

RNA Structure and folding

Overview:

The main functional biomolecules in cells are polymers – DNA, RNA and proteins

For RNA and Proteins, the specific sequence of the polymer dictates its final structure

Can we predict the final structure of RNA or protein given just the sequence information?

Can we design any biomolecular structure that we want?

The world of RNA

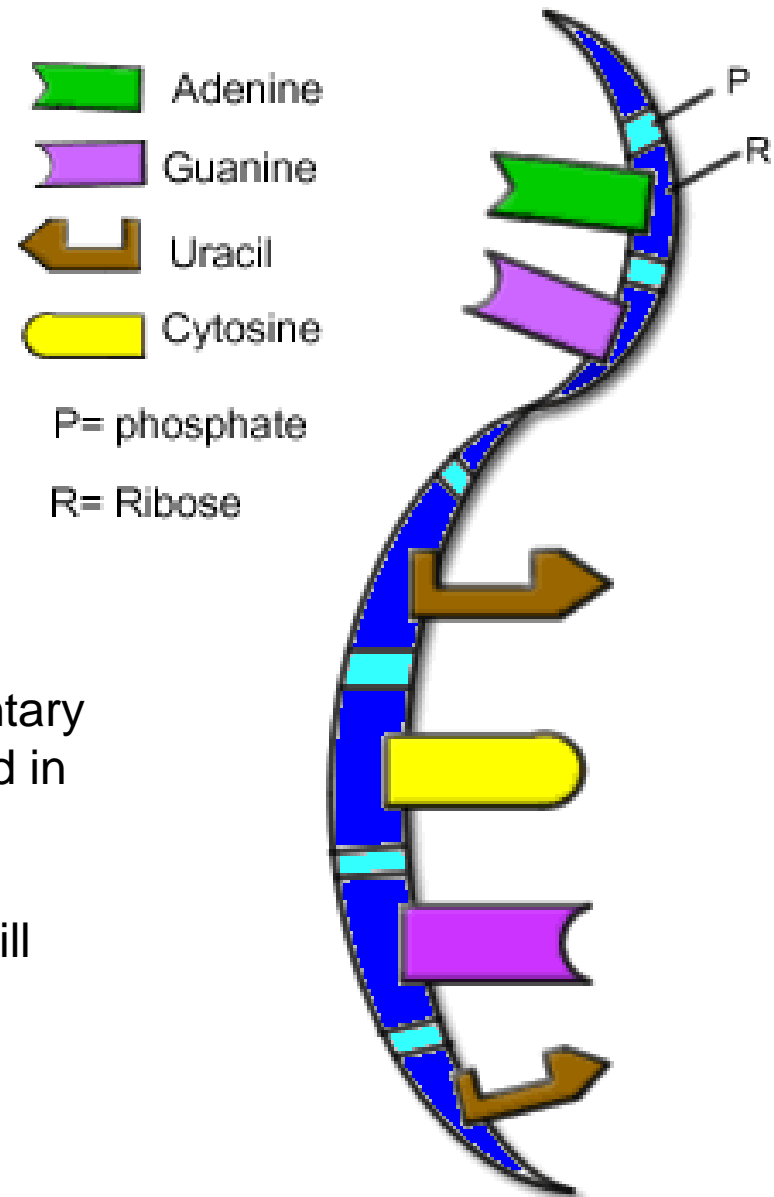
RNA is a linear polymer built from 4 possible monomers: A, G, U, C

These monomers can form complimentary interactions to form base-pairings:

A with U and C with G

Most often RNA is found as a single-stranded polymer that via base-pairing with complementary regions within its own sequence, is able to fold in on itself

RNA has a wide variety of functions that we will now explore



RNA function in cell

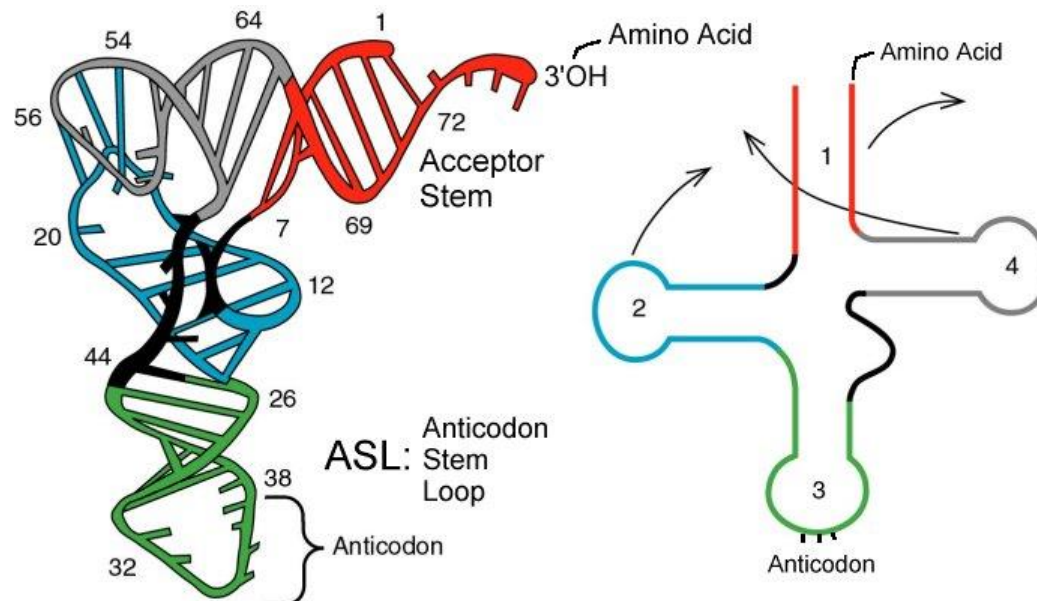
mRNA = messenger RNA

RNA's main function in the cell is to act as a messenger molecule in the process of making a protein

