

# Eine RNA-Menagerie: miRNAs und andere kodierende und nichtkodierende RNAs

Peter N. Robinson

Institut für medizinische Genetik  
Charité Universitätsmedizin Berlin

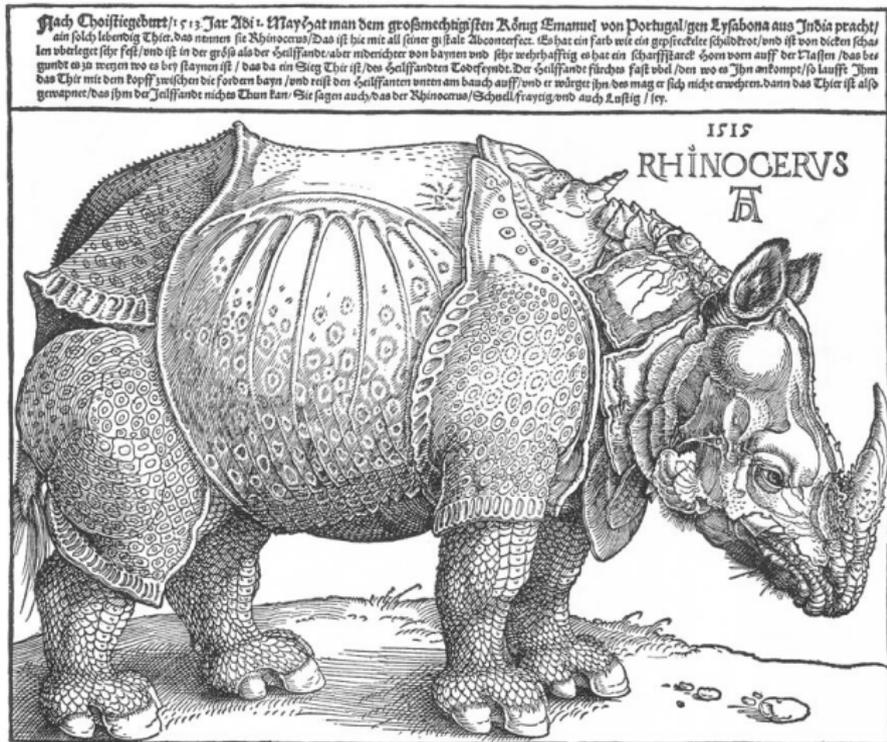
18. Dezember 2015

# Outline

- 1 Eine RNA-Menagerie
- 2 mRNA, tRNA, rRNA, snRNA, snoRNA
- 3 RNA-Struktur
- 4 RNA: Sekundärstruktur
- 5 miRNAs

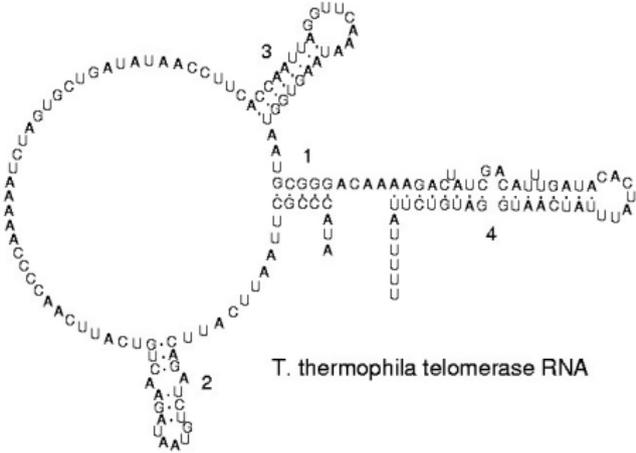
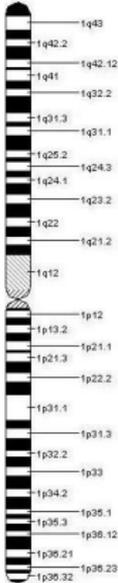
# Eine RNA-Menagerie

- mRNA
- tRNA
- rRNA
- snRNA
- snoRNA
- miRNA
- XIST-RNA
- piRNA





# RNA vs. DNA (2): Länge

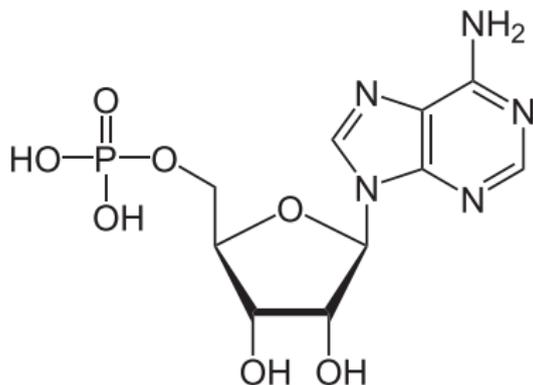


- DNA: Millionen von Basenpaaren

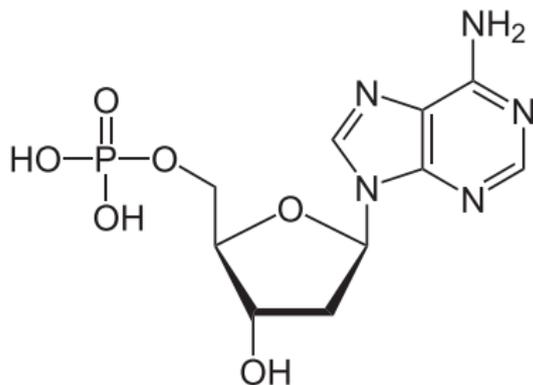
- RNA: ~20 bis mehrere tausend Nukleotide

Bildquelle: Wikipedia

# RNA vs. DNA (3): Biochemie



Adenosinmonophosphat (Wikipedia)



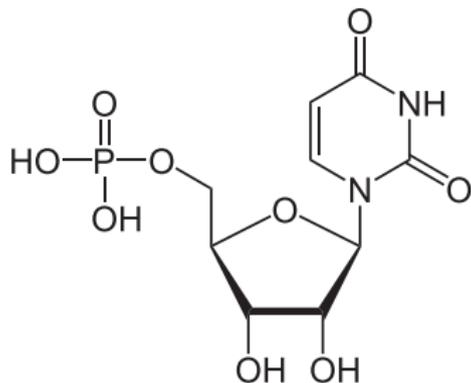
Desoxyadenosinmonophosphat (Wikipedia)

- RNA: Ribose

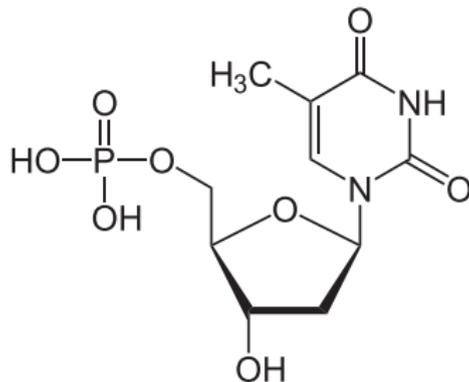
- DNA: 2-Desoxyribose

*Die zusätzliche 2-Hydroxylgruppe macht die RNA weniger stabil als die DNA, da sie leichter hydrolysiert werden kann*

# RNA vs. DNA (4): Biochemie



Uridinmonophosphat (Wikipedia)



Desoxythymidinmonophosphat (Wikipedia)

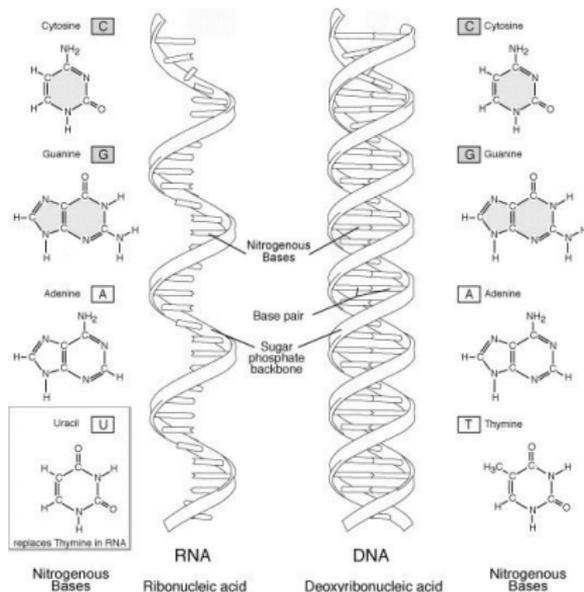
- RNA: Uracil

- DNA: Thymin

*In der RNA ist nicht Thymin (T) sondern Uracil (U) zu Adenin (A) komplementär*

# RNA vs. DNA (5): Biologische Rollen

- DNA
  - ▶ Trägerin der Erbinformation.
- RNA
- Im Gegensatz zur DNA spielt die Struktur der RNA bei deren Funktion eine wesentliche Rolle
- 3D-Struktur aus mehreren kürzeren Helices, ähnlich wie Proteine
- Katalyse wie bei Enzyme
- Zahlreiche unterschiedliche Funktionen ...



