Demystifying the RNA World

Spring 2011
RNA Secondary Structure
Outline

• Chemical Composition of RNA
• Biologically Important Non-covalent Interactions
• Base-pairing in RNA (and DNA)
• The Hyperchromic Effect
• Introduction to RNA Thermodynamics
• Turner Rules
Chemical Composition of RNA

1. Nitrogenous Bases

2. The Pentoses of Nucleotides

3. Nucleosides are Formed by Joining Nitrogenous Base to a Sugar

4. Nucleotides - Nucleoside Phosphates

5. Nucleic Acids are Polynucleotides
1. Nitrogenous Bases

- **Pyrimidines**
  - Cytosine (DNA, RNA)
  - Uracil (RNA)
  - Thymine (DNA)

- **Purines**
  - Adenine (DNA, RNA)
  - Guanine (DNA, RNA)
**DNA & RNA Differences?**

*Why does DNA contain thymine instead of uracil?*

Cytosine spontaneously deaminates to form uracil. Repair enzymes recognize these mutations as “normal” and replace these Us with Cs. But how would the repair enzymes distinguish natural U from mutant U? Nature solves this dilemma by using thymine (5-methyl-U) in place of uracil.
2. Pentoses of Nucleotides

- D-ribose (in RNA)
- 2-deoxy-D-ribose (in DNA)
- The difference - 2'-OH vs 2'-H
- This difference affects secondary structure and stability.
3. Nucleosides - Linkage of a base to a sugar

- Base is linked via a glycosidic bond

- Named by adding -idine to the pyrimidine root name or -osine to the purine root name

- Sugars make nucleosides more water-soluble than free bases

- Conformation can be syn or anti

- "Nucleotide phosphate" is redundant!
- A nucleotide consists of a sugar, a phosphate, and one of the nitrogenous bases. (nucleoside + phosphate = nucleotide)
5. Nucleic Acids - Polynucleotides

- Polymers linked 3' to 5' by phosphodiester bridges
  - RNA: pApCpGpU or ACGU
  - DNA: pTpGpCpA or TGCA

- Each phosphate has a negative charge

- Ribonucleic acid and deoxyribonucleic acid
- Sequence is **always** read 5' to 3'
- In terms of genetic information, this corresponds to "N- to C-terminus" in proteins
### Dihedral angle review

<table>
<thead>
<tr>
<th>Torsion angle</th>
<th>Atoms involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$(n-1)O_3'-P-O_5'-C_5'$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$P-O_5'-C_5'-C_4'$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$O_5'-C_5'-C_4'-C_3'$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$C_5'-C_4'-C_3'-O_4'$</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>$C_4'-C_3'-O_3'-P$</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>$C_3'-O_3'-P-O_5'(n+1)$</td>
</tr>
<tr>
<td>$\chi$</td>
<td>$O_4'-C_1'-N_1-C_2$ (pyrimidines)</td>
</tr>
<tr>
<td>$\nu_0$</td>
<td>$O_4'-C_1'-N_5-C_4$ (purines)</td>
</tr>
<tr>
<td>$\nu_1$</td>
<td>$C_4'-O_4'-C_1'-C_2'$</td>
</tr>
<tr>
<td>$\nu_2$</td>
<td>$O_4'-C_1'-C_2'-C_3'$</td>
</tr>
<tr>
<td>$\nu_3$</td>
<td>$C_1'-C_2'-C_3'-C_4'$</td>
</tr>
<tr>
<td>$\nu_4$</td>
<td>$C_2'-C_3'-C_4'-O_4'$</td>
</tr>
</tbody>
</table>

*Atoms designated $(n-1)$ and $(n+1)$ belong to adjacent units.*
The sequence of bases along a nucleic acid chain carries genetic information.

The chain of sugars linked by phosphodiester bonds is referred to as the backbone of the nucleic acid.
Outline

• Chemical Composition of RNA

• Biologically Important Non-covalent Interactions

• Base-pairing in RNA (and DNA)

• The Hyperchromic Effect

• Introduction to RNA Thermodynamics

• Turner Rules
Molecular interactions within cells are mediated predominantly by non-covalent interactions:

(i) Ionic (electrostatic) interactions,
(ii) Hydrogen bonds,
(iii) van der Waals interactions,
(iv) Hydrophobic interactions

These interactions are related, and are involved in stabilizing the three dimensional structures of macromolecules (proteins, nucleic acids)