DNA (gene) → mRNA → Protein

tRNA = transfer RNA

these are used in translation to recognize the 64 codons. There is one tRNA for each codon, each representing one of the 20 amino acids



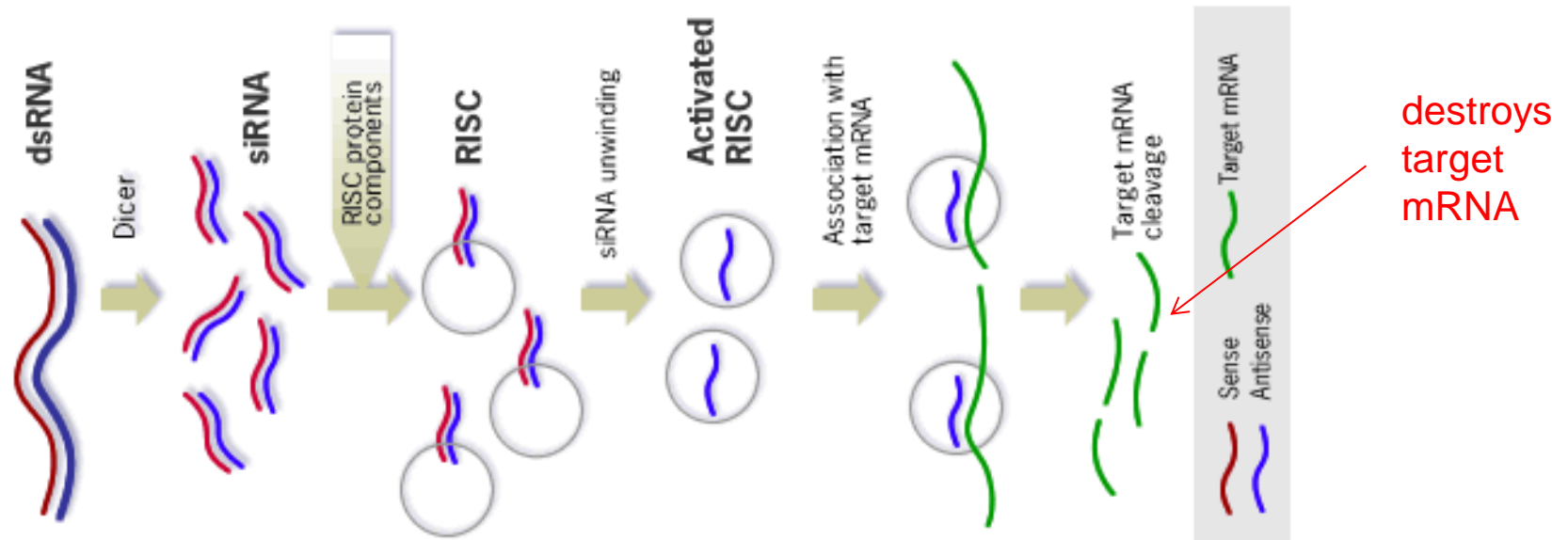
RNA function in cell

rRNA = ribosomal RNA

these are RNAs that get incorporated into the ribosome to give it part of its function.

microRNA or siRNA

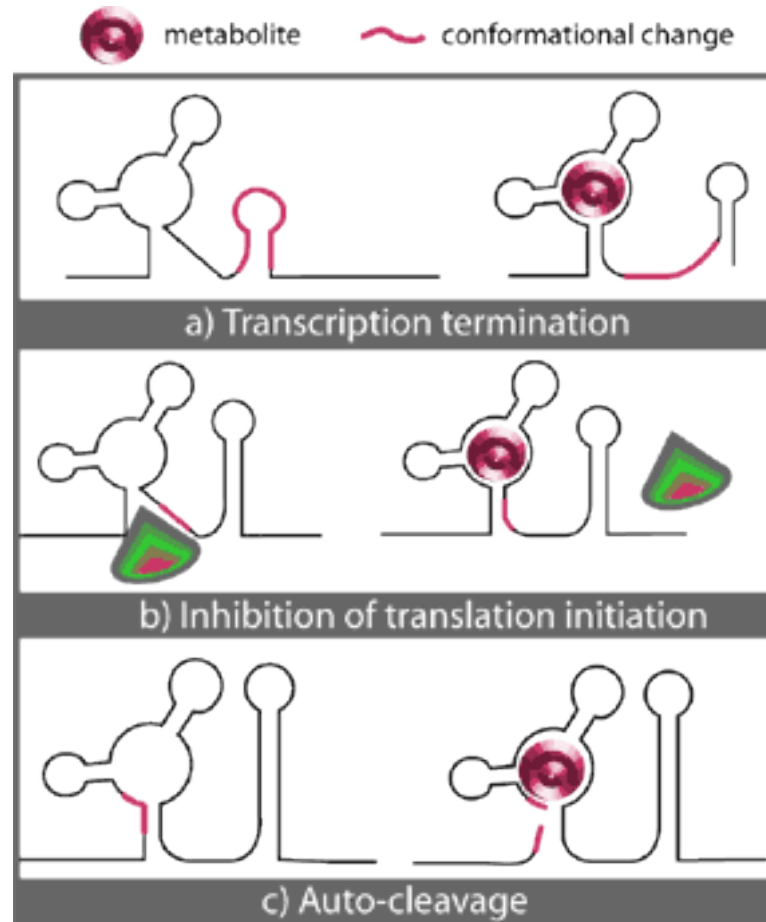
these are relatively recent discovered form of RNA that is used to regulate gene expression – so called RNA interference (won Nobel prize)



