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



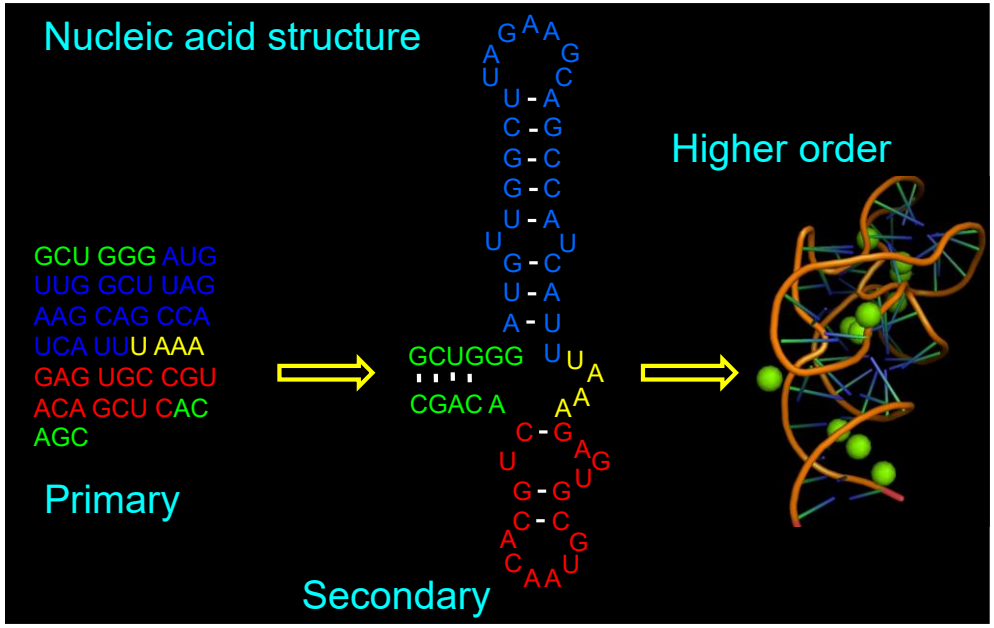
An open mind and the right experiment might just lead you to the next big discovery! So let's get to it!

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

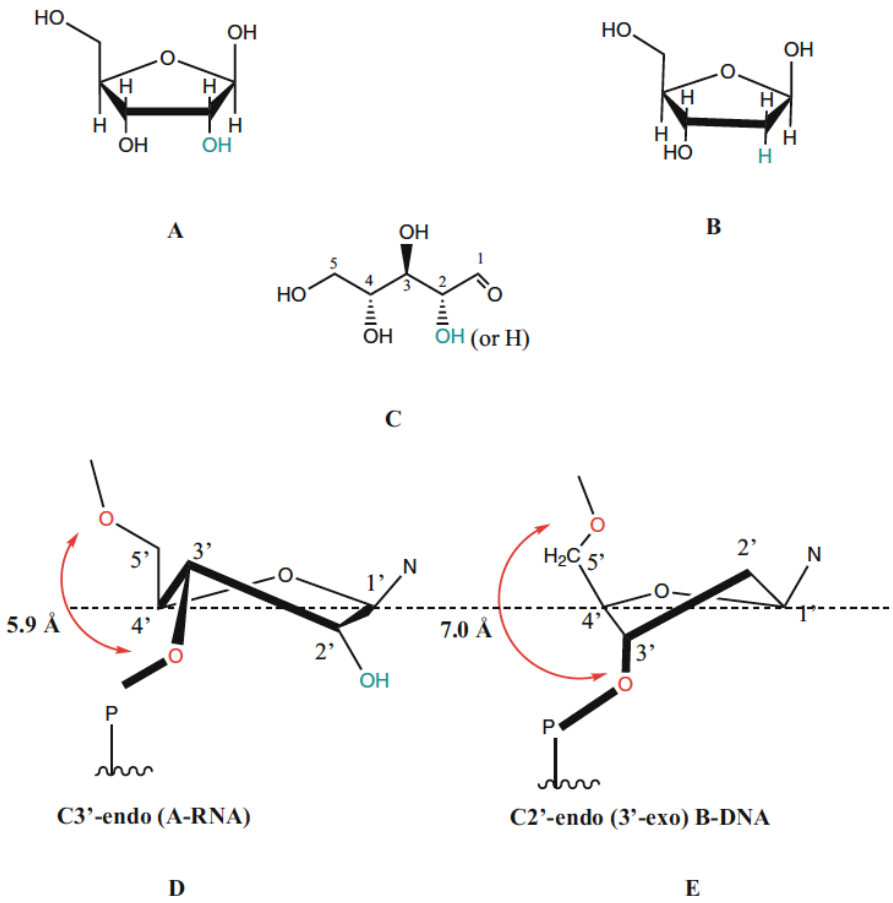


Fig. 1.1 Ribose and 2'-Deoxy-ribose Sugars. Planar ring structures of the ribose (**a**), deoxyribose sugars (**b**), along with the straight chain form of ribose molecule (**c**) are shown. The most stable conformation is C3'-endo for the ribose (**d**) and C2'-endo for the deoxyribose (**e**)

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.

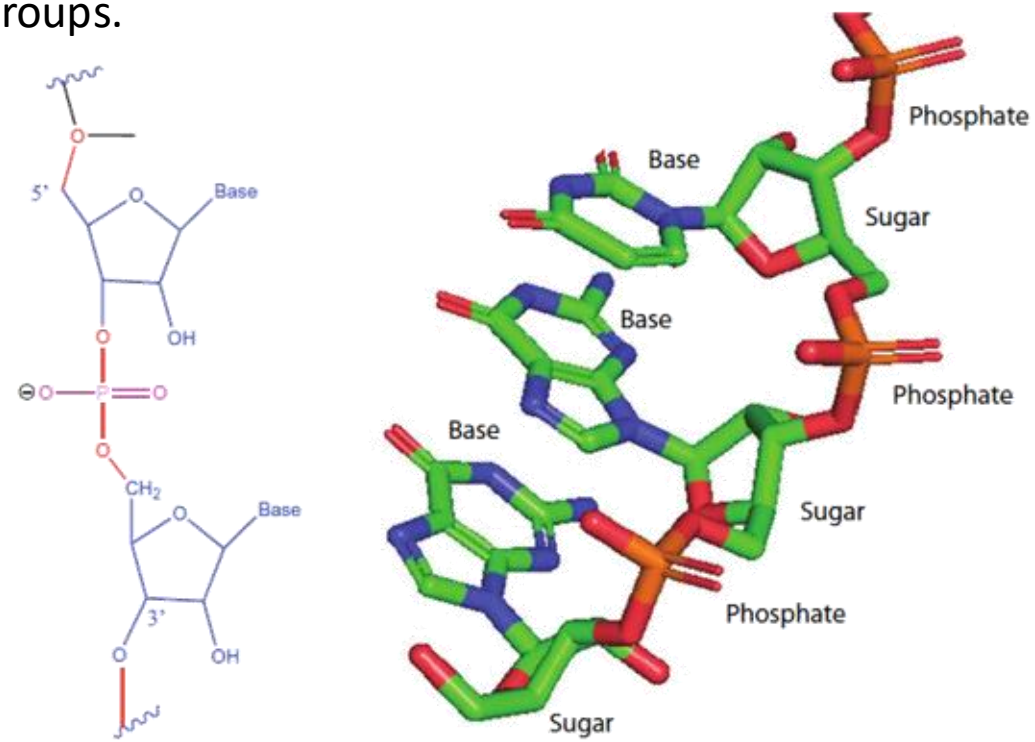
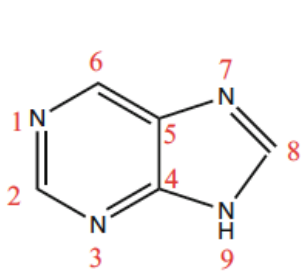


Fig. 1.2 Phosphodiester Linkage. The 5′- and 3′-hydroxyl groups of sugars form ester linkages with the phosphate groups to make the polymer. The crystal structure of RNA (1.5 Å resolution) is shown. The sugars and phosphates form the phosphodiester linkages which are considered the backbone of nucleic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

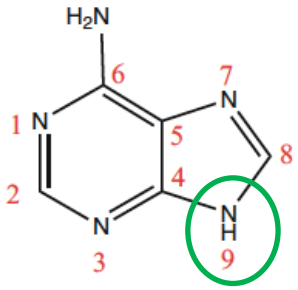
Each phosphate group has one **negative charge**. Nucleic acids have an overall negative charge due to the charges on the backbone phosphates (Fig. 1.2).

RNA: composition and base pairing

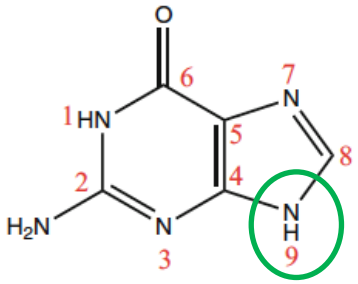
The **nucleobases, or bases**, contain nitrogen in the heterocyclic aromatic ring structures and are basic. The bases are not directly linked to each other but via the sugar connected to phosphate. A base is attached at **1' position on the sugar** through a C1'-N linkage (Fig. 1.2).



Purine

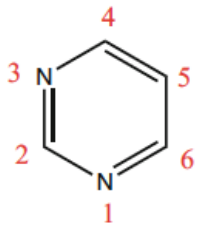


Adenine, A

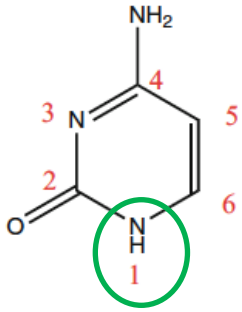


Guanine, G

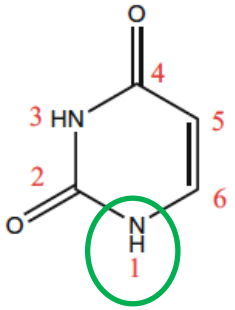
Nucleobases bind to sugars through the nitrogen at position **9** in purines, and position **1** in pyrimidines.



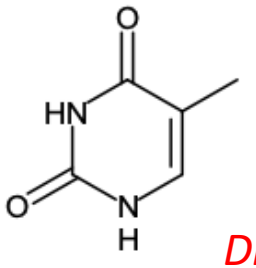
Pyrimidine



Cytosine, C



Uracil, U



thymine

DNA uses T!!!

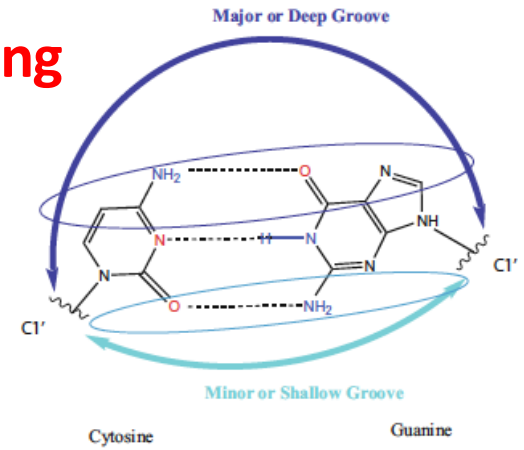
Fig. 1.3 Nucleobases in RNA. The purine (R) bases are guanine (G) and adenine (A). The pyrimidine (Y) bases are cytosine and uracil

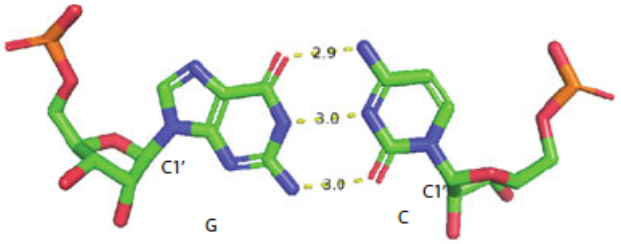
RNA: composition and base pairing

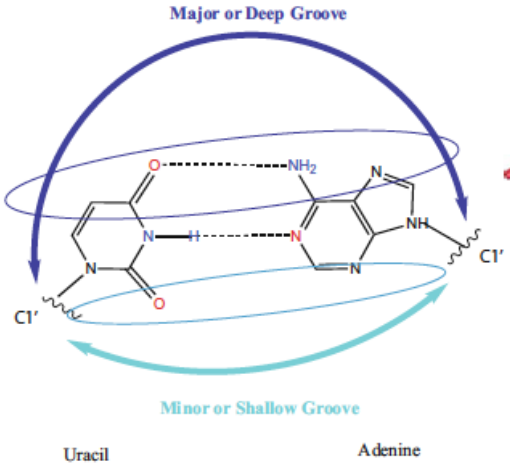
Canonical RNA base pairing

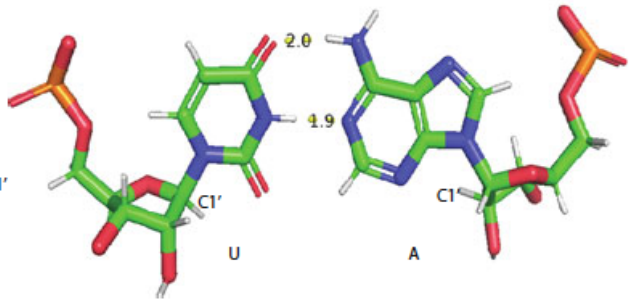
These two base pairs occupy space in an identical manner and are said to be **isosteric**.

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.







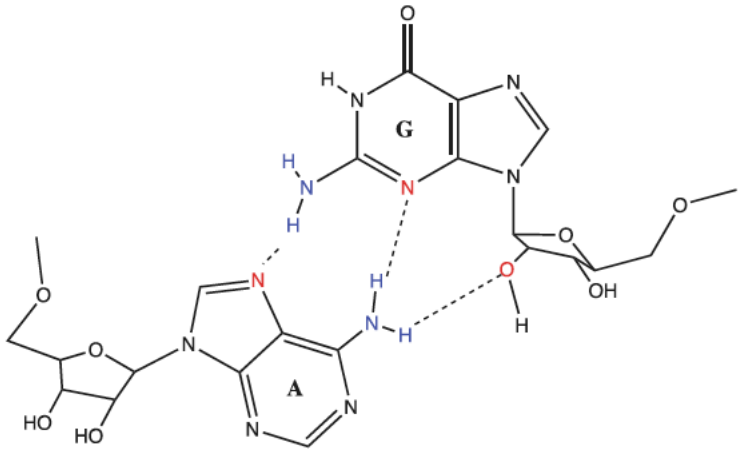


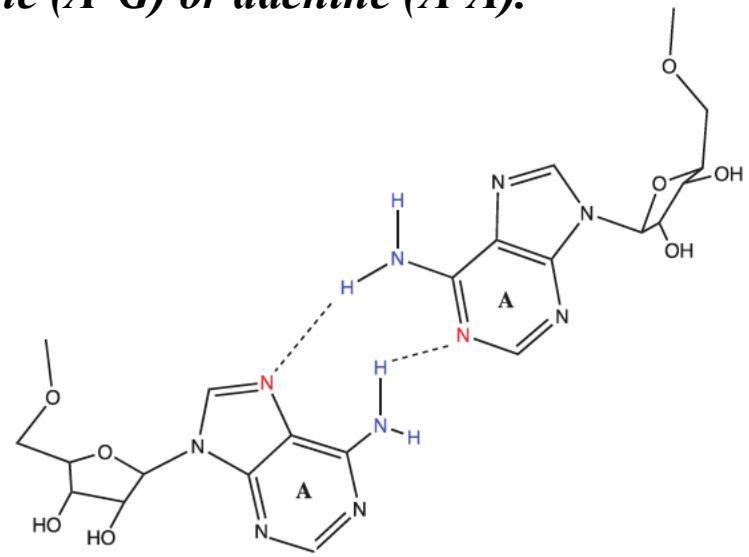
RNA: composition and base pairing

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).***



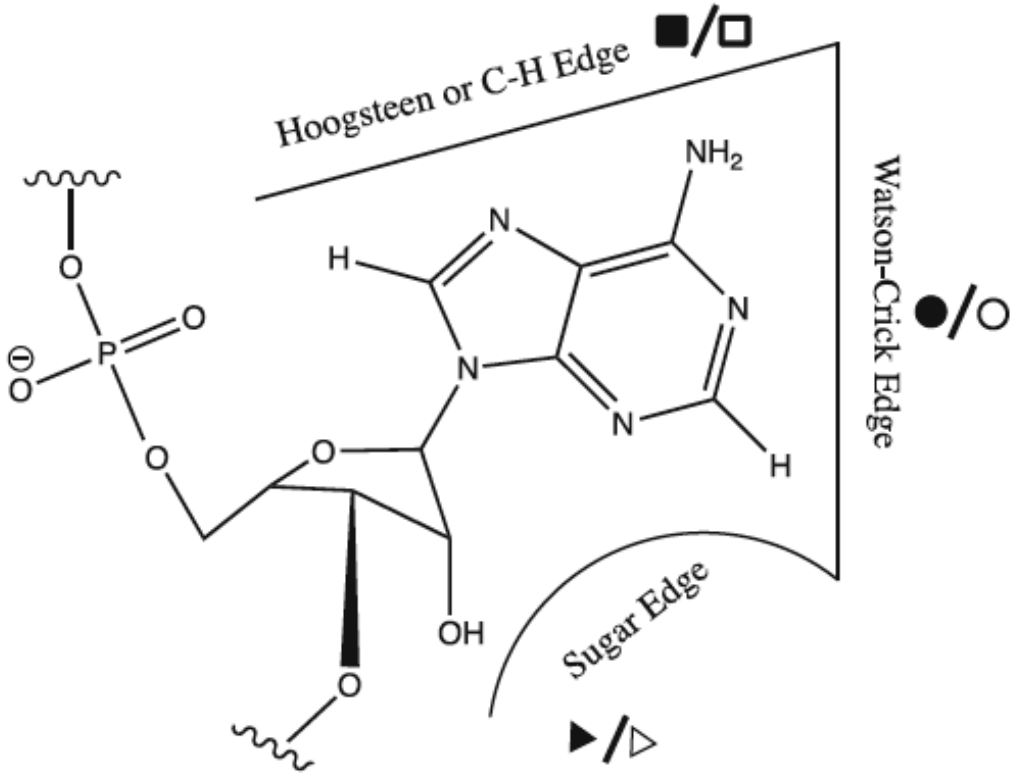


Some **commonly seen noncanonical base pairs** are:
Adenine•Cytosine; Adenine•Guanine; Adenine•Adenine; Uracil•Guanine; Uracil•Cytosine; Uracil•Uracil.