Living systems are composed of ≥ 90% water

⇒ .: properties and interactions of water play a fundamental role in molecular interactions
(i) Ionic (Electrostatic) Interactions

Coulomb’s law

\[ F \propto \frac{q_1 q_2}{r^2} \]

\( F = \text{Electrostatic force between two electric charges, } q_1 \text{ and } q_2 \)

\( r = \text{distance between the charges} \)

\[ F = \frac{k q_1 q_2}{D r^2} \]

where

\( k = \text{proportionality constant (1389 kJ mol}^{-1} \text{ or 332 kcal mol}^{-1}) \)

\( D = \text{dielectric constant of the medium} \)

The electrostatic interaction energy (potential energy), \( E \), is given by

\[ E = \frac{k q_1 q_2}{D r} \]
(ii) Hydrogen Bonds

A hydrogen atom that can interact simultaneously with two electronegative atoms is said to form a hydrogen bond.

- Linear (strong)
  \[ \theta = 180^\circ \]
  - \( \Delta H = 2 \text{ kcal/mole} \)
  - \( \Delta H = 3 \text{ kcal/mole} \)
  - \( \Delta H = 5 \text{ kcal/mole} \)
  - \( \Delta H = 7 \text{ kcal/mole} \)
  - About 12-30 kJ/mole
(iii) van der Waals interactions

van der Waal’s interactions include a number of nonspecific, short-range forces between uncharged molecules

- London dispersion forces
- Dipole-dipole interactions
- Dipole-induced dipole interactions
- Steric repulsions
London-van der Waals interaction

6-12 potential: Energy of attraction depends on $r^{-6}$
Energy of repulsion depends on $r^{-12}$
Where $r =$ distance between the two atoms

$U = \frac{B}{r^{12}} - \frac{A}{r^{6}}$

Short-range, strong interactions
Difficult to measure precisely

Region of attraction
Region of repulsion
Sum of van der Waals radii
van der Waals contact distance

Responsible for the dense packing within proteins
(iv) Hydrophobic Interactions

This type of interaction is contributed mainly by the solvent water and arises due to its unique properties. Hydrophobic interactions are primarily responsible for the separation of nonpolar solutes and water.

Aggregation of nonpolar solute molecules in water minimizes their surface area exposed to water, and hence reduces the entropy loss of water molecules when a nonpolar solute is introduced into water.

\[ \therefore \text{The hydrophobic effect is an entropy-driven process} \]
Two Important Points About Weak Forces (weak relative to covalent bonds)

Biomolecular recognition is mediated by weak chemical forces.

Weak forces restrict organisms to a narrow range of environmental conditions.

Energies Of Non-covalent Interactions

- Ionic bonds: 20 kJ/mole
- Hydrogen bonds: 4 - 20 kJ/mole
- van der Waals: 0.4 - 4.0 kJ/mole
- Hydrophobic interactions: <40 kJ/mole
How do the four non-covalent interactions work together to drive the association of two strands of DNA (or RNA) to form a double helix?
Ionic interactions

Each phosphate group in a DNA (RNA) strand carries a negative charge.

Thus, unfavorable electrostatic interactions take place when DNA (RNA) strands come together.

The strength of these repulsive forces are mediated by the distances between the phosphate groups in the double helix, the high dielectric constant of water, and the presence of ionic species in solution.

This is the main reason why monovalent AND/OR divalent ions in solution stabilize all DNA and RNA structures.
Hydrogen bonds

Hydrogen bonds are important in determining the specific base pairs in a double helix, however, in solution DNA (RNA) single strands form hydrogen bonds with $\text{H}_2\text{O}$.

Because the number of hydrogen bonds broken and formed are the same, hydrogen bond formation DOES NOT contribute substantially to driving the process of double helix formation.
van der Waals interactions

Within a double helix, base pairs are parallel and stacked in close proximity.

The typical base separation between base planes in DNA is 3.4 Å and distances between closely approaching atoms approx. 3.6 Å.

This separation distance corresponds nicely with the van der Waals contact distance.

Base stacking and van der Waals interactions are nearly optimal in a double-helical structure.
Hydrophobic interactions

More complete base stacking moves the non-polar surfaces of the bases out of water into contact with each other.

Thus, many weak interactions contribute to the overall energetics of the process, some favorably and some unfavorably.

And…

van der Waals and hydrophobic interactions drive the formation of DNA (and RNA) double helices.
What about the importance of Hydrogen bonds though?