many developmental genes are regulated by miRNAs in your genome

RNA function in cell

Riboswitches – many mRNA molecules can detect metabolites by binding them and changing the structure of the mRNA = regulation

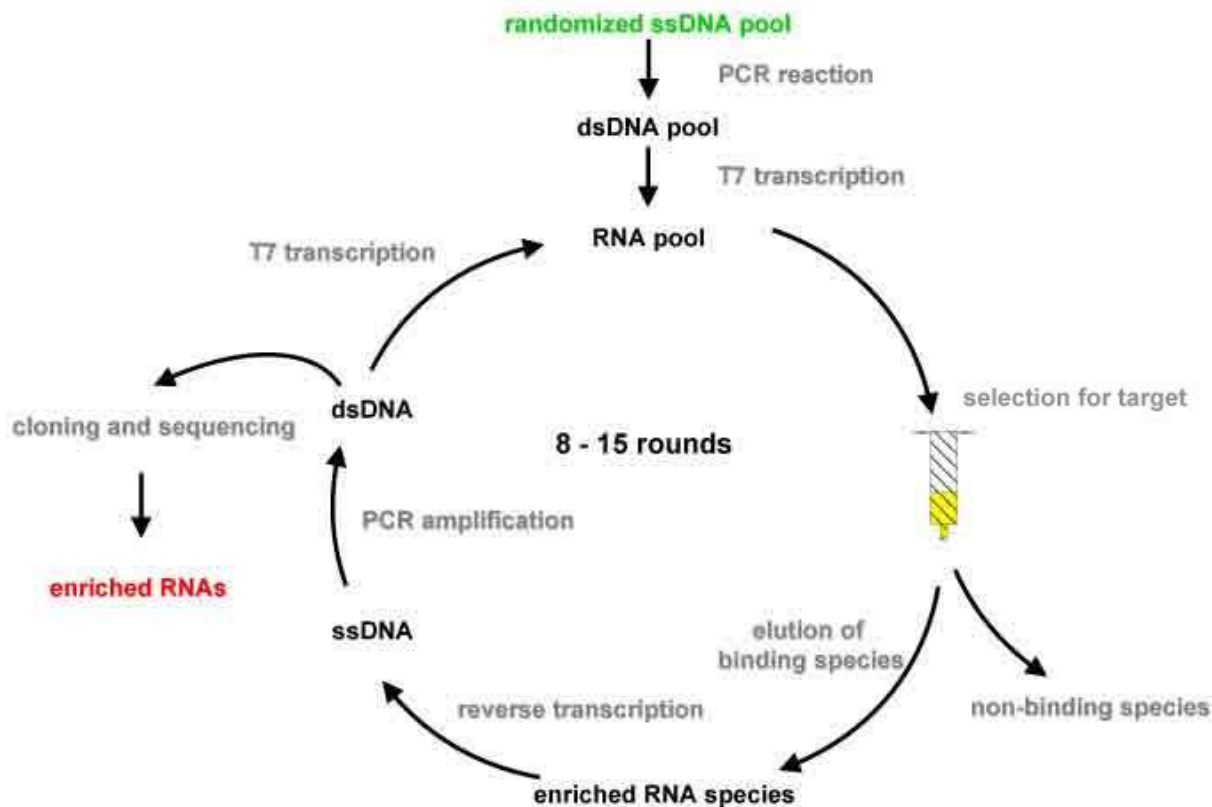


RNA function in cell

Ribozymes – like enzymes (catalytic proteins) except made from RNA. Able to catalyze reactions.

in-vitro selection experiments – can select RNA molecules out of a random library of sequences to catalyze a specific chemical reaction

