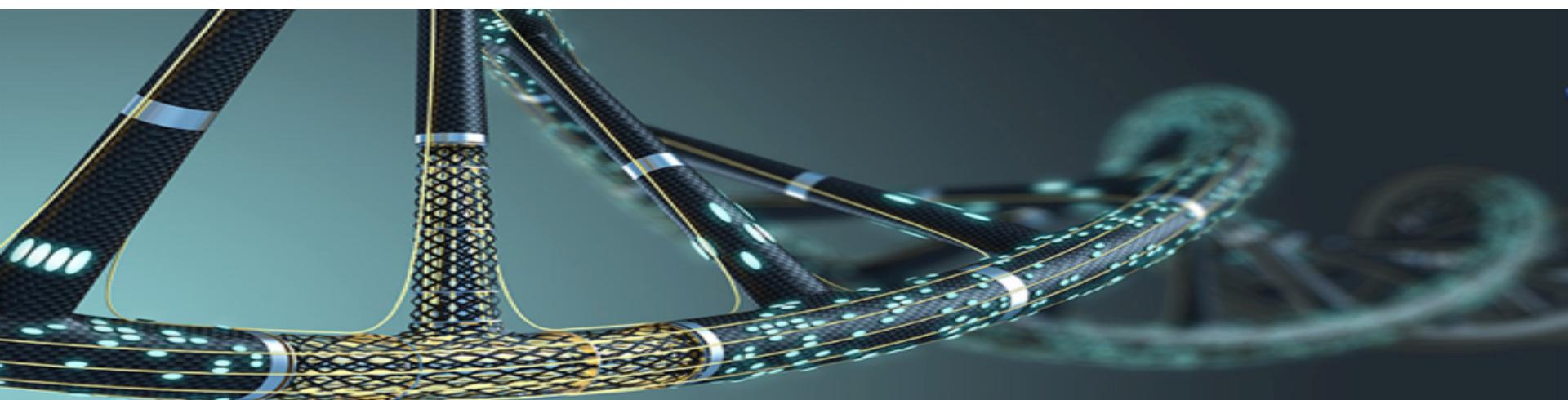




Майнор по биоинформатике

Лекция 14

Мария Попцова

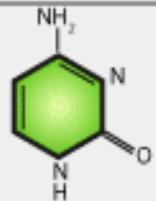


Вторичные структуры РНК.
Методы предсказания вторичной
структур РНК.

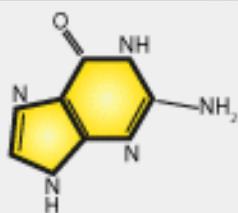
Структурное выравнивание.
База данных RFam

DIFFERENCE BETWEEN DNA AND RNA

CYTOSINE C



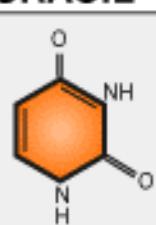
GUANINE G



ADENINE A



URACIL U



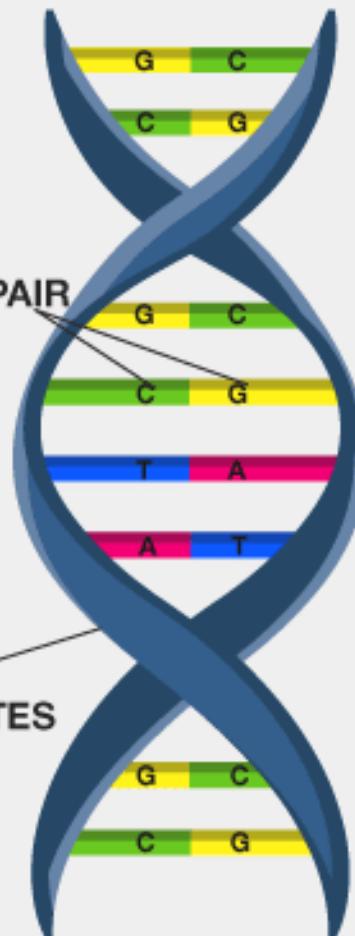
NUCLEOBASES



RNA
RIBONUCLEIC ACID

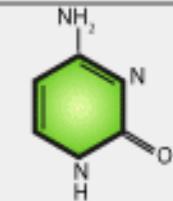
BASE PAIR

VS

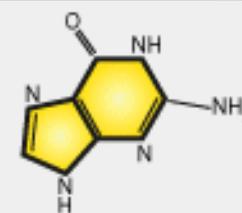


DNA
DEOXYRIBONUCLEIC ACID

CYTOSINE C



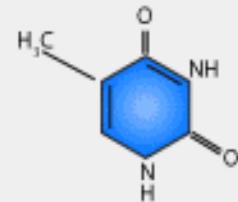
GUANINE G



ADENINE A



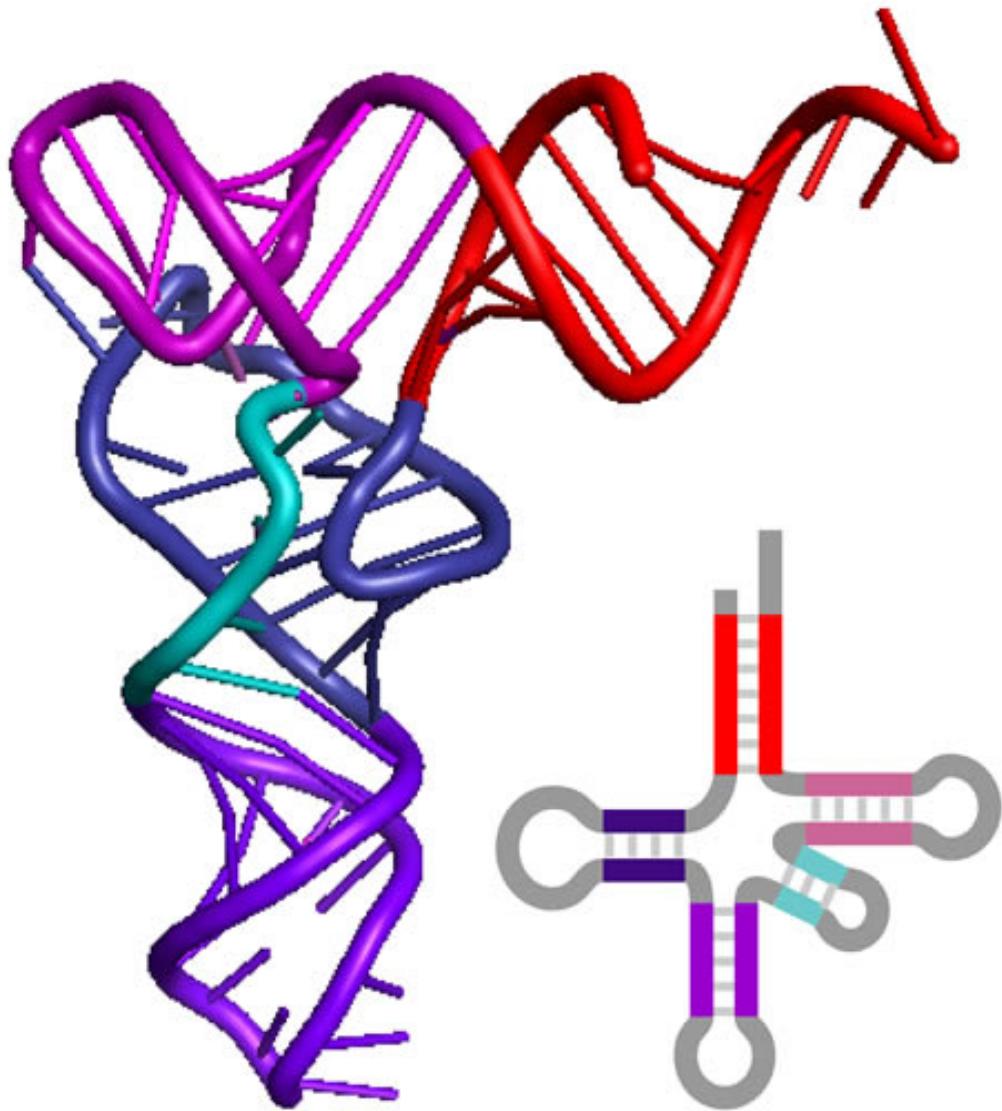
THYMINE T



RNA objects in the cell

- mRNA
 - All protein-coding genes genes
- RNA genes
 - tRNA
 - rRNA
- Regulatory RNA
 - small interfering RNA (siRNA)
 - micro RNA (miRNA)
 - small nucleolar RNA (snoRNA)

transfer RNA (tRNA)

