We have learned much about nucleic acid since their discovery in the 1820s, yet a great deal remains unknown.



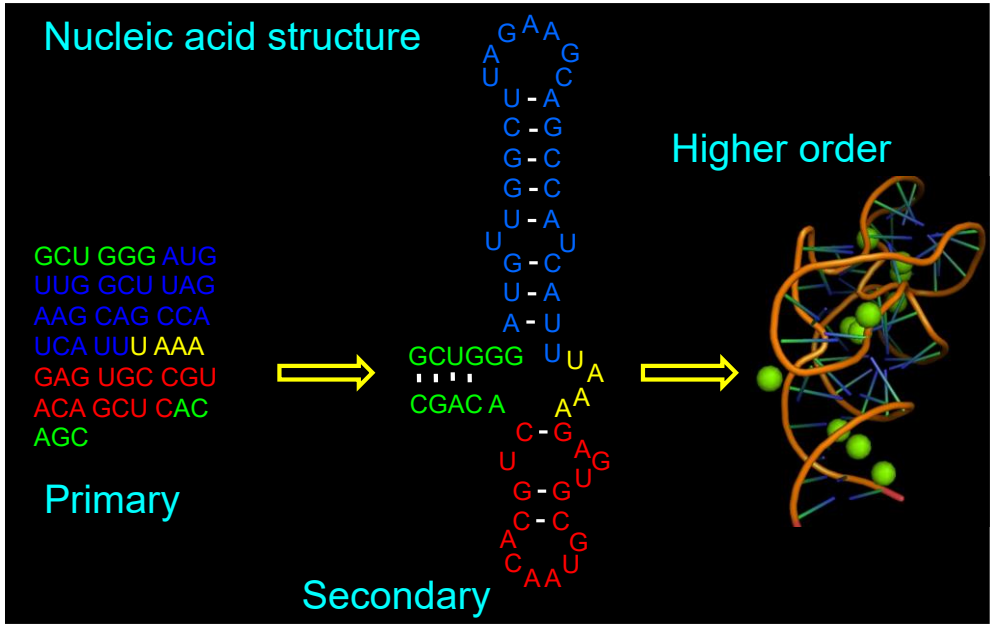
An open mind and the right experiment might just lead you to the next big discovery! So let's get to it!

RNA: composition and base pairing

Nucleic acids are polymers made up of repeating units of **sugars** that are connected via **phosphate molecules**, with each sugar bound to **one nucleobase**.

The nucleobases are connected to the sugar–phosphate backbone in a particular order.

The **sequence** of nucleobases is the **primary structure** of a nucleic acid.



RNA: composition and base pairing

Sugar

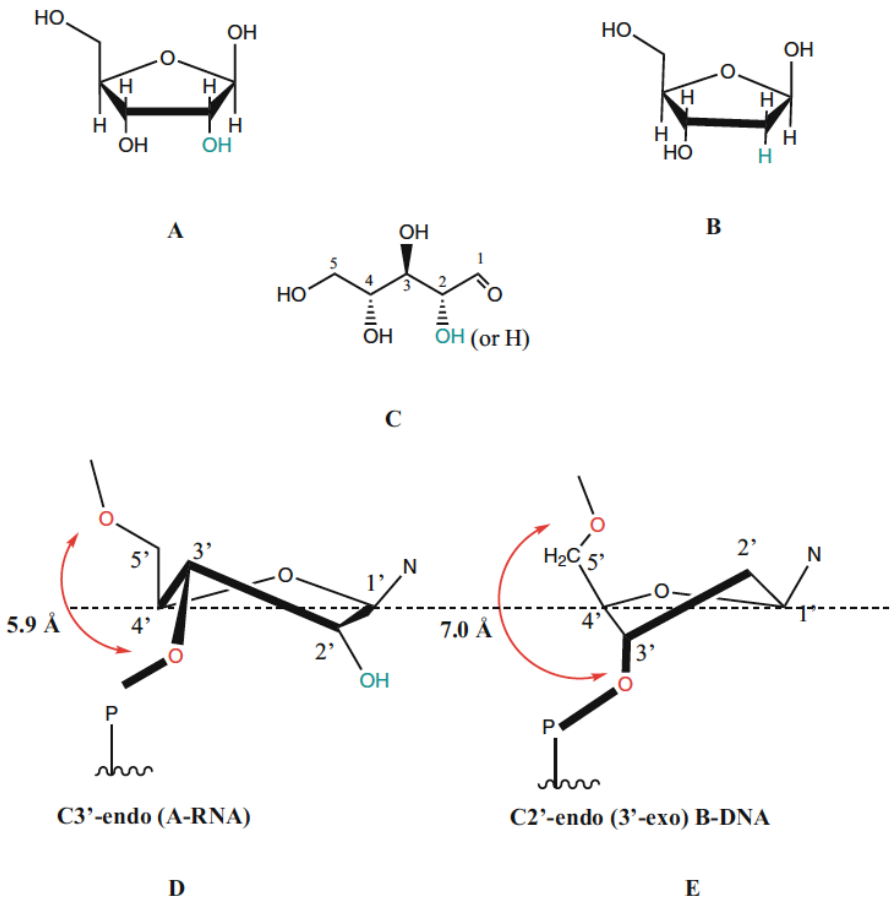


Fig. 1.1 Ribose and 2'-Deoxy-ribose Sugars. Planar ring structures of the ribose (**a**), deoxyribose sugars (**b**), along with the straight chain form of ribose molecule (**c**) are shown. The most stable conformation is C3'-endo for the ribose (**d**) and C2'-endo for the deoxyribose (**e**)

Carbon 5' is outside the ring and the oxygen of 4'O is now part of the ring structure.

Various ring conformations are possible as the molecular structures are dynamic.

The most stable conformations of ribose involve displacing at least one atom out of the plane defined by C4'-O-C1' atoms (Fig. 1.1d, e) [1].

The most stable conformation for **ribose sugar** is **3'-endo**, where C3' lies above the plane made by C4'-O-C1' atoms (Figure 1.1d).

When the 2'-position has a hydrogen (instead of the hydroxyl) group attached to it (**deoxy**), then the preferred conformation is **2'-endo** (Figure 1.1e).

RNA: composition and base pairing

The phosphate groups connect the sugars to each other via **two phosphodiester linkages** (Fig. 1.2). The ester linkage is formed by losing a water molecule between an acid and alcohol functional groups.