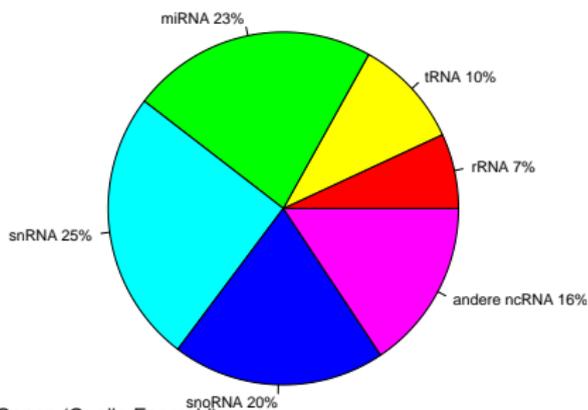
Wikipedia

# Outline

- 1 Eine RNA-Menagerie
- 2 mRNA, tRNA, rRNA, snRNA, snoRNA**
- 3 RNA-Struktur
- 4 RNA: Sekundärstruktur
- 5 miRNAs

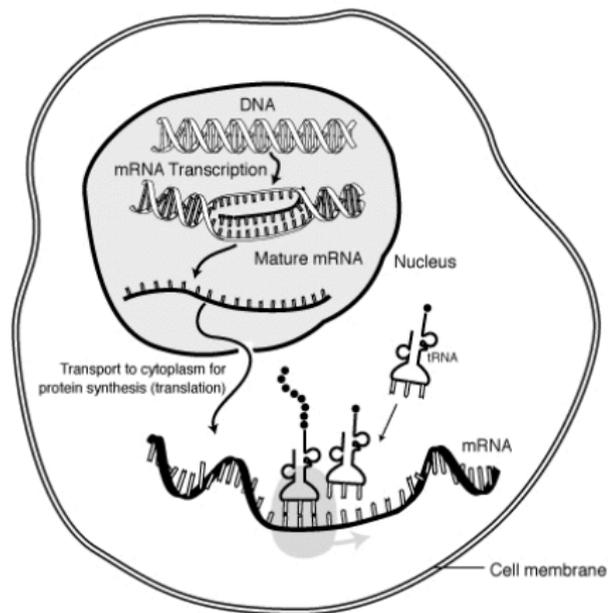
# Eine RNA-Menagerie

- Zahlreiche Klassen von RNA
- Es folgt zunächst ein Überblick über mRNA, tRNA, rRNA, snRNA, die Sie bereits kennen (sollten)



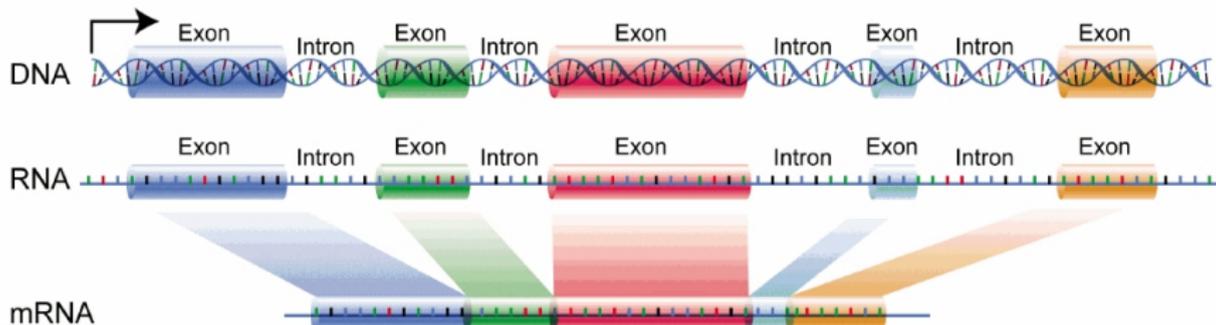
- Verteilung von ca. 13,000 RNA Genen (Quelle Ensembl)

# mRNA



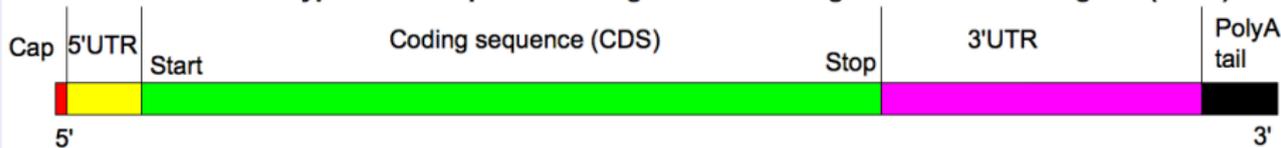
Wikipedia commons

## mRNA-Spleißen



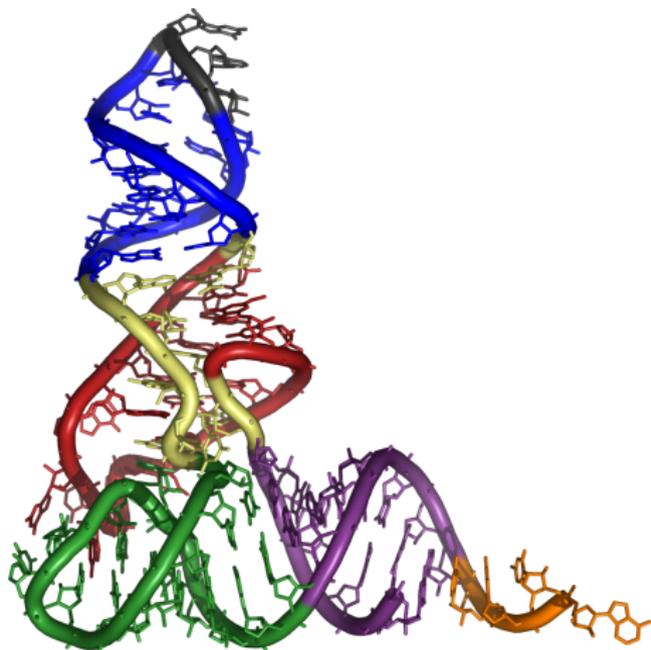
## Struktur einer reifen mRNA

The structure of a typical human protein coding mRNA including the untranslated regions (UTRs)



# tRNA

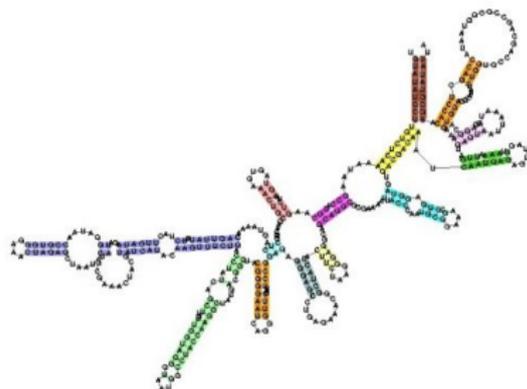
- ~ 85 Nukleotide lang
- Struktur etwa wie der Buchstabe "L"
- Drei-Nukleotid Anticodon auf der Spitze des L bindet an komplementäres Codon in mRNA
- Der "Fuß" des L bindet an eine der 20 Aminosäuren



Anticodonarm: blau, Anticodon: schwarz (Wikipedia commons)