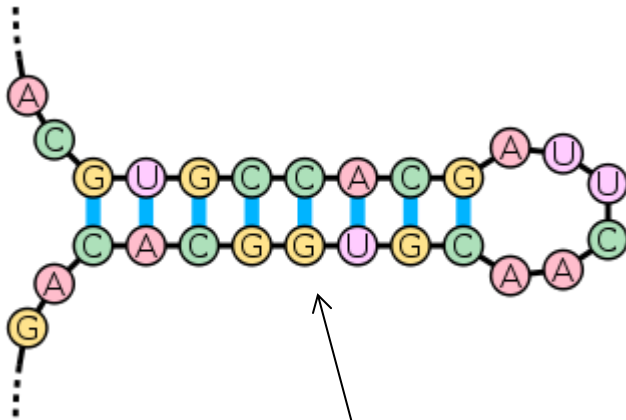
The RNA SELEX Process



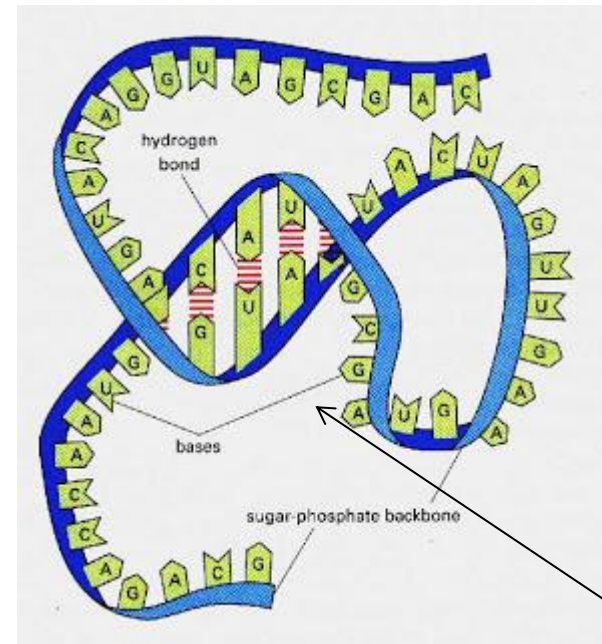
RNA structure:

A polymer of RNA first folds by forming complimentary base pairings: G = C and A = U

The simplest form of RNA structure is a hairpin loop that in 3D looks like a double helix



2ndary structure

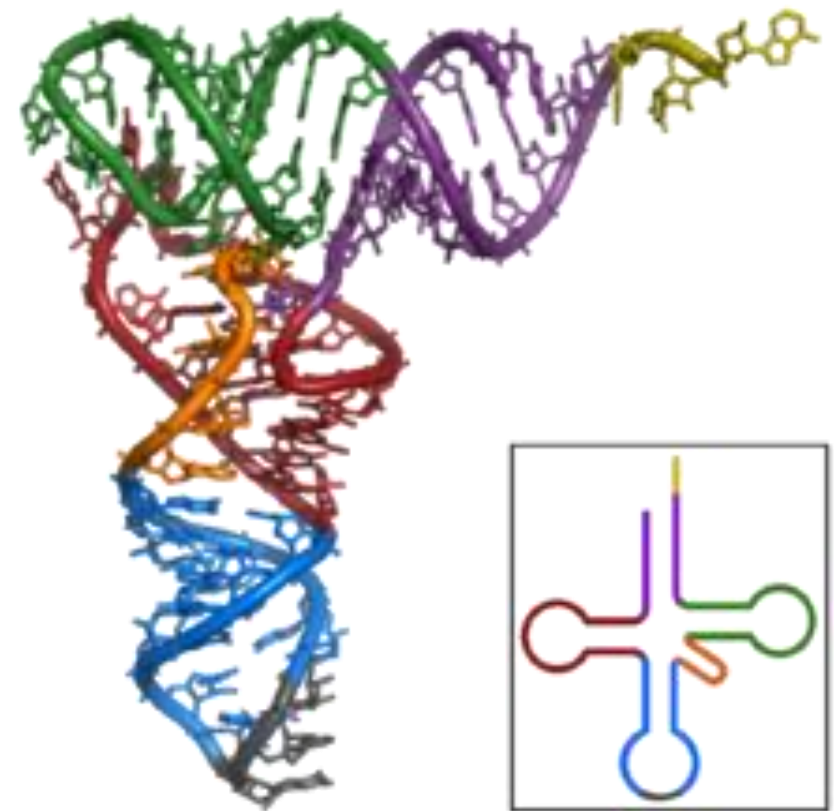
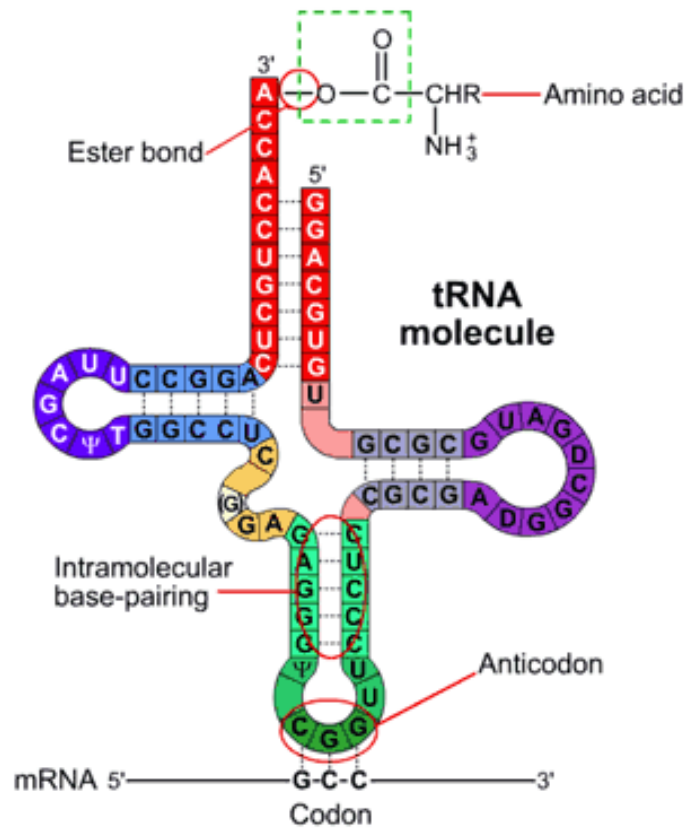