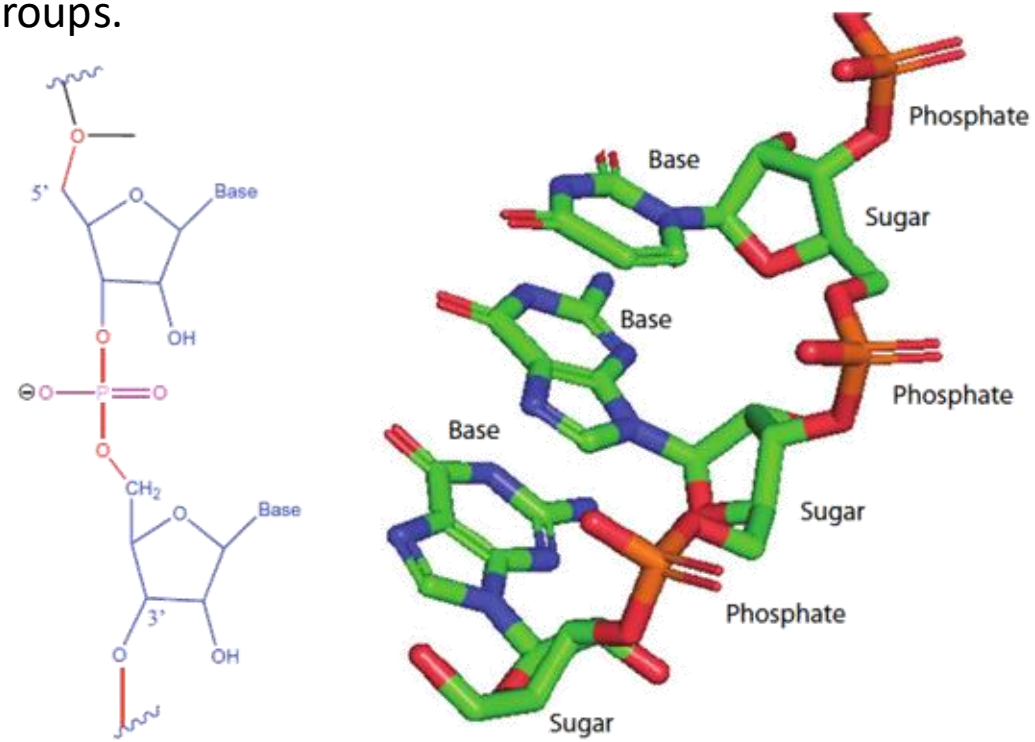
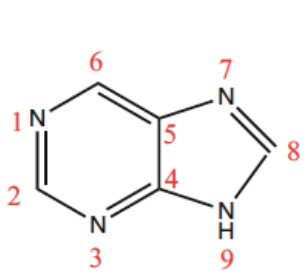


Fig. 1.2 Phosphodiester Linkage. The 5′- and 3′-hydroxyl groups of sugars form ester linkages with the phosphate groups to make the polymer. The crystal structure of RNA (1.5 Å resolution) is shown. The sugars and phosphates form the phosphodiester linkages which are considered the backbone of nucleic acids. Figure made using PDB file 2G91 in PyMol

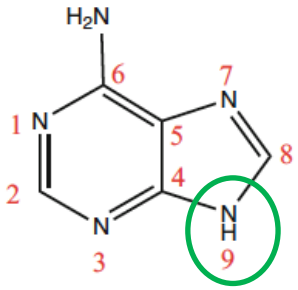
Each phosphate group has one **negative charge**. Nucleic acids have an overall negative charge due to the charges on the backbone phosphates (Fig. 1.2).

RNA: composition and base pairing

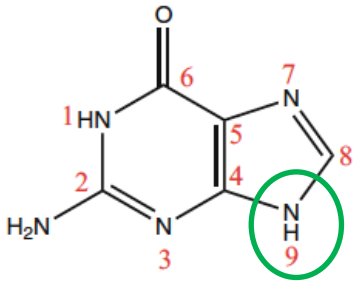
The **nucleobases, or bases**, contain nitrogen in the heterocyclic aromatic ring structures and are basic. The bases are not directly linked to each other but via the sugar connected to phosphate. A base is attached at **1' position on the sugar** through a C1'-N linkage (Fig. 1.2).



Purine

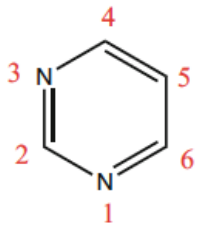


Adenine, A

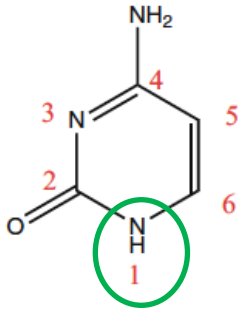


Guanine, G

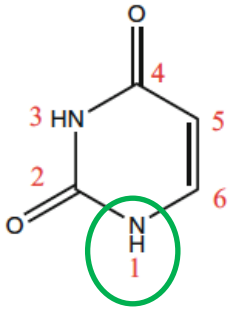
Nucleobases bind to sugars through the nitrogen at position **9** in purines, and position **1** in pyrimidines.



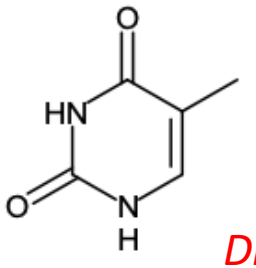
Pyrimidine



Cytosine, C



Uracil, U



thymine

DNA uses T!!!

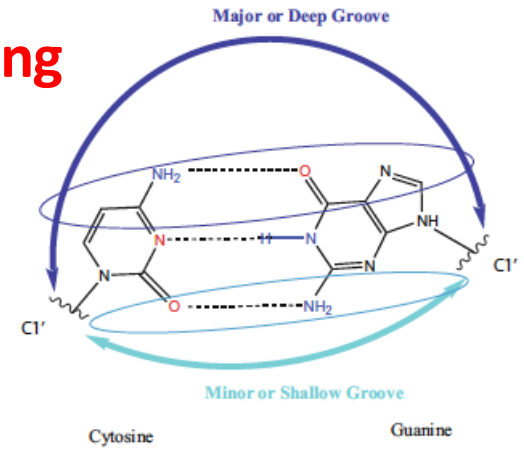
Fig. 1.3 Nucleobases in RNA. The purine (R) bases are guanine (G) and adenine (A). The pyrimidine (Y) bases are cytosine and uracil

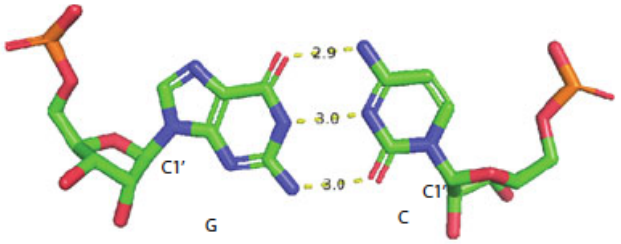
RNA: composition and base pairing

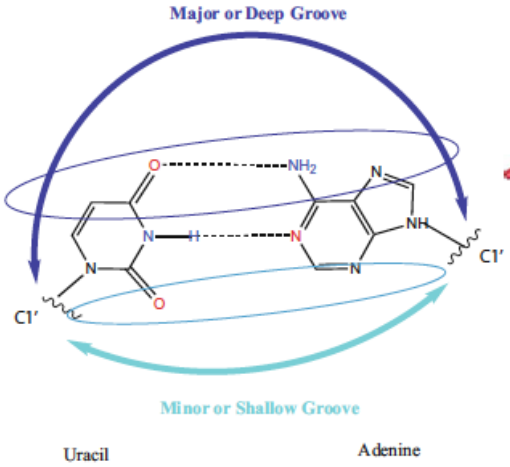
Canonical RNA base pairing

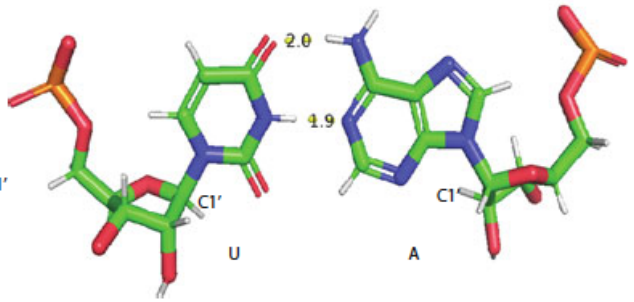
These two base pairs occupy space in an identical manner and are said to be **isosteric**.

The four combinations (G-C/C-G/A-U/U-A) can substitute for each other to form the **regular A-form** helical structures in RNA without distortions.