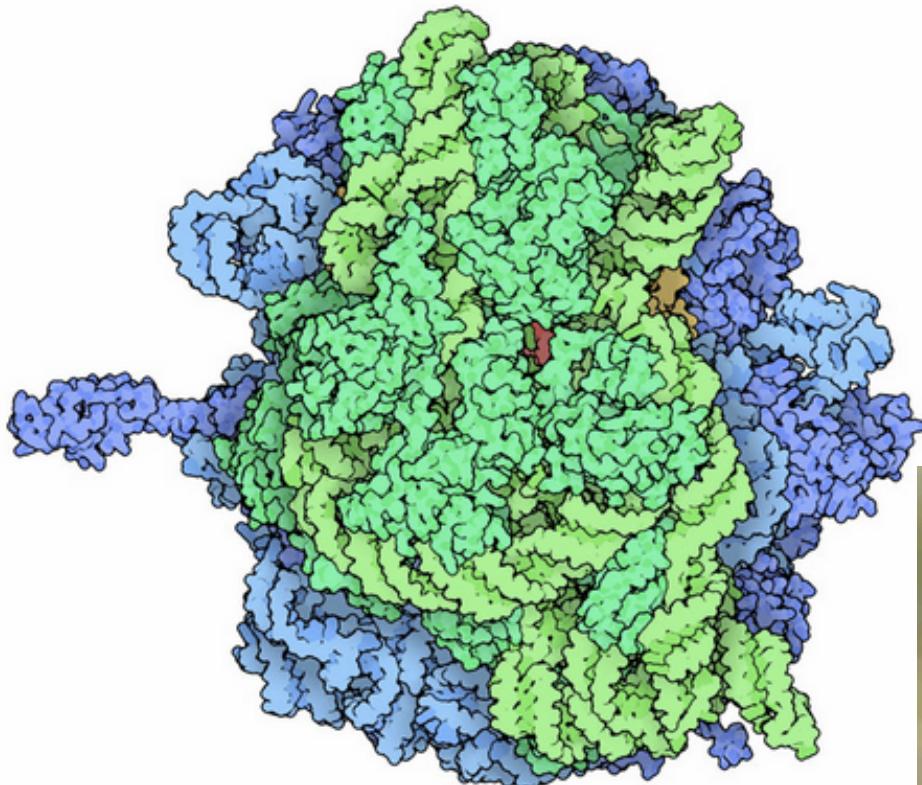


61 different tRNA

	U	C	A	G	
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA TER UAG TER	UGU Cys UGC Cys UGA TER UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

Hydrophobic - Imino Hydrophobic - Aliphatic Polar - Neutral
Hydrophobic - Aromatic Polar - Acid Polar - Basic

Ribosome (rRNA+proteins)