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

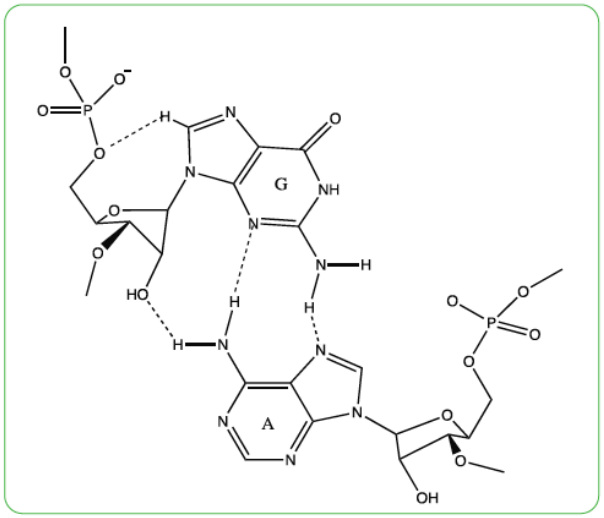
Non-Canonical RNA base pairing



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!

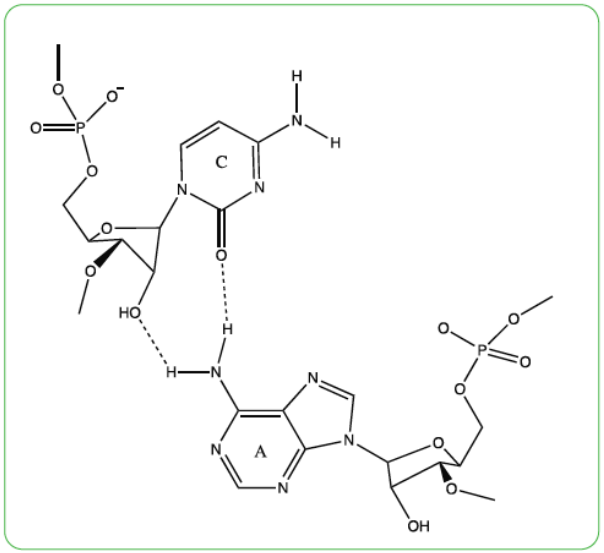
RNA: composition and base pairing

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.



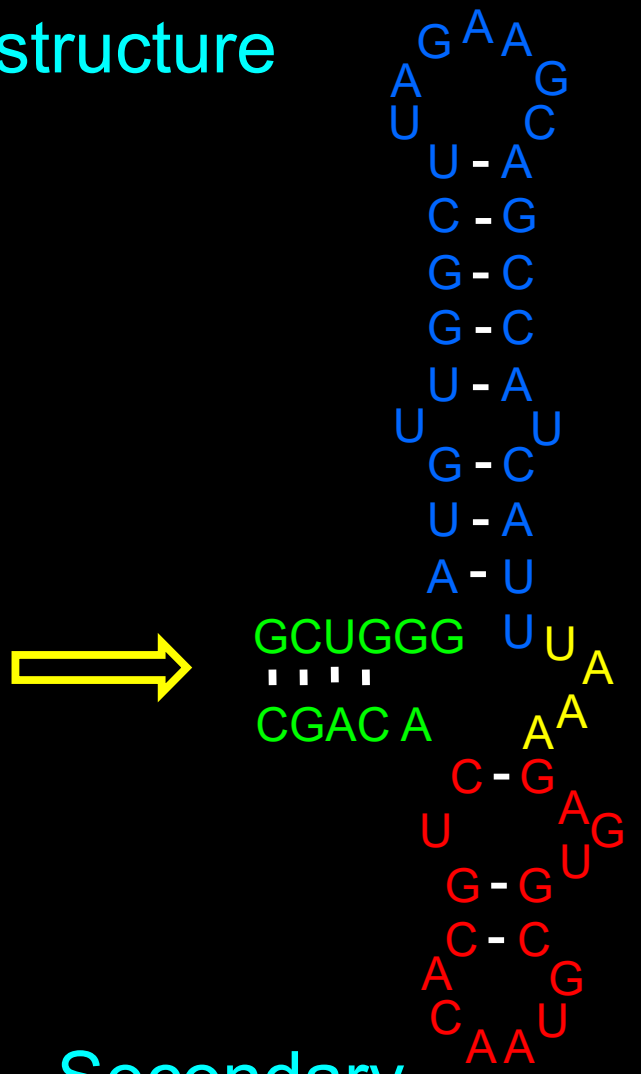
A is using the Hoogsteen edge and **C** the sugar edge.

Architecture of RNA

Nucleic acid structure

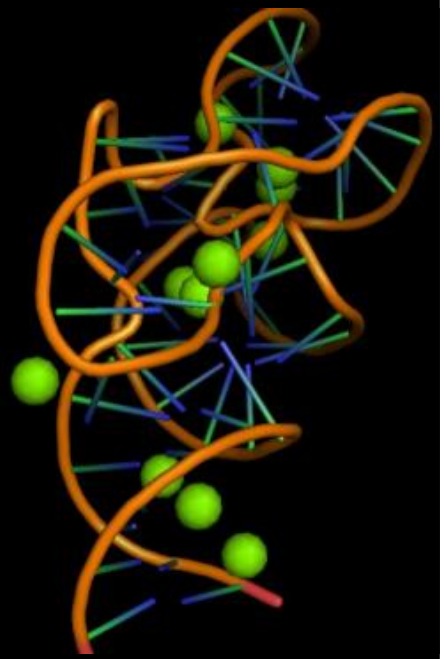
GCU GGG AUG
UUG GCU UAG
AAG CAG CCA
UCA UUU AAA
GAG UGC CGU
ACA GCU CAC
AGC

Primary



Secondary

Higher order



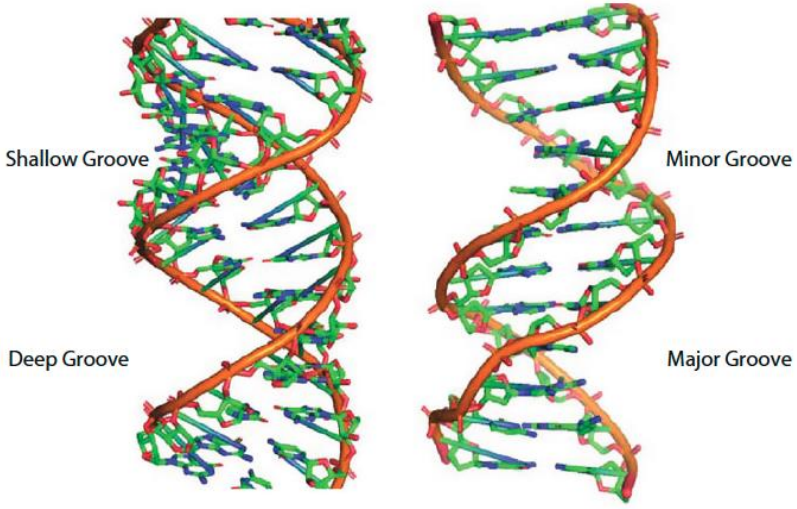
Architecture of RNA

Helix

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

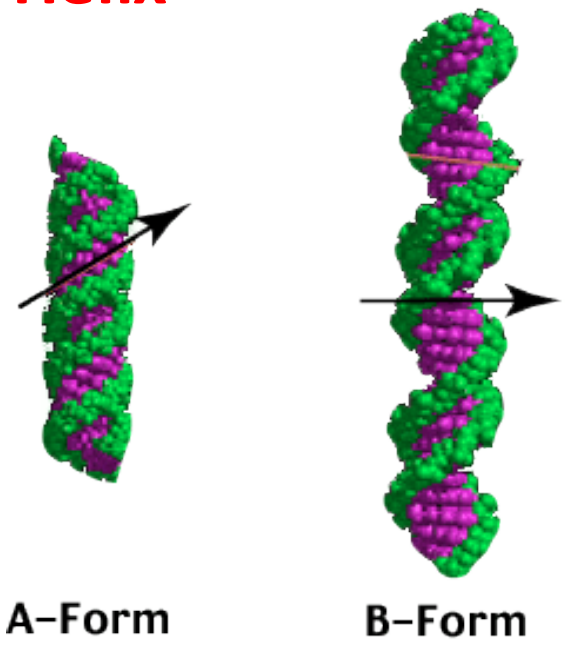
Differences between the A-and B-helical forms occur due to
- the differences in sugar pucker
- additional hydrogen bonding due to the 2'-OH in RNA.

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

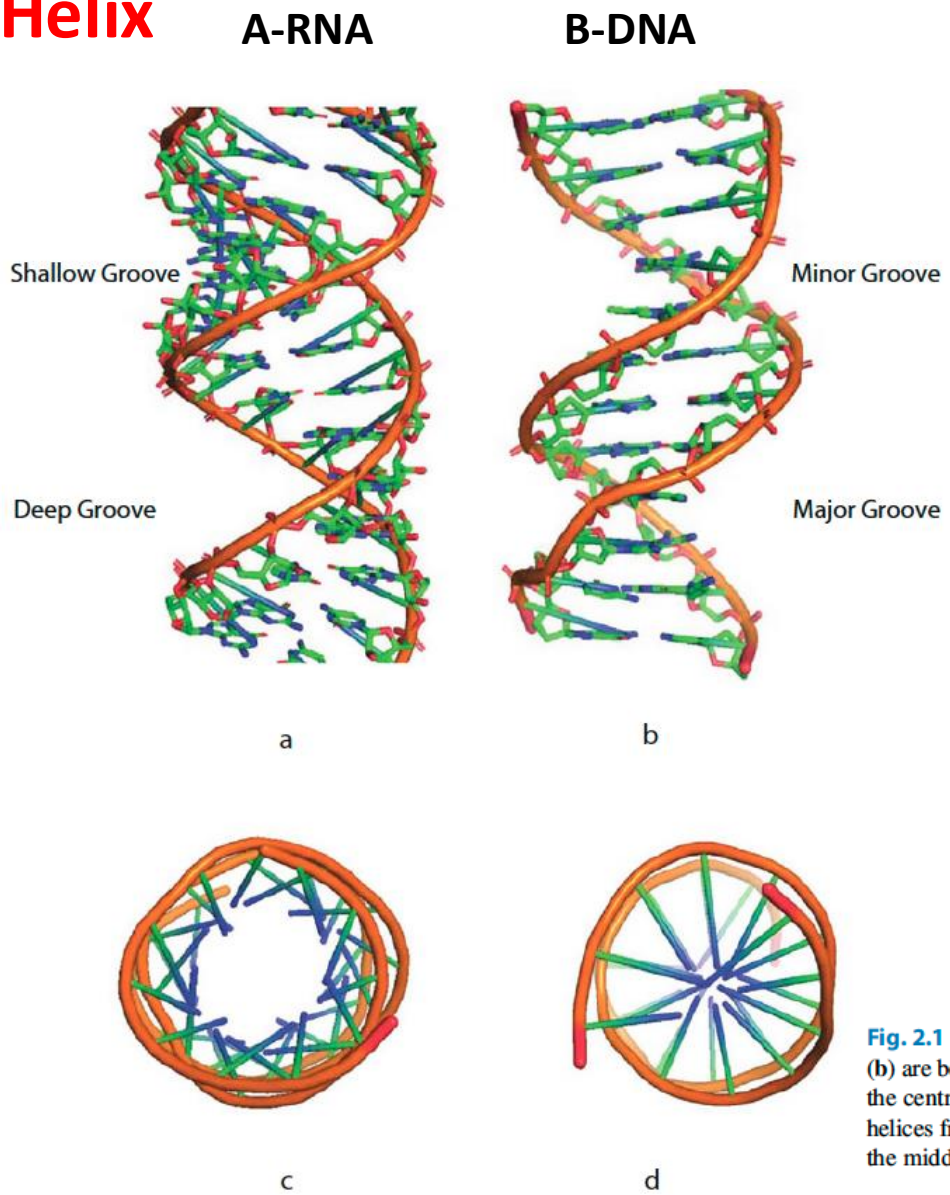
A RNA single strand folds to form short A-form helical regions interspersed by other structural elements. DNA helix is mostly formed by two complementary strands coming together to form a continuous B-form helix. RNA helix is shorter, wider, and more compact, with base pairs that are titled relative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

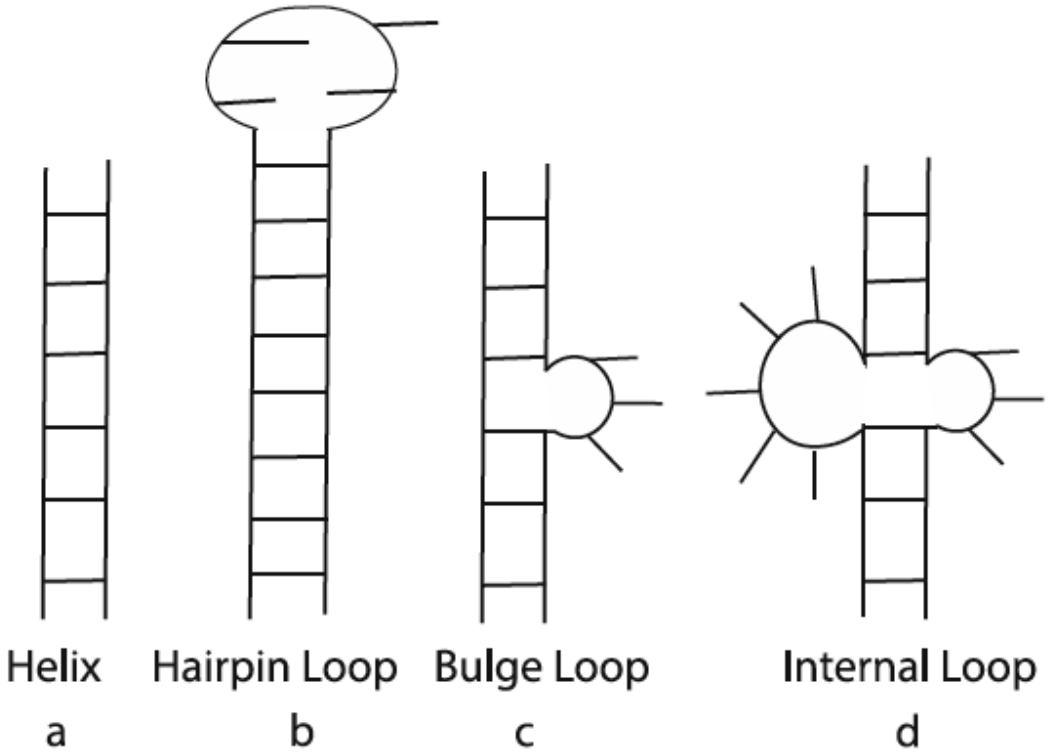
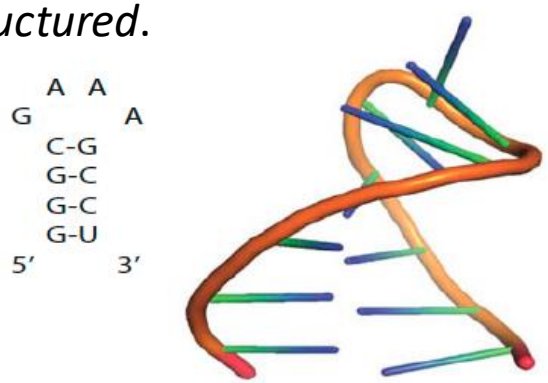


Fig. 2.2 Some common secondary structures in RNA. (**a**) The double helix is the region of the standard A-form helix. When it contains non-canonical base pairs as well, it is referred to as paired region or stem (P). (**b**) The hairpin loops (L) are the turning points in the RNA strand that allows 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

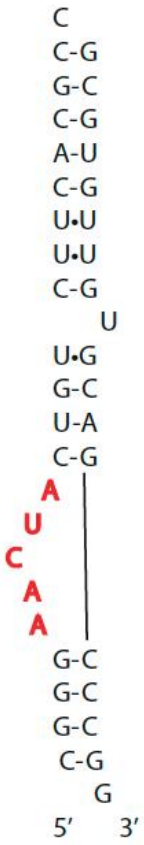
Bulges

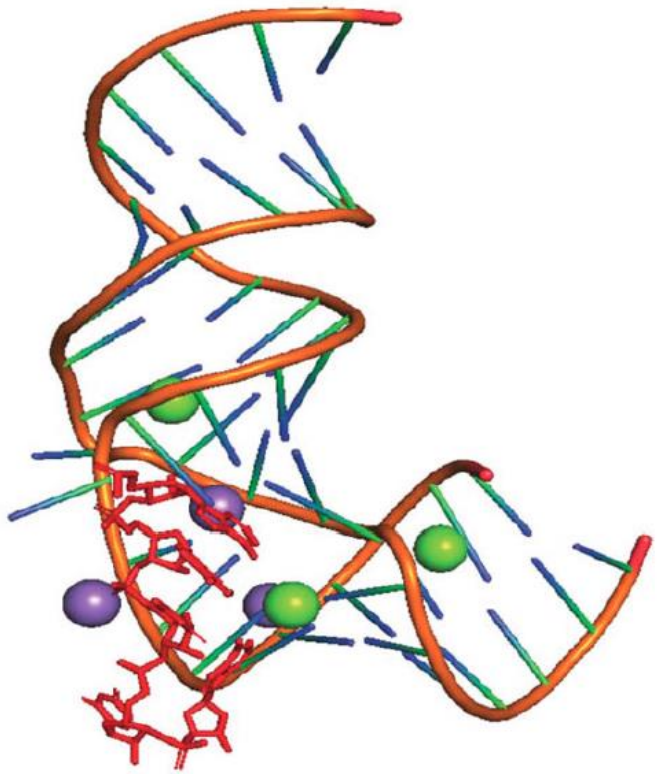
A bulge loop is an **asymmetrical loop where one strand contains additional bases**.