# rRNA

- ribosomale RNA
- zusammen mit den ribosomalen Proteinen am Aufbau und der enzymatischen Aktivität des Ribosoms und damit an der Proteinsynthese beteiligt.
- 60S Untereinheit (28S,6,8S, 5S rRNA) und 40S Untereinheit (18S rRNA)



5' Domäne der kleinen rRNA (Wikipedia commons)

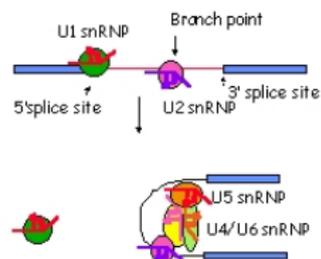
## Das Spleißosom

- small nuclear RNA
- Immer mit spezifischen Protein assoziiert: *small nuclear ribonucleoproteins* (snRNP)
- Spleißen
- Regulation von Transkriptionsfaktoren (7SK RNA)
- Aufrechterhaltung der Telomere

### Roles of snRNPs

- **U1 snRNP** binds 5' splice site
- **U2 snRNP** binds to branch point
- **U4/U6 snRNP**, snRNAs are base paired. U6 is catalytic
- **U5 snRNP** contacts the 5' splice site
  - forms tri-snRNP complex with U4/U6

### Assembly of the spliceosome



# Outline

- 1 Eine RNA-Menagerie
- 2 mRNA, tRNA, rRNA, snRNA, snoRNA
- 3 RNA-Struktur**
- 4 RNA: Sekundärstruktur
- 5 miRNAs

# Ebenen der RNA-Struktur

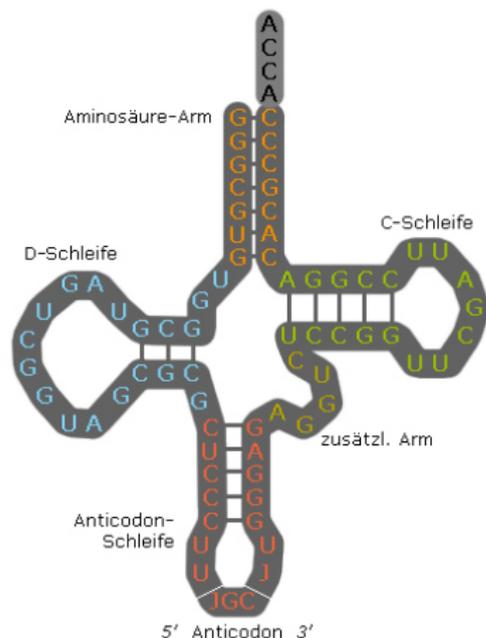
- Primärstruktur: Die Nukleotidsequenz
  - ▶ z.B. die Sequenz CUCUCGGUAAGCUUAGGUACCA
- Sekundärstruktur: Paare von Nukleotiden, welche eine Wasserstoffbrückenbildung miteinander eingehen
- Hairpin- und Stemloop-Strukturen, Helixstrukturen sowohl Einzelstrang- als Doppelstrangbereiche.
- Tertiärstruktur (3D)
- Quartärstruktur (Beziehung zu anderen RNAs/Proteinen im Komplex)



# RNA-Sekundärstruktur

- G–C: drei Wasserstoffbrücken
- A–U: zwei Wasserstoffbrücken
- G–U: Eine Wasserstoffbrücke ("wobble pair" → Wackelpaar)

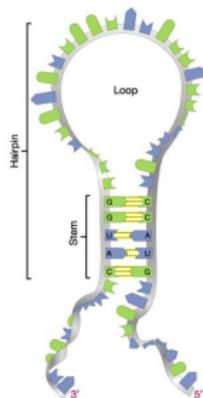
KLEEBLATTSTRUKTUR DER tRNA



Wikipedia

# RNA-Sekundärstruktur

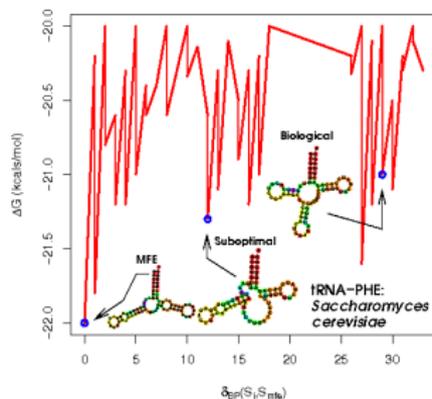
- RNA-Sekundärstruktur:  
häufige Motive wie Hairpin,  
Helix, stem loop, bulge loop,  
interior loop, multiple loop
- RNA-Strukturbestimmung  
experimentell schwierig, daher  
ein wichtiges Thema für die  
Bioinformatik



Hairpin loop (Haarnadel)

# Strukturvorhersage durch Maximierung von Basenpaarungen

- Minimieren der freien Energie
- Wasserstoffbrücken sind eine Schlüsselkomponente der RNA-Stabilität
- Viele Algorithmen versuchen daher, die Anzahl der Wasserstoffbrücken zu maximieren



S. cerevisiae tRNA-PHE: Energien alternativer Strukturen (Wikipedia)

# Strukturvorhersage durch Maximierung von Basenpaarungen

## Primary Proximity Constraint

Bilden Nukleotide  $i$  und  $j$  eine Wasserstoffbrücke, dann  $|i - j| > 3$

Diese Bedingung ergibt sich aus der Tatsache, dass eine RNA-Kette nicht ausreichend flexibel ist, damit sich eine Wasserstoffbrücke zwischen eng benachbarten Nukleotiden bilden könnte.

# Strukturvorhersage durch Maximierung von Basenpaarungen

## Nesting Constraint

Sind  $(i, j)$  und  $(p, q)$  zwei Wasserstoffbrücken (Paare von Nukleotiden), wobei  $i < p < j$ , dann gilt  $q < j$

Diese "Schachtelungsbedingung" verbietet überkreuzte Wasserstoffbrücken, erlaubt dagegen geschachtelte Wasserstoffbrücken. Überkreuzte Wasserstoffbrücken, so genannte *Pseudoknoten*, kommen relativ selten vor. Algorithmen, welche Pseudoknoten zulassen, sind wesentlich weniger effizient als solche, die sie verbieten.

# Strukturvorhersage durch Maximierung von Basenpaarungen

- Beispiel: UUGACAUCG
- Ziel: die Sekundärstruktur mit der maximalen Anzahl an Basenpaaren finden, wobei zwischen zwei paarenden Basen mindestens eine ungepaarte stehen soll



# RNA-Struktur: Klammern und Punkte

- Wir können die RNA-Struktur als Strings mit balanzierten Klammern und Punkten mit der entsprechenden Schachtelebene (nesting level) darstellen

UUGACAUCG	UUGACAUCG
(..) (...)	(.(.....))
011001110	011222210

UUGACAUCG     2+3=5

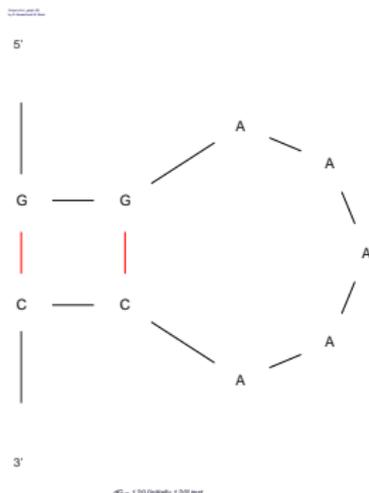
UUGACAUCG     1+3=4

# RNA-Struktur: Klammern und Punkte

- Die Sekundärstruktur kann von der Sequenz und Klammerndarstellung ermittelt werden
- Beispiel

GGAAAAACC

((.....))