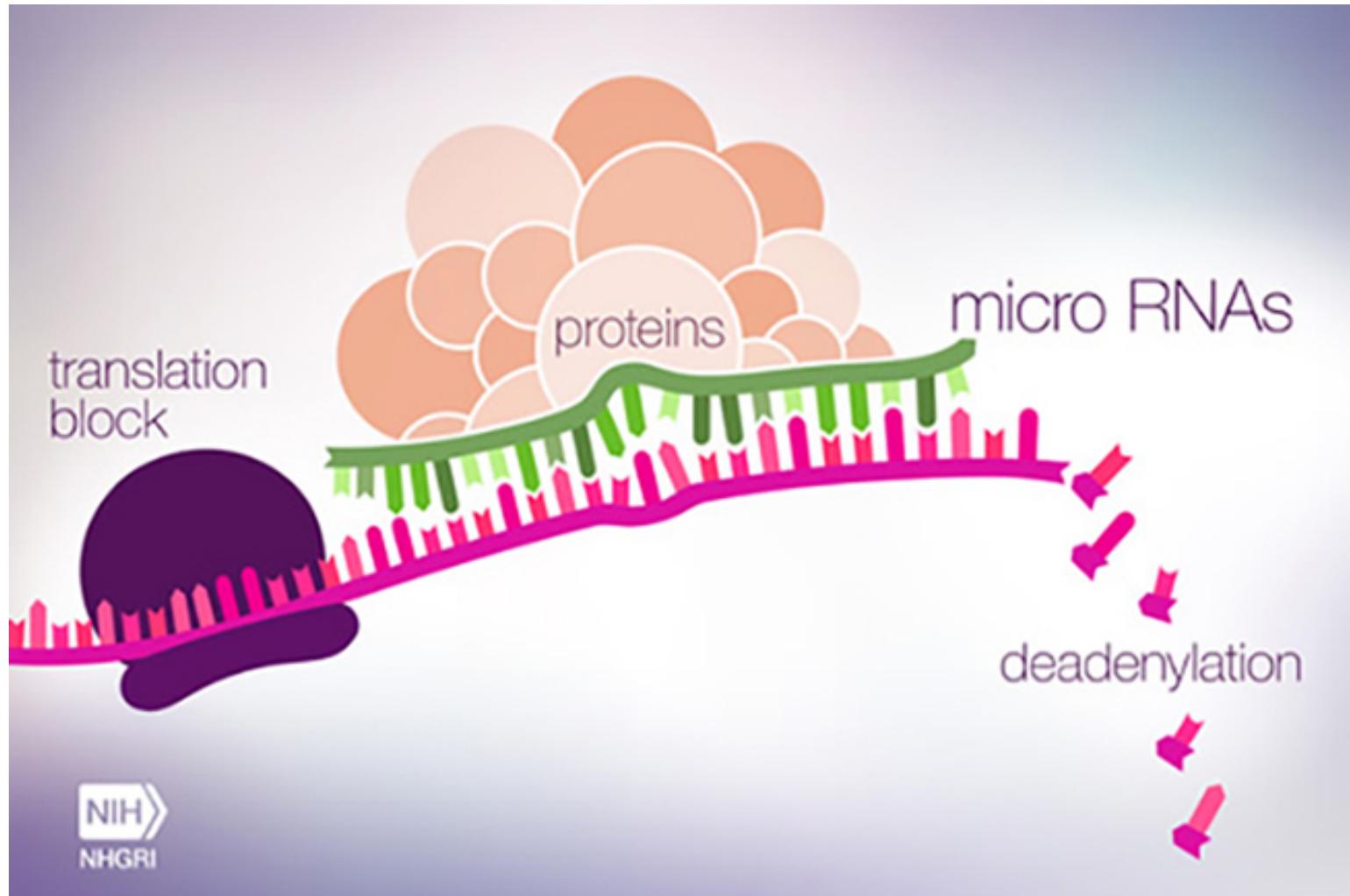
mRNA



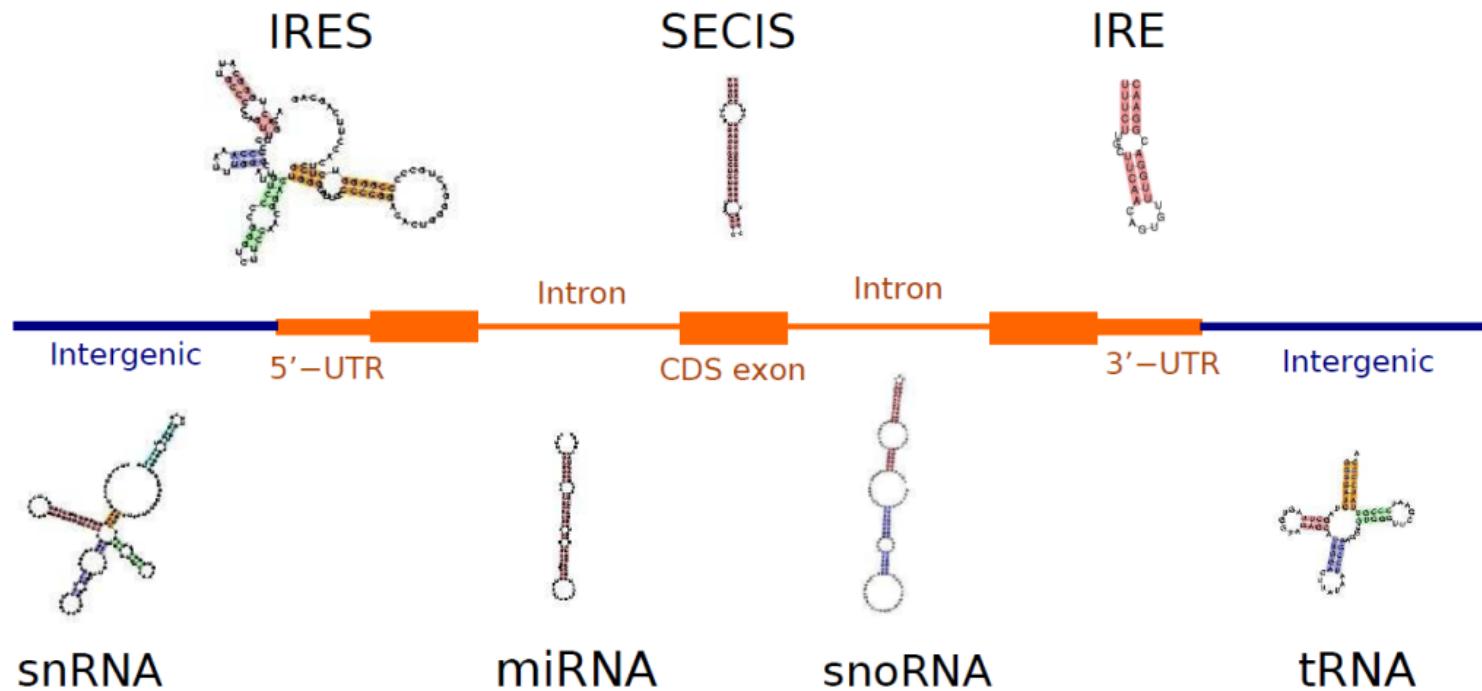
The small sub unit of the ribosome positions the mRNA so that it

Bacterial ribosome, with the small subunit in green and the large subunit in blue.

Regulatory micro RNA



Structured RNAs: examples



The Nobel Prize in Physiology or Medicine 2006

"for their discovery of RNA interference - gene silencing by double-stranded RNA."



Photo: L. Cicero
Andrew Z. Fire

Prize share: 1/2



Photo: J. Mottern
Craig C. Mello

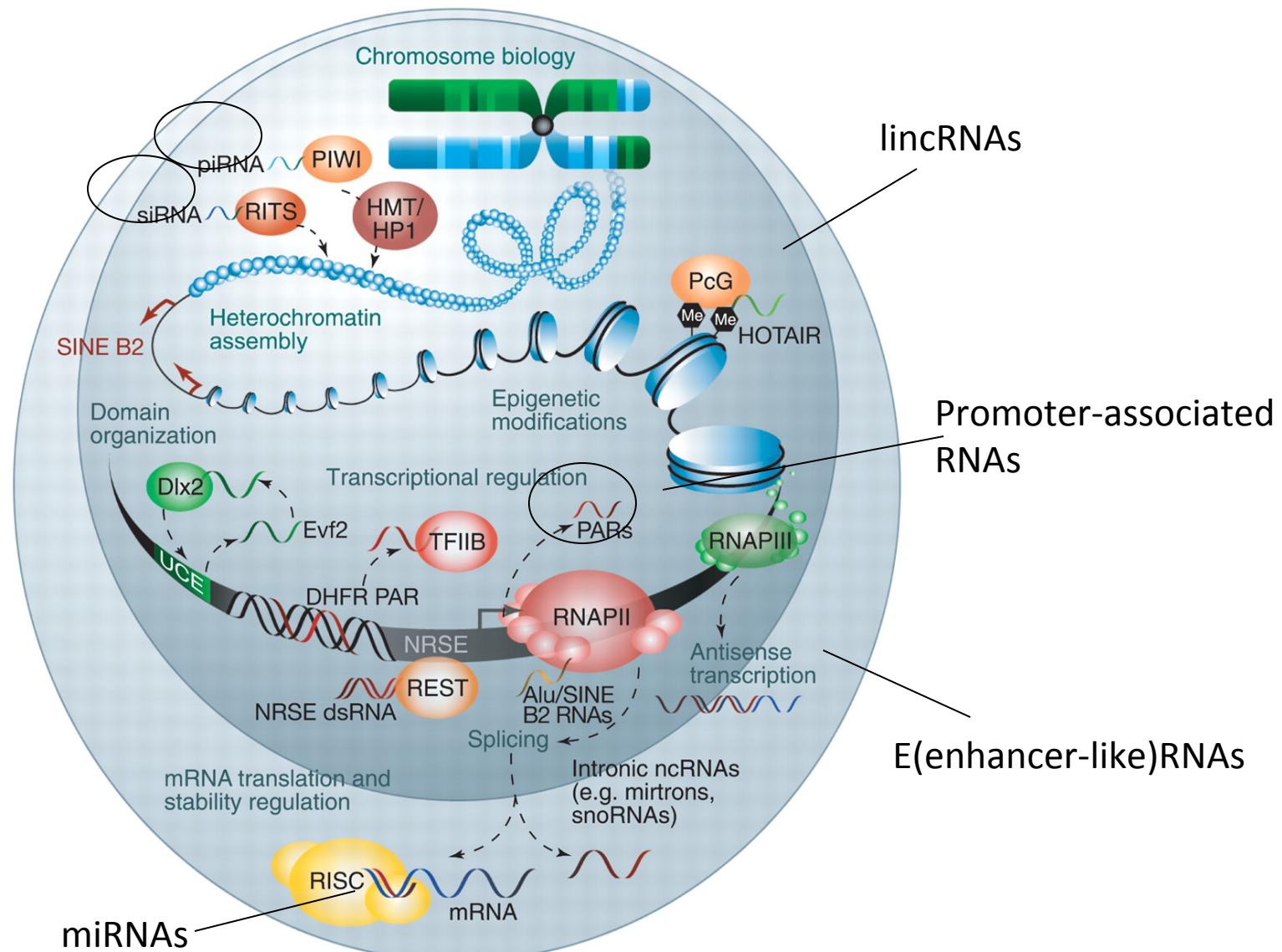
Prize share: 1/2

RNA interference (RNAi)



The Eukaryotic Genome as an RNA machine

The ‘RNA world’



RNA folding prediction algorithms

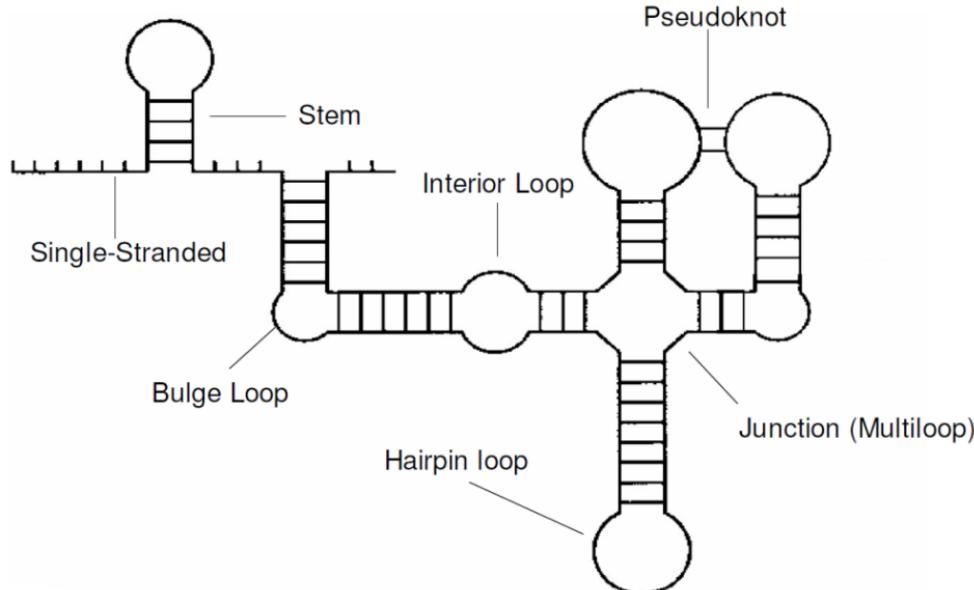
- Approximation: prediction of RNA secondary structure

```
RNAfold < trna.fa
```