- **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





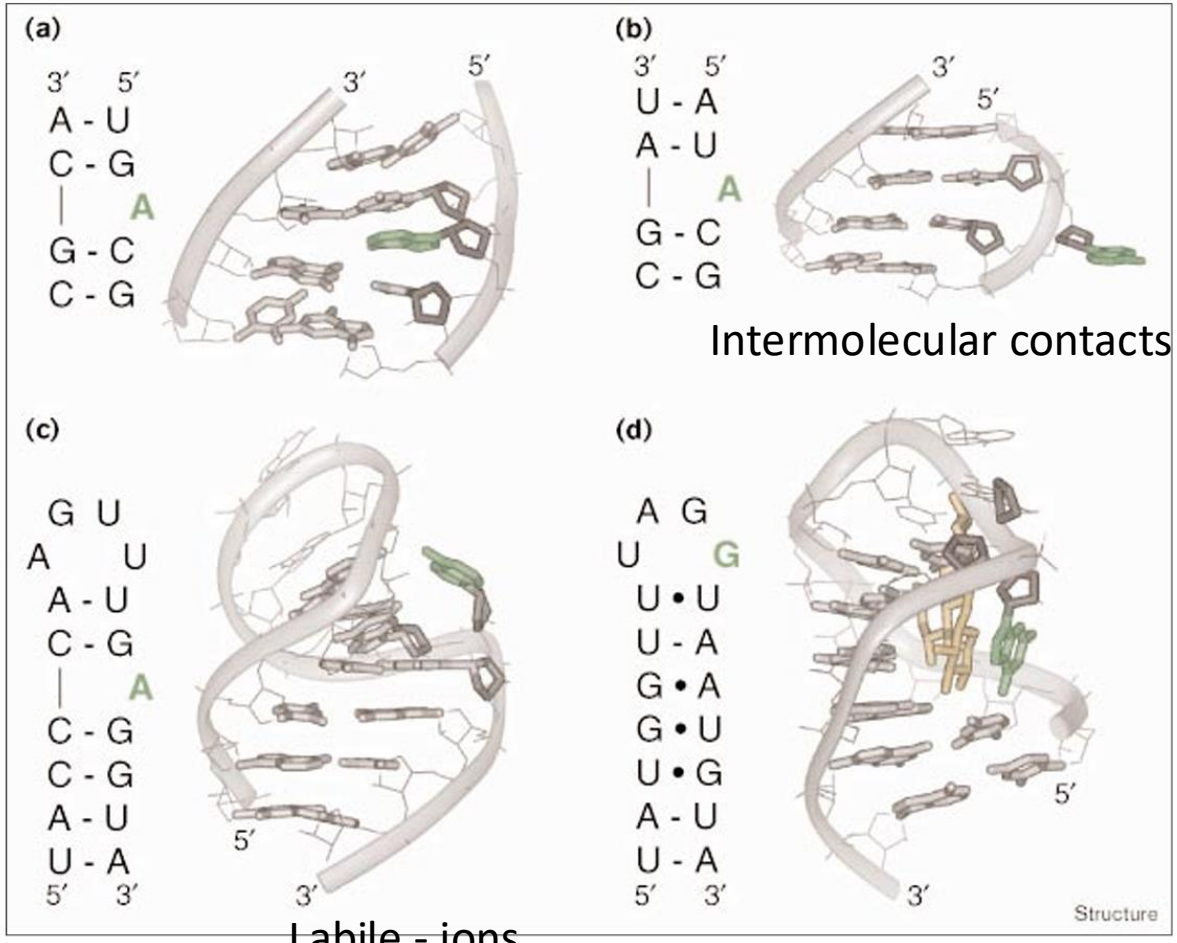
Metal ions stabilize different conformations

RNA bulges as architectural and recognition motifs

Thomas Hermann and Dinshaw J Patel

Bulges

Single unpaired nucleotide



Intermolecular contacts

Labile - ions

Bulges in RNA architectures. In the simplest case, RNA bulges consist of a single unpaired nucleotide within one strand of a base-paired region. **(a)** The bulged residue can be accommodated by stacking between the adjacent base pairs, inducing a considerable kink in the helix axis, as found in the stem of a 24-nucleotide hairpin loop which is the binding site of a bacteriophage coat protein [7]. **(b)** Alternatively, the unpaired nucleotide may be pushed out of the stem, giving rise to a quasi-continuous double helix, as observed in the crystal structure of an A-form duplex containing an adenosine bulge [8]. **(c)** Packing of the unpaired base into one of the helix grooves contributes to bulge stabilization, as seen in a spliced leader RNA hairpin containing a bulged adenosine that packs into the shallow groove of the stem [9]. **(d)** In some RNA complexes with peptides or small molecules, as in the tobramycin aptamer RNA complex shown here [21], bulged nucleotides participate as 'flap' residues closing over the ligand-binding site. Bulged residues are shown in green. In (d), tobramycin is shown in orange.

Architecture of RNA

Internal loops

An **internal loop** is a region of noncanonical base pairing that occurs within the helix; these regions can be **symmetrical or nonsymmetrical**.

The symmetrical loops often form noncanonical base paired regions and cause distortions in the A-form helical backbone.

These ***distortions*** to the A-form structure then serve as recognition sites for proteins.

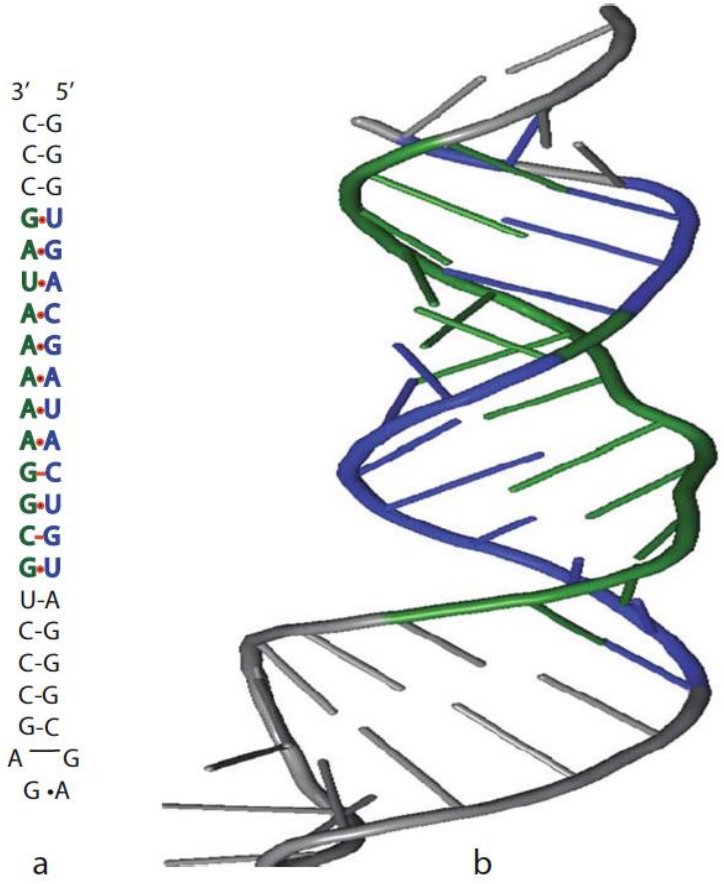


Fig. 2.5 Symmetrical Internal Loop Distort RNA Backbone. (**a**) The secondary structure of loop E motif has several noncanonical base pairs in a symmetrical internal loop. (**b**) NMR structure of loop E forms a distorted helical structure. The figure was made using PDB file 1MNX using PyMol

Architecture of RNA

Internal loops

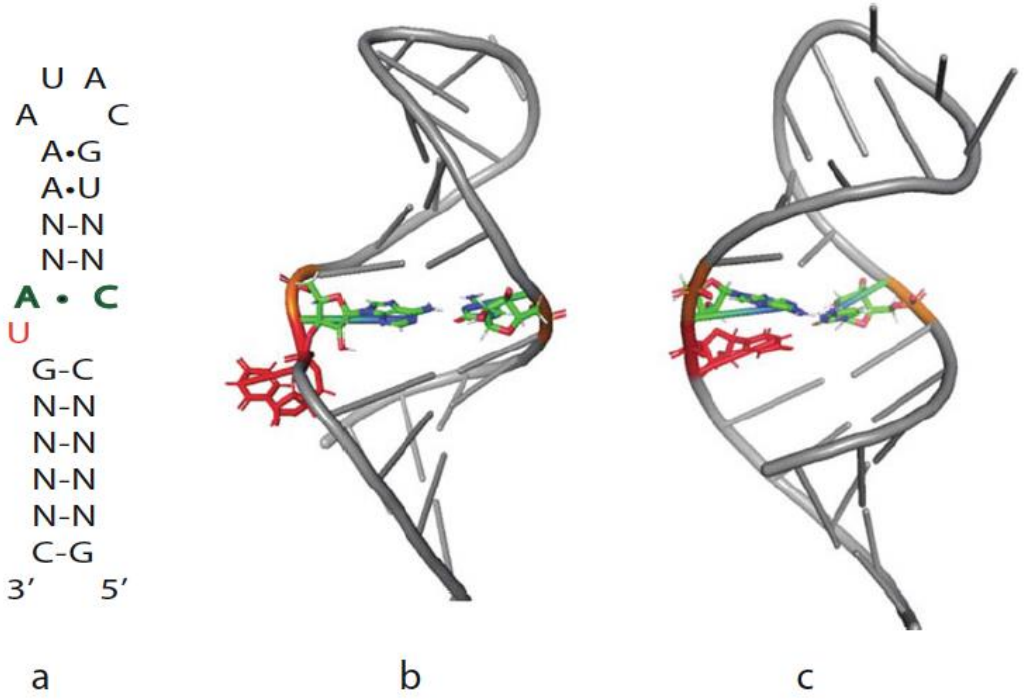


Fig. 2.6 Asymmetrical Internal Loop. (**a**) The secondary structure of the U6 spliceosomal RNA has a 1 x 2 internal loop. This RNA adopts (at least) two different conformations in two different pHs (5.7 and 7). (**b**) In the NMR structure of U6 at pH 5.7, the A•C pair forms and U (red) is unstacked 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.