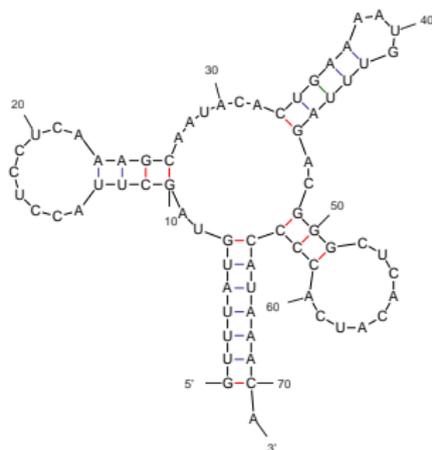


# Übung für zu Hause

- die DNA-Sequenz für das menschliche Mitochondriengenom aufrufen (Accessionnummer: 17981852)
- `http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&id=17981852`
- in der FEATURES-Liste nach dem Gen für tRNA-Phe suchen (Position 579–649), die Sequenz für dieses Gen im neuen Fenster aufrufen
- Display auf FASTA einstellen, die FASTA-formatierte Sequenz kopieren
- Die Sequenz mit dem mfold-Programm analysieren
- `http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi`

# Übung für zu Hause

© 2000 Pearson Education, Inc.



dG = -10.44 [initially -11.40] tRNA-Phe

GUUUUAGUAGCUUACCUCUCAAAGCAAUACACUGAAAAUGUUUAGACGGGCUCACAUACACCCCAUAAACA  
(((((((.....)))))).....((((.....))))..((((.....)))))))).

- Wieviele Strukturen werden vorhergesagt? Worin unterscheiden sie sich?

# Outline

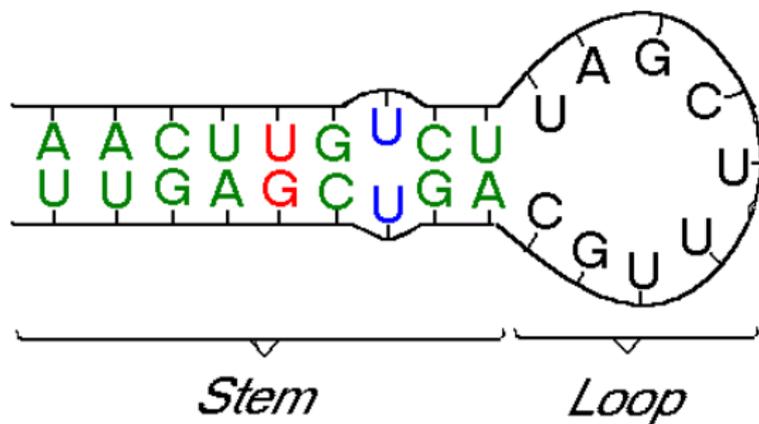
- 1 Eine RNA-Menagerie
- 2 mRNA, tRNA, rRNA, snRNA, snoRNA
- 3 RNA-Struktur
- 4 RNA: Sekundärstruktur**
- 5 miRNAs

# Bioinformatik der RNA-Faltung

- Zahlreiche Algorithmen
- Dynamic programming
- Freie Energie
- Hier stellen wir einen vereinfachten DP-Algorithmus vor (Nussinov)

# Bioinformatik der RNA-Faltung

- Dominiert wird eine RNA Struktur von den Basenpaaren die sich zwischen komplementären Basen bilden.
- Die meisten dieser Basenpaarungen sind Watson-Crick Basenpaare.
- "Palindrome" häufig



- *Watson-Crick pairs*
- *UG pairs*
- *Mismatch*

# Bioinformatik der RNA-Faltung

- Eine vereinfachte Version des Nussinov-Algorithmus versucht, die Anzahl der gepaarten Basen zu maximieren
- Unser Score: +1 für Basenpaar, 0 für alles Andere
- Wir betrachten eine RNA-Sequenz  $1, 2, \dots, n$

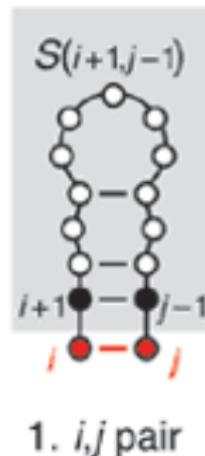
$S(i, j)$  Max. Score für die Subsequenz  $i, i + 1, \dots, j$ .

- $S(i, j)$  kann rekursiv berechnet werden (Dynamic programming)

# RNA-Faltung: DP (1)

- Falls  $i, j$  ein WC-Baasenpaar sind

$$S^1(i, j) = 1 + S(i + 1, j - 1)$$



# RNA-Faltung: DP (2)

- Falls  $i$  ungepaart bleibt

$$S^2(i, j) = S(i + 1, j)$$

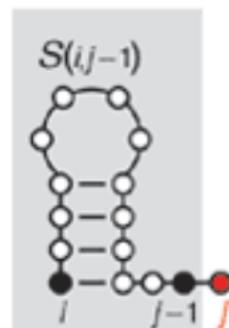


2.  $i$  unpaired

# RNA-Faltung: DP (3)

- Falls  $j$  ungepaart bleibt

$$S^3(i, j) = S(i, j-1)$$

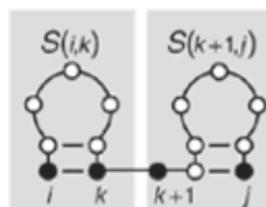


3.  $j$  unpaired

# RNA-Faltung: DP (4)

- Falls  $i, j$  jeweils mit anderen Nukleotiden gepaart sind, handelt es sich um eine Bifurkation, die Struktur  $S(i, j)$  besteht dann aus den Strukturen für zwei Subsequenzen  $i, \dots, k$  und  $k + 1, \dots, j$ :

$$S^4(i, j) = \max_{i < k < j} S(i, k) + S(k + 1, j)$$



4. Bifurcation

# RNA-Faltung: Dynamic programming

- Falls  $i$  und  $j$  also ein Basenpaar bilden, wird dem Score ein Punkt hinzugefügt, ganz egal was die Struktur der Subsequenz  $i+1, \dots, j-1$  ist
- Daher müssen wir den Score für  $S(i+1, j-1)$  nicht neu berechnen
- Ähnliche Argumente gelten für die anderen drei Möglichkeiten
- Der optimale Score  $S(i, j)$  ist daher lediglich das Maximum der vier Optionen

$$S(i, j) = \max \begin{cases} 1 + S(i+1, j-1) \\ S(i+1, j) \\ S(i, j-1) \\ \max_{i < k < j} S(i, k) + S(k+1, j) \end{cases}$$

# RNA-Faltung: Dynamic programming

- Realistischere Algorithmen betrachten Stems, Haarnadelstrukturen, Bulges, innere Schleifen, auch Pseudoknoten



Grafik: Steffen P, Giegerich R (2005) *BMC Bioinformatics* 6:224.

# Initialization

```
1:  $s = \text{newMatrix}(n, n)$ 
2: for  $i \leftarrow 2$  to  $n$  do
3:    $s[i, i-1] \leftarrow 0$ 
4: end for
5: for  $i \leftarrow 1$  to  $n$  do
6:    $s[i, i] \leftarrow 0$ 
7: end for
```

- DP-Matrix initialisieren
- Nur die Hälfte der Matrix wird verwendet
- Wir analysieren die Faltung der Sequence GGGAAAUCC

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0								
G <sub>2</sub>	0	0							
G <sub>3</sub>		0	0						
A <sub>4</sub>			0	0					
A <sub>5</sub>				0	0				
A <sub>6</sub>					0	0			
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

# Füllen der Matrix (1)

```
1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for
```

- Die For-Schleifen besuchen sukzessive die Nebendiagonalen
- Die DP-Rekursion prüft vier Möglichkeiten (a,b,c,d):

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{i < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	?							
G <sub>2</sub>	0	0							
G <sub>3</sub>		0	0						
A <sub>4</sub>			0	0					
A <sub>5</sub>				0	0				
A <sub>6</sub>					0	0			
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

## Füllen der Matrix (2)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for
    
```

- Die DP-Rekursion prüft vier Möglichkeiten (a,b,c,d)

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{j < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	?							
G <sub>2</sub>	0	0							
G <sub>3</sub>		0	0						
A <sub>4</sub>			0	0					
A <sub>5</sub>				0	0				
A <sub>6</sub>					0	0			
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

- Option a) prüft, ob eine Basenpaarung der Nukleotiden  $i$  und  $j$  gibt ( $\delta(i,j)$ ).
- Prüfe, ob die Basenpaarung  $i,j$  zusammen mit dem vorausgegangenen Alignment in  $s[i+1,j-1]$  (in blau) einen maximalen Score hat
- $\delta(2,1)$  ergibt 0, weil  $G_2$  und  $G_1$  nicht paaren können.
- $\delta(i,j) + s[i+1,j-1] = \delta(1,2) + s[2,1] = 0 + 0 = 0$ .

# Füllen der Matrix (3)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for
    
```

- Die DP-Rekursion prüft vier Möglichkeiten (a,b,c,d)

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{i < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	?							
G <sub>2</sub>	0	0							
G <sub>3</sub>		0	0						
A <sub>4</sub>			0	0					
A <sub>5</sub>				0	0				
A <sub>6</sub>					0	0			
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

- Option b) prüft, ob das vorausgegangene Alignment  $s[i+1,j]$  (in blau) zusammen mit einem ungepaarten Nukleotid  $i$  einen maximalen Score hat
- $s[i+1,j] = s[2,2] = 0$ .