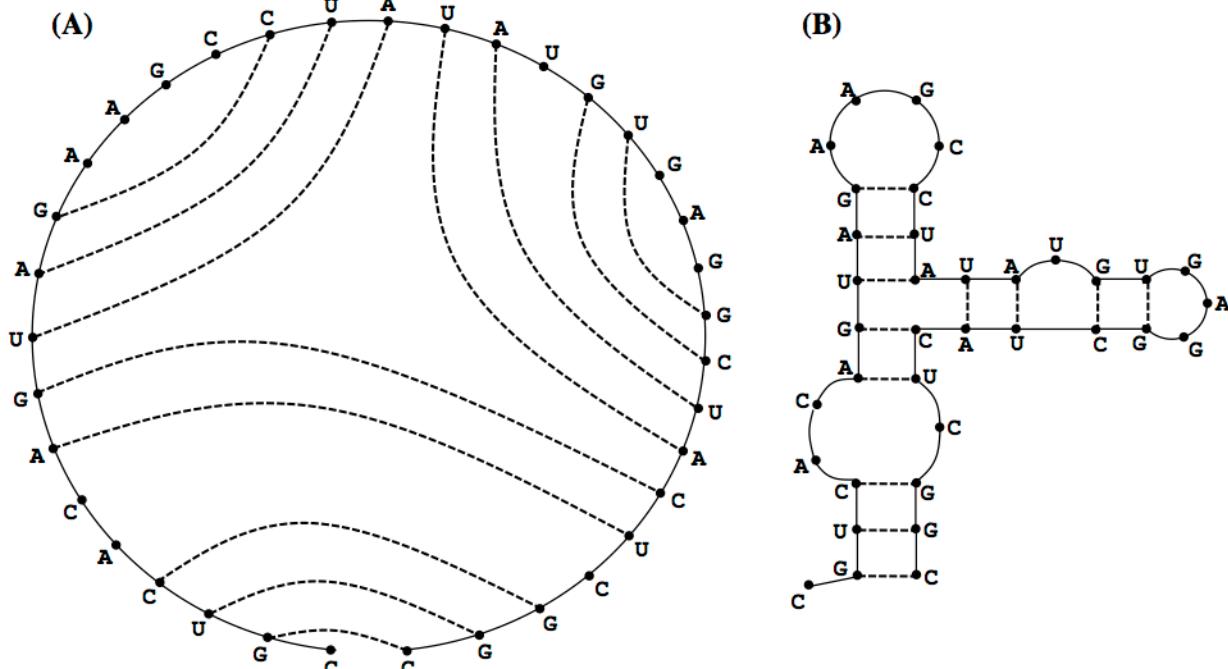
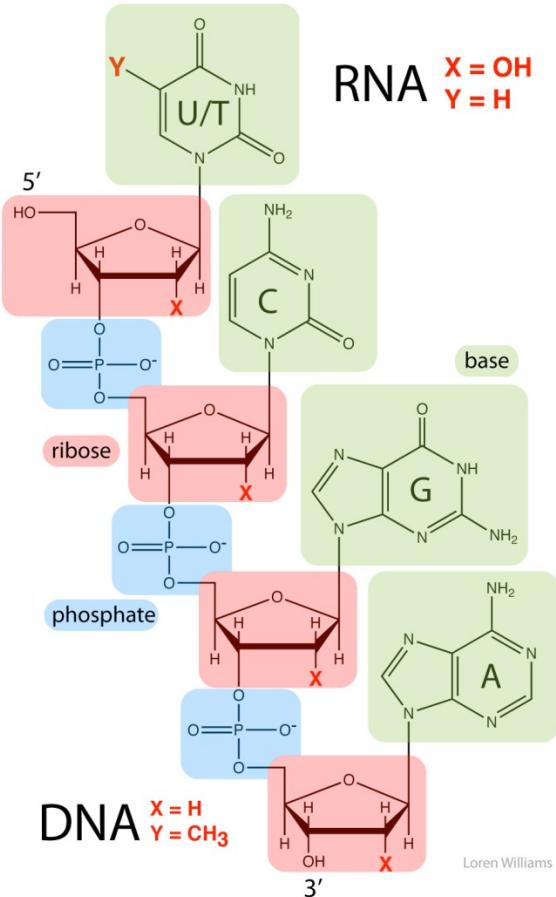
```
>AF041468
```

```
GGGGGUAUAGCUCAGUUGGUAGAGCGCUGCCUUUGCACGGCAGAUGUCAGGGGUUCGAGUCCCCUUACCUCA  
((((((..(((.....))))(((((.....))))))).....((((.....)))))))....)).  
-31.10 kcal/mol
```

RNA secondary structure elements



RNA backbone



Secondary structure: set of base pairs which can be mapped into a plane

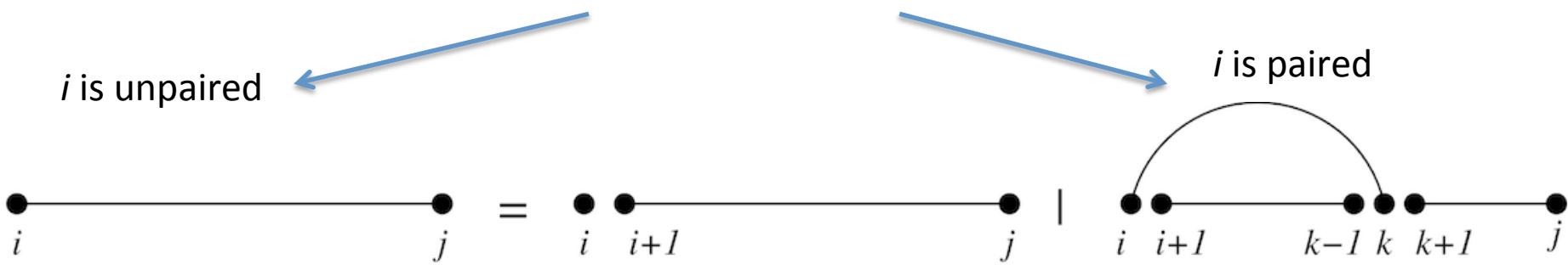
Concepts of Folding RNA Sequences

- RNA is thermodynamically folded in the most stable structure
- More pairs – more stability
- Task is to find the maximum number of base pairs for an RNA sequence.

Nussinov Algorithm

- The idea is to keep track of the number of base pairs of any sub-sequence starting at some position, say i , and ending at position j .