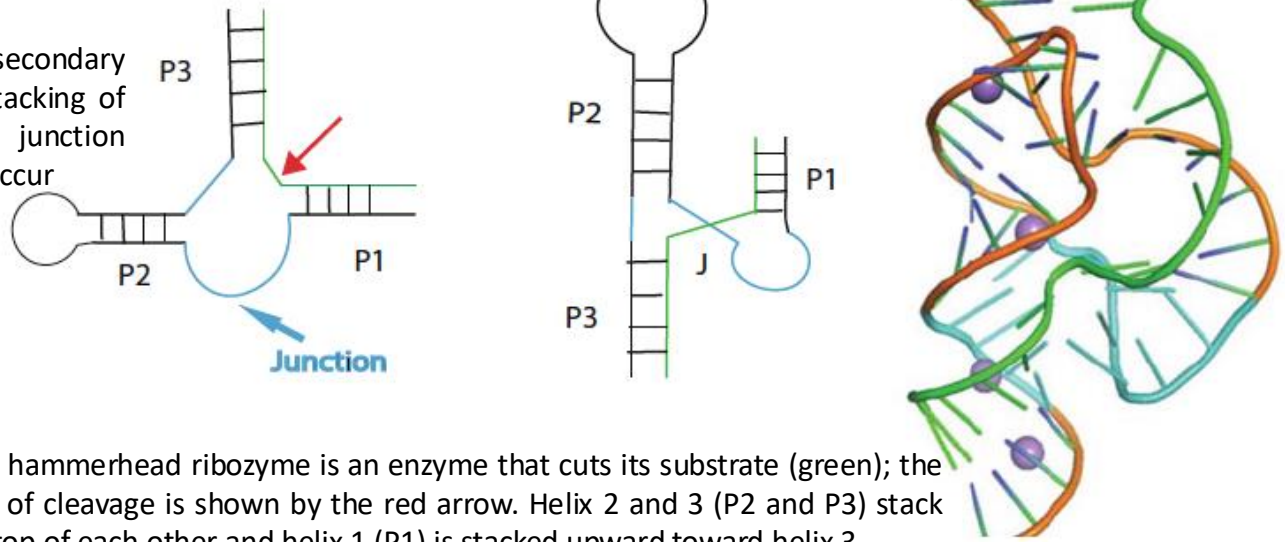
Architecture of RNA

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

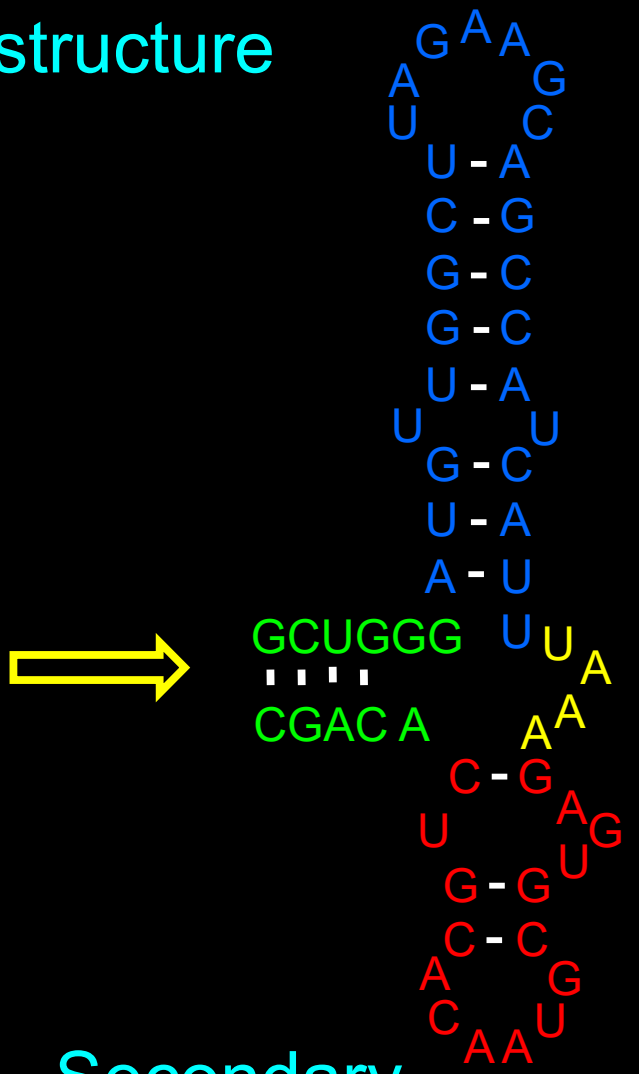
a b c

Architecture of RNA

Nucleic acid structure

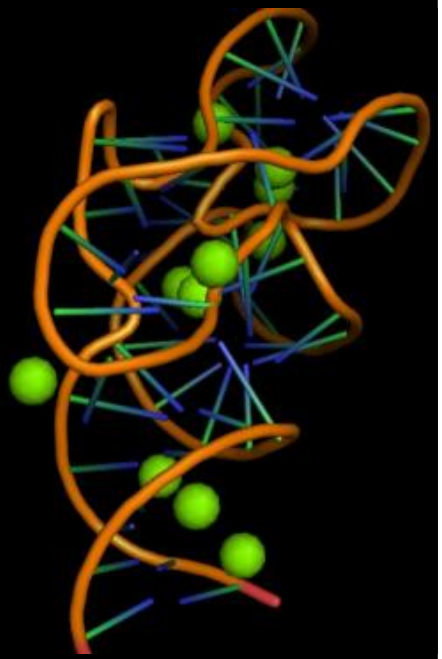
GCU GGG AUG
UUG GCU UAG
AAG CAG CCA
UCA UUU AAA
GAG UGC CGU
ACA GCU CAC
AGC

Primary



Secondary

Higher order

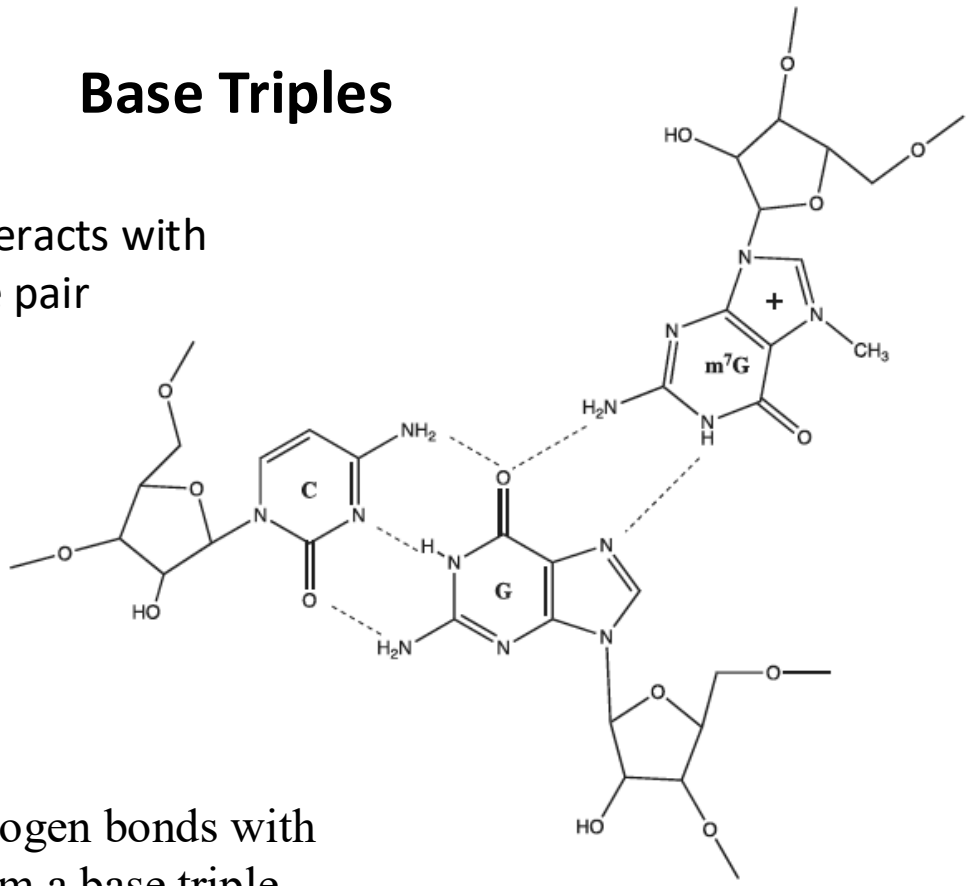


Architecture of RNA

Ternary interactions

Base Triples

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



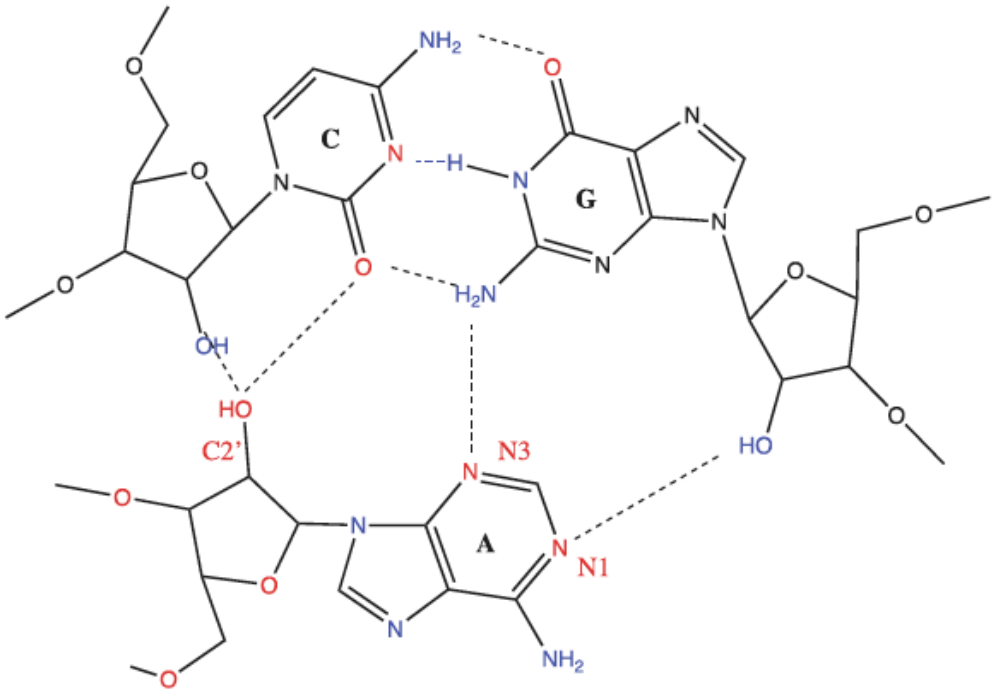
A C-G canonical base pair hydrogen bonds with m^7G (N7methyl guanine) to form a base triple.

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.

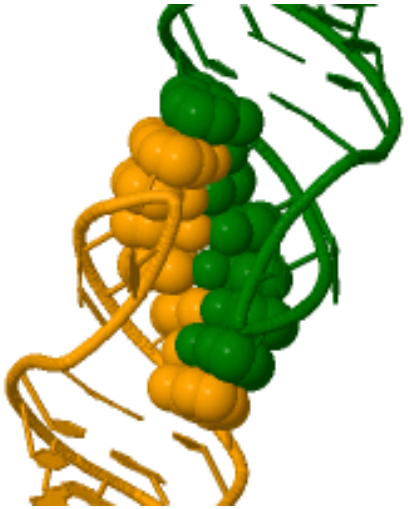
It is a tertiary interaction that allows **packing of the RNA structure**.

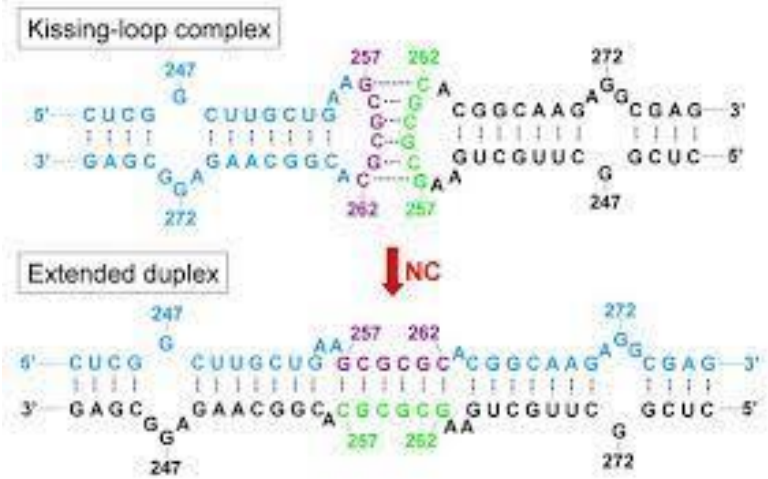
Architecture of RNA

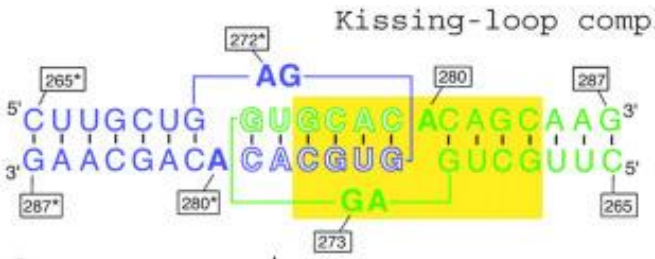
Ternary interactions

Kissing Loops

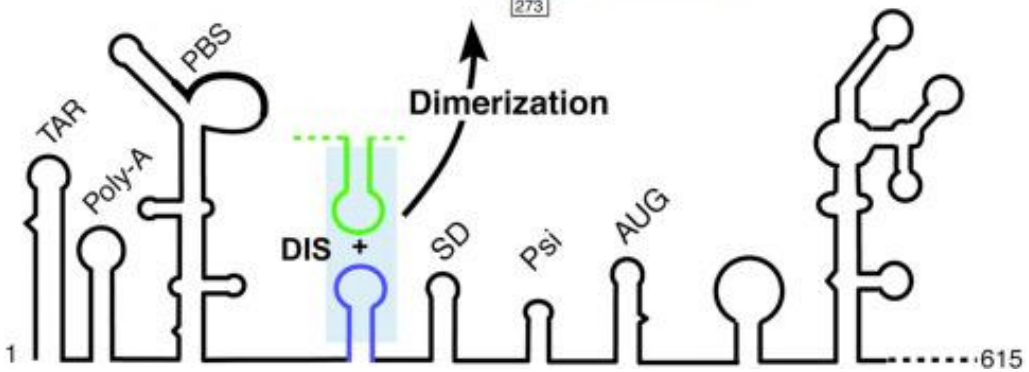
Base pairs form between two loops







HIV-1 DIS Kissing Loop

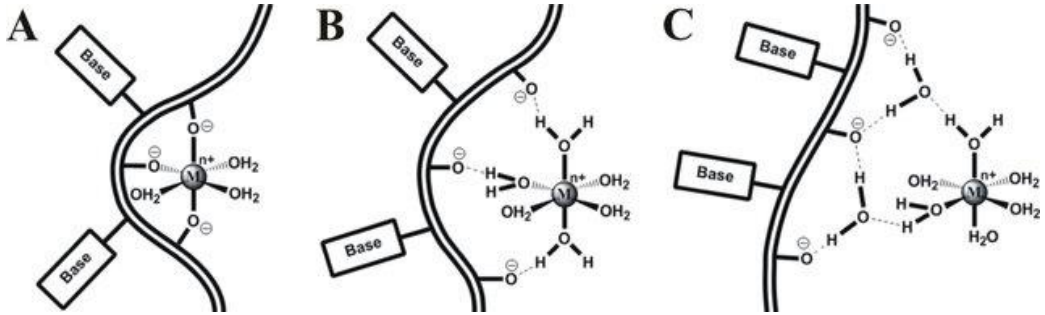


Metal ions and RNA

RNA molecules fold up to form structures when **repulsive forces** between negatively charged phosphate groups are countered by positively charged ions.

Inside the cell, **potassium and magnesium** ions are the most abundant.

These ions bind to the RNA backbone (**diffuse binding, C**) and to particular sites on RNA (**site-specific binding, A and B**) where the charges on the phosphates are closer together.



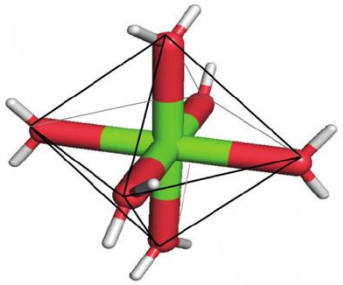
In addition, metal ions, in particular magnesium ions, are integral part of RNA's **catalytic function.**

Magnesium and potassium inside the cell

MONOVALENT IONS

- K^+ is the most abundant
- 140 mM in the cells
- charge neutralization
- screen of phosphate negative charges
- diffuse layer around RNA

DIVALENT IONS



6 H_2O
Octahedral geometry

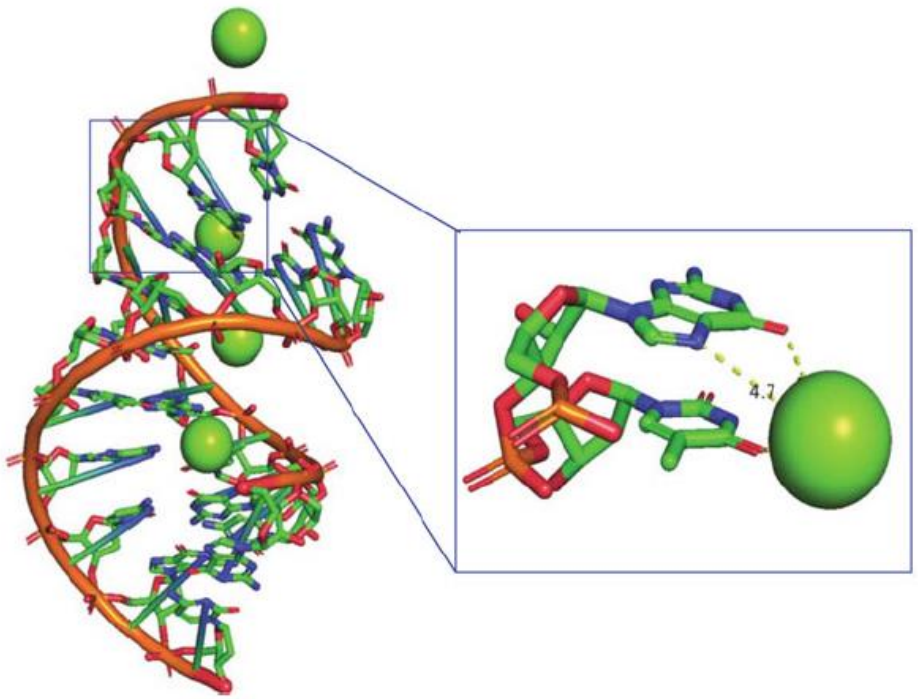
- $MgCl_2$ and $MgSO_4$ highly soluble at neutral pH
- 1 mM in the cytoplasm
- necessary for proper RNA folding
- involved in catalytic processes
- high charge density
- entropically favorite ($1Mg^{2+}$ vs $2K^+$)

Metal ion-RNA interactions

Magnesium cations interact with the **RNA backbone and base residues**.

Preferred ligands are anionic phosphate oxygens and some electronegative atoms of bases.

Mg^{2+}-RNA interaction is primarily seen in the deep groove of A-form helices.



Diffuse binding provides charge screening that overcomes electrostatic repulsion between RNA backbone segments and stabilizes the entire molecule.

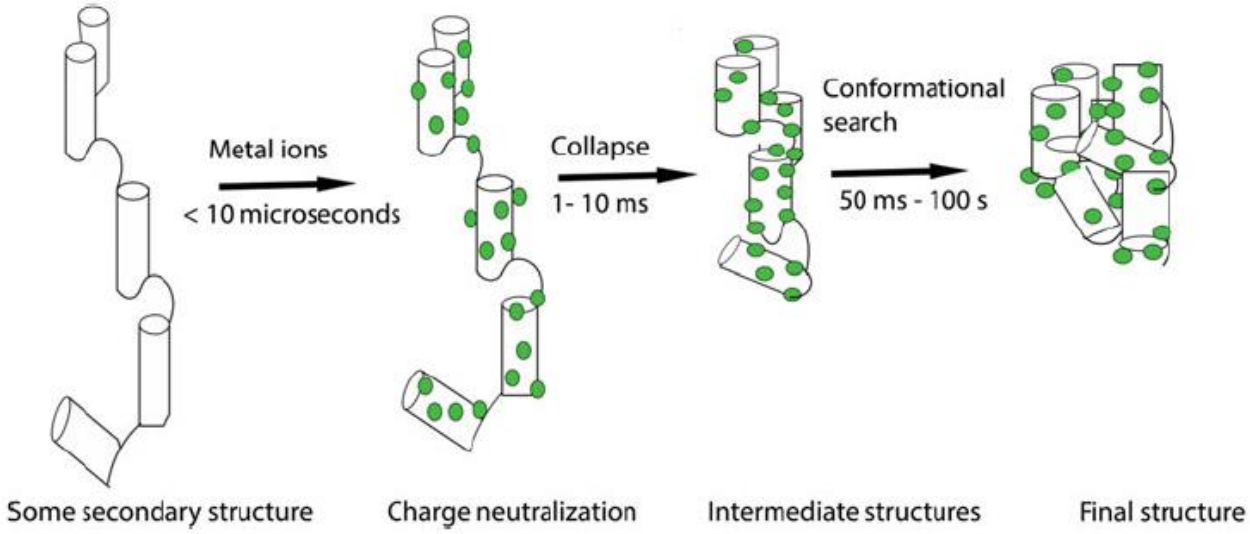
Metal ions in RNA folding

Most functional RNA molecules adopt stable, tightly folded globular conformations.

A suggested mechanism for RNA folding is the **two-stage model**.

1. **secondary structure first** – activated by monovalent and divalent ions
2. **RNA tertiary structure form** – 3D and intramolecular networks

Mg^{2+} stabilizes and increases the rigidity of certain RNA conformations that are very unfavorable in the presence of monovalent ions alone, even at high concentrations



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

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

RNA: composition and base pairing

DNA and RNA properties prediction

https://eurofinsgenomics.com/en/resources/tools/oligo-analysis-tool/

http://www.oligoevaluator.com/LoginServlet

IDT OligoAnalyzer (account)

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

cTAR DNA: 5’-GGC AGA GAG CTC CCA GGC TCA GAT CTG CC-3’

TBA1 DNA: 5’-GGT TGG TGT GGT TGG-3’

TBA2 DNA: 5’-AGT CCG TGG TAG GGC AGG TTG GGG TGA CT-3’

RNA1: 5’-UAG GGG GAA GCU UUG G-3’

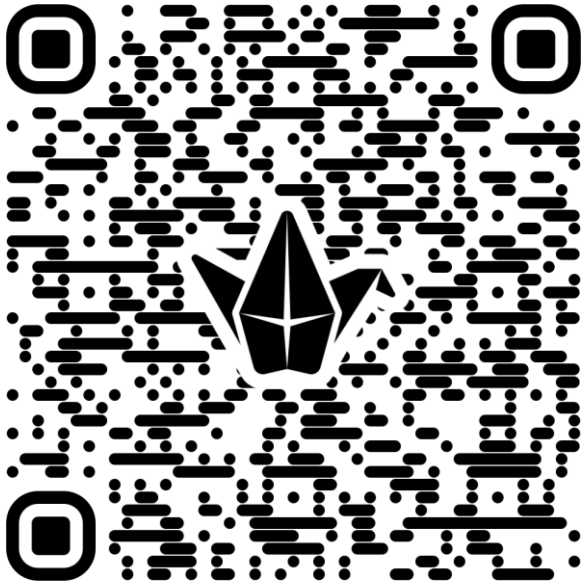
RNA: composition and base pairing

1. Check base numbers and MW of the oligo

2. Calculate Tm and Ɛ

3. Check self dimer/secondary structure

4. Perform points 1 and 2 for the counterpart of 1 DNA and 1 RNA oligo, and annotate the differences



Tool A

oligo	nt	MW (g/mol)	Tm (°C)	ε (L/mol*cm)
TAR RNA	29-mer	9290.62	73.7	288 200
cTAR DNA	29-mer	8888.75	73.7	289 050
TBA1 DNA	15-mer	4726.07	48	164 400
TBA2 DNA	29-mer	9085.87	72.3	302 450
RNA1	16-mer	5217.21	51.7	185 300

TAR DNA	29-mer	8910.74	72.3	302 450
TBA1 RNA	15-mer	4881.95	48	169 650

Suggested tool:

IDT OligoAnalyzer (account)

RNA: composition and base pairing

- MW (RNA vs DNA) is different because of different sugars and bases

- Ɛ different but slightly similar if only C-G-A

- only the bases absorb at 260 nm

- oligo absorbance is different from the sum of base composition

- Tm (duplex vs folded ssOLIGO)