# Füllen der Matrix (4)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for
    
```

- Die DP-Rekursion prüft vier Möglichkeiten (a,b,c,d)

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{j < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	?							
G <sub>2</sub>	0	0							
G <sub>3</sub>		0	0						
A <sub>4</sub>			0	0					
A <sub>5</sub>				0	0				
A <sub>6</sub>					0	0			
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

- Option c) prüft, ob das vorausgegangene Alignment  $s[i, j-1]$  (in blau) zusammen mit einem ungepaarten Nukleotid  $j$  einen maximalen Score hat
- $s[i, j-1] = s[1, 1] = 0$ .

# Füllen der Matrix (5)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for
    
```

- Die DP-Rekursion prüft vier Möglichkeiten (a,b,c,d)

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{i < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0							
G <sub>2</sub>	0	0							
G <sub>3</sub>		0	0						
A <sub>4</sub>			0	0					
A <sub>5</sub>				0	0				
A <sub>6</sub>					0	0			
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

- Option d) ist in diesem Schritt nicht möglich, weil es bei  $i = 1, j = 2$  keine Ganzzahl  $k$  gibt mit  $i < k < j$ .
- $s[2, 1] = \max(a, b, c) = \max(0, 0, 0) = 0$

# Füllen der Matrix (6)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for

```

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{i < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0							
G <sub>2</sub>	0	0	0						
G <sub>3</sub>		0	0	0					
A <sub>4</sub>			0	0	0				
A <sub>5</sub>				0	0	0			
A <sub>6</sub>					0	0	?		
U <sub>7</sub>						0	0		
C <sub>8</sub>							0	0	
C <sub>9</sub>								0	0

- Die folgenden Schritte sind ähnlich, bis wir Zelle (6,7) erreichen
- $\delta(6,7) = 1$ , weil A<sub>6</sub> mit U<sub>7</sub> eine Basenpaarung bilden kann
- Daher erreicht  $\delta(6,7) + s[7,6]$  den maximalen Score von 1

# Füllen der Matrix (7)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i, j] ← max(a, b, c, d)
6:   end for
7: end for

```

$$S(i, j) = \max \begin{cases} \text{a) } \delta(i, j) + S(i+1, j-1) \\ \text{b) } S(i+1, j) \\ \text{c) } S(i, j-1) \\ \text{d) } \max_{j < k < j} S(i, k) + S(k+1, j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0	0						
G <sub>2</sub>	0	0	0	0					
G <sub>3</sub>		0	0	0	0				
A <sub>4</sub>			0	0	0	0			
A <sub>5</sub>				0	0	0	?		
A <sub>6</sub>					0	0	1		
U <sub>7</sub>						0	0	0	
C <sub>8</sub>							0	0	0
C <sub>9</sub>								0	0

- Zelle (6, 7) bekommt den Wert 1 und wir fahren analog fort
- Die nächste "interessante" Zelle ist (5, 7)
- Es gibt zwei Möglichkeiten, den maximalen Score für  $s[5, 7]$  zu erreichen
  - 1 Neue Basenpaarung  $A_5$  und  $U_7$  und Score von  $s[6, 6]$  übernehmen (Option a)
  - 2 das  $A_5$  ungepaart lassen und Score von Zelle  $s[6, 7]$  übernehmen (Option b)
- Da  $s[i, j]$  nur den maximalen Score einer optimalen Struktur der Subsequence  $s_{i...j}$  speichert, ist es in diesem Schritt unwesentlich, von welcher Option der besten Alignment-Score kam

# Füllen der Matrix (8)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i, j] ← max(a, b, c, d)
6:   end for
7: end for

```

$$S(i, j) = \max \begin{cases} \text{a) } \delta(i, j) + S(i+1, j-1) \\ \text{b) } S(i+1, j) \\ \text{c) } S(i, j-1) \\ \text{d) } \max_{i < k < j} S(i, k) + S(k+1, j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0	0	0	0	0			
G <sub>2</sub>	0	0	0	0	0	0	1		
G <sub>3</sub>		0	0	0	0	0	1	?	
A <sub>4</sub>			0	0	0	0	1	1	
A <sub>5</sub>				0	0	0	1	1	1
A <sub>6</sub>					0	0	1	1	1
U <sub>7</sub>						0	0	0	0
C <sub>8</sub>							0	0	0
C <sub>9</sub>								0	0

- Hier sieht man die bis zur Zelle (3, 8) befüllte Matrix
- Option a: Paarung A<sub>3</sub> und U<sub>8</sub> zusammen mit Score für beste Substruktur der Subsequenz s<sub>4...7</sub> (1+1).
- Option b: Base i (A<sub>3</sub>) ungepaart zusammen mit Score für beste Substruktur der Subsequenz s<sub>4...8</sub> (1).
- Option c: Base j (U<sub>8</sub>) ungepaart zusammen mit Score für beste Substruktur der Subsequenz s<sub>3...7</sub> (1).
- Option d: Summe der besten Scores der Subsequenzen s<sub>3...4</sub> und s<sub>5...8</sub> (0+1) bzw. s<sub>3...5</sub> und s<sub>6...8</sub> (0+1) oder s<sub>3...6</sub> und s<sub>7...8</sub> (0+0)
- Die Wahl fällt auf Option a und wir setzen s[3, 8] = 2.

# Füllen der Matrix (9)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i, j] ← max(a, b, c, d)
6:   end for
7: end for
    
```

$$S(i, j) = \max \begin{cases} \text{a) } \delta(i, j) + S(i+1, j-1) \\ \text{b) } S(i+1, j) \\ \text{c) } S(i, j-1) \\ \text{d) } \max_{j < k < j} S(i, k) + S(k+1, j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0	0	0	0	0	1	2	
G <sub>2</sub>	0	0	0	0	0	0	1	2	?
G <sub>3</sub>		0	0	0	0	0	1	2	2
A <sub>4</sub>			0	0	0	0	1	1	1
A <sub>5</sub>				0	0	0	1	1	1
A <sub>6</sub>					0	0	1	1	1
U <sub>7</sub>						0	0	0	0
C <sub>8</sub>							0	0	0
C <sub>9</sub>								0	0

- Wir fahren bis Zelle (2, 9) fort
- Der optimale Score wird durch eine Basenpaarung G<sub>2</sub> mit C<sub>9</sub> zusammen mit Score für beste Substruktur der Subsequenz s<sub>3...8</sub> (1+2) erreicht.

# Füllen der Matrix (10)

```

1: for s ← 2 to n do
2:   i ← 0
3:   for j ← s to n do
4:     i ← i + 1
5:     s[i,j] ← max(a, b, c, d)
6:   end for
7: end for

```

$$S(i,j) = \max \begin{cases} \text{a) } \delta(i,j) + S(i+1,j-1) \\ \text{b) } S(i+1,j) \\ \text{c) } S(i,j-1) \\ \text{d) } \max_{i < k < j} S(i,k) + S(k+1,j) \end{cases}$$

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0	0	0	0	0	1	2	?
G <sub>2</sub>	0	0	0	0	0	0	1	2	3
G <sub>3</sub>		0	0	0	0	0	1	2	2
A <sub>4</sub>			0	0	0	0	1	1	1
A <sub>5</sub>				0	0	0	1	1	1
A <sub>6</sub>					0	0	1	1	1
U <sub>7</sub>						0	0	0	0
C <sub>8</sub>							0	0	0
C <sub>9</sub>								0	0

- Die letzte Zell (1,9) gibt den Score einer optimalen Struktur für die ganze Sequenz  $s_{1...9}$  an
- Der optimale Score kann hier durch Option a oder b erreicht werden (Überprüfen Sie!)