Decomposition of RNA secondary structures
for the Nussinov algorithm



E_{ij} as the maximum number of base pairs (or optimal energy) for a secondary structure on $i..j$

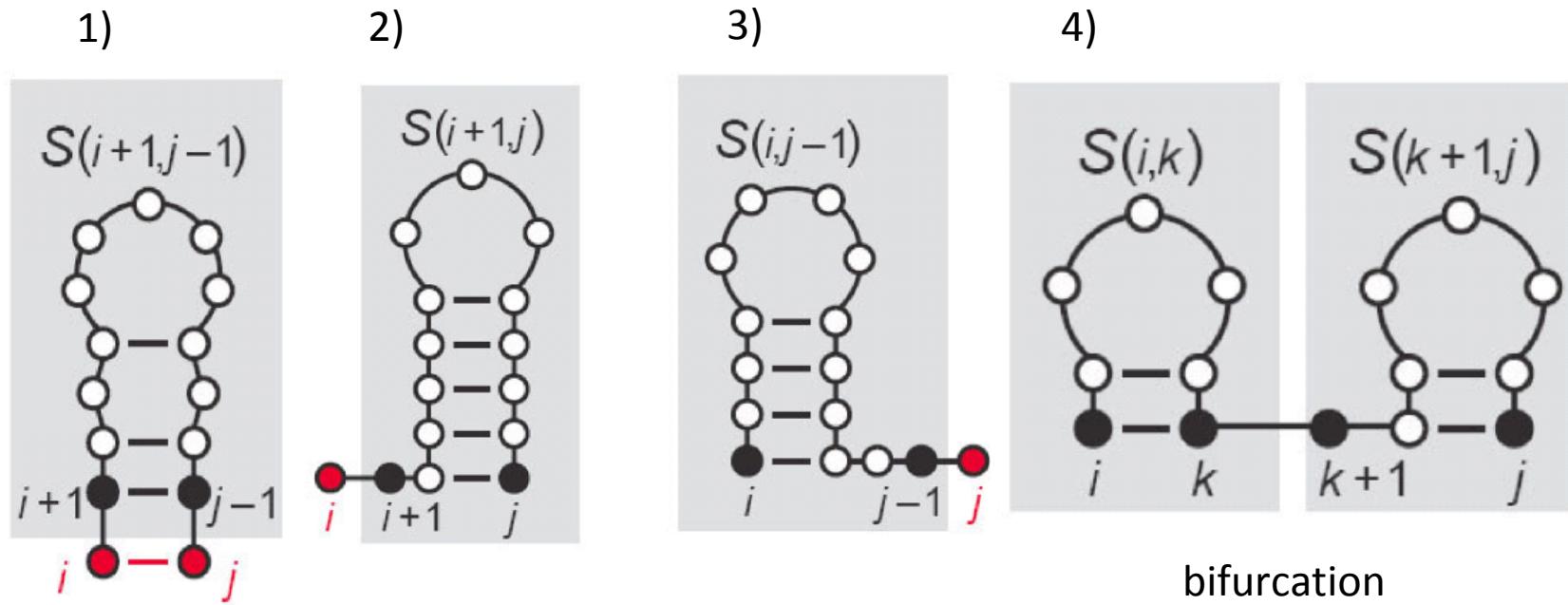
$$E_{ij} = \max \left\{ \begin{array}{l} E_{i+1,j} \\ \max_{k, (i,k) \text{ pairs}} \{ E_{i+1,k-1} + E_{k+1,j} + \beta_{ik} \} \end{array} \right\},$$

1 or 0

Nussinov Algorithm

$$S(i, j) = \max \begin{cases} S(i + 1, j - 1) + w(i, j) & (1) \\ S(i + 1, j) & (2) \\ S(i, j - 1) & (3) \\ \max_{i < k < j} S(i, k) + S(k + 1, j) & (4) \end{cases}$$

$w(i, j) = \begin{cases} 1 & i, j \text{ are complementary} \\ 0 & \text{otherwise} \end{cases}$



Recursion

$\rightarrow j$

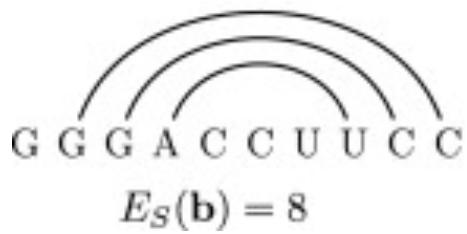
Fill up the table (DP matrix) -- diagonal by diagonal

	G	G	G	A	A	A	U	C	C
G	0	0	0	0					
G	0	0	0	0	0				
G		0	0	0	0	0			
A			0	0	0	0	?		
A				0	0	0	1		
A					0	0	1	1	
U						0	0	0	0
C							0	0	0
C								0	0

$$S(i, j) = \max \begin{cases} S(i + 1, j - 1) + w(i, j) & (1) \\ S(i + 1, j) & (2) \\ S(i, j - 1) & (3) \\ \max_{i < k < j} S(i, k) + S(k + 1, j) & (4) \end{cases}$$

$w(i, j) = \begin{cases} 1 & i, j \text{ are complementary} \\ 0 & \text{otherwise} \end{cases}$

	G	G	G	A	C	C	U	U	C	C
G	0	0	0	0	3	3	4	4	6	8
G		0	0	0	0	3	3	3	5	8
G			0	0	0	0	1	2	5	5
A				0	0	0	0	2	2	2
C					0	0	0	0	0	0
C						0	0	0	0	0
U							0	0	0	0
U								0	0	0
C									0	0
C										0



	G	G	G	A	C	C	U	U	C	C
G	0	0	0	0	3	3	4	4	6	8
G		0	0	0	0	3	3	3	5	8
G			0	0	0	0	1	2	5	5
A				0	0	0	0	2	2	2
C					0	0	0	0	0	0
C						0	0	0	0	0
U							0	0	0	0
U								0	0	0
C									0	0
C										0



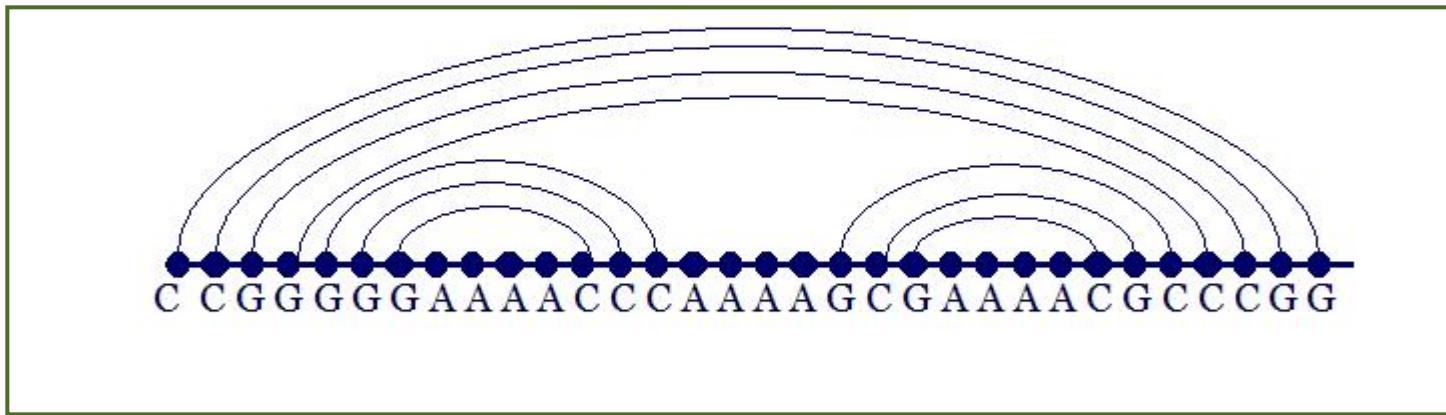
Free Energy Minimization

Idea:

- Overcome the main drawback of Nussinov's algorithm: non-realism of base pair maximization!
- Define an energy model for RNA that can be parameterized by experimentally measured energies
- Devise an algorithm that minimizes the free energy of RNA according to this model
- Algorithm (by Zuker) will be similar to Nussinov's algorithm

Free energy model

Free energy of a structure is the sum of all interactions energies

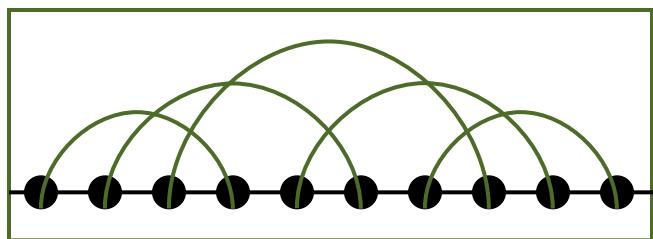


$$\text{Free Energy}(E) = E(\text{CG}) + E(\text{CG}) + \dots$$

Each interaction energy can be calculated thermodynamically

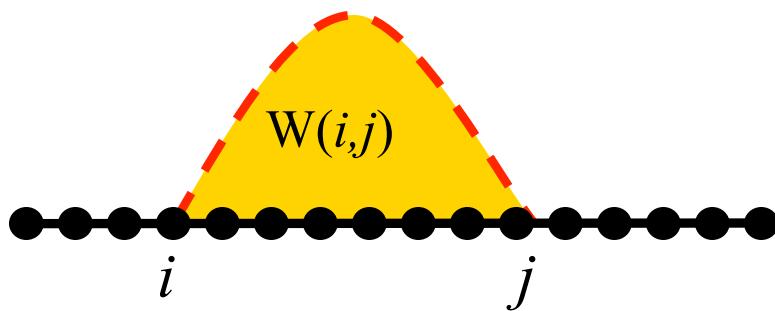
Why is MFE secondary structure prediction hard?