double helix

RNA structure representation

The 2ndary structure of RNA is a particular pairing of its complimentary bases

RNA tertiary structure is the final 3D fold of the polymer



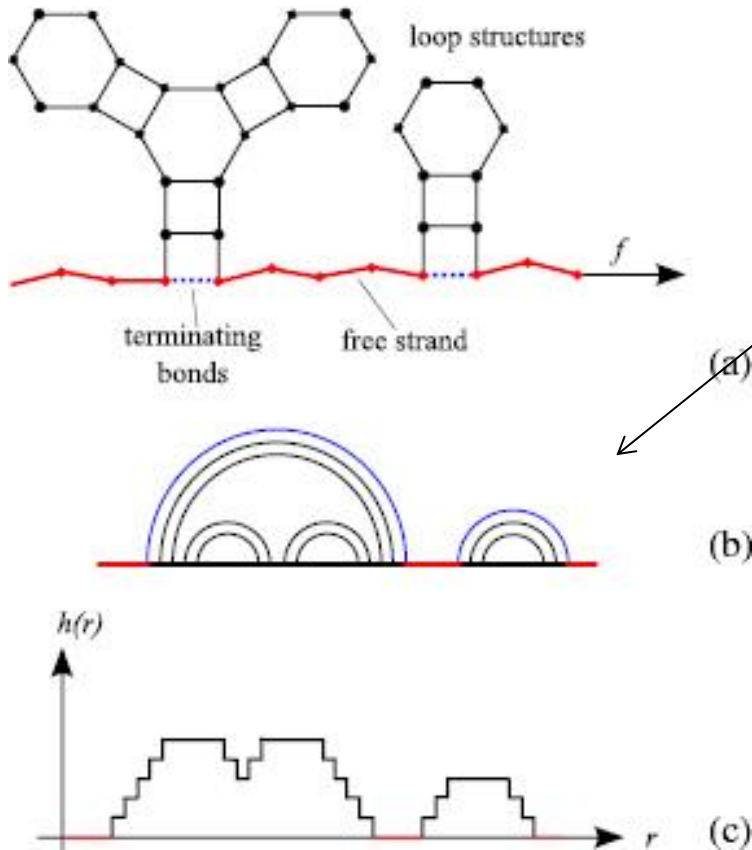
2ndary structure possesses 3 hairpin loops

Tertiary structure

RNA structure representation

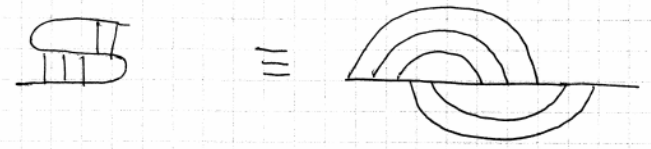
We will be interested in studying the formation of RNA secondary structure – tertiary is too hard of a problem

We need ways to represent the structure in diagrams or strings for use in calculations



Rainbow diagram – shows pairings as loops

no loops are allowed to cross = pseudo-knot



these occur but are computationally hard to deal with

String representation:

$(((((\dots))(\dots))))\dots((\dots))\dots$

1,1,1,1,0,0,-1,-1,1,1,0,0,-1,-1, ...

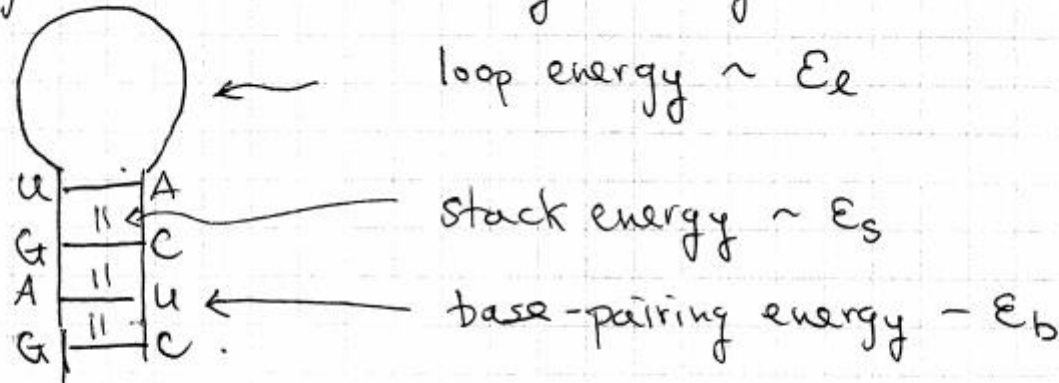
nice property: sum of the string = 0

RNA folding

A biomolecules function depends on it's structure. Can we predict the most probable 2ndary structure of an RNA molecule by just knowing it's sequence?