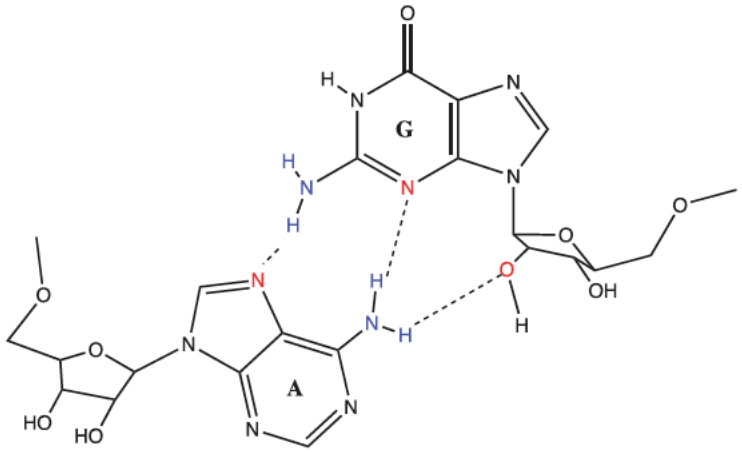


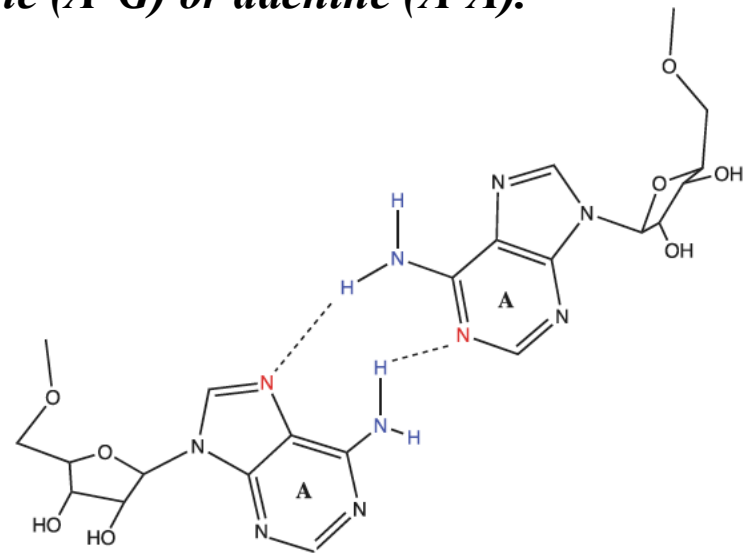
RNA: composition and base pairing

Non-Canonical RNA base pairing

In RNA, many different non-canonical interactions are possible between bases

Example: ***Adenine can hydrogen bond to guanine (A•G) or adenine (A•A).***



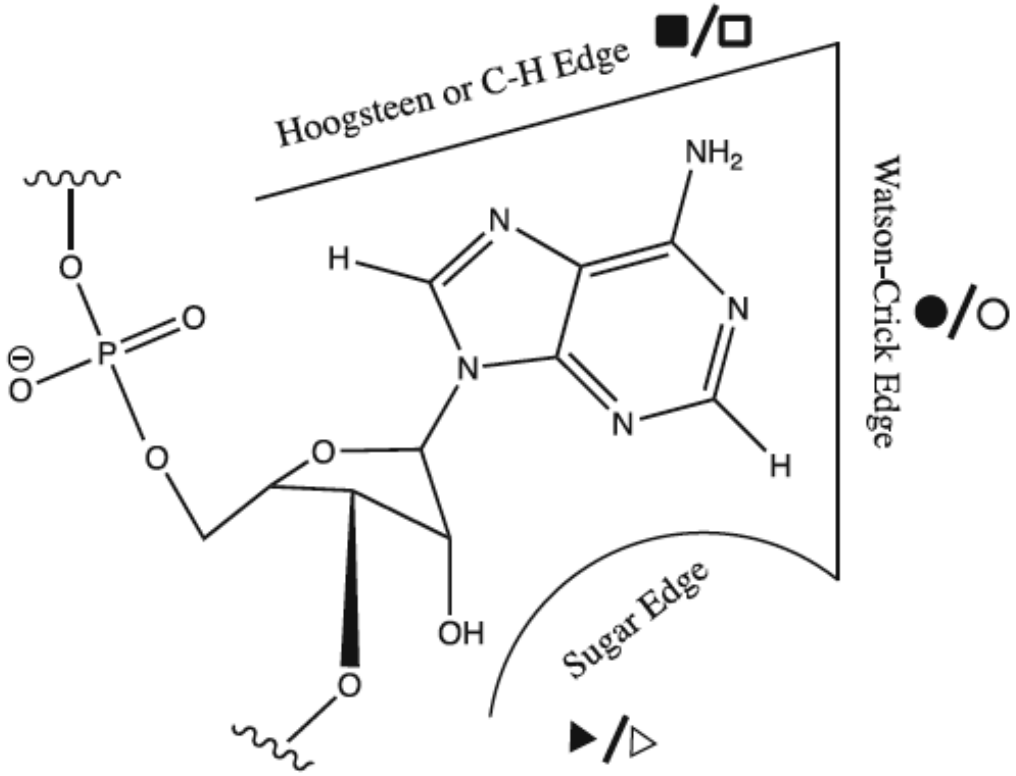


Some **commonly seen noncanonical base pairs** are:
Adenine•Cytosine; Adenine•Guanine; Adenine•Adenine; Uracil•Guanine; Uracil•Cytosine; Uracil•Uracil.

Noncanonical base pairings do not all occur at the same frequency within RNA molecules. **The most common noncanonical base pairs are the G•U and G•A**.

RNA: composition and base pairing

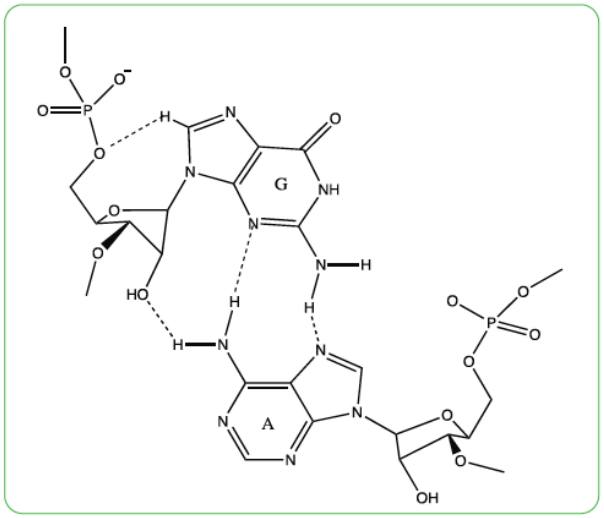
Non-Canonical RNA base pairing



A nucleobase presents **three distinct edges** with which it can form hydrogen bonds. These are designated as: the Watson–Crick (WC,W or canonical) edge, the Hoogsteen edge (HE, H), and the sugar edge (SE, S). **6 combinations** of edge-to-edge interactions!

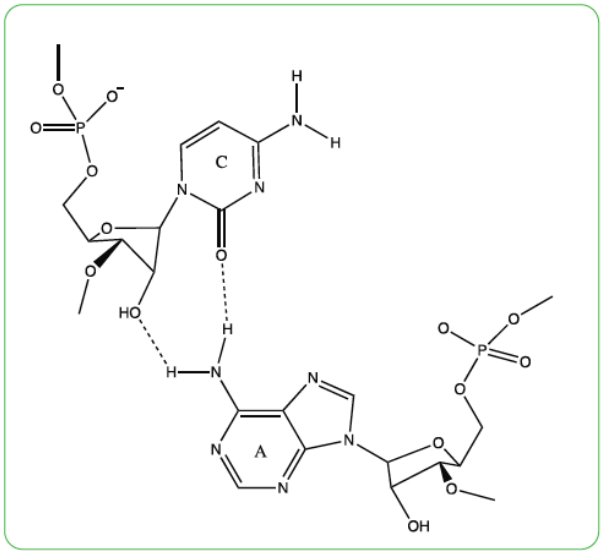
RNA: composition and base pairing

Non-Canonical RNA base pairing



A is using the Hoogsteen edge and **G** the sugar edge.