- MFE structure can be found by calculating free energy of all possible structures
- BUT the number of potential structures grows exponentially with the number, n , of bases



RNA folding with Dynamic programming (Zucker and Steigler)

- $W(i,j)$: MFE structure of substrand from i to j



1) Energy minimization method

What are the assumptions?

Native tertiary structure or "fold" of an RNA molecule is (one of) its "lowest" free energy configuration(s)

Gibbs free energy = ΔG in kcal/mol at 37°C

= equilibrium stability of structure

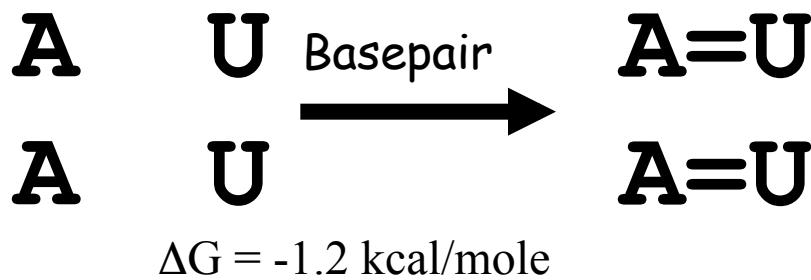
lower values (negative) are more favorable

Is this assumption valid?

in vivo? - this may not hold, but we don't really know

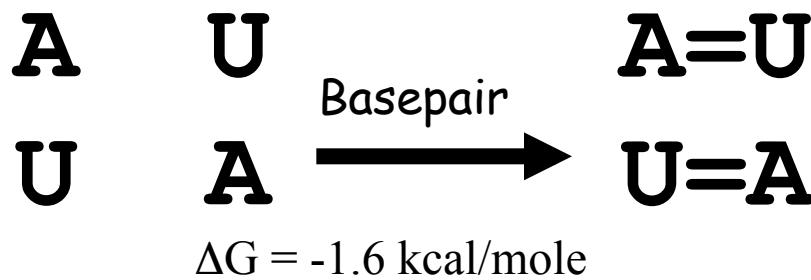
Free energy minimization

What are the rules?



What gives here?

Why 1.2 vs 1.6?



Energy minimization calculations: *Base-stacking is critical*

AA UU	-1.2	CG GC	-3.0
AU or UA UA AU	-1.6	GC CG	-4.3
AG, AC, CA, GA UC, UG, GU, CU	-2.1	GU UG	-0.3
CC GG	-4.8	XG, GX YU, UY	0

- Tinocco et al.

Nearest-neighbor parameters

Most methods for free energy minimization
use nearest-neighbor parameters (derived from
experiment) for predicting stability of an RNA secondary structure
(in terms of ΔG at 37°C)

& most available software packages use
the **same set of parameters:**

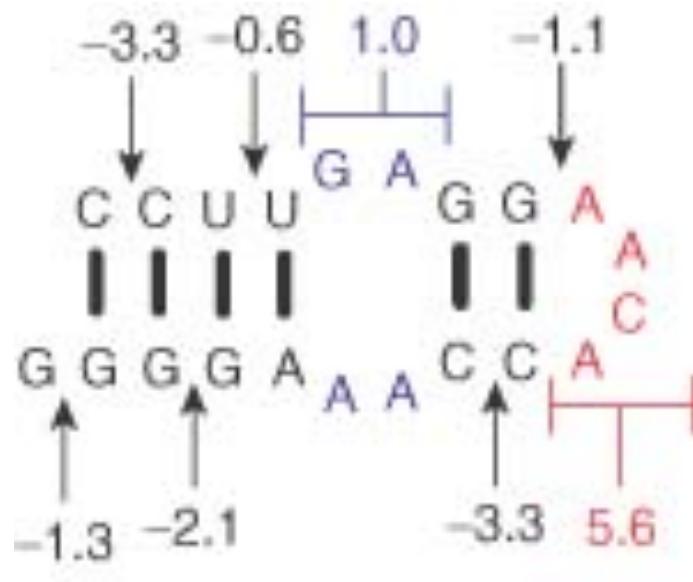
Mathews, Sabina, Zuker & Turner, 1999

Energy minimization - calculations:

Total free energy of a specific conformation for a specific RNA molecule
= sum of incremental energy terms for:

- helical stacking
(sequence dependent)
- loop initiation
- unpaired stacking

(favorable "increments" are < 0)



$$\begin{aligned} \text{DG}_{37}^{\circ} = & -1.3 - 3.3 - 2.1 - 0.6 \\ & + 1.0 - 3.3 - 1.1 + 5.6 \\ = & -5.1 \text{ kcal/mol} \end{aligned}$$

Fig 6.3
Baxevanis &
Ouellette 2005

ZUKER ALGORITHM

Initialisation: (for $j - i \leq m$)

$$W_{ij} = 0$$

Recursion: (for $i < j - m$)

$$W_{ij} = \min \begin{cases} W_{ij-1} & \text{--- } j \text{ unpaired} \\ \min_{i \leq k < j-m} W_{ik-1} + W_{k+1j-1} + E(???) & \text{--- } j \text{ paired} \end{cases}$$

But how many possible conformations for a single RNA molecule?

Huge number:

Zuker estimates $(1.8)^N$ possible secondary structures for a sequence of N nucleotides

for 100 nts (small RNA...) =

3×10^{25} structures!

Solution? Not exhaustive enumeration...

- Dynamic programming

$O(N^3)$ in time

$O(N^2)$ in space/storage

iff pseudoknots excluded, otherwise:

$O(N^6)$, time

$O(N^4)$, space

Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information

Michael Zuker and Patrick Stiegler⁺

Division of Biological Sciences, National Research Council of Canada, Ottawa K1A 0R6, Canada

Received 5 November 1980

ABSTRACT

This paper presents a new computer method for folding an RNA molecule that finds a conformation of minimum free energy using published values of stacking and destabilizing energies. It is based on a dynamic programming algorithm from applied mathematics, and is much more efficient, faster, and can fold larger molecules than procedures which have appeared up to now in the biological literature. Its power is demonstrated in the folding of a 459 nucleotide immunoglobulin $\gamma 1$ heavy chain messenger RNA fragment. We go beyond the basic method to show how to incorporate additional information into the algorithm. This includes data on chemical reactivity and enzyme susceptibility. We illustrate this with the folding of two large fragments from the 16S ribosomal RNA of *Escherichia coli*.