Ans: yes! enumerate all the possible structures (states), each has an energy, then use Boltzmann distribution to determine the probabilities

Energies involved in forming 2ndary structure



Turns out $E_s \gg E_b$ — so stacking drives folding

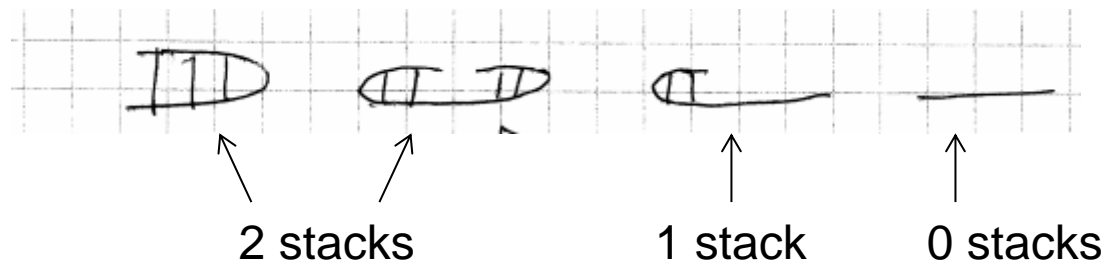
Consecutive base-pairs is a stack — a lone base-pair is not

Simple model for RNA folding

Since stacking energy dominates, ignore the contribution of lone base-pairs and only consider the energy that comes from forming stacks

Each stack lowers the energy of the structure by, $-|\epsilon_S|$

A structure that has n stacks has an energy of $E = -n |\epsilon_S|$



For small RNA sequences, we can enumerate all possible structures that possess stacks (do not draw structures that have lone base-pairs that are not part of a stack)

Calculate their energy (and possible entropy for the unpaired bases).

Ground state is the lowest (free) energy structure

For probabilities use Boltzman: $P(\text{structure}) = \exp(-E_{\text{structure}}/kT)/Z$

RNA structure prediction in the real world

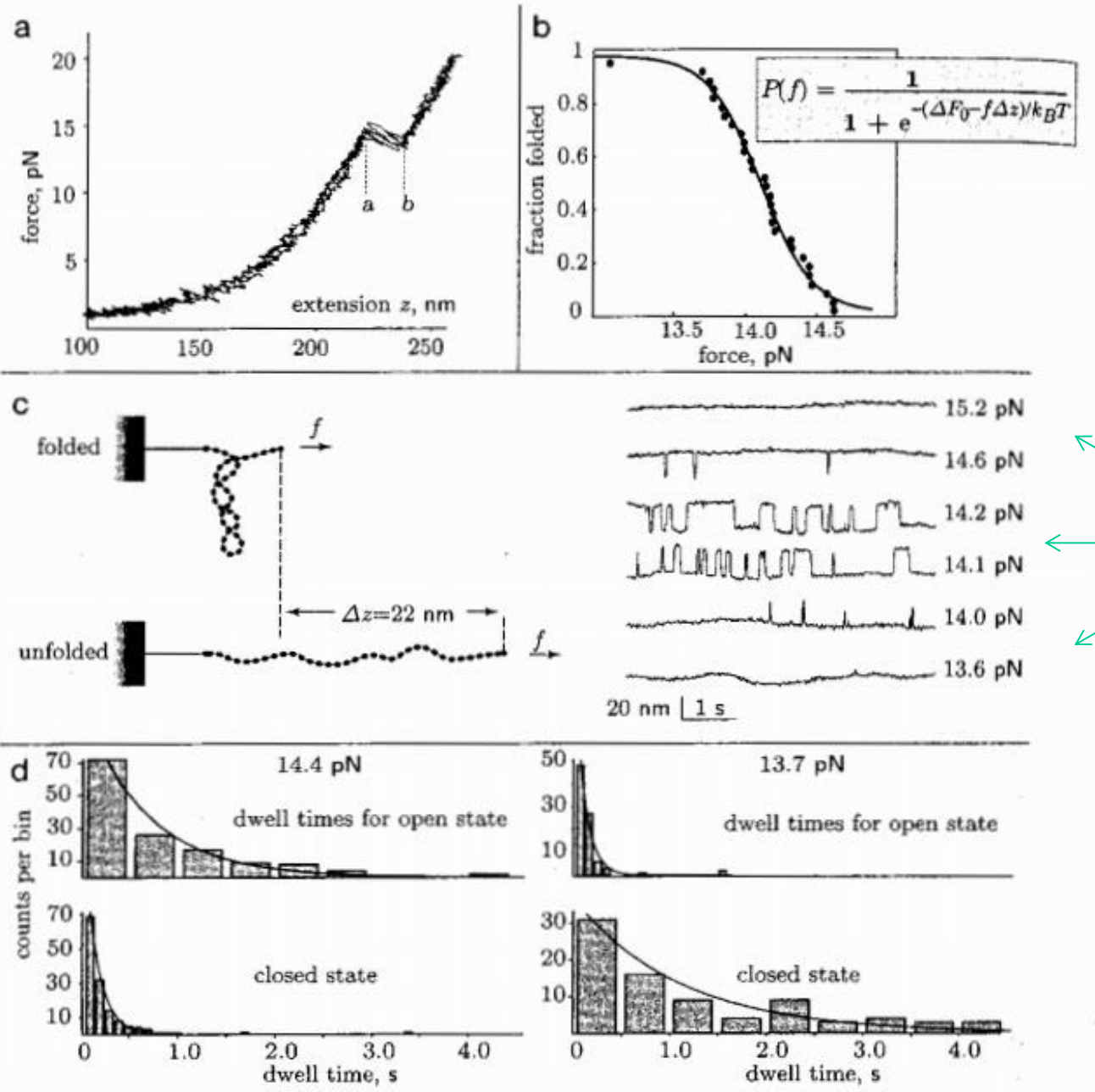
In reality, we can not draw all these structures by hand. Use a computer to enumerate possible structure sequences and calculate the energy of the sequence on each structure

