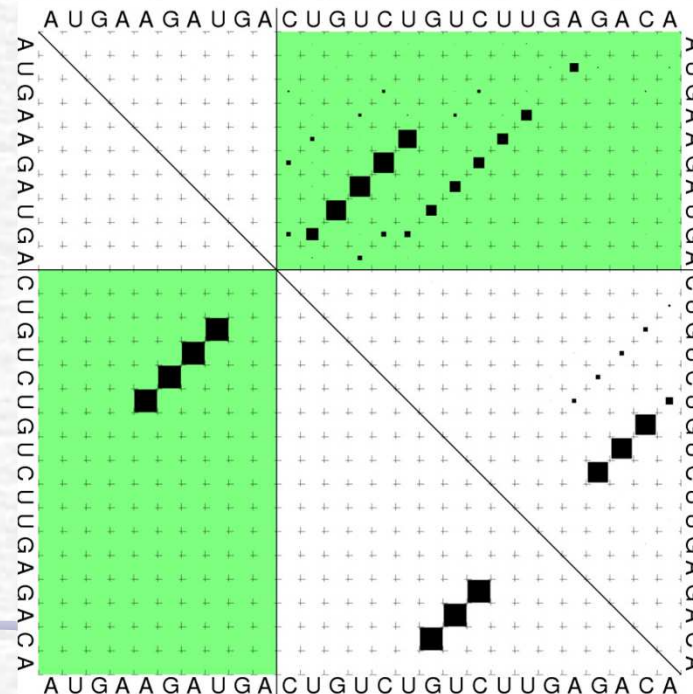
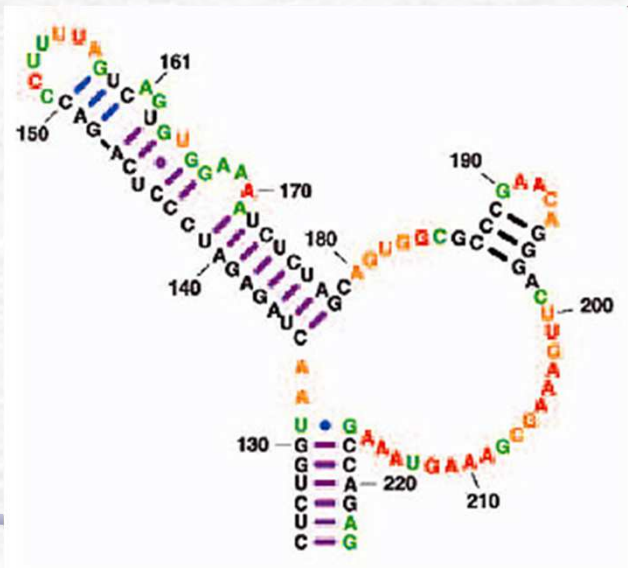


Kissing hairpins

RNA

Sekundární strukturu lze reprezentovat ve 2D:

- Stem-loop diagram
- Dot plot
- Nussinův diagram

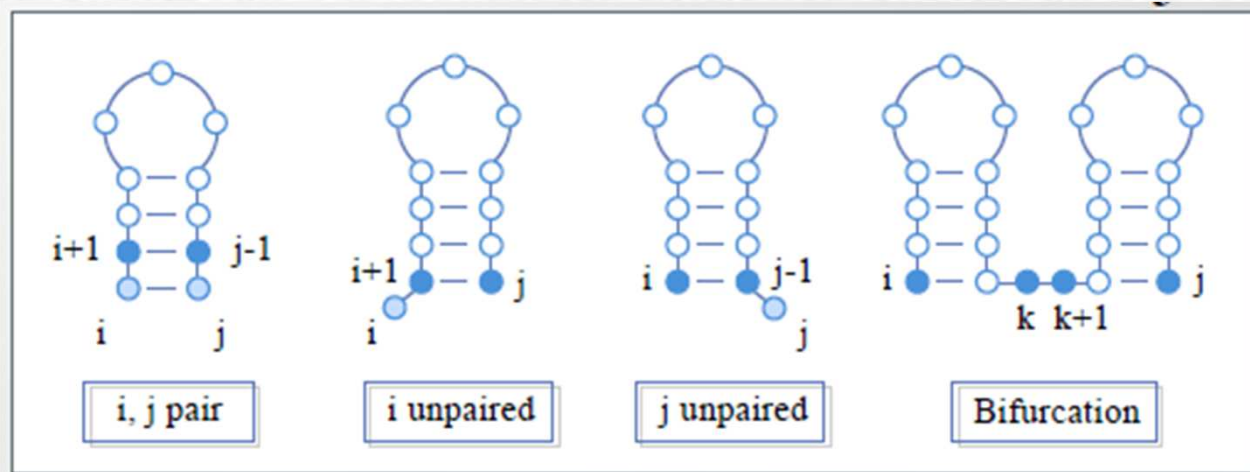


Predikce sekundární struktury RNA

- **Cíl:** předpovědět, které nukleotidy budou spolu vytvářet vodíkové vazby, kde budou jaké smyčky a jaká větvení (vytvořit 2D diagram)
- NP-úplný problém
- Různé přístupy:
 - Maximalizace počtu nukleotidových párů
 - Využití volné energie
 - Analýza kovariance sekvencí
 - Hledání homologních sekvencí

Nusinnův algoritmus

- ✓ Hledá strukturu s maximálním počtem spárovaných bází
- ✓ Princip dynamického programování
- ✓ Předpoklad: již známe sekundární strukturu pro část sekvence
- ✓ Hledáme, jak k této struktuře přidat další dvě báze (z každé strany jednu) – 4 možnosti



Nusinnův algoritmus

1. Konstrukce a inicializace matice
 - Záhlaví řádků a sloupců cílová sekvence
 - Hlavní diagonála + diagonála pod ní => hodnota 0
2. Vyplnění matice (pravé horní poloviny)
 - pomocí rekurzivního algoritmu
 - Výběr maximální hodnoty
3. Zpětný průchod maticí
 - sekundární struktura RNA
 - Využití informace z vyplňování matice

Konstrukce a inicializace

→ j

i ↓

	A	A	A	G	C	C	C	U	U
A	0								
A	0	0							
A	X	0	0						
G	X	X	0	0					
C	X	X	X	0	0				
C	X	X	X	X	0	0			
C	X	X	X	X	X	0	0		
U	X	X	X	X	X	X	0	0	
U	X	X	X	X	X	X	X	0	0

Vyplnění matice


- Od buňky (1,2) po diagonále

$$G(i, j) = \text{Max} \begin{cases} G(i+1, j) \\ G(i, j-1) \\ G(i+1, j-1) + d(i, j) \\ \text{Max}[G(i, k) + G(k+1, j)] \forall k = i+1 \dots j-1 \end{cases}$$

- $d(i, j)$ hodnota 0 nebo 1, podle toho, zda báze na pozicích i a j se mohou párovat

Vyplnění matice

	A	A	A	G	C	C	C	U	U
A	0	0							
A	0	0	0						
A	X	0	0	0					
G	X	X	0	0	1				
C	X	X	X	0	0				
C	X	X	X	X	0	0			
C	X	X	X	X	X	0	0		
U	X	X	X	X	X	X	0	0	
U	X	X	X	X	X	X	X	0	0



Vyplnění matice

	A	A	A	G	C	C	C	U	U
A	0	0	0	0	1	1	1	2	3
A	0	0	0	0	1	1	1	2	3
A	X	0	0	0	1	1	1	2	2
G	X	X	0	0	1	1	1	1	1
C	X	X	X	0	0	0	0	0	0
C	X	X	X	X	0	0	0	0	0
C	X	X	X	X	X	0	0	0	0
U	X	X	X	X	X	X	0	0	0
U	x	X	X	X	X	X	x	0	0