Hydrogen bonds are important in determining specificity!

If you bring two bases together that cannot form hydrogen bonds, the hydrogen bonds with water must be broken, but no new ones are formed, which would not be favorable.

i.e. Formation of helices with non-complementary base pairs is unfavorable.

The van der Waals and hydrophobic forces drive nucleic acids to form double helices, and the hydrogen bonds determine which strands of nucleic acid come together to form helices.
Outline

• Chemical Composition of RNA
• Biologically Important Non-covalent Interactions
  • Base-pairing in RNA (and DNA)
• The Hyperchromic Effect
• Introduction to RNA Thermodynamics
• Turner Rules
Watson-Crick base pairing

The two different base pairs, AT and CG, have essentially the same shape, which allows them to fit nicely into the regular structure of DNA.
The rules of base pairing (or **nucleotide** pairing) are:

**A** with **T (U)**: the **purine** adenine (A) always pairs with the **pyrimidines** thymine (T) and/or uracil (U)

**C** with **G**: the **pyrimidine** cytosine (C) always pairs with the **purine** guanine (G)

This is consistent with there not being enough space (20 Å) for two purines to fit within the helix and too much space for two pyrimidines to get close enough to each other to form hydrogen bonds between them.
G·U wobble base pairs are the most common and highly conserved non-Watson–Crick base pairs in RNA.

G·U pairs have been found in virtually every class of functional RNA, and have been shown to play many essential roles that are based upon the unique chemical and structural properties of the wobble pair.
Surface electrostatic potential maps of an isolated Watson-Crick GC base pair, G•U wobble base pair, and Watson-Crick AU base pair

“...the negativity at the major groove of G•U wobble base pairs is determined by the combined effect of the base atoms and the sugar-phosphate backbone, which is impacted by stacking pattern and groove width as a result of base sequence.”

(yellow is the most negative, and green is the most positive)

So we can have A-U, G-C, and G•U, but what about A•C?

Actually, yes you can BUT…

The adenine must be protonated in order to form a $A^+•C$ base pair using the Watson-Crick face of the base.
Okay, okay, we can have A-U, G-C, and G•U, AND A•C?

BUT, the whole purine-pyrimidine rule for base-pairing still applies… right?

No, actually not really…
Purine-Purine and Pyrimidine-Pyrimidine Base Pairs are found too…
And don’t get me started on base triples…

Check out the Non-Canonical Base-Pair Database:
http://prion.bchs.uh.edu/bp_type/
The names of the edges of the bases

Hoogsteen – blue
Sugar – green
Watson Crick – red

And why do you care…?
Cis vs. Trans Orientations of the Glycosidic Bonds

Interacting Edges

Glycosidic Bond Orientations

Cis orientation of the Glycosidic Bonds

Trans orientation of the Glycosidic Bonds
“There are more things in heaven and earth (and base pairing), Horatio, than are dreamt of in your (Watson-Crick) philosophy.” (Hamlet, Shakespeare)
• Chemical Composition of RNA
• Biologically Important Non-covalent Interactions
• Base-pairing in RNA (and DNA)
• The Hyperchromic Effect
• Introduction to RNA Thermodynamics
• Turner Rules
The Double Helix can be Reversibly Melted (DNA or RNA)
The observed absorbance of an RNA or DNA sample increases (typically between ~5-30%) as the RNA or DNA is denatured (the hyperchromic effect)
Optical Properties of DNA and RNA

The rings of the bases are made up of alternating single and double bonds (conjugated ring system).

Without getting into Physical Chemistry too much, the particle in the box, etc. you may remember that conjugated ring systems absorb light with a wavelength dependence on the number of conjugated bonds...

UV absorption of DNA and RNA results from $n \rightarrow \pi$ and $\pi \rightarrow \pi^*$ transitions of the conjugated ring systems of the nucleobases.
Optical properties of RNA
the hyperchromic effect

Short Version:
The hyperchromic effect derives from changes in the molar absorptivity of the nucleic acid bases, $\varepsilon$, due to stacking interactions when folded.

Long Version:
The transition dipole moment of the absorbing base interacts with the light-induced dipoles of neighboring bases.

This interaction depends on the separation and orientation of the bases.

When bases are stacked ~parallel and the transition dipole moments of adjacent bases are oriented more or less “head-to-head” (as they are in a helix), the probability of photon absorbance by a base is reduced due to the light induced dipoles of the neighboring bases.