Реализации алгоритма Цукера

- RNAfold
- Mfold
- RNAStructure

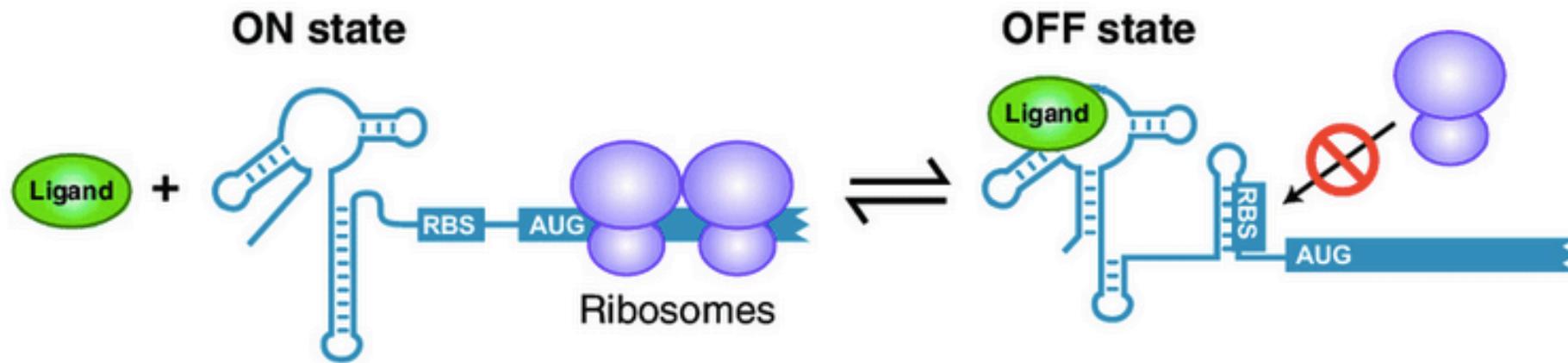
Riboswitches

- **Riboswitches** are cis-regulatory conserved non-coding structural RNA sensors that are present in the 5' untranslated regions (UTRs) of the bacterial mRNA, bind to specific ligands and control the expression of downstream genes.

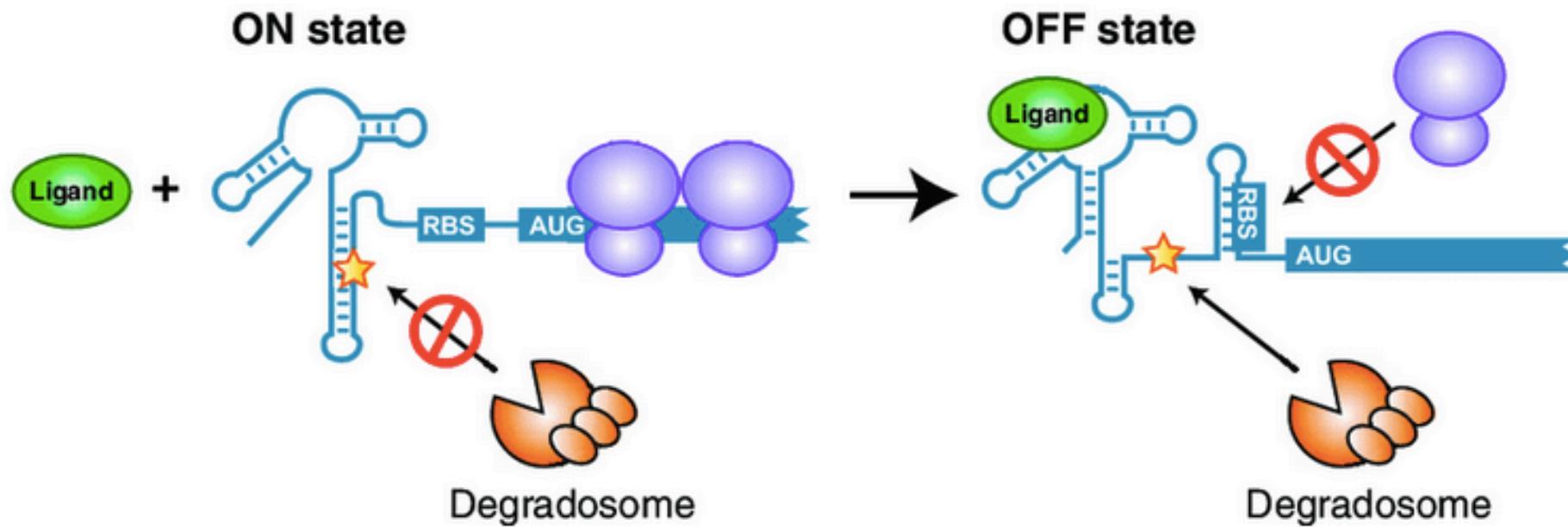
Examples of riboswitches

- Cobalamin riboswitch (also B12-element)
 - binds either adenosylcobalamin (the coenzyme form of vitamin B12) or aquocobalamin to regulate cobalamin biosynthesis
- Lysine riboswitch (also L-box)
 - binds lysine to regulate lysine biosynthesis
- Purine riboswitches
 - binds purines to regulate purine metabolism and transport.

A Non-Nucleolytic Repression Mechanism (*thiM* and *btuB* riboswitches)



B Nucleolytic Repression Mechanism (*lysC* riboswitch)



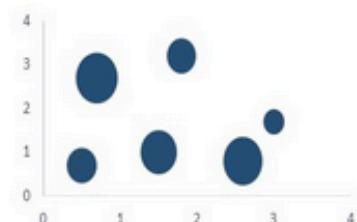


RiboD - A Database for Prokaryotic Riboswitches

[About](#)[Search](#)[Advanced Search](#)[Genomes](#)[Predict](#)[About us](#)

Riboswitches are cis-regulatory conserved non-coding structural RNA sensors that are present in the 5' untranslated regions (UTRs) of the bacterial mRNA, bind to specific ligands and control the expression of downstream genes. **RiboD** is a database of prokaryotic riboswitches that provides a comprehensive list of computationally predicted riboswitches, riboswitch regulated genes and operons from the sequenced prokaryotic genomes available in the [RefSeq](#). Currently, the RiboD database contains riboswitches for 1777 prokaryotic genomes and covers 31 metabolite and ion sensing riboswitch classes. The database supports a number of search capabilities to facilitate easy access to and utilization of the information associated with riboswitches as per user needs. Users can search for riboswitches based on riboswitch class, riboswitch-regulated genes, annotated biological processes/pathways of riboswitch-regulated genes and phylum. All of the listed computationally identified tandem riboswitches can be found by selecting the “Tandem Riboswitches” option under the search tab. In the advanced search option, the user can specify multiple fields to refine the search query and extract the specific information. The user can download the required information generated from each search query in excel file format. RiboD is the first database of prokaryotic riboswitches which provides a