A-G and A-C non-canonical base pairs are isosteric.



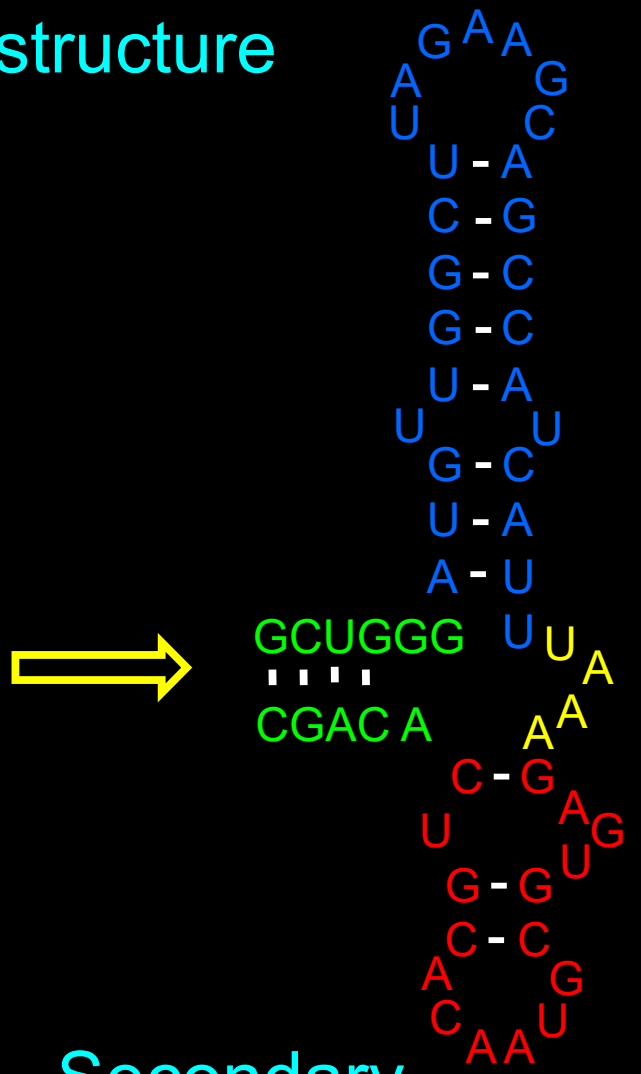
A is using the Hoogsteen edge and **C** the sugar edge.

Architecture of RNA

Nucleic acid structure

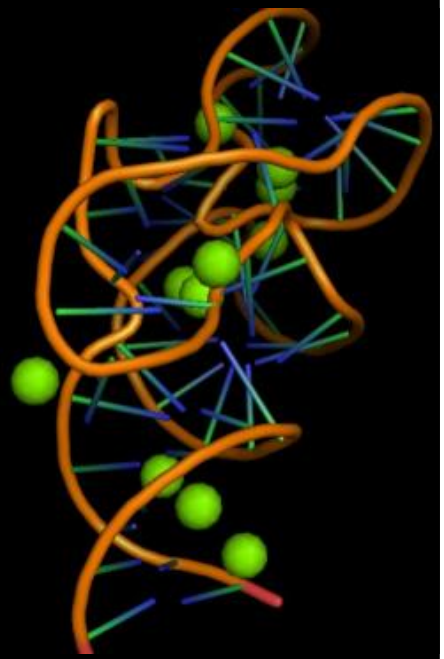
GCU GGG AUG
UUG GCU UAG
AAG CAG CCA
UCA UUU AAA
GAG UGC CGU
ACA GCU CAC
AGC

Primary



Secondary

Higher order



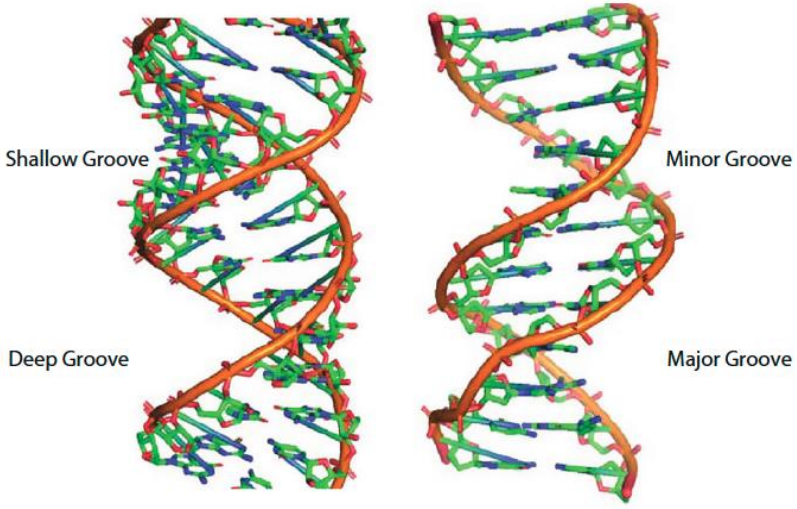
Architecture of RNA

Helix

RNA forms the **A-form** helical structures which are different from the **B-form** helical structures in DNA in significant ways.

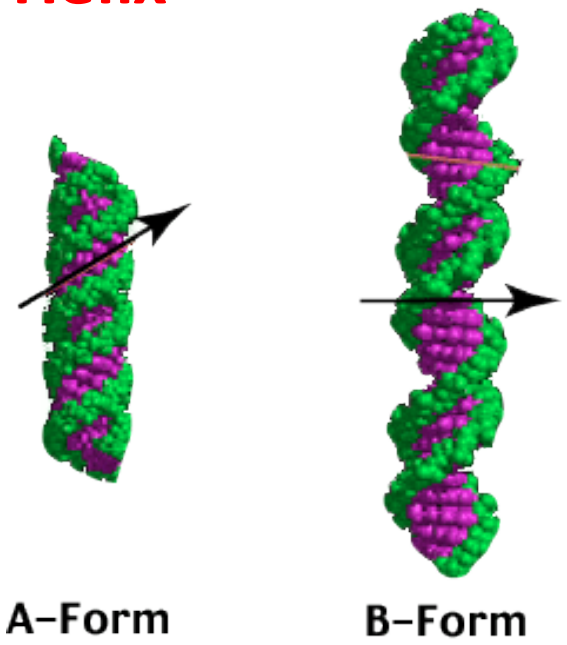
Differences between the A-and B-helical forms occur due to
- the differences in sugar pucker
- additional hydrogen bonding due to the 2'-OH in RNA.

Both A- and B-form helices are *right-handed* and are stabilized by *stacking and hydrogen bonding* interactions between bases to form base pairs.