# Traceback

- Der Traceback beginnt immer in der rechten oberen Ecke der Matrix und verfolgt den Pfad einer optimalen Struktur
- In diesem Fall ist der Traceback relativ einfach, weil keine verzweigte RNA-Struktur vorliegt
- Im einfachen Fall kann man beginnend mit Zelle  $(1, N)$  die Scores der drei umgebenden Zellen (links, schräg links unten und unten) vergleichen und so schlussfolgern, wie der Score in der aktuellen Zelle zustande gekommen sein muss und so eine optimale Struktur finden.
- Oft (wie hier) sind viele Pfade mit demselben Score möglich
- Siehe das Durbin-Buch für Einzelheiten zum allgemeinen Traceback-Algorithmus.

	G <sub>1</sub>	G <sub>2</sub>	G <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	U <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>
G <sub>1</sub>	0	0	0	0	0	0	1	2	③
G <sub>2</sub>	0	0	0	0	0	0	1	2	③
G <sub>3</sub>		0	0	0	0	0	1	②	2
A <sub>4</sub>			0	0	0	0	①	1	1
A <sub>5</sub>				0	0	①	1	1	1
A <sub>6</sub>					0	①	1	1	1
U <sub>7</sub>						0	0	0	0
C <sub>8</sub>							0	0	0
C <sub>9</sub>								0	0

GGGAAAUCC

. ( ( ( . . ) ) )

# Aufgabe

Bestimmen Sie nun eine optimale Sekundärstruktur für die RNA-Sequenz AACGCUUGA mit dem Nussinov-Algorithmus

# Outline

- 1 Eine RNA-Menagerie
- 2 mRNA, tRNA, rRNA, snRNA, snoRNA
- 3 RNA-Struktur
- 4 RNA: Sekundärstruktur
- 5 miRNAs**

# miRNAs

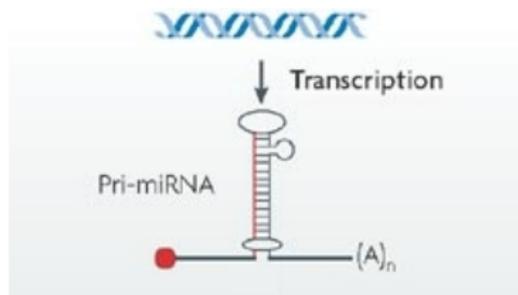
- Wir werden uns in der verbleibenden Zeit mit micro-RNAs (miRNAs) beschäftigen
- miRNAs sind 1993 in *C. elegans* entdeckt worden
- Die große Bedeutung von miRNAs in einer Reihe von biologischen Prozessen auch bei Säugern ist wohl seit Anfang des Jahrtausends nach und nach klar geworden, zahlreiche Aspekte des miRNA-Metabolismus sind noch nicht geklärt
- Wichtiges Thema für die Bioinformatik: Beitrag der miRNAs zur Genregulation verstehen

# miRNAs

- Sehr kurze RNA-Moleküle ( $\sim 22$  nt)
- Antisense-Regulatoren anderer Gene
- miRNAs entstehen aus Vorstufen mit  $\sim 70$  nt, welche eine umgekehrt komplementäre enthalten, die die Bildung einer Haarnadelstruktur ermöglicht
- Mindestens tausend miRNAs beim Menschen
- eine miRNA reguliert i.d.R. bis zu ein paar hundert proteinkodierende Gene
- Grundsätzlich eine *negative* Regulation

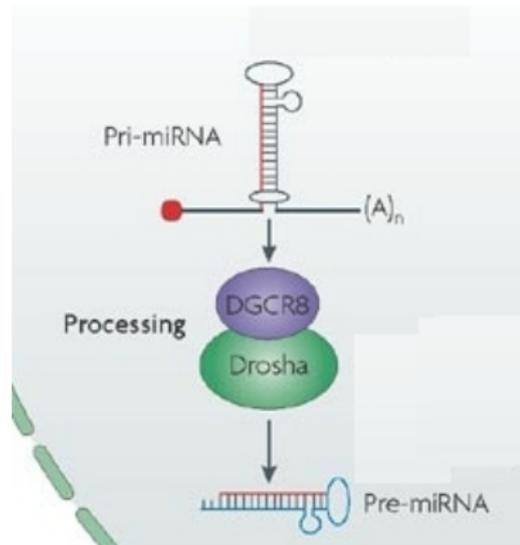
# Biogenese der miRNAs (1)

- miRNA-Vorstufen (**pri-miRNAs**, primary microRNAs) werden als unabhängige miRNA-Gene transkribiert oder stellen in anderen Fällen Abschnitte von Introns proteinkodierender Gene dar
- eine pri-miRNA kann Sequenzen mehrerer miRNAs enthalten
- pri-miRNAs falten als Haarnadelstrukturen mit imperfekter Basenpaarung



## Biogenese der miRNAs (2)

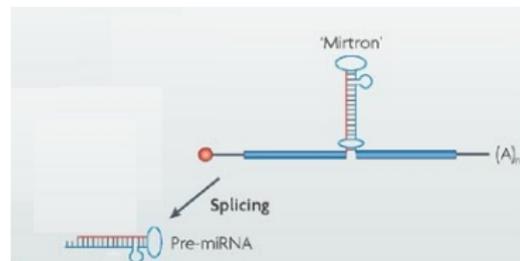
- pri-miRNAs werden dann durch die Endonuclease **Drosha** verarbeitet\*
- Das Ergebnis sind  $\sim 70$  nt Haarnadeln namens **prä-miRNAs**



\*) Drosha bindet an das Produkt des DiGeorge syndrome critical region Gen 8 (*DGCR8*)

## Biogenese der miRNAs (3)

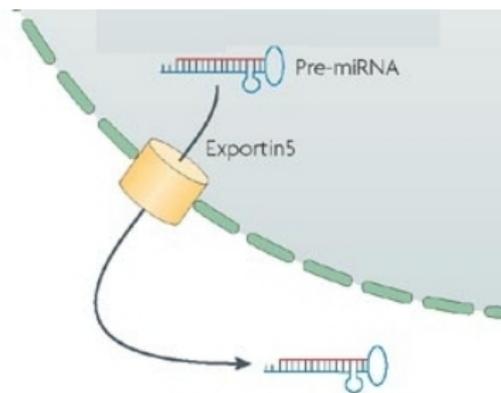
- In einigen Fällen entstammen die Prä-miRNA-Sequenzen herausgespleißten Introns<sup>†</sup>
- Mirtrons benötigen daher Drosha-DGCR8 nicht



<sup>†</sup>) Mirtrons kommen bei *Caenorhabditis elegans*, *D. melanogaster* und Säugetieren vor.

## Biogenese der miRNAs (4)

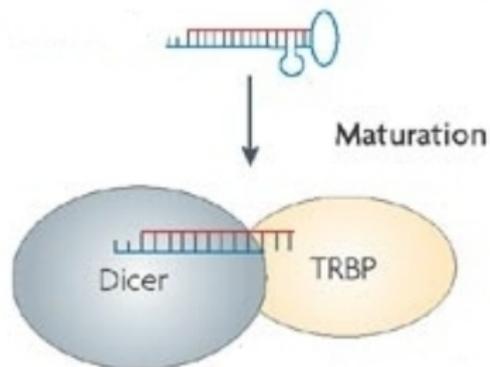
- prä-miRNAs werden durch Exportin-5 ins Zytoplasma transportiert



# Biogenese der miRNAs (5)

- Im Zytoplasma werden die prä-miRNAs durch **Dicer**\* gespalten
- Das Ergebnis ist ein  $\sim 20$  bp miRNA-Duplex

\*) Dicer bildet einen Komplex mit TAR RNA binding protein (TRBP).