This results in a lower extinction coefficient for helical DNA or RNA than for the single stranded or unfolded form.
Absorbance properties of the nucleotides

Due to the absorbance properties of the individual nucleotides, (particularly the shoulder at 280 nm for G)

Unfolding of RNA can be monitored at both 260 and 280 nm to determine the approximate base pair composition for the unfolding process.
In addition, the magnitude of the hyperchromic effect at different wavelengths is RNA sequence dependent.

Monitoring at 260 nm reports on both the presence of A-U and G-C base pairs, while monitoring at 280 nm reports almost exclusively on the presence of G-C base pairs.
Outline

• Chemical Composition of RNA

• Biologically Important Non-covalent Interactions

• Base-pairing in RNA (and DNA)

• The Hyperchromic Effect

• Introduction to RNA Thermodynamics

• Turner Rules
Thermodynamics of Biological Systems

Thermodynamics is concerned with the bulk behavior of substances.

Thermodynamics describes the relationships among the various forms of energy and how energy affects matter on the macroscopic level.

Equilibrium thermodynamics can indicate whether a process will occur spontaneously, but cannot predict how fast it will go.

It can be used to study the equilibrium positions and energy changes involved in biochemical reactions.
The First Law

Law of conservation of energy – defines the “internal energy” of the system

• The total energy of a system and its surroundings is constant

• Energy can neither be created or destroyed.

Energy can take different forms:

**Heat** – the manifestation of the kinetic energy associated with the random motion of molecules

**Potential energy** – released upon the occurrence of some process

**Work** – defined as force times the distance moved under its influence (organized motion)
The First Law

Law of conservation of energy – defines “internal energy” of the system

U (or E) is the internal energy - a function that keeps track of heat transfer and work expenditure in the system

U is independent of path (“State function”)

\[ \Delta U = \Delta q - \Delta w \]

- q is heat absorbed BY the system
- w is work done BY the system ON the surroundings
For heat absorbed at constant $P$, we define a new energy function called “Enthalpy”

A better function for constant pressure

$$\Delta H = E + PV$$

If $P$ is constant, $H = q_P$

(Typically true for most biochemical processes)

$H$ is the heat absorbed at constant $P$

Since volume is approx. constant for biochemical reactions

(in solution)

∴ $H$ is approx. the same as $E$

$\Delta H$ is the easily measured heat that is generated or absorbed by a biochemical process
The Second Law

The total entropy of the system plus that of its surroundings always increases.
“The universe tends towards maximum disorder”

Simply stated, entropy is a measure of the disorder or randomness of a system

Systems tend to proceed from ordered to disordered states

The entropy change for (system + surroundings) is unchanged in reversible processes and positive for irreversible processes

All processes proceed toward equilibrium - i.e., minimum potential energy
Entropy

A measure of disorder
An ordered state is low entropy
A disordered state is high entropy

\[ dS \geq dq/T \]
\[ dS_{\text{reversible}} = dq/T \]

In general, for any constant energy process (\(\Delta U=0\)), a spontaneous process is characterized by \(\Delta S>0\).
Free Energy
(Gibbs free energy)

The free energy, G, is defined as
\[ G = H - TS \]

(It follows from the second law that \( dq = dH \) and \( dH - TdS < 0 \) at constant pressure – for reactions to proceed as shown)

For any process at constant P and T (as in most biochemical reactions):
\[ \Delta G = \Delta H - T\Delta S \]

If \( \Delta G = 0 \), reaction is at equilibrium
If \( \Delta G < 0 \), reaction proceeds as written
How do the principles of thermodynamics apply to the formation of the DNA double helix?
Before helix formation both strands are free to rotate and translate in solution.

Thus, double helix formation appears to result in an increase in order for the system.

Therefore, based on the Second Law of Thermodynamics, double helix formation would require the release of heat to increase the entropy of the surroundings.

For these strands at 25 °C (pH 7.0) in 1 M NaCl, 250 kJ/mole of heat is released to the surroundings.
We can collect Abs vs. T data, convert to dA/dT for convenience, and see the 260 nm /280 nm ratio for individual unfolding events.