RNA: composition and base pairing

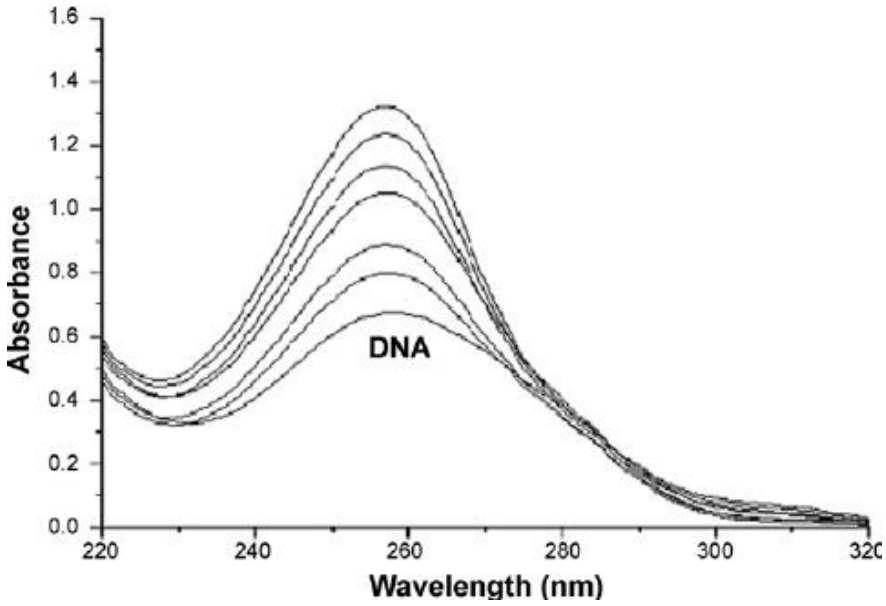
Spettrofotometro



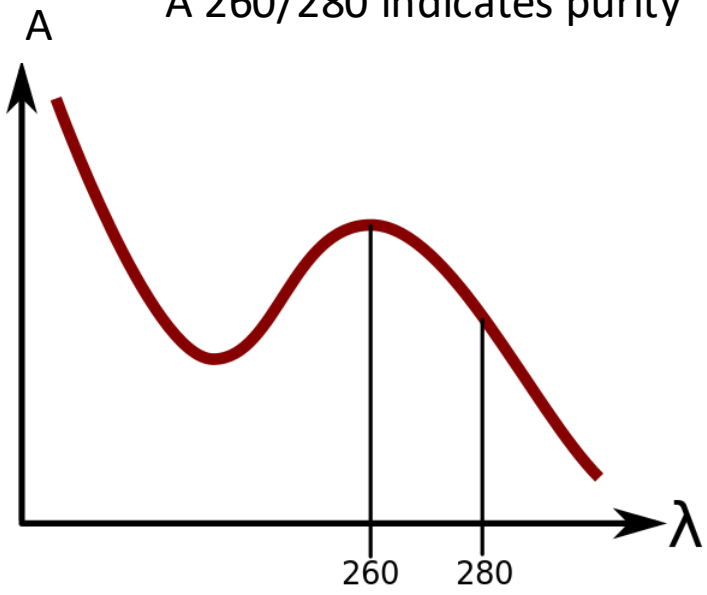
$$A = \varepsilon c l$$

Nanodrop





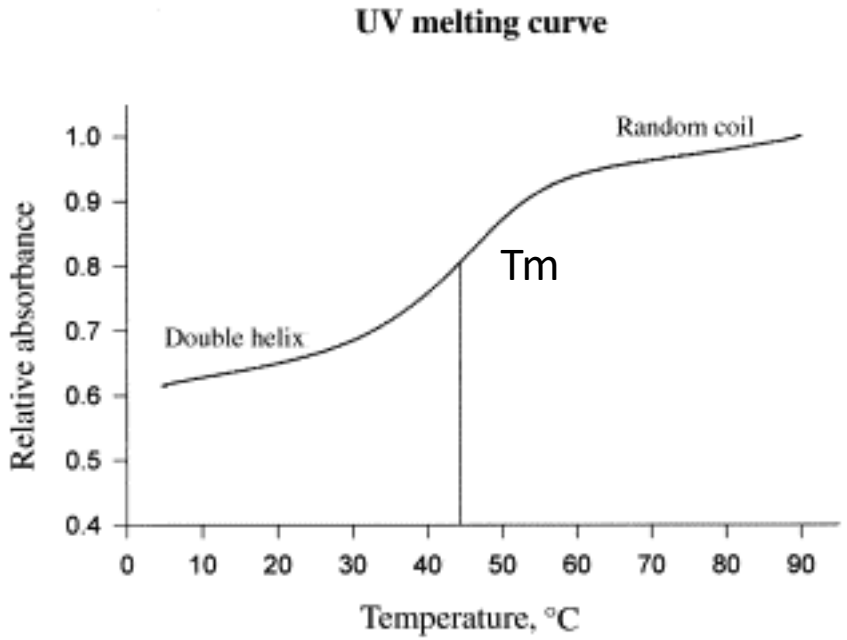
A 260/280 indicates purity

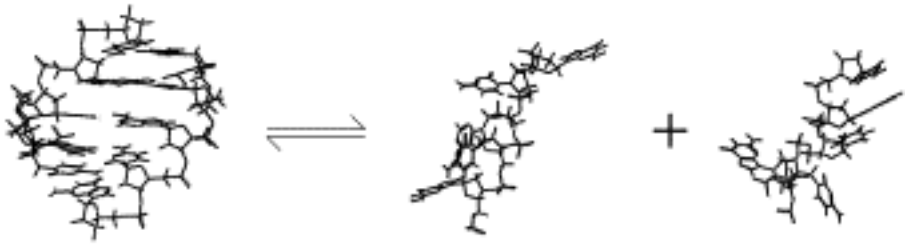


RNA: composition and base pairing

dsOLIGO absorb differently from the ssOLIGOS

Tm (duplex)

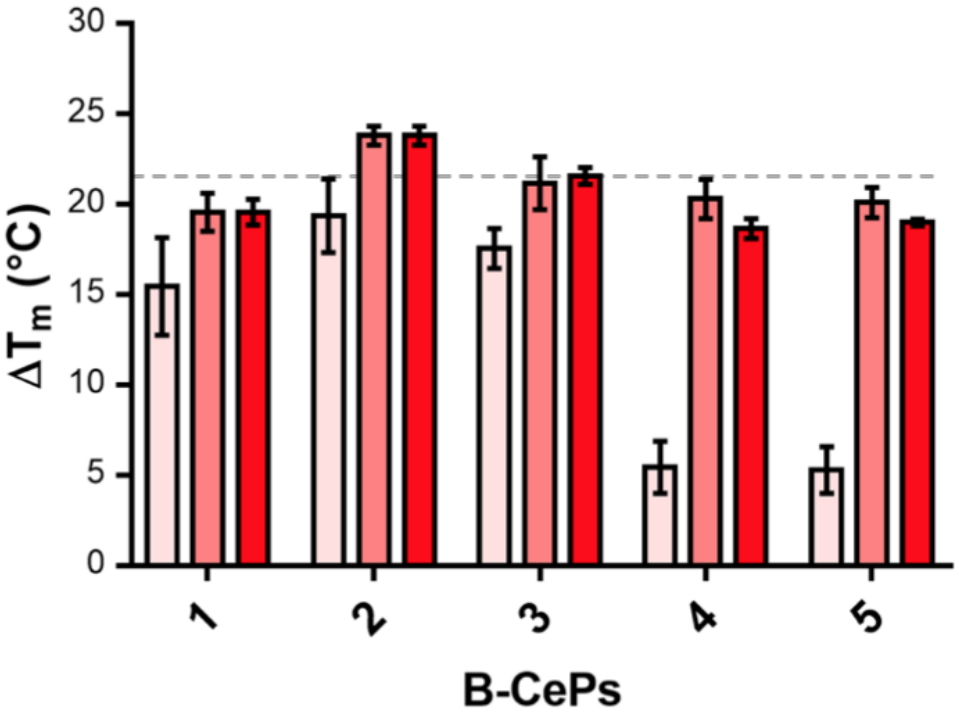


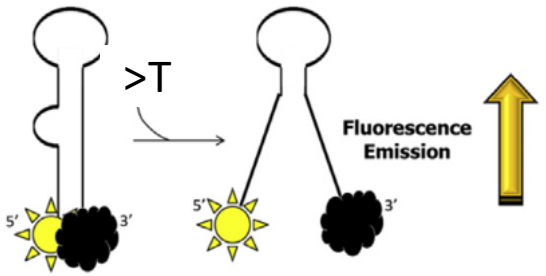


RNA: composition and base pairing

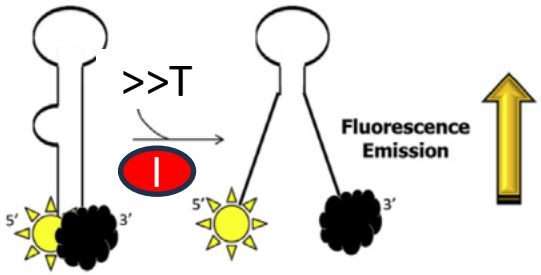
Tm (folded ssOLIGO)

Stabilization of TAR hairpin





Tm1 = 22 ° C



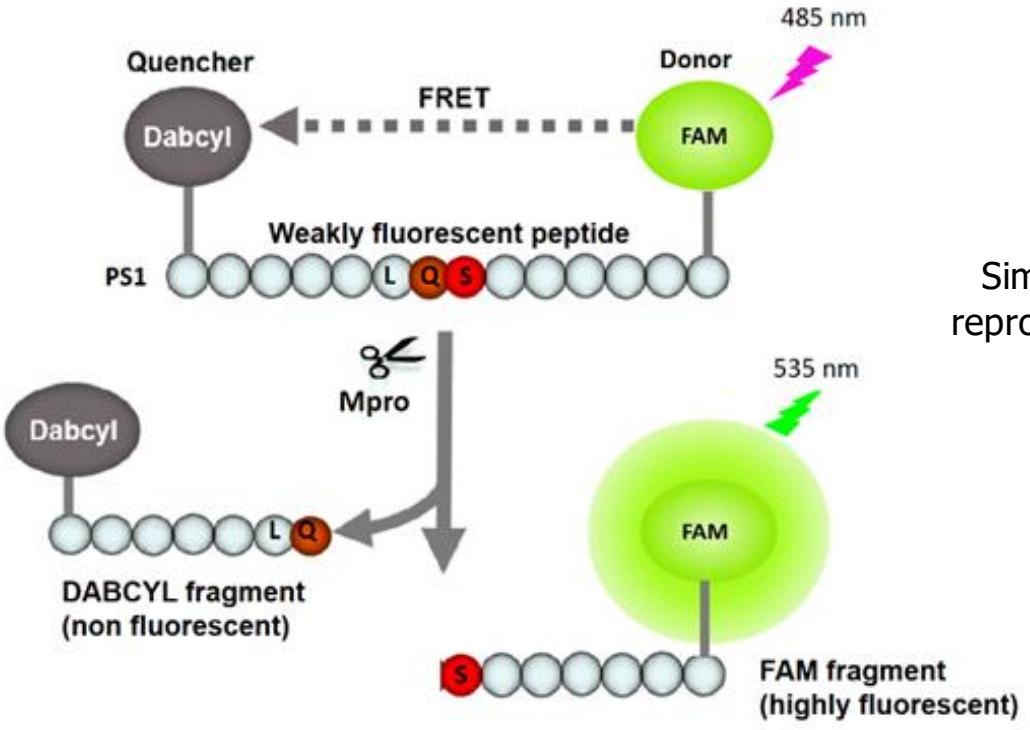
Tm2 = 37 ° C

ΔTm = Tm2 – Tm1

FRET: Fluorescence resonance energy transfer

FRET is a **distance-dependent transfer of energy between two chromophores**, a donor and an acceptor molecule. When in close proximity, electronically excited donor chromophores can transfer energy to acceptor chromophores. As a result, the transfer of energy quenches the fluorescence intensity of the donor and reduces its excited-state lifetime while increasing the emission intensity of the acceptor chromophore. It is an advantageous technique for investigating various biological phenomena that produce changes in molecular proximity.

Quencher: acceptor molecule that accept donor energy but does not emit fluorescence





Simple, fast and high reproducible FRET-based assay

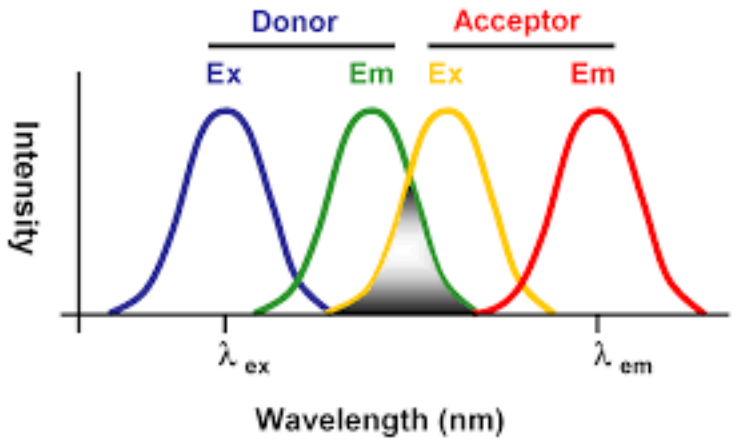
FRET: Fluorescence resonance energy transfer

Conditions for FRET:

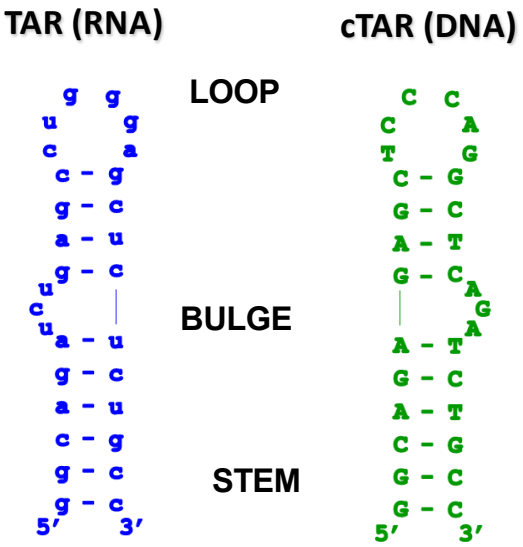
1.The donor and acceptor molecules must be in **close proximity to one another**, typically 10-100 Å (1-10 nm).

FRET efficiency (E) is defined by the equation $E = R_0^6/(R_0^6 + r^6)$, where R_0 is the Förster radius, and r is the actual distance between the donor and acceptor molecules. The Förster radius is the distance at which 50% of the excitation energy is transferred from the donor to the acceptor, and the R_0 value usually lies between 10-100 Å.

1.The absorption or excitation spectrum of the acceptor must overlap the fluorescence emission spectrum of the donor. The degree to which they overlap is referred to as the spectral overlap integral. The greater the degree of overlap, the greater the likelihood FRET is to occur.



RNA: composition and base pairing



5′-ggcagaucugagccugggagcucucugcc-3′
3′-CCGTCTAGACTCGGACCCTCGAGAGACGG-5′

Hybrid TAR/cTAR

Folding by snap cooling

- 5’ 95 ° C
- On ice

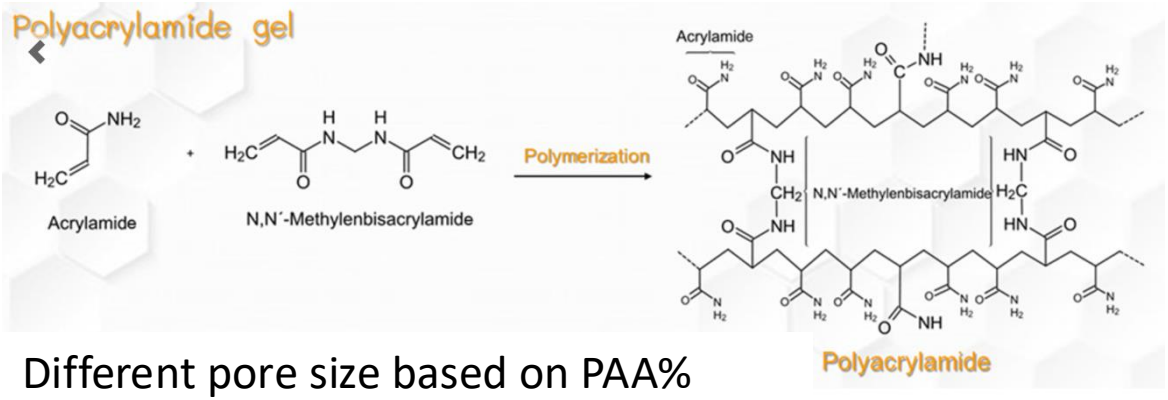
Strutture cineticamente più stabili

Slow annealing

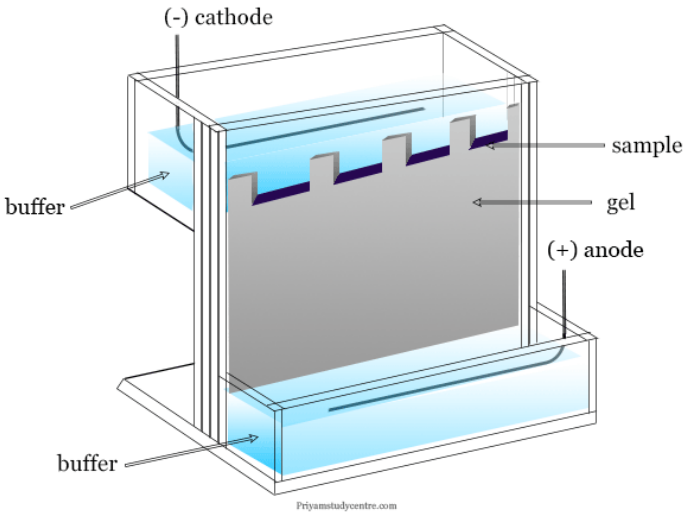
- 5’ 95 ° C
- Slowly cool down to RT

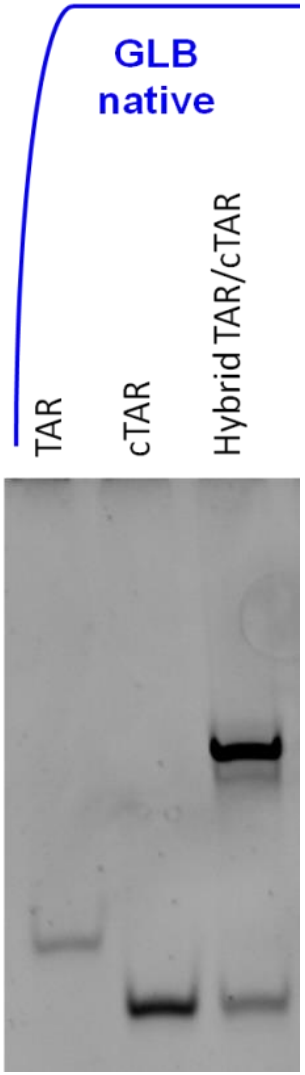
Strutture termodinamicamente più stabili