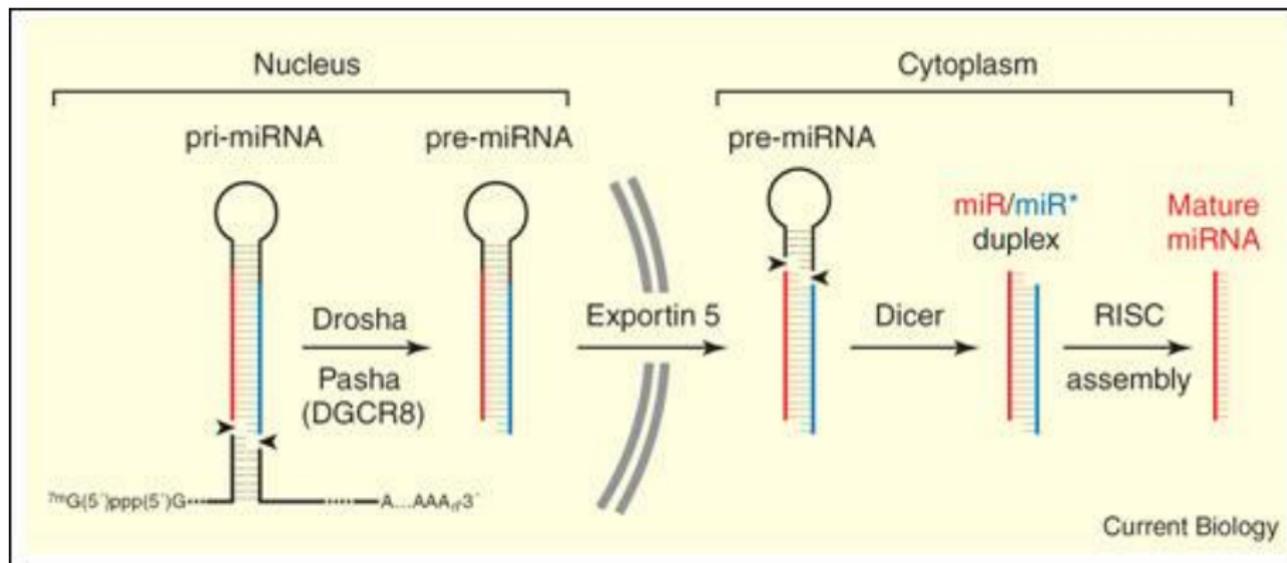
## Biogenese der miRNAs (6)

- Nach der Verarbeitung durch Dicer werden die miRNAs eingefügt in einen Ribonukleoproteinkomplex namens **miRNA-induzierte Silencing-Komplex (miRISCs)**
- Der Aufbau des miRISC ist an die prä-miRNA-Verarbeitung gekoppelt
- Ein Strang des Duplex bildet die reife miRNA, der andere wird abgebaut
- Die wichtigsten Proteinbestandteile des miRISC sind Proteine der **Argonaut-Familie**



# Biogenese der miRNAs (7)

- Zusammengefasst...



# Wie steuern miRNAs die Genexpression?

- miRNAs steuern die Expression von jeweils bis zu einigen hundert Genen posttranskriptionell
- Verminderung der mRNA-Stabilität
- Verminderung der mRNA-Translation

# Wie steuern miRNAs die Genexpression?

- Basenpaarung an die 3'-UTR der Ziel-mRNAs
- Perfekte Basenpaarung in der Saatregion (Nukleotide 2–8 der miRNA)
- Die Saatregion initiiert die miRNA-mRNA-Assoziation
- Fehlpaarung in der mittleren Region
- (Imperfekte) Basenpaarung in der 3'-Region der miRNA

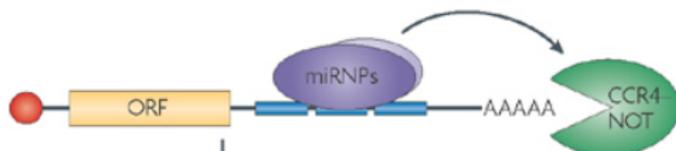
# Wie steuern miRNAs die Genexpression?

- miRNA-Bindungsstellen sind in der 3'-UTR der Ziel-mRNA gelegen
- Mehrfache Bindungsstellen sind in der Regel für eine wirksame Repression der Genexpression benötigt
- Synergistische Wirkung insbesondere von Bindungsstellen, die nah beieinander gelegen sind (10–40 nt)

# Wie steuern miRNAs die Genexpression?

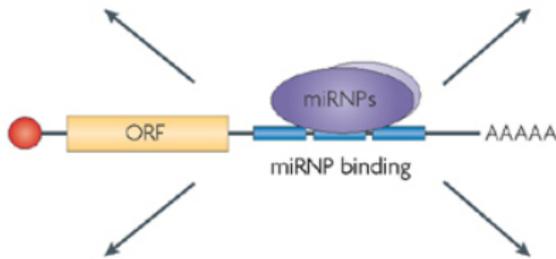
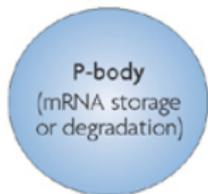
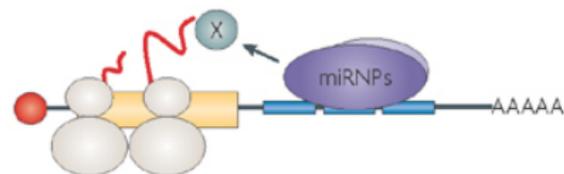
**Deadenylation**

(followed by decapping and degradation)



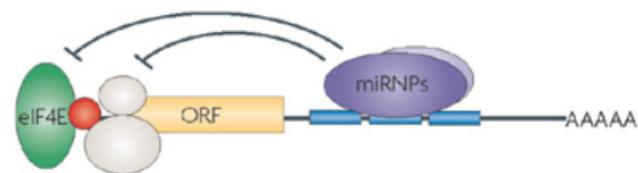
**Proteolysis**

(degradation of nascent peptide)



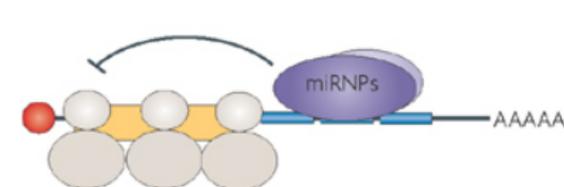
**Initiation block**

(repressed cap recognition or 60S joining)



**Elongation block**

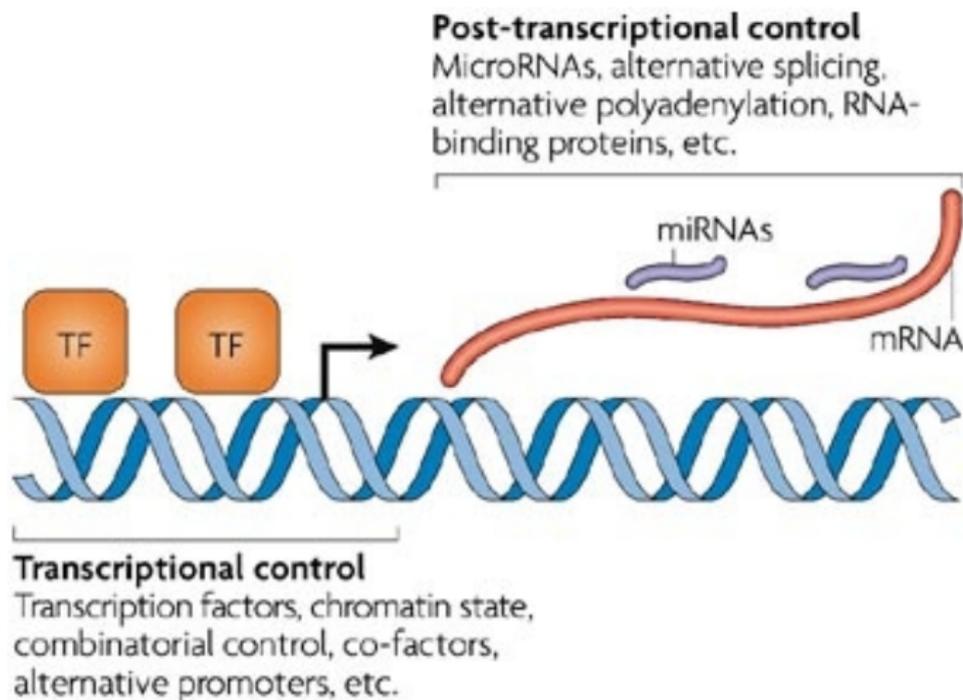
(slowed elongation or ribosome 'drop-off')



Filipowicz W et al. (2008) *Nat Rev Genet.* 9:102–14.