Data Statistics

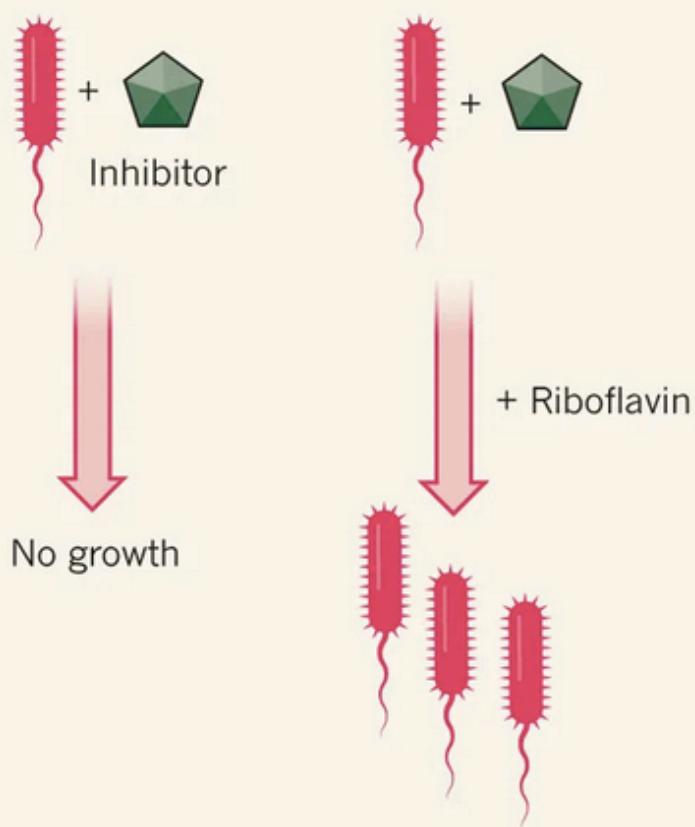


Frequently Asked Questions

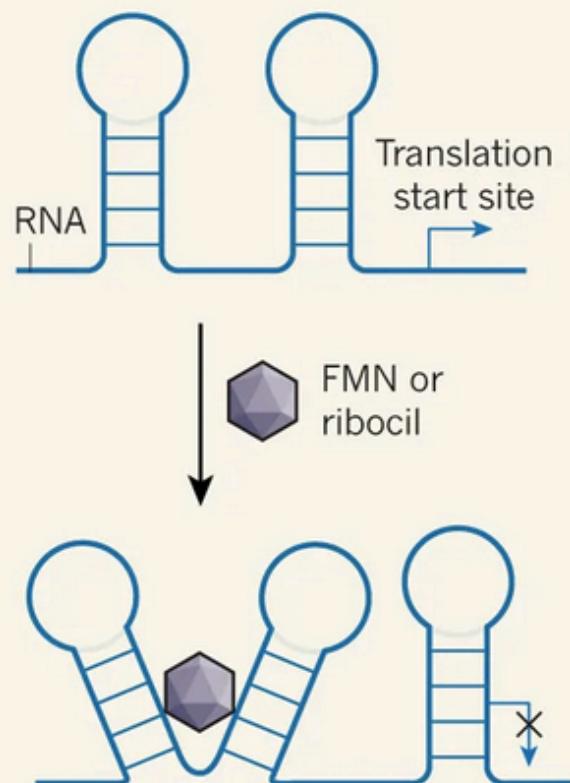
FAQ

Riboswitches and antibiotics

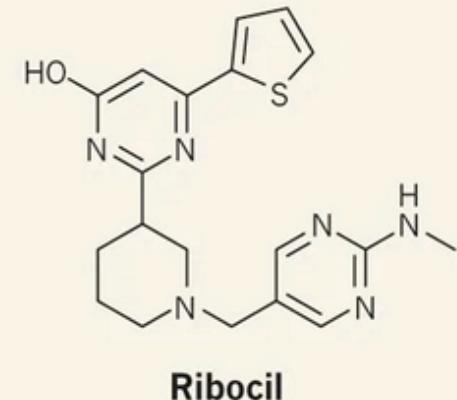
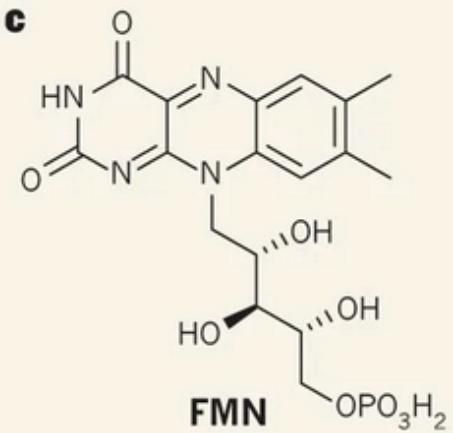
a Screen for pathway inhibitors



b Riboswitch target



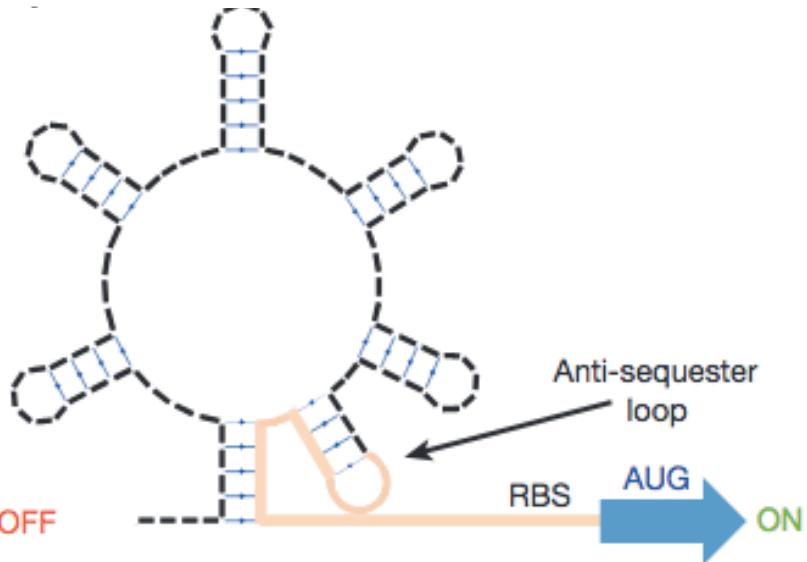
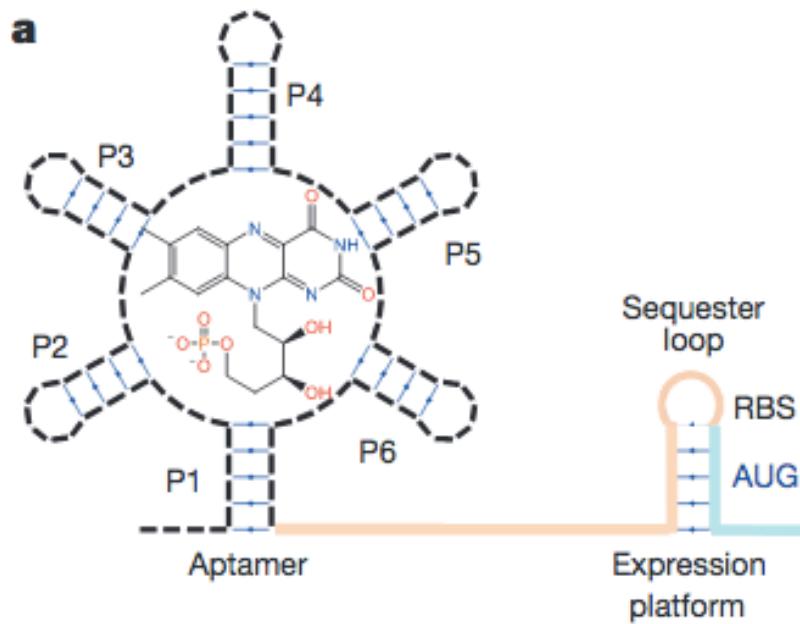
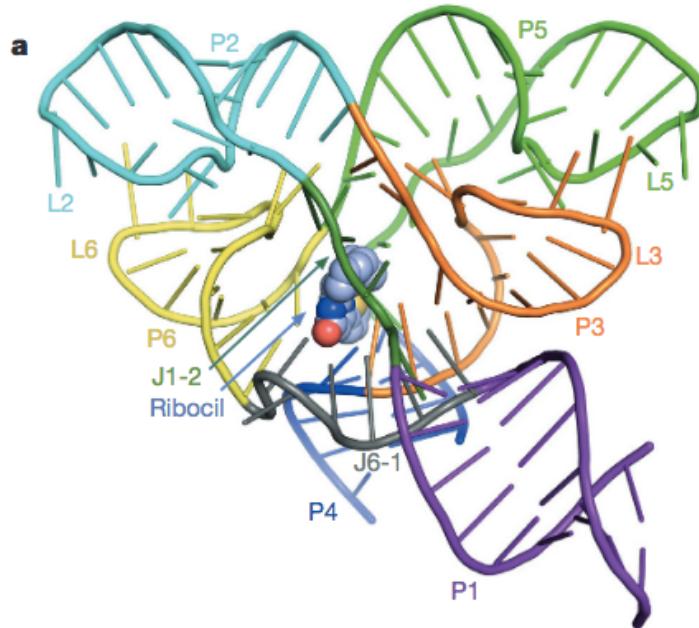
c



Selective small-molecule inhibition of an RNA structural element

Merck Research Laboratories

672 | NATURE | VOL 526 | 29 OCTOBER 2015



Nobel prizes for antibiotics

- 1939 - Герхарду Домагку за открытие антибактериального эффекта пронтозила (Prontosil)
- 1945 - Александру Флемингу, впервые выделившему пенициллин, и Ховарду Флори с Эрнстом Чейном, получившим его в чистом виде.
- 1952 - Зельману Ваксману за открытие стрептомицина (им был предложен термин «антибиотики» в 1942 году).