Zpětný průchod maticí

	A	A	A	G	C	C	C	U	U
A	0	0	0	0	1	1	1	2	<u>3</u>
A	0	0	0	0	1	1	1	2	<u>3</u>
A	X	0	0	0	1	1	1	<u>2</u>	2
G	X	X	0	0	1	1	<u>1</u>	1	1
C	X	X	X	0	0	<u>0</u>	0	0	0
C	X	X	X	X	0	<u>0</u>	0	0	0
C	X	X	X	X	X	0	0	0	0
U	X	X	X	X	X	X	0	0	0
U	x	X	x	x	x	x	x	0	0

Zpětný průchod maticí

	A	C	U	G	A	G	U	C	C	A	A	G	G
A	0	0	1	1	1	2	3	3	3	3	3	4	5
C		0	0	1	1	2	2	2	2	3	3	4	4
U			0	0	1	1	1	2	2	3	3	3	3
G				0	0	0	1	2	2	2	2	3	3
A					0	0	1	1	1	1	1	2	3
G						0	0	1	1	1	1	2	2
U							0	0	0	1	1	1	2
C								0	0	0	0	1	2
C									0	0	0	1	1
A										0	0	0	0
A											0	0	0
G												0	0
G													0

Nusinnův algoritmus

Nedostatky:

- Smyčky s nulovým počtem residuí
 - Fyzikálně nemožné
 - Minimální počet bází ve smyčce: 3
- Stejná váha jednotlivých bázových párů
 - Ve skutečnosti různé bázové páry – různá stabilita
- Neuvažuje se vliv sousedních bázových párů

Zuckerův algoritmus

- Odstraňuje nedostatky Nussinova algoritmu
 - Modifikace 2. a 3. kroku Nussinova algoritmu
- Využití jednoduchých energetických pravidel
 - Gibbsova volná energie pro všechny báze páry
 - různým báze páram přidělena různá váha
 - Destabilizační vliv smyček
- MFOLD

Analýza kovariance sekvencí

- Analýza **záměn nukleotidů**, které **neovlivní schéma párování bází** (neovlivní sekundární ani terciární strukturu a tedy ani funkci)
- Tzn. jestliže zjistíme, že při záměně jednoho nukleotidu, dochází zároveň k záměně dalšího nukleotidu na jiném místě sekvence, je pravděpodobné, že tyto nukleotidy spolu budou interagovat
- Barevné značení pozic zarovnání dvou sekvencí
 - hledání regionů stejně označených pozic (pozice beze změny, transition, transversion)
 - vyžaduje vytvoření a inspekci mnoha zarovnání
- Využití vzájemné informace mezi sloupci vícenásobného zarovnání sekvencí (teorie informace)
 - mezi každou dvojicí sloupců
 - 0 – žádná vazba mezi sloupci
 - 1 – první sloupec perfektně určuje nukleotid v druhém sloupci

AUGUAGCGGA

GAGCAGGGGA

* * * * * * * *

CSFG

- Stochastic Context Free Grammars
- Pomohou určit, do které rodiny RNA nová sekvence patří (identifikují možné struktury)
- Pomocí trénovací množiny se určí pravděpodobnosti jednotlivých pravidel

$$S \rightarrow aSu \mid uSa \mid cSg \mid gSc \mid aS \mid cS \mid uS \mid gS \mid Sa \mid Sc \mid Sg \mid Su \mid SS \mid \varepsilon$$
$$\begin{aligned} S \rightarrow SS \rightarrow aSuS \rightarrow acSguS \rightarrow acuSaguS \rightarrow acugSaguS \rightarrow \\ acugaguS \rightarrow acugagucSg \rightarrow acugaguccSgg \rightarrow acugaguccSagg \\ \rightarrow acugaguccaSagg \rightarrow acugaguccaagg \end{aligned}$$