RNA: composition and base pairing



Different pore size based on PAA%





PAA 20%, TBE 1X
SybrGreen staining

TAR, cTAR 1uM

RNA: composition and base pairing

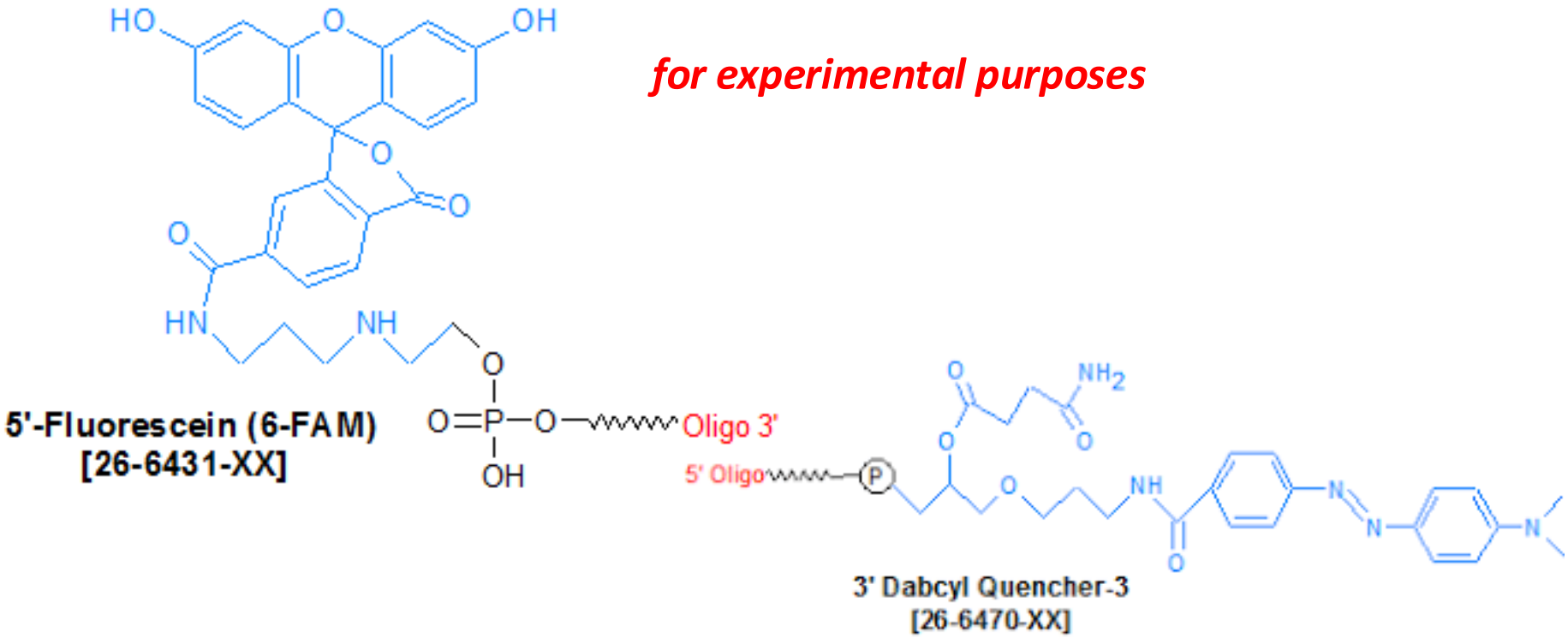
RNA modifications

Naturally occurring

Artificially introduced

for therapeutic purposes

for experimental purposes



5'-Fluorescein (6-FAM)
[26-6431-XX]

3' Dabcyl Quencher-3
[26-6470-XX]

RNA: composition and base pairing

RNA modifications

Naturally occurring

The Chemical Diversity of RNA Modifications

R. Jordan Ontiveros[1,†], **Julian Stoute**[1,†], **Kathy Fange Liu**[1,*]
[1]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Enzyme-mediated chemical modifications on RNA species across all domains of life have been documented over more than 50 years. A large repertoire of **over 100 distinct modifications** have been described, with the first report of modified nucleotides from highly abundant RNAs dating **back to 1960**.

These ***post-transcriptional*** **modifications are enzymatically installed and removed in a site-specific manner by *writer and eraser proteins*** respectively, while ***reader proteins*** can interpret modifications and transduce the signal for downstream functions. These enzymes utilize a myriad of reactions, such as methylations, deaminations, thiolations, glycosylations, isomerizations, etc..

Advances in mass spectrometry have enabled the detection and characterization of modifications in relatively lowly expressed RNA species, revitalizing the study of RNA modifications and shaping the field of **epitranscriptomics**.

REVIEW ARTICLE

WILEY

Analyzing RNA posttranscriptional modifications to decipher the epitranscriptomic code

L. Deng[1] | J. Kumar[1] | R. Rose[2] | W. McIntyre[1] | Daniele Fabris[1]

As the diverse nature and scale of RNA functions progressively emerged, the spotlight turned toward the **over 170 types of RNA post-transcriptional modifications (rPTMs)** present in natural RNA.

It is generally accepted that modifications **help compensate for the limited chemical diversity afforded by the four canonical ribonucleotides** and have a determinant influence on the **higher-order structure and recognition properties** of parent RNAs.

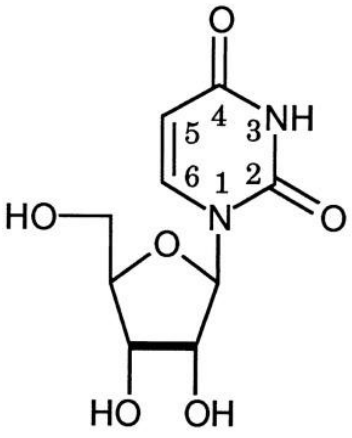
Their **reversible nature** makes them ideal signals for **turning on/off the biological functions** enacted by the parent RNAs. The elucidation of such functions has been traditionally hampered by the dearth of analytical approaches for detecting their presence, quantifying their levels, and identifying their sequence positions in parent strands.

Over 170 different covalent variants of the canonical ribonucleotides have been discovered.

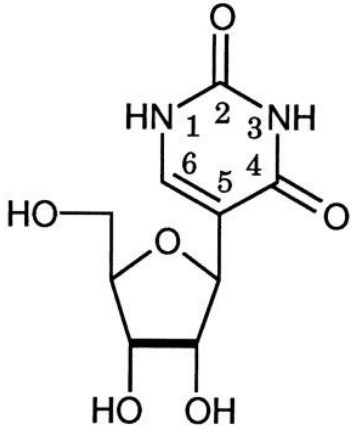
These variants originate from the activity of specialized enzymes (known as ***writers***) that recognize target RNA strands and catalyze their posttranscriptional modification.

General classification in well-defined groups:

1. **Pseudouridine** that, unique among known rPTMs, is formed by an *isomerization process*, such as that of uridine's N-glycosidic bond into a C-glycosidic form.

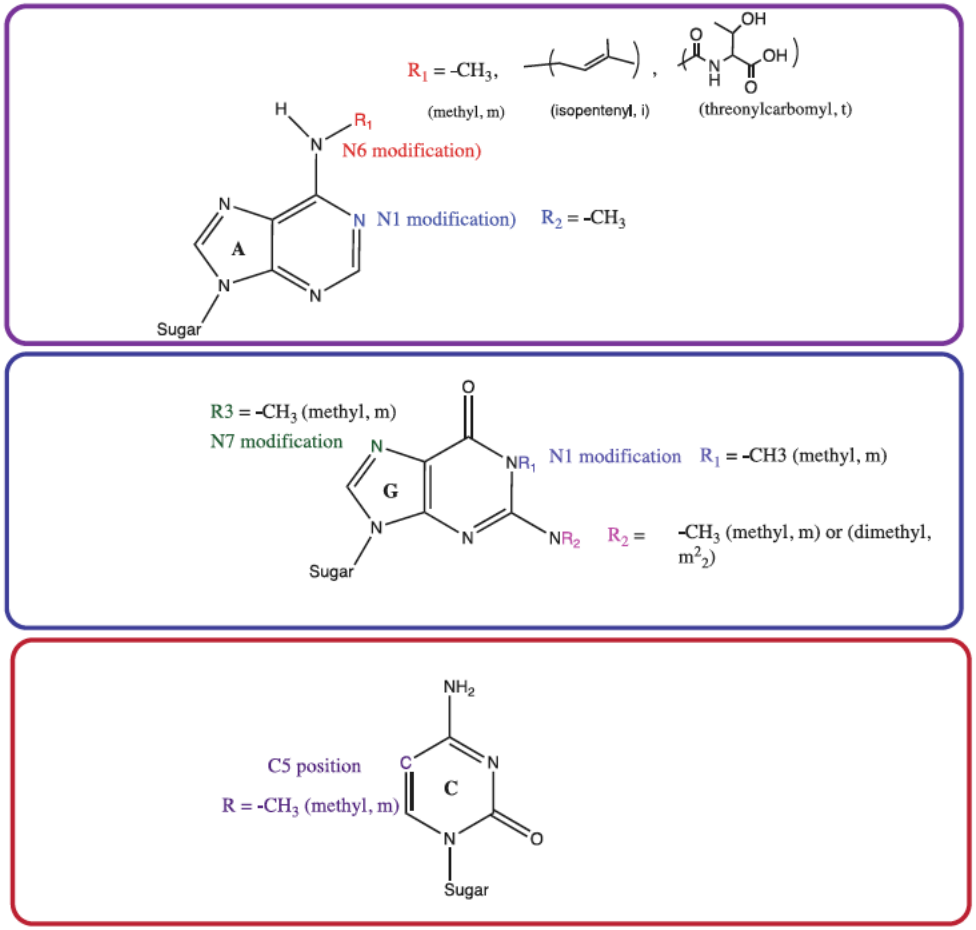


uridine, U

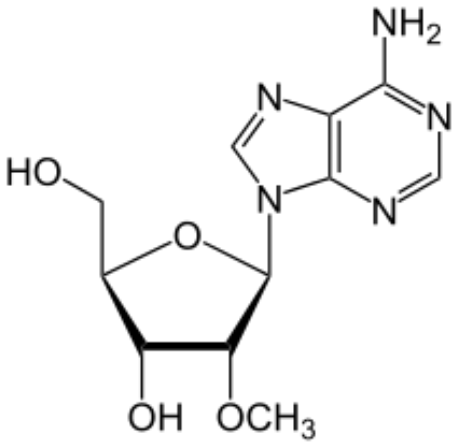


pseudouridine, Ψ

2. The second contains ribonucleotide variants produced by rather simple **alterations of the nitrogenous bases**, such as *methylation, deamination, reduction, thiolation, or alkylation*.



3. A separate class is generated by **methylation of the 2′-hydroxyl group of ribose**.



4. The final class consists of rPTMs with **multiple modifications** or introduction of sizable modifying groups.

At the molecular level, these alterations are capable of affecting the ability to establish intra- and intermolecular interactions responsible for stabilizing higher-order structure and specific interactions with other nucleic acids and protein factors

Techniques for rPTMs analysis

- ✓ 2D-TLC
- ✓ HPLC and UV/NMR
- ✓ LC-MS
- ✓ ESI-MS
- ✓ CE
- ✓ CE-MS

Identification and quantification

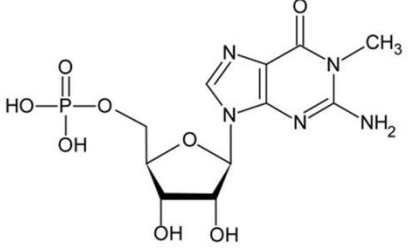
- ✓ NGS
- ✓ MS

Identification of rPTMs sequence position

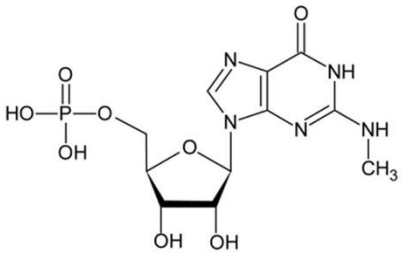
Techniques for rPTMs analysis

Classic approaches involved

- isolation of the RNA of interest,
- enzymatic hydrolysis into individual nucleosides/nucleotides,
- followed by analytical separation procedures

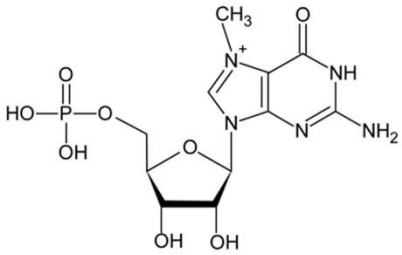


1-methylguanosine monophosphate
(m¹G)



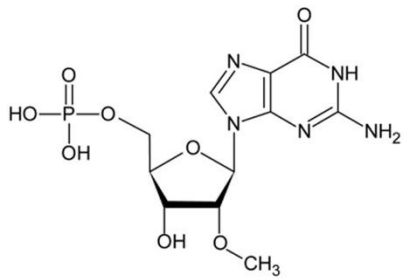
N2-methylguanosine monophosphate
(m²G)

Separation and analysis can be severely hampered by the fact that rPTMs stem from a limited number of precursor types and frequently display **isomeric structures with identical functional groups and chemical properties**.



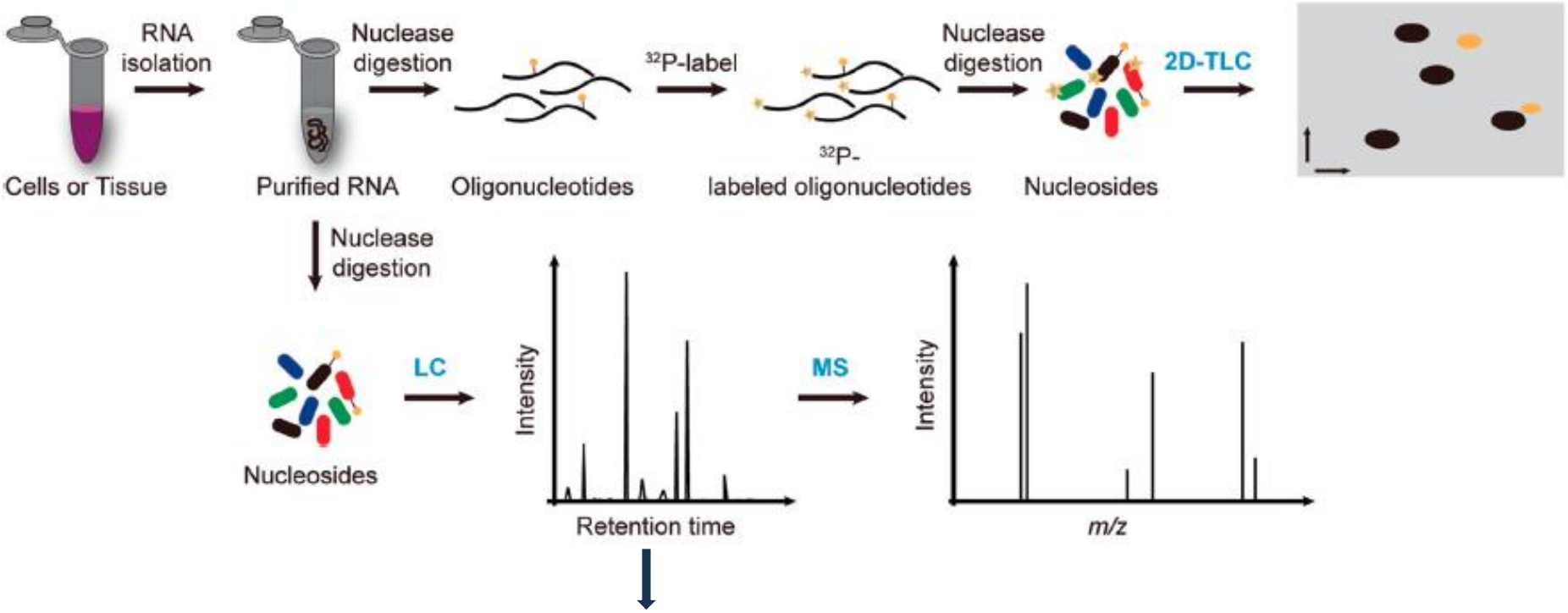
7-methylguanosine monophosphate
(m⁷G)

The very similar physicochemical properties afforded by this group of isomers call for the utilization of **high-resolution separation techniques** to achieve proper isolation.



2′-O-methylguanosine monophosphate
(Gm)

Flowchart of 2D-Thin Layer Chromatography (2D-TLC) and Liquid Chromatography-Mass Spectrometry (LC-MS)



Spectroscopic techniques

2D-TLC

Based on distinct mobilities of different nucleotides in the solvent.

Workflow: isolated RNA is first partially digested into oligonucleotides using RNAses and then labeled with ^{32}P using T4 polynucleotide kinase (T4 PNK). 5'-^{32}P-NMP is acquired by nuclease P1 digestion and further separated by 2D-TLC. The nucleotides can be determined by **assignment** to the standards by comparing their retardation factor (Rf) values. **Quantification** of the nucleotides is achieved by measuring the radioactivity of the corresponding spots in the TLC plate.