Nature Reviews | Genetics

# miRNAs vs. Transkriptionsfaktoren



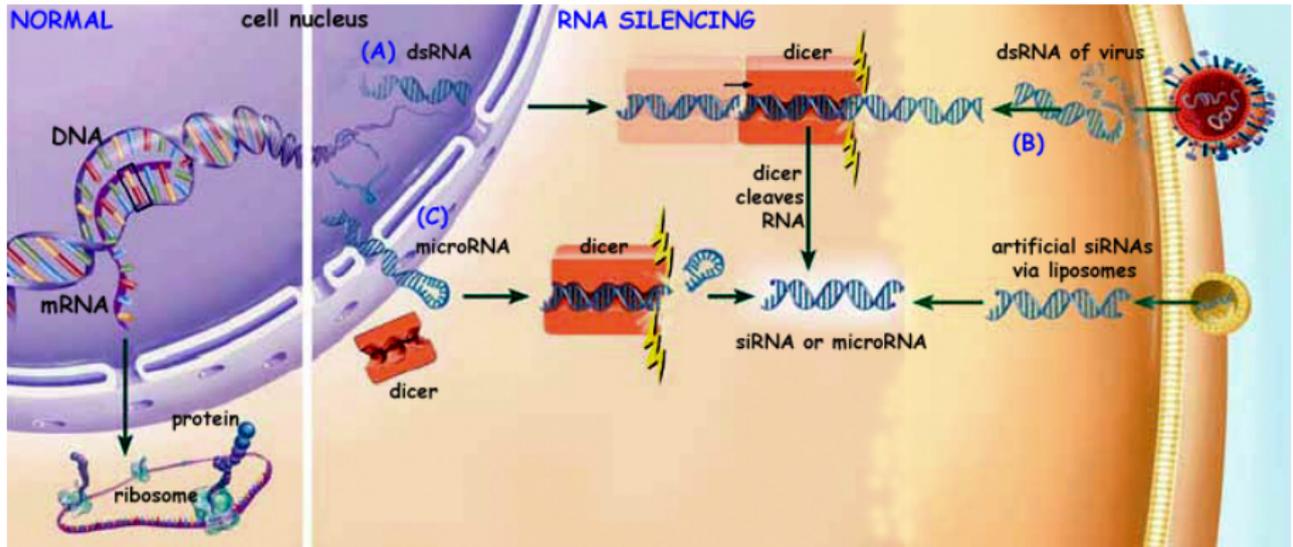
Nature Reviews | Genetics

# miRNA-Funktionen

miRNA-Gene stellen ca. 1–2% aller Gene bei Eukaryonten dar. Die Funktionen der meisten miRNAs sind noch unbekannt.

- Regulatoren des zeitlichen Ablaufs der Entwicklung der Larvenstadien (*lin-4*, *lin-7*, *C. elegans*)
- Links-rechts-Asymmetrie der Chemorezeptorexpression (*Isy-6*, *C. elegans*)
- Apoptose, Fettstoffwechsel (*miR-14*, *D. melanogaster*)
- Hämatopoietische Differenzierung (*miR-181a*, Maus)
- Spaltung von Hox-B8-Transkripten (*miR-196*, Maus)
- Rolle bei Krebs, neurologischen Krankheiten, . . . , ?

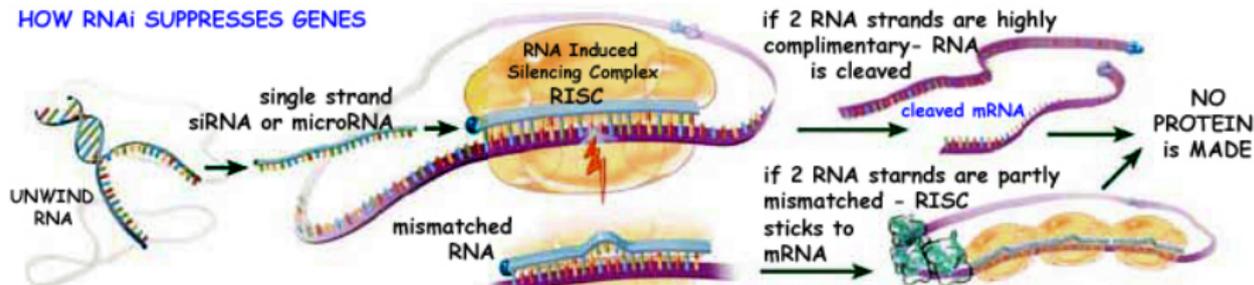
# miRNA vs. siRNA



- (A) dsRNA von Transposons, (B) Viren (C) miRNAs werden von Dicer prozessiert

# miRNA vs. siRNA

## HOW RNAi SUPPRESSES GENES

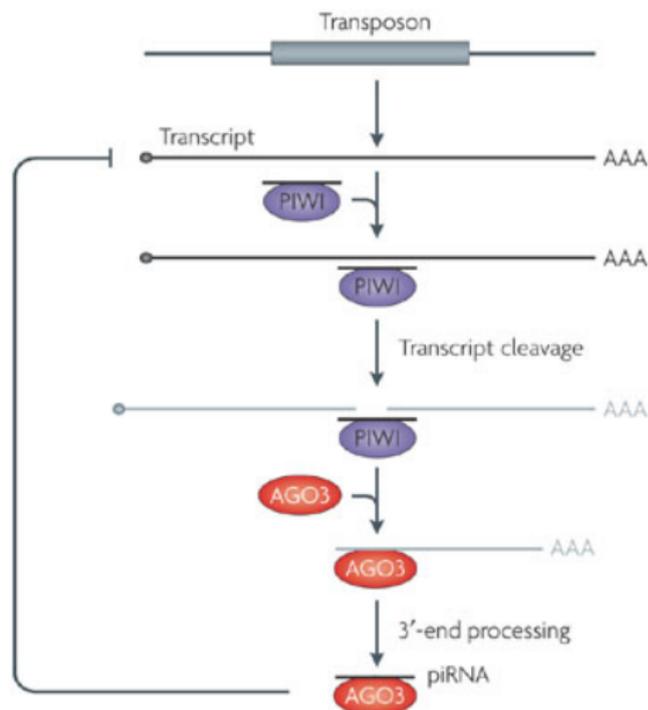


- siRNA/miRNA bilden den RNA-induzierten Silencing-Komplex (RISC)
- Ist die Bindung an die Ziel-mRNA (nahezu) perfekt, wird die Ziel-mRNA gespalten (eher der Fall bei siRNA)
- Ist die Bindung nicht perfekt, wird die mRNA destabilisiert bzw. die Translation gehemmt (eher der Fall bei miRNA)

# piRNAs

- Piwi-Unterfamilie der Argonautenproteine
- Ziel: Transposons, Retroelemente
- Spaltung der Ziel-mRNA erzeugt eine neue piRNA
- Rolle vor allem im Keimgewebe, um neue Insertionen von Transposons zu verhindern

● Chapman, Carrington, Nature Genetics Reviews 2007



# zum Schluss

- Email: peter.robinson@charite.de

## weiterführende Literatur

- Filipowicz W, Bhattacharyya SN, Sonenberg N. (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet.* **9**:102-14.
- Eddy SR (2004) How do RNA folding algorithms work? *Nature Biotech.* **22**:1457-1458
- Durbin R, Eddy SR, Krogh A, Mitchison G (1998) Biological Sequence Analysis Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press (Kapitel 10)