Architecture of RNA

Helix



	A-form helix	B-form helix
Handedness	Right	Right
Base pairs per turn	11	10
Rotation/base pair (twist angle)	33°	36°
Inclination of base pair to the axis	+19°	-6°
Rise/base pair	2.6 Å	3.4 Å
Pitch/turn	28.2 Å	33.2 Å
Glycosyl angle	Anti	Anti
Sugar pucker	C3'-endo	C2'-endo
Diameter	23 Å	20 Å

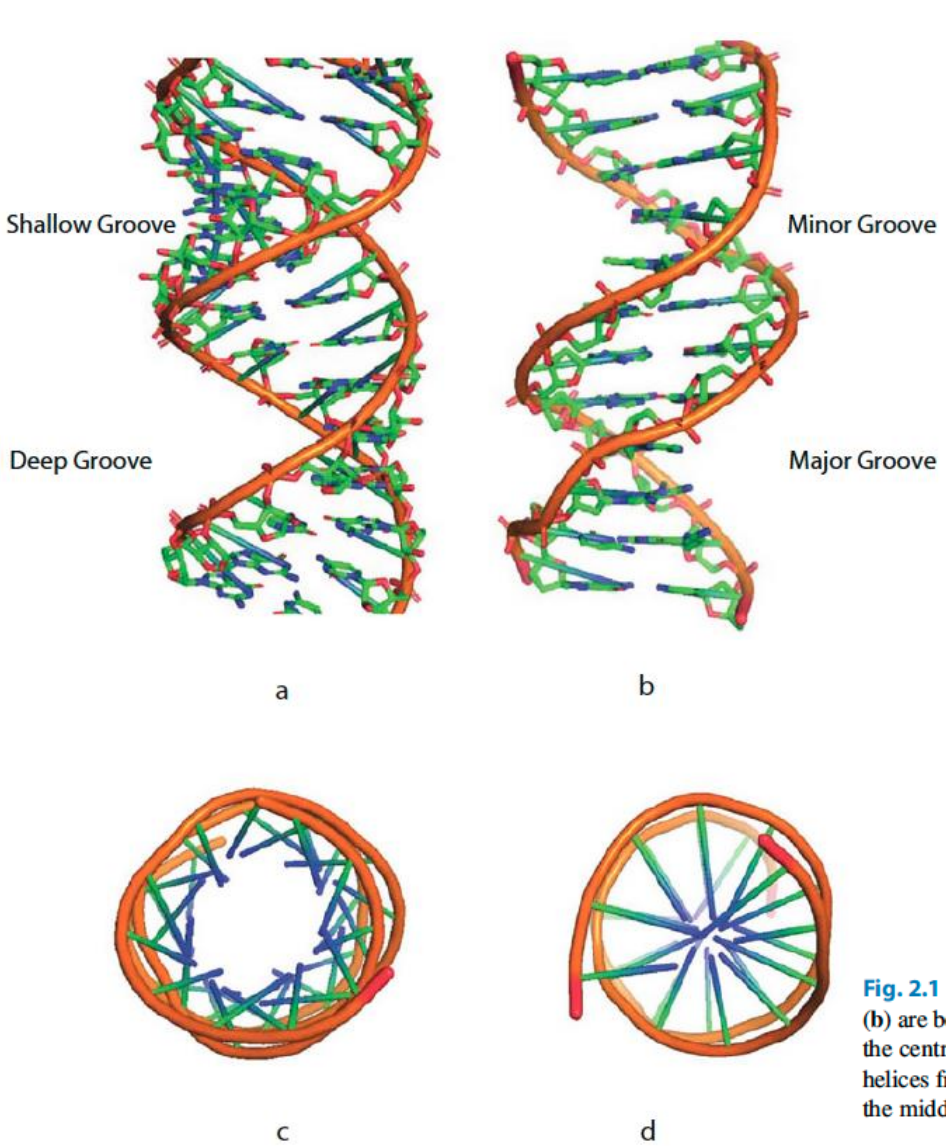
A RNA single strand folds to form short A-form helical regions interspersed by other structural elements. DNA helix is mostly formed by two complementary strands coming together to form a continuous B-form helix. RNA helix is shorter, wider, and more compact, with base pairs that are titled relative to the central axis

An A-form RNA helix is **23 Å in diameter with 11 base pairs per helical turn**; each turn is 28 Å; the bases are at a **~ 19° angle to the helical axis** (tilted steps on a ladder) with 5.9 Å phosphate-to-phosphate distance (7 in the B-form) due to a **more rigid 3'-endo** conformation of the sugars.

Helical regions in RNA secondary structure are marked by **P** to denote paired regions. Paired regions include canonical or non-canonical base pairs and are also called **stems**.

Architecture of RNA

Helix



A-form helices have a **deep groove** that is narrower than the major groove of B-DNA; it also has a **shallow groove** that is wider than the B-DNA's minor groove.

In A-form helices, the richness of the deep groove hydrogen bond donors and acceptors is inaccessible to proteins. **The narrow groove provides proteins access** to the base ligands from the sugar edge side.

Fig. 2.1 A-form RNA and B-form DNA helices. The RNA A-form helix (**a**) and DNA B-form helix (**b**) are both right-handed helices. The number of base pair per turn, the angle of base pairs relative to the central helical axis, and the grooves are different in RNA and DNA helices. Looking down the helices from the top—A-form in (**c**) and B-form in (**d**)—show the differences in base pair packing in the middle of the helices

Architecture of RNA

Non-helical regions

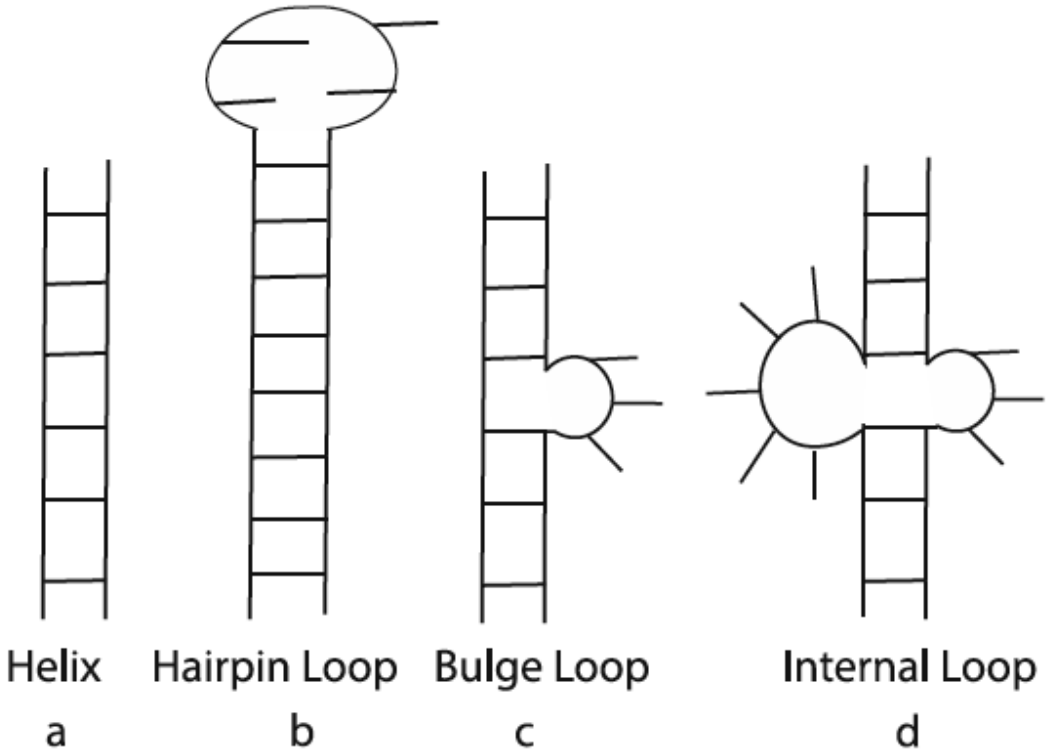
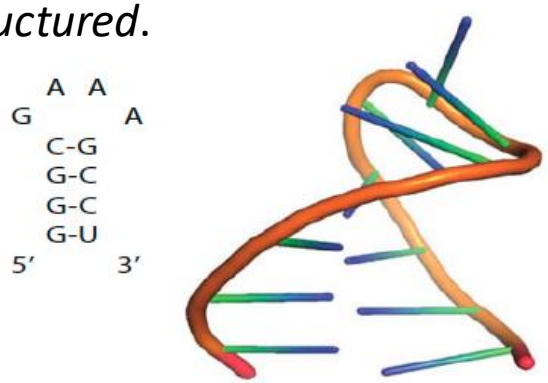


Fig. 2.2 Some common secondary structures in RNA. (**a**) The double helix is the region of the standard A-form helix. When it contains non-canonical base pairs as well, it is referred to as paired region or stem (P). (**b**) The hairpin loops (L) are the turning points in the RNA strand that allows for double-helical structures to form a stem. (**c**) The bulge loops form when one strand is longer than the other. A three-nucleotide bulge loop is shown above. (**d**) The internal loops form when both strands have noncanonical base pairs in a helix; these loops can be either symmetric (n x n) or asymmetric (n x m) based on the number of bases on each strand. An asymmetric 4 x 3 internal loop is shown above. Both bulge- and internal-loop are also two-way junctions (J)

Architecture of RNA

Hairpin Loops/Stem Loops (SL)

Hairpin loops or stem loops (SL) form when **the single stranded RNA folds upon itself** using a hairpin turn to form a short loop. The nucleotides in the hairpin loop form *noncanonical interactions and are often structured*.