This allows us to observe experimentally which helical stem unfolds first in the unfolding pathway of a particular RNA.
The two curves (at 260 and 280 nm) are then fit simultaneously to a model of sequential, interacting, **two-state** transitions with each transition having 4 parameters:

- $t_m$ - the melting temperature
- $\Delta H$ - the enthalpy of unfolding
- $A_{260}$ - the amplitude of the 260 nm hyperchromic effect
- $A_{280}$ - the amplitude of the 280 nm hyperchromic effect

The ratio of $A_{260}$ and $A_{280}$ for each transition in the molecule as well as for selected mutants allows the assignment of the specific unfolding pathway.
Since the stability of the molecule is independent of $\lambda_{obs}$, therefore, the same $tm$ and $\Delta H$ will apply to both the 260 and 280 nm data, although the intensities of the transitions will vary.

$tm$ and $\Delta H$ are the thermodynamic parameters derived directly from the melting data.

$$\Delta G = \Delta H - T\Delta S \quad \Delta G = -RT \ln K_{eq}$$

Due to the definition of melting temperature, $tm$, $\Delta G$ must = 0 and $K_{eq} = 1$ at the melting temperature.

$tm = \text{temp} \text{ at which half the molecules have unfolded}$

two state assumption : $F \leftrightarrow U$ so $K_{eq} = [U]/[F]$ at $tm$:

$[U] = [F]$, thus $K_{eq} = 1$ and since $\ln(1) = 0$

$\Delta G = 0 \text{ at } tm !!!$

$\Delta G = 0 = \Delta H - tm\Delta S$ \quad and $\Delta S = \Delta H/tm$

Finally, $\Delta G_{(37 \text{ C})} = \Delta H - (310.15)(\Delta H/tm) \text{ tm in Kelvin}$

Melting data in the form of a melting profile ($dA/dT$ vs. temperature) is fit to a model of sequential, interacting, two-state transitions.
Assumptions

- Individual transitions display two-state unfolding behavior.

- The $\Delta H$ and $\Delta S$ terms are temperature independent ($\Delta C_p$ is zero).

The two-state assumption is generally only applicable for relatively short oligonucleotides (less than 20 continuous base pairs).

Even short oligonucleotides can display non-two-state behavior.
Outline

• Chemical Composition of RNA
• Biologically Important Non-covalent Interactions
• Base-pairing in RNA (and DNA)
• The Hyperchromic Effect
• Introduction to RNA Thermodynamics
• Turner Rules
The Turner Rules

For over a decade, the laboratory of Professor D.H. Turner at the University of Rochester (here in New York!!!) as well as a number of other laboratories (including Tinoco’s group) has been estimating nearest neighbor parameters for RNA based on melting studies of synthetically constructed oligoribonucleotides.

Stacking Energies

Energy rules were first derived for stems containing canonical base pairs: Watson-Crick (WC) base pairs and G-U wobble pairs.

In the sample helix at the right, the total free energy is given by the addition of 7 free energy terms, 1 for each pair of adjacent base pairs. This includes energy contributions for both base pair stacking and hydrogen bonding.

Such nearest neighbor rules work very well for WC base pairs, and satisfactorily for single G-U pairs surrounded by WC pairs. They break down for 2 or more consecutive G-U pairs and for non-canonical pairs.

For a stack of base pairs:

5’-WX-3’
3’-ZY-5’

the corresponding energy at 37° was determined and can be calculated for other temperatures.
The Turner Rules

Hairpin Loop Energies

Hairpin loop free energies are the sum of up to 3 terms.

1. A purely entropic term that depends on the loop size (the number of single stranded bases in the loop). For loops larger than 30, an extra term, $1.75RT\ln(\text{size}/30)$, is added, where $R$ is the universal gas constant and $T$ is absolute temperature.

2. There is a favorable stacking interaction between the closing base pair of the hairpin loop and the adjacent mismatched pair. These energies are given in special hairpin loop terminal stacking energy tables.

3. Certain tetraloops have special bonus energies attached to them.

Interior and Bulge Loop Energies

Interior and bulge loops are closed by 2 base pairs.

Interior loop energies are the sum of up to 3 terms.

1. As with hairpin loops, there is a purely entropic term for interior loop energies that depends on the loop size. For loops larger than 30, an extra term, $1.75RT\ln(\text{size}/30)$, is added.

2. As with hairpin loops, there are special terminal stacking energies for the mismatched base pairs adjacent to both closing base pairs.

3. For non-symmetric interior loops, there is an asymmetric loop penalty.
The Turner rules are a set of experimentally determined parameters which allow us to predict the stability of RNA secondary structures.

These rules and data from other sources are used by all of the existing structure prediction programs (as far as I know) which use energy minimization to calculate the lowest energy theoretical secondary structure for an RNA sequence.