<http://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>

Real-world RNA secondary structure prediction uses energies for base-pairing, stacking, looping and forming pseudo-knots

Wide range of applications from predicting mRNA secondary structure, the locations of miRNAs in a genome, designing PCR templates

Application to RNA folding



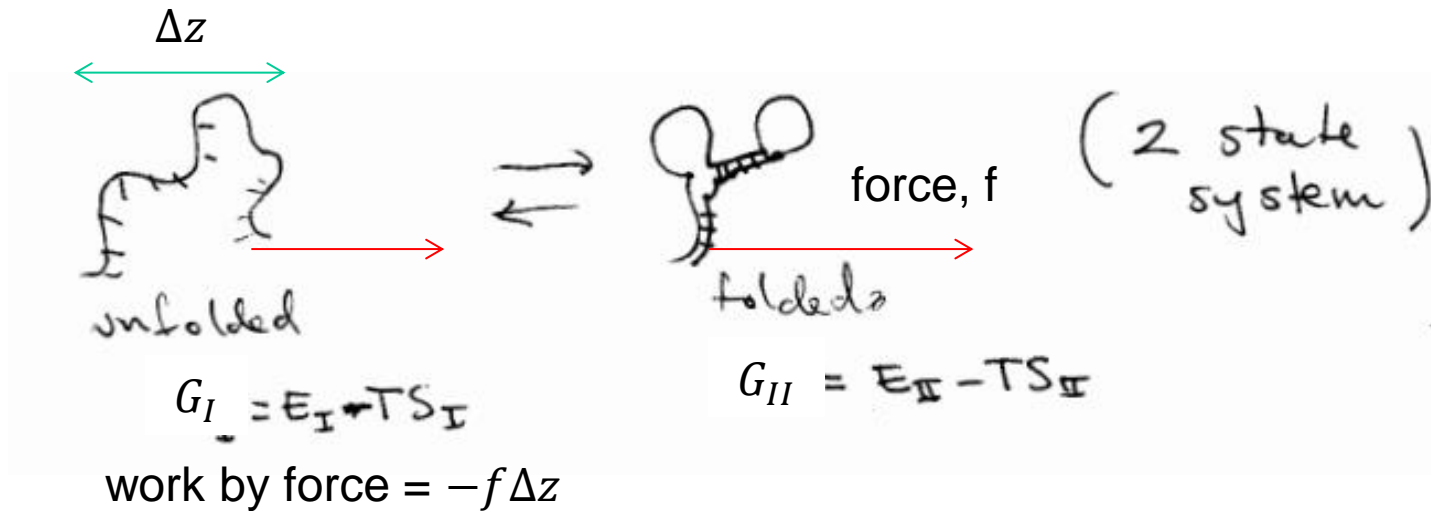
Using an optical tweezer, the Bustamante lab looked at how a single RNA molecule unfolded under different applied forces

different forces

Measured the probability of being folded

Measured dwell times

Application to RNA folding



The optical trap does work when the RNA unfolds a distance, Δz . So the free energy difference between the unfolded and folded in the presence of a force is:

$$\Delta G = G_I - G_{II} - f\Delta z = \Delta G_0 - f\Delta z$$

So the probability of being folded is:

$$P_{II} = \frac{\exp\left(-\frac{G_{II}}{kT}\right)}{\exp\left(-\frac{G_{II}}{kT}\right) + \exp\left(-\frac{G_I - f\Delta z}{kT}\right)} = \frac{1}{1 + \exp\left(-\frac{\Delta G_0 - f\Delta z}{kT}\right)}$$

Dwell times:

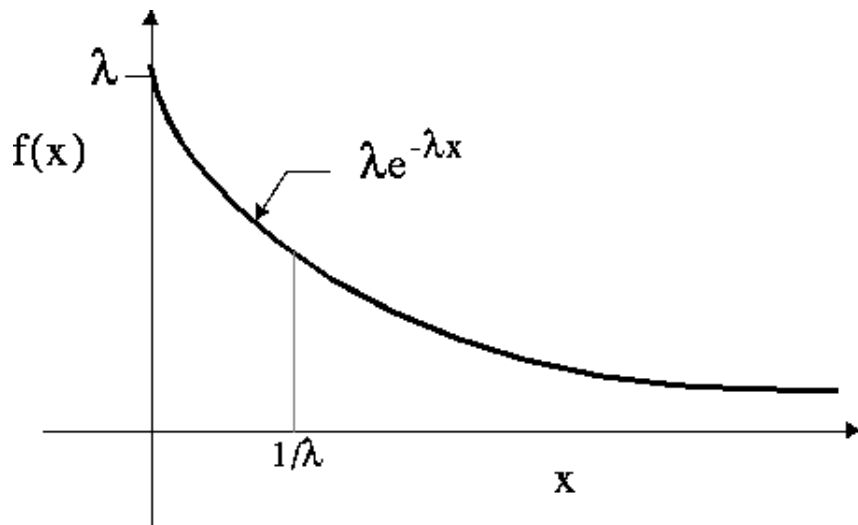


Figure 6. Exponential pdf

So dwell times are distributed exponentially

$$P_{2 \rightarrow 1}(t) = k_+ e^{-k_+ t}$$

Many reactions occur quickly, but some are very slow

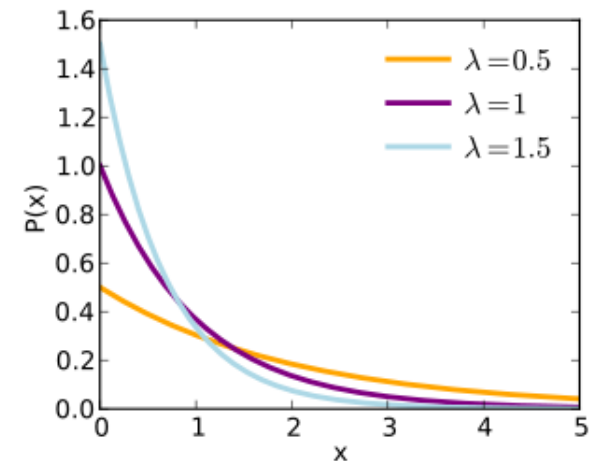
The energy barrier affects the rate. From Boltzmann

rate ($2 \rightarrow 1$) = (attempt rate) \times (probability of jumping barrier from $1 \rightarrow 2$)

or

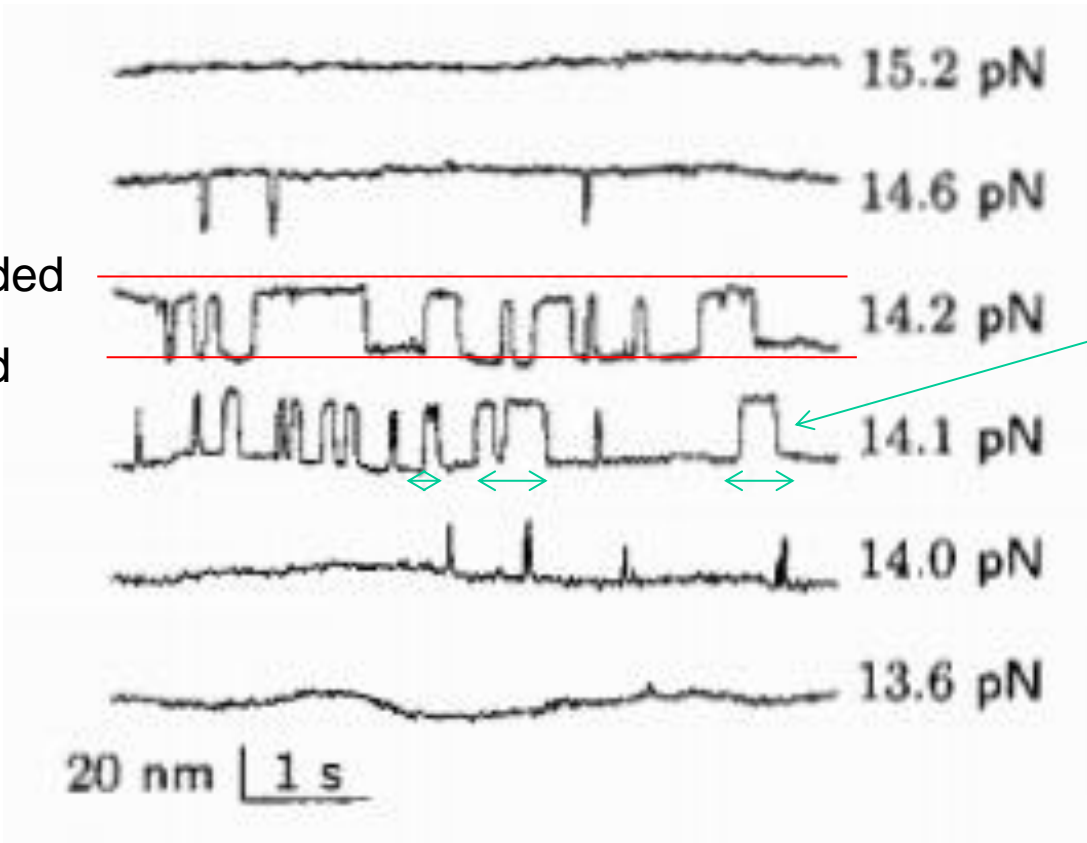
$$k_+ = v \exp\left(-\frac{E_B}{k_B T}\right)$$

so larger barrier \rightarrow slower rate and longer dwell times



RNA folding dwell times:

unfolded
folded



record the duration of time spent in a given state

histogram these values