PROS

- very sensitive
- requires only a small amount of RNA (50-200 ng)
- applied to abundant and non-abundant RNAs
- it can also detect RNA modifications in specific tRNA or snRNA sequences, which can be isolated by gel purification or hybridization methods
- simple instrumentation

DRAWBACKS

- requirement of a radioactive reagent
- bias caused by RNase digestion and discrepant ^{32}P labeling efficiency
- need of standards

HPLC and LC-MS

Based on nucleosides polarity and polarity, m/z, specific ions

Workflow: RNA or oligonucleotides are completely digested to nucleosides and separated by reverse column chromatography.

Then, these nucleosides are ionized (**MS**) and further fragmented (**MS/MS**) into specific product ions via mass spectrometry.

Integration of **retention time, mass-to-charge ratio (m/z) and product ions** are capable of determining a certain nucleoside. In addition, **quantification** of nucleosides can be achieved through the external standard curve in the same batch.

PROS
- no radioactive reagent
- extremely high sensitive (detection limit that can reach the low femtomolar range)
- requires only a small amount of RNA (50 ng)
- applied to abundant and non-abundant RNAs
- detection of low abundant modifications

DRAWBACKS
- expensive HPLC and MS
- no sequence info

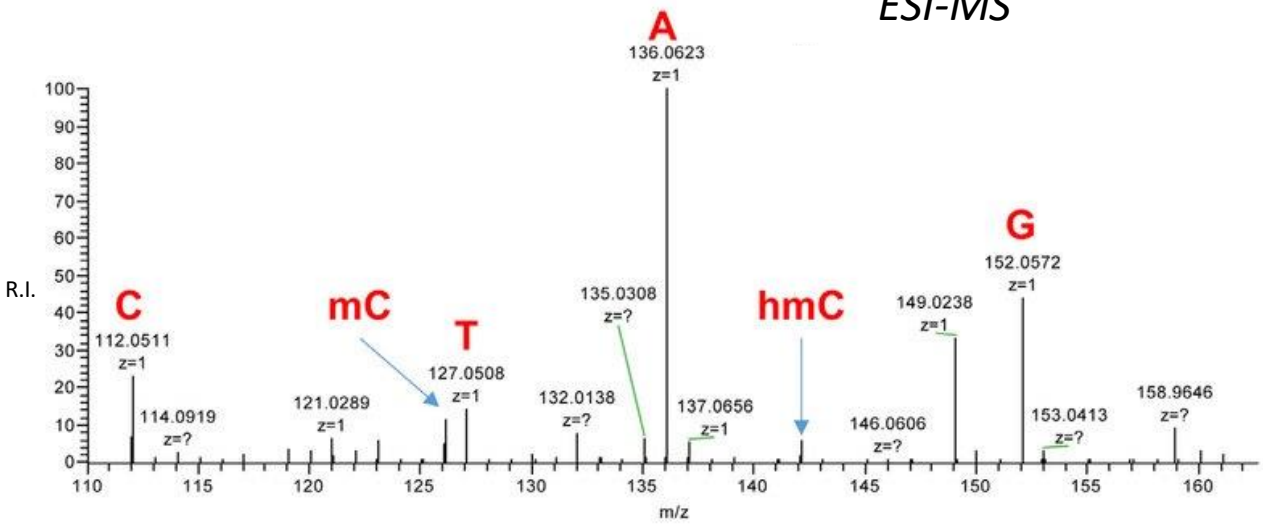
ESI-MS

- Preserves the integrity of the N-Glycosidic bond

- directly interface with separation methods

- no interferences of MALDI matrix

ESI-MS



MS/MS

- unique fragmentation data
- MS^n specific position of the modifier

negative

mG m/z= 376.25

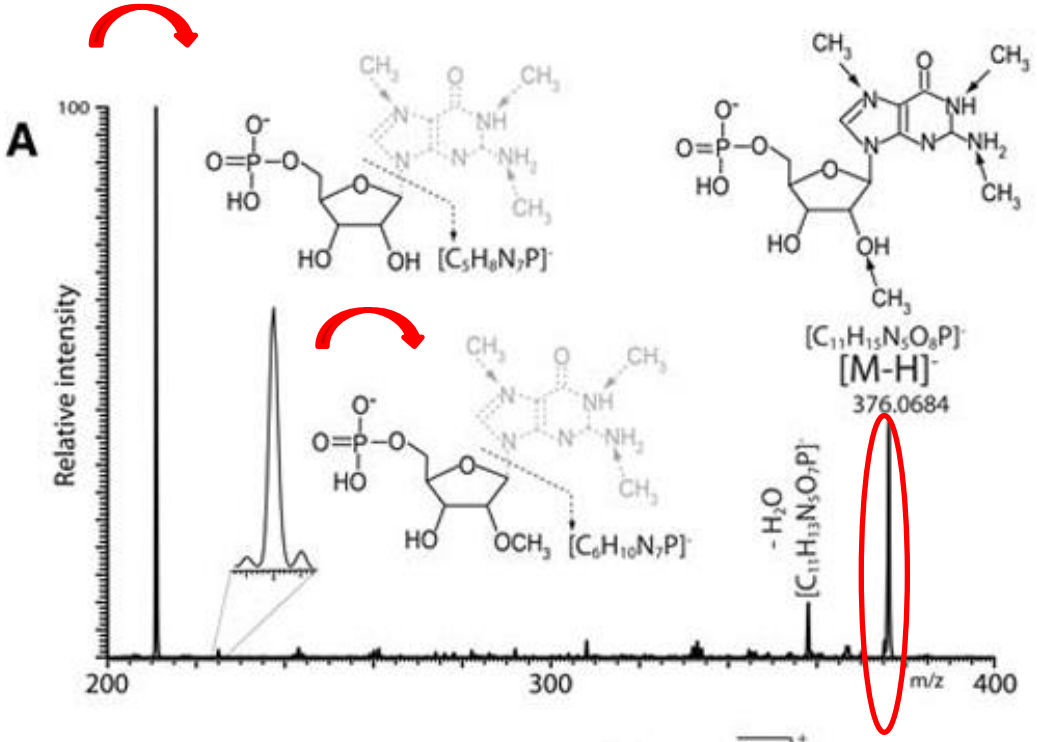
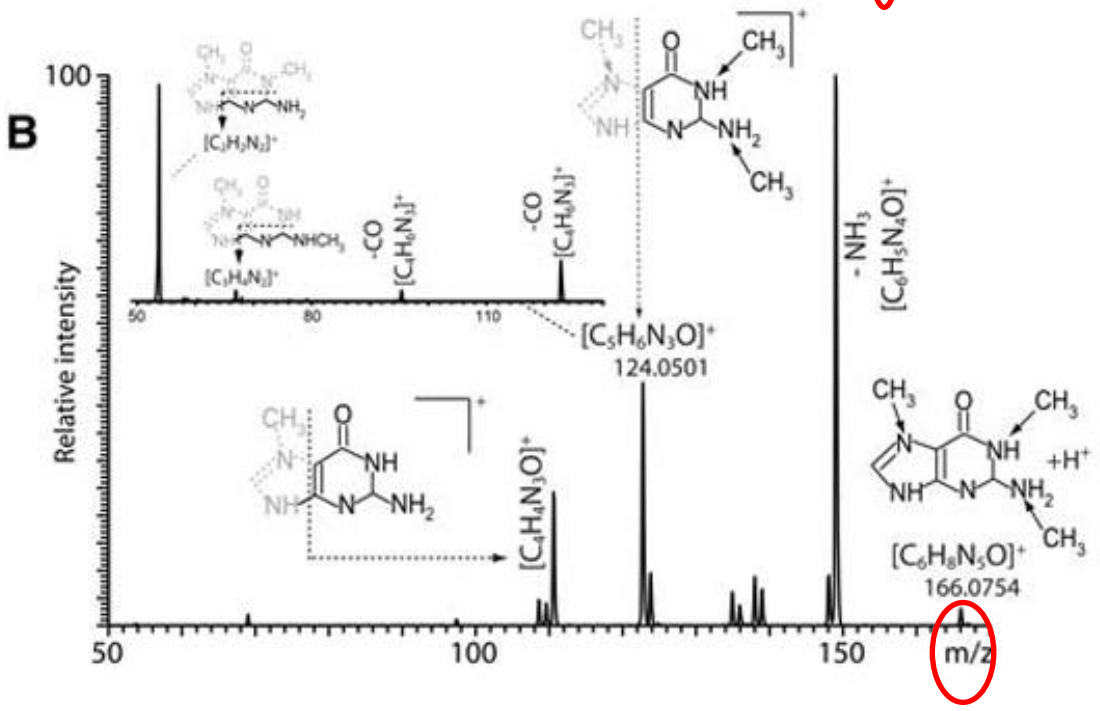


FIGURE 2 (A) Anionic MS/MS spectrum of methyl-guanosine isomers present in *Saccharomyces cerevisiae* total RNA digest, which was obtained by activating the corresponding ion at m/z 376. 25 on an ion trap analyzer. (B) Cationic MS^3 spectrum obtained by activating m/z 378 → 166 →. The inset displays the MS^4 spectrum obtained by activating m/z 378 → 166 → 124 →. Solid arrows indicate possible methylated positions; dashed arrows suggest putative cleavages. The intrinsic challenge of discerning multiple isobaric species can be resolved by implementing consecutive MS^n steps that are capable of providing the conclusive identification of each species in the sample. Figure reproduced with permission from reference (Rose et al., 2015).

positive

mG m/z= 378



Techniques for rPTMs analysis

- ✓ 2D-TLC
- ✓ HPLC and UV/NMR
- ✓ LC-MS
- ✓ ESI-MS + MS/MS
- ✓ CE
- ✓ CE-MS

Identification and quantification

- ✓ NGS
- ✓ MS

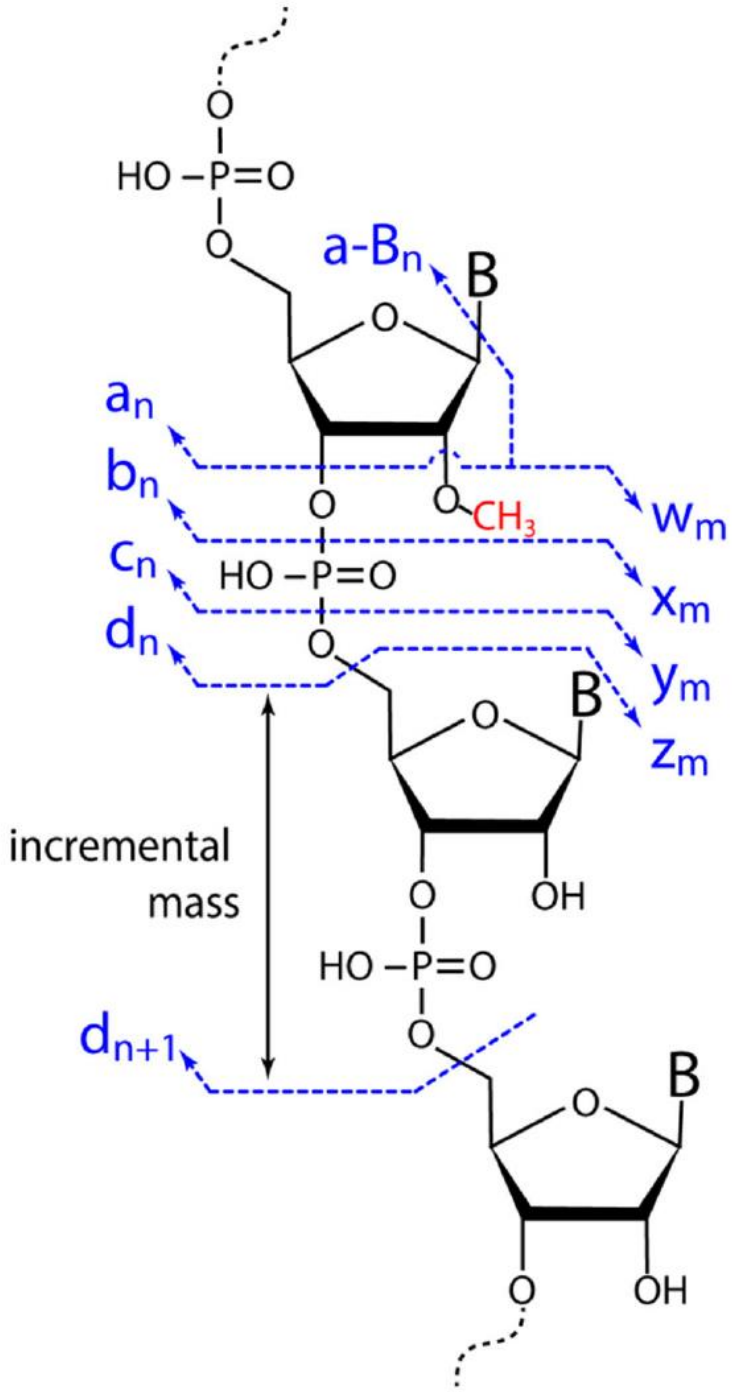
Identification of rPTMs sequence position

MS-based characterization of modified RNA strands

MS-based mapping of rPTMs

Nucleic acids sequencing is based on **gas-phase fragmentation** during tandem MS.

Gas-phase fragmentation induces **cleavage of the phosphodiester linkage** between consecutive nucleotides, leading to characteristic **ion series**.



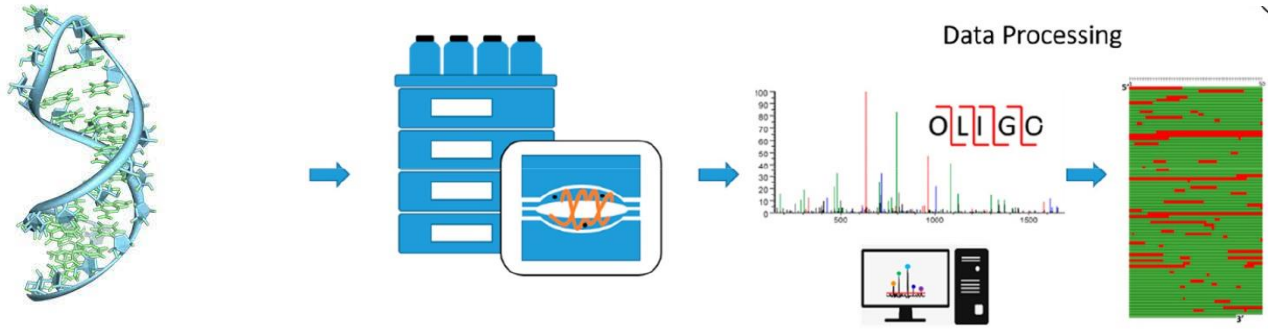
SCHEME 11 Types of fragment generated by gas-phase activation of oligo-ribonucleotides. Ion series containing the oligo's 5′- or 3′-end are respectively labeled a, a-B, b, c, d, or w, x, y, z (blue) depending on the bond cleaved by the process. The suffixes indicate the position of the nucleotide counting from either the 5′- or 3′-end. The mass difference between consecutive fragments in each series corresponds to the incremental mass of the intervening nucleotide. The identity and location of a modified variant can be unambiguously recognized from the unique spacing between consecutive series fragments. [Color figure can be viewed at wileyonlinelibrary.com]

MS-based mapping of rPTMs

Ionisation techniques:
- MALDI-MS
- ESI-MS (Direct injection or LC-MS)