The loops can be of various sizes. **Four nucleotide loops** (tetraloop) occur more frequently. Among all tetraloops, **GNRA** occur more often in RNA (N: Any nucleotide; R: Purine). Different size hairpin loops may adopt similar 3D structures.

The hairpin loops are involved in many long-range interactions important for RNA folding and its stability. Stem loops play an important role in **interactions with proteins** and influence diverse biological processes, including rates of transcription, transcription termination, and susceptibility to nucleases.

Architecture of RNA

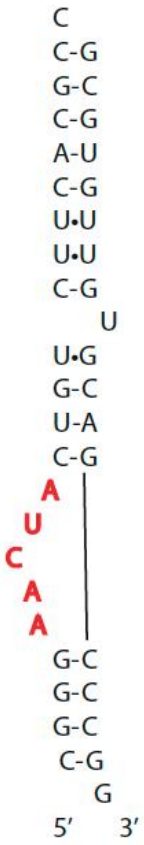
Bulges

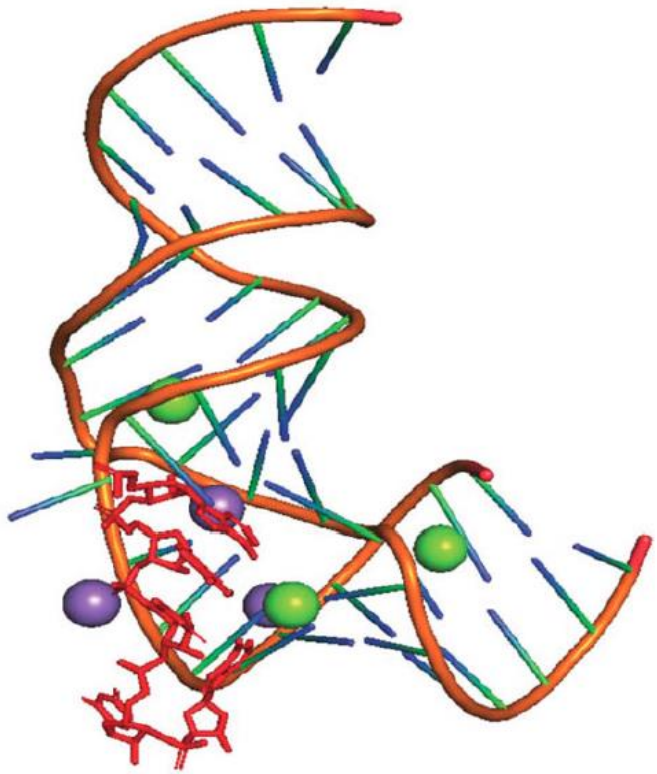
A bulge loop is an **asymmetrical loop where one strand contains additional bases**.

- **nucleotides can be bulged out of the helix**
The overall geometry remains close to an A-form helix, thus increasing RNA's stability by co-axial stacking

- **nucleotides can be stacked within the helix**
When the loop nucleotides are stacked within the helix, it causes the RNA to bend.

- Additional interactions
- altered dimensions
- recognition site for proteins





Metal ions stabilize different conformations

Architecture of RNA

Internal loops

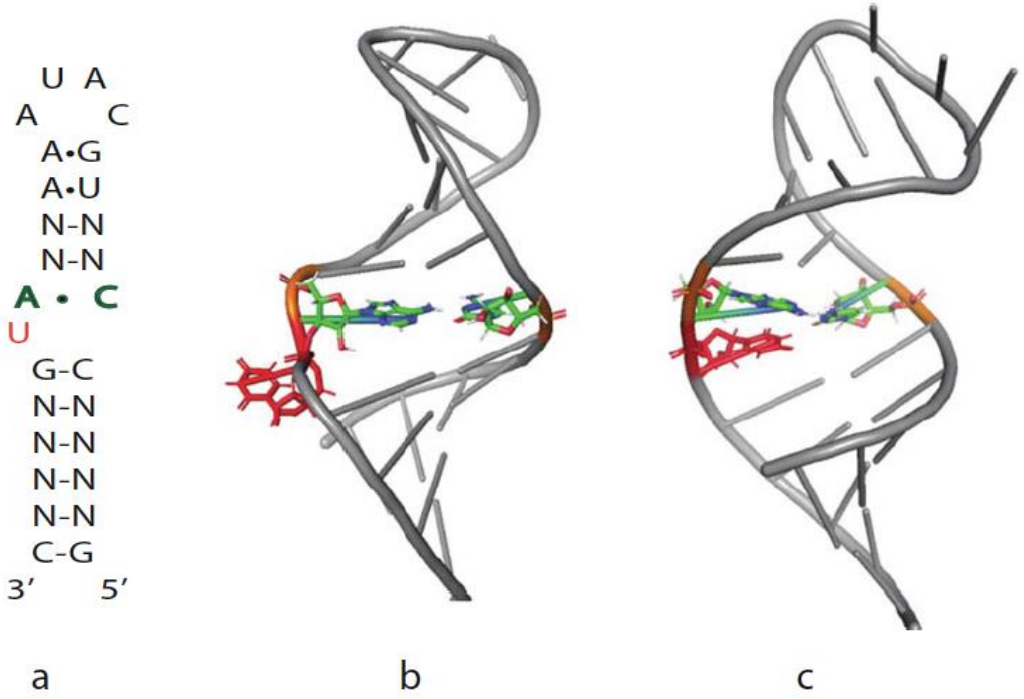


Fig. 2.6 Asymmetrical Internal Loop. (**a**) The secondary structure of the U6 spliceosomal RNA has a 1 x 2 internal loop. This RNA adopts (at least) two different conformations in two different pHs (5.7 and 7). (**b**) In the NMR structure of U6 at pH 5.7, the A•C pair forms and U (red) is unstacked out of the helix. (**c**) In the NMR structure at pH 7, all the nucleotides of 1 x 2 internal loop are stacked into the helical structure, bending the top stem (note the position of the top loop in **b** versus **c**). Figure made using PDB files 1SYZ and 2KF0 using PyMol

The non-helical secondary structural elements change the backbone of RNA in unique ways, and hence, serve as **recognition sites for interactions** between RNA–RNA, RNA–protein, and RNA–small molecules.

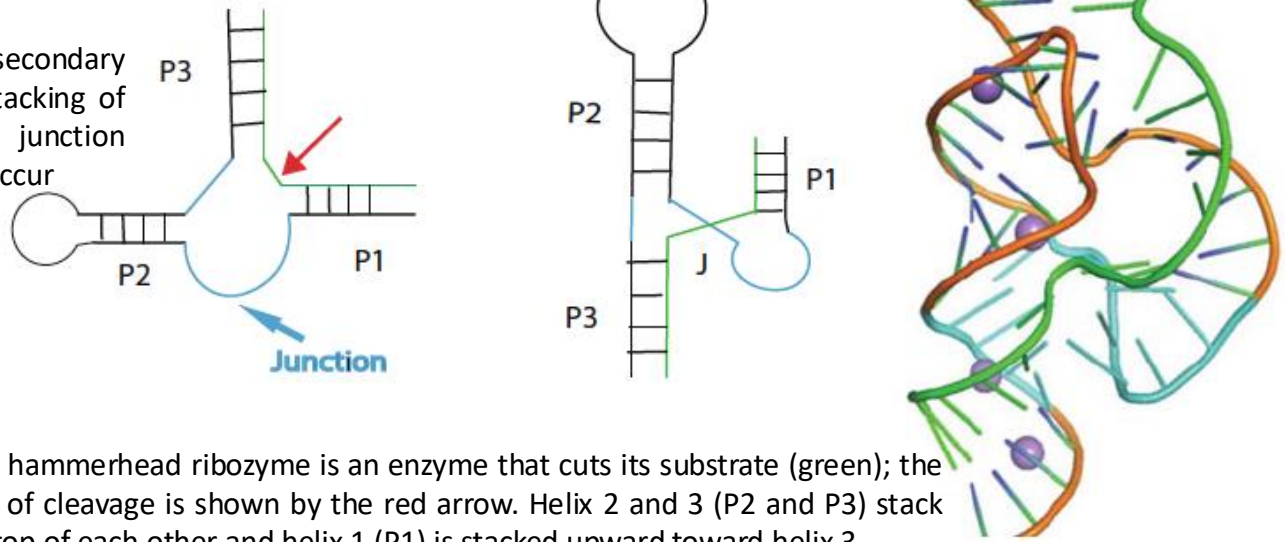
Architecture of RNA

Junctions

A junction is **a region where multiple helices intersect**; the junction region is marked by J and the numbers correspond to the paired regions that link to it. **Internal loop and bulge** loops are two-way junctions. **Higher order junctions** are seen in many large RNA structures, where three, four, or five helices are seen connected to a junction region.

The secondary structures are rewritten when the 3D structural information (b) becomes available

Originally proposed secondary structure (a) before stacking of helices around the junction region was known to occur



The hammerhead ribozyme is an enzyme that cuts its substrate (green); the site of cleavage is shown by the red arrow. Helix 2 and 3 (P2 and P3) stack on top of each other and helix 1 (P1) is stacked upward toward helix 3

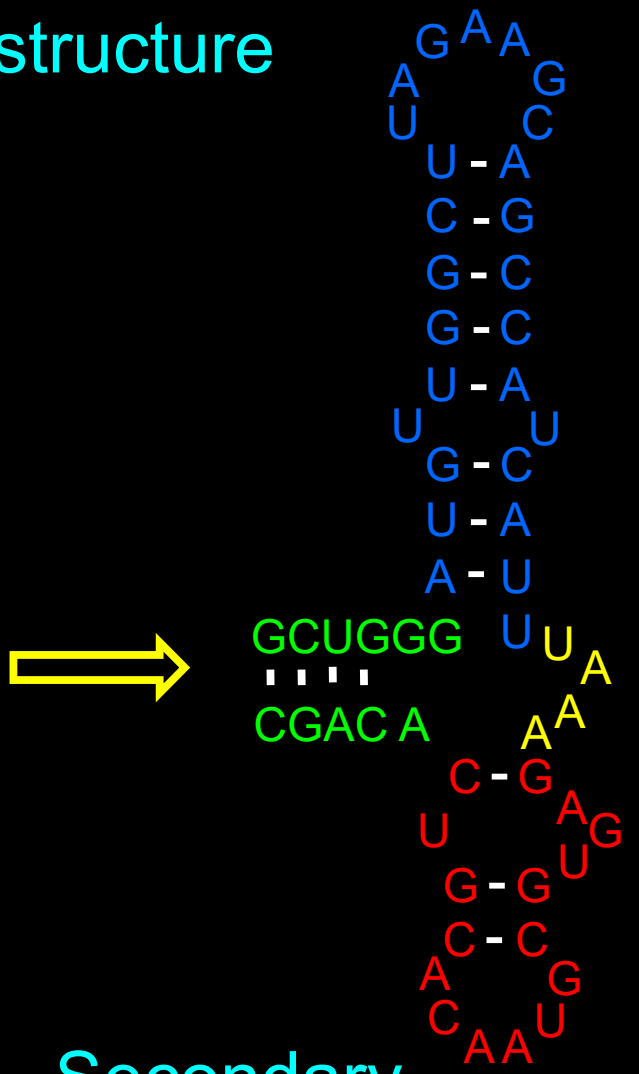
a b c

Architecture of RNA

Nucleic acid structure

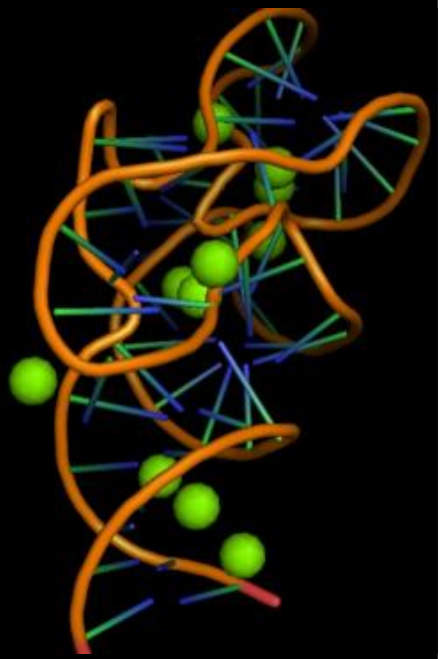
GCU GGG AUG
UUG GCU UAG
AAG CAG CCA
UCA UUU AAA
GAG UGC CGU
ACA GCU CAC
AGC

Primary



Secondary

Higher order

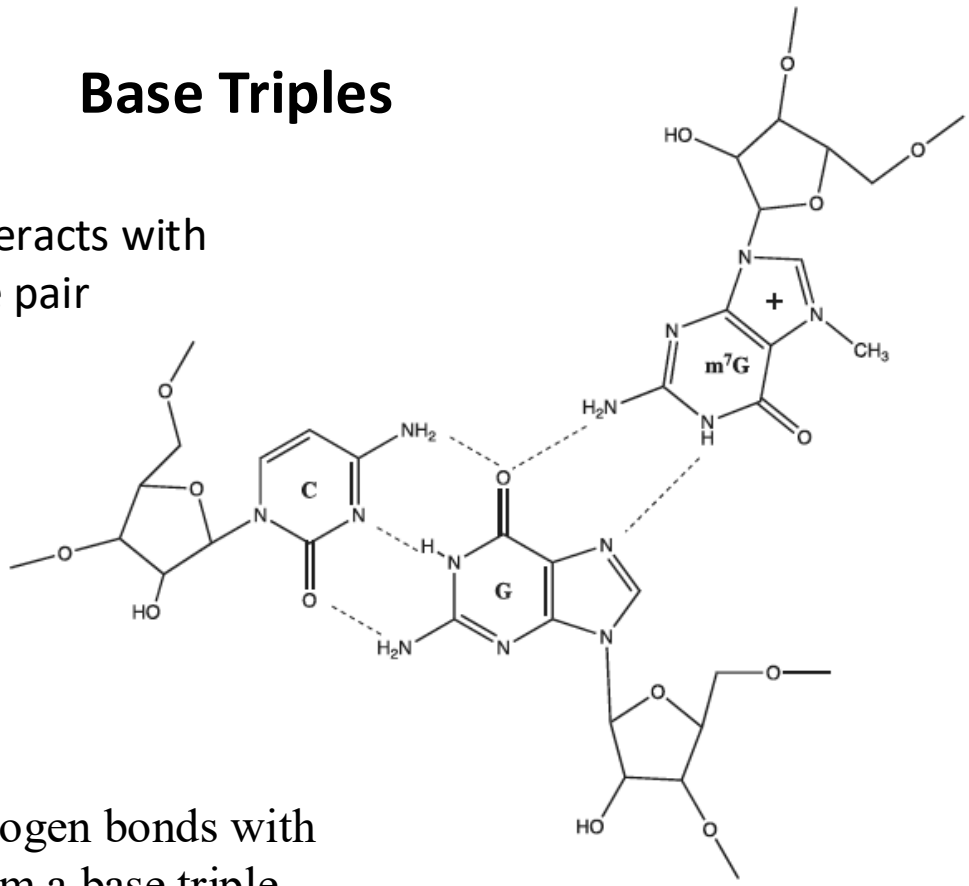


Architecture of RNA

Ternary interactions

Base Triples

Base triples form when a base interacts with an already existing canonical base pair



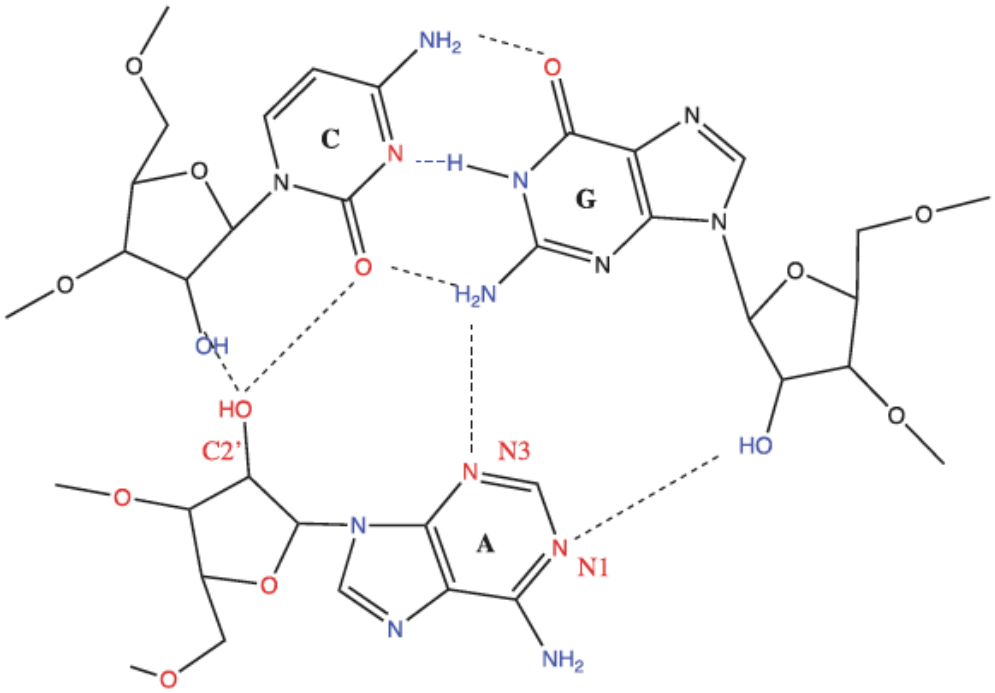
A C-G canonical base pair hydrogen bonds with m^7G (N7methyl guanine) to form a base triple.

Watson-Crick + Hoogsteen hydrogen bonds

Ternary interactions

A-Minor interactions

Adenines can insert themselves into the shallow groove of another base pair



When the unpaired **adenosines utilize its O2', N3, and N1** to interact with the shallow (minor) groove of a canonically paired RNA helix, forms an A-minor interaction.

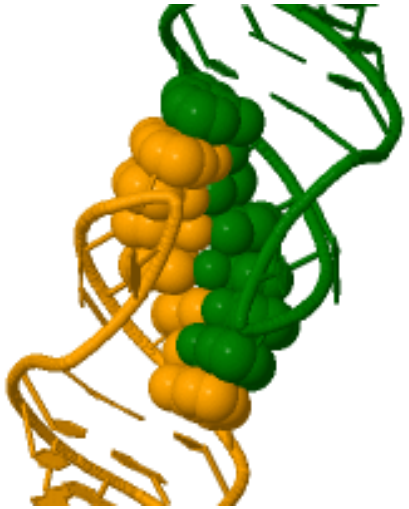
It is a tertiary interaction that allows **packing of the RNA structure**.

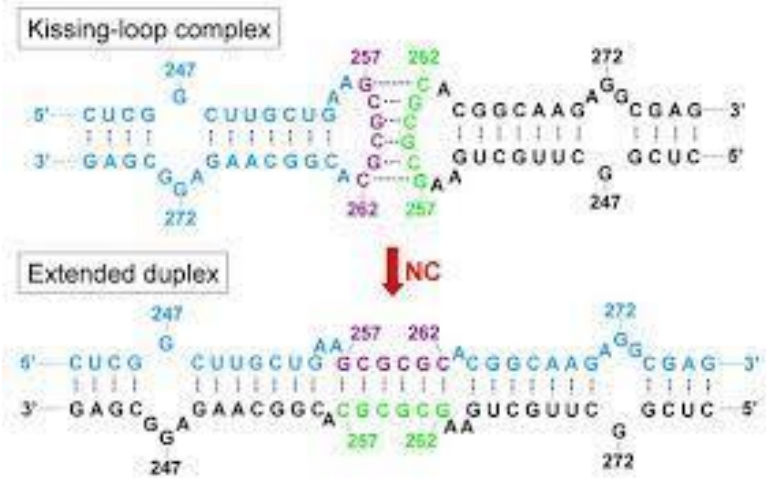
Architecture of RNA

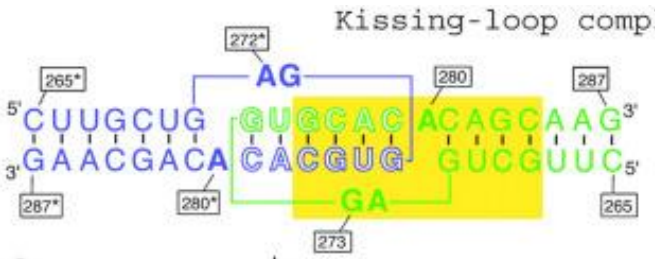
Ternary interactions

Kissing Loops

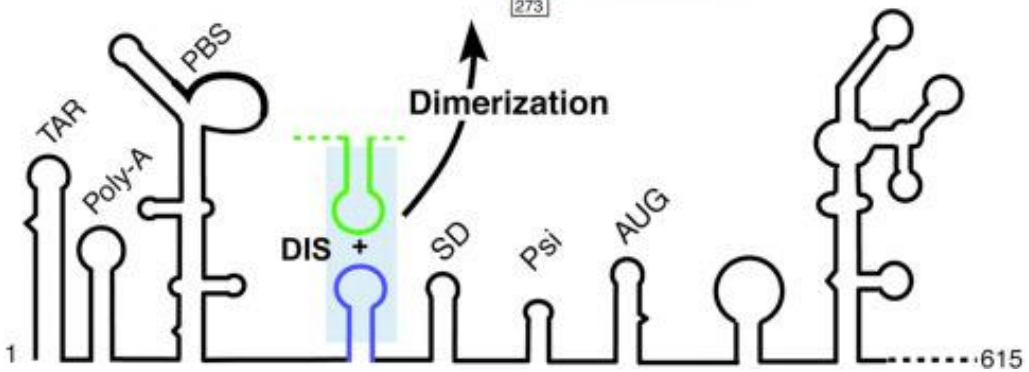
Base pairs form between two loops







HIV-1 DIS Kissing Loop



TAR RNA: 5’-GGC AGA UCU GAG CCU GGG AGC UCU CUG CC-3’

Which info can we get from the primary structure????