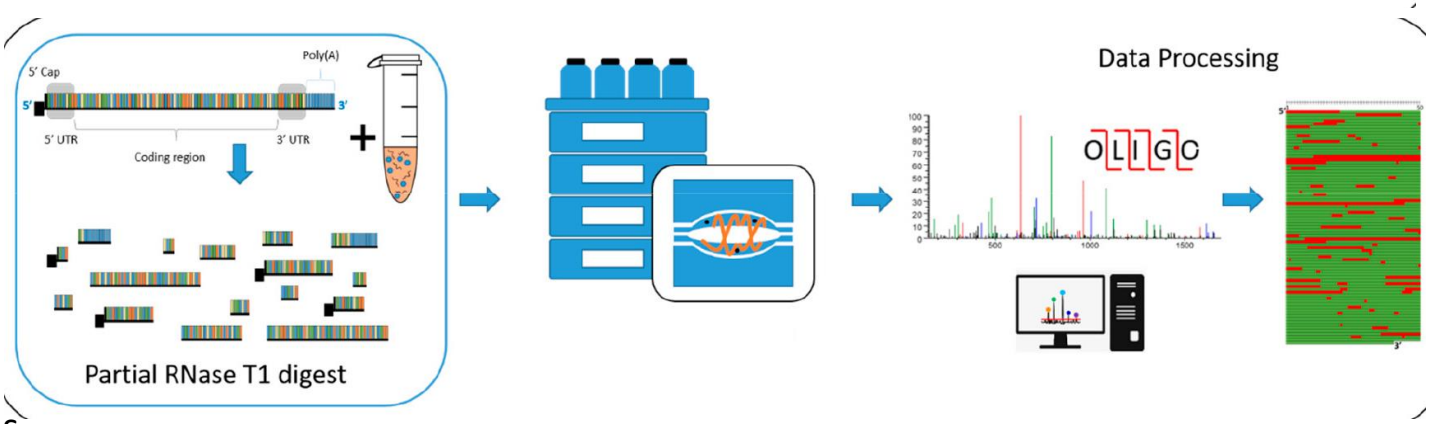
Top-down

RNA up to 50 nt



Bottom-up

RNA >50 nt



- different RNAses
- different T or times

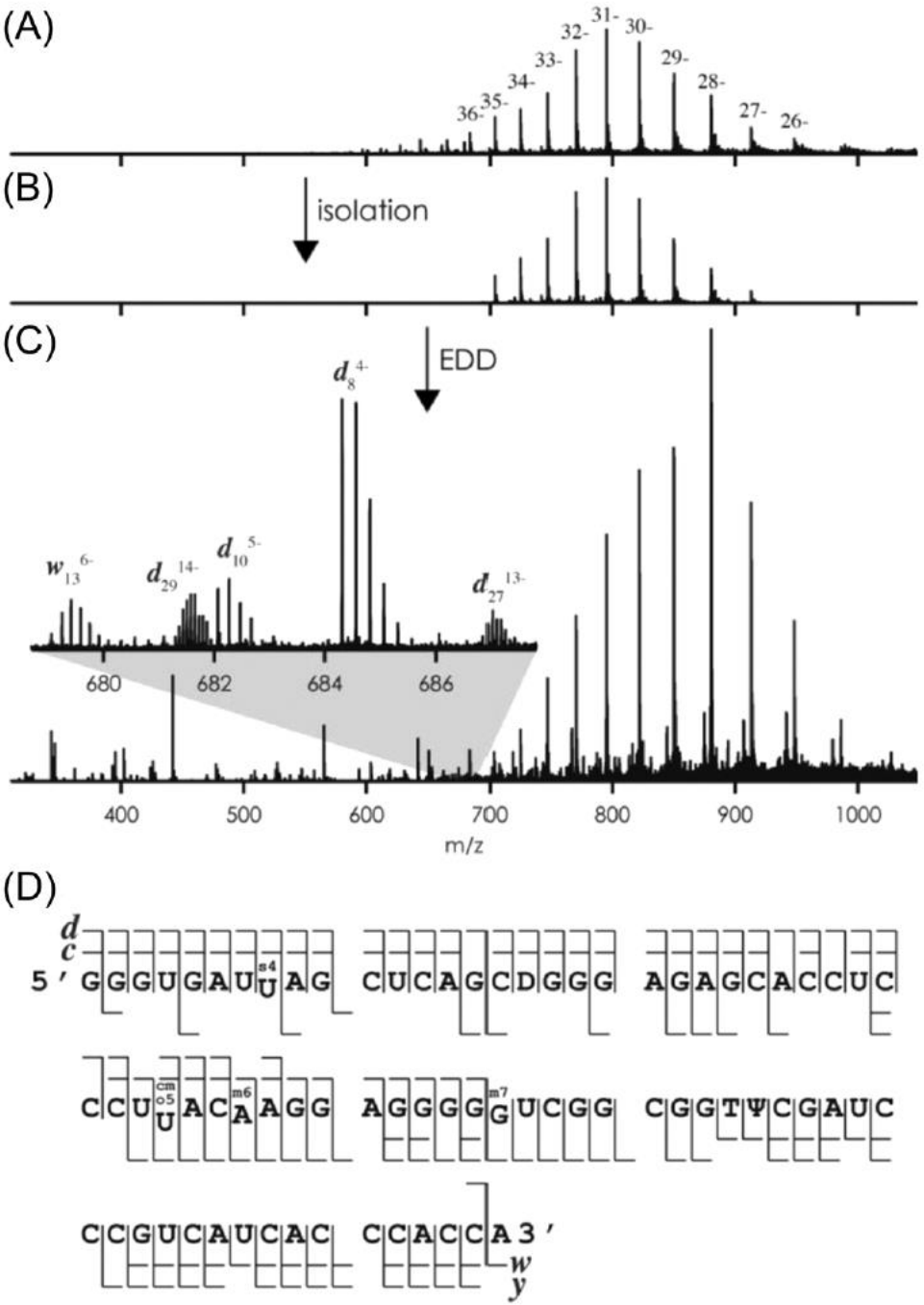


FIGURE 4 (A) Representative ESI-MS spectrum obtained from a 2 mM solution of $tRNA^{Val}$ in 1:1 water/methanol, 10 mM piperidine, and 10 mM quinuclidine; (B) isolation of ions with charge state ranging from −27 to −35; (C) dissociation spectrum produced by electron detachment dissociation (EDD); and (D) fragmentation map obtained by combining the EDD data observed in panel (C) with those obtained in parallel experiments by CID (not shown). The combination afforded full sequence coverage for the 75-nt species. Adapted with permission from reference (Taucher & Breuker, 2012).

MS^n used to provide not only sequence position, but also to distinguish isomeric mods.

RNA modifications

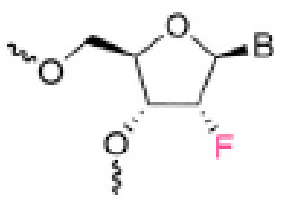
Artificially introduced

RNA has been artificially modified to alter its chemical and thermodynamic stability. Modifications that stabilize the RNA in the cell make it suitable for in vivo use, for example for RNA-based therapeutics.

The 2'-OH is often artificially modified to 2'-deoxy, 2'-O methyl, or 2'-fluoro.

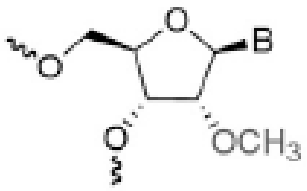
Alteration of the 2'-OH to deoxy or 2'-O methyl **increases RNA's stability and resistance to degradation by endonucleases**.

- *Improves binding affinity to target RNA*
- *Does not improve stability or PK*
- *Useful in RISC mechanisms*
- *Nucleoside metabolites incorporated into host DNA and RNA*
- *Can cause degradation of paraspeckle proteins*



2'-Fluoro

- *Improves PK and stability*
- *Only a modest improvement in potency*
- *Modest reduction in immune stimulation*



2'-O-methyl

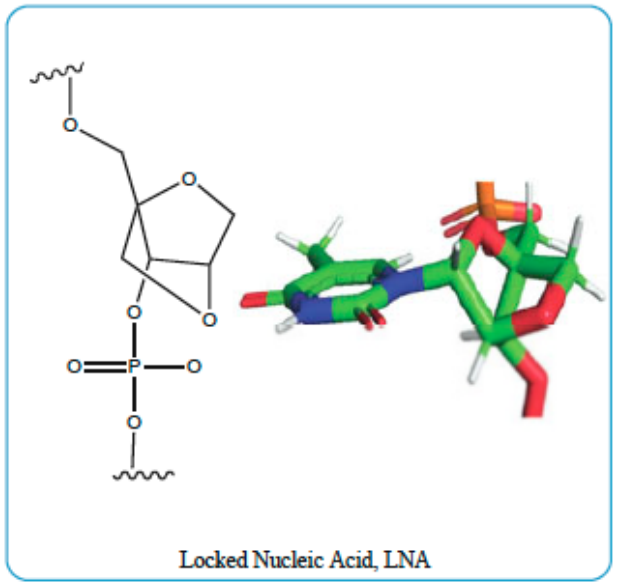
RNA modifications

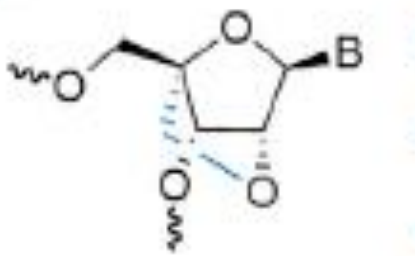
Artificially introduced

Sugar modification such as **2'-4' methylene bridge** makes *RNA resistant to nucleases while maintaining its solubility in aqueous solution*.

This modification is called a **locked nucleic acid (LNA).**

The entropic constraint imposed by this modification *increases its binding affinity* for the complementary strand.





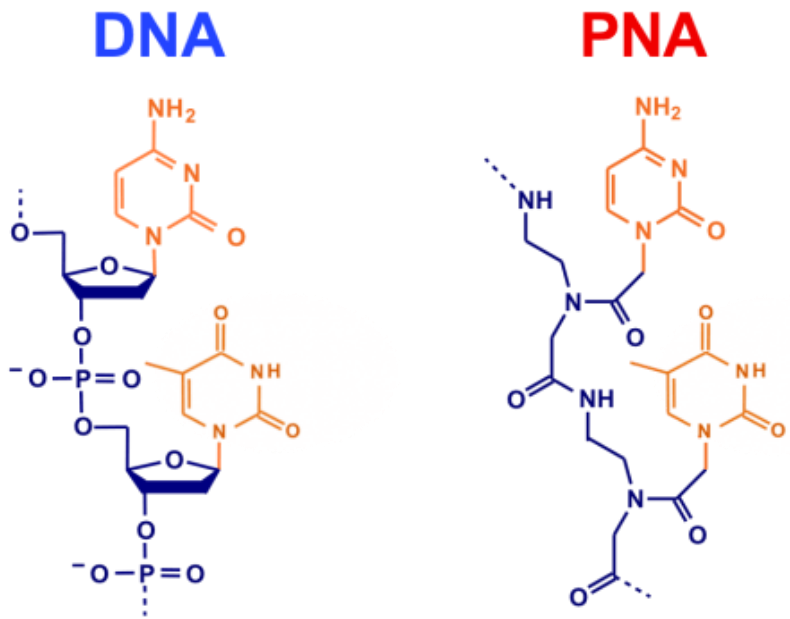
- *Greatly improves binding affinity, improves potency*
- *Increased sequence dependent toxicity risk*
- *Hepatotoxicity and renal toxicity observed in clinical trials*

Peptide Nucleic Acid (PNA)

The backbone of RNA has been modified using repeating units of **N-(2-aminoethyl) glycine linker** to which bases are attached via a **methyl carbonyl linker**. Now instead of the naturally occurring sugar-phosphate backbone, **a peptide backbone is created**.

This modification is of particular significance **as it alters the charge** of the RNA backbone from negative to neutral; thus, there is **no electrostatic repulsion** upon hybridization, which increases the stability of PNA: RNA strands over RNA:RNA or RNA:DNA hybridization.

The PNA backbone is resistant to nucleases, thus providing a good substitute for RNA-based drugs.



Happy B-day DNA!!!
February 28, 1953



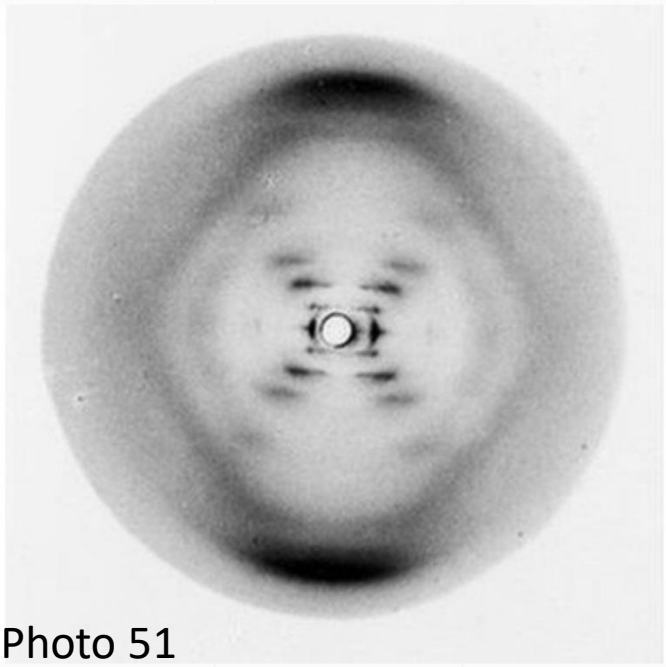


Photo 51

«the sectret of life»

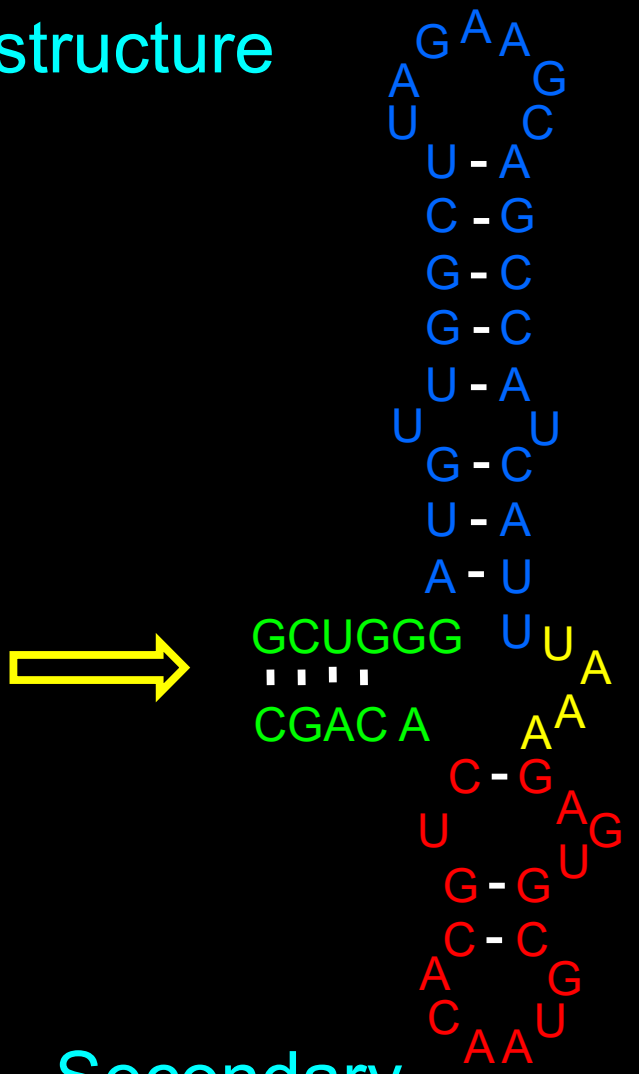


Architecture of RNA

Nucleic acid structure

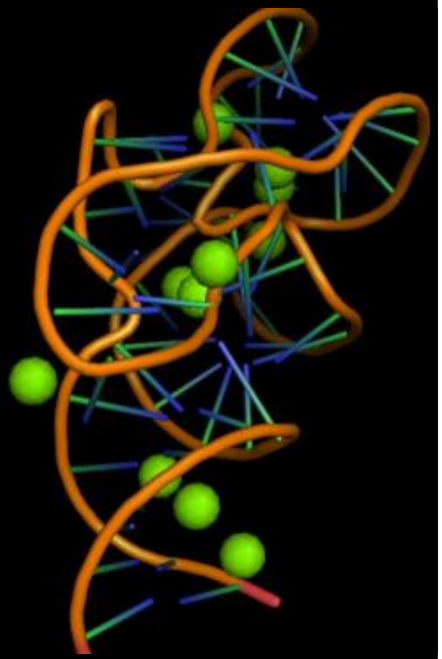
GCU GGG AUG
UUG GCU UAG
AAG CAG CCA
UCA UUU AAA
GAG UGC CGU
ACA GCU CAC
AGC

Primary



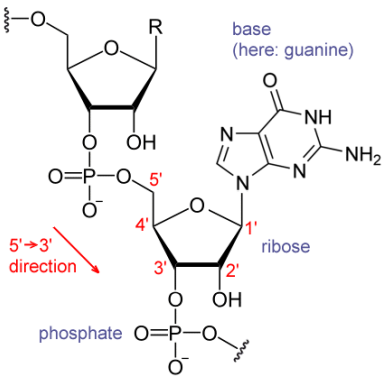
Secondary

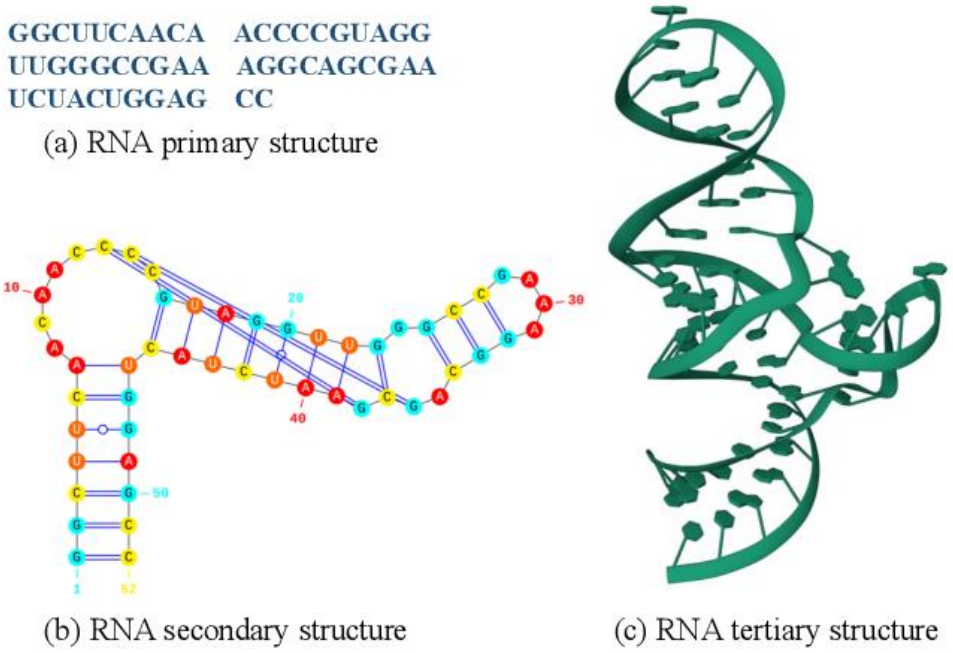
Higher order



Architecture of RNA

The **primary structure** of a nucleic acid is the sequence of nucleobases written in 5' to 3' direction.





The folding of an RNA is sequence-dependent. The sequence of RNA dictates its **secondary and tertiary structures**.

The flexibility of RNA to form many different isoenergetic structures determines its cellular functions.

How to study 2D RNA structure ???

Experimental techniques such as chemical or enzymatic probing, in conjunction with deep-sequencing techniques, enable the study of the RNA secondary structure.

- RNAses sequencing/footprinting + PAGE

- structure specific RNAses + MS

- SHAPE

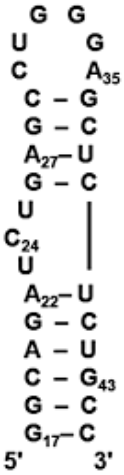
RNAses sequencing/footprinting + PAGE

- base- and structure-specific RNAses are employed

- high-resolution PAGE with radioactively/fluorescently labelled RNA are needed

- allow for sequencing (2D info) and footprinting (RNA-RNA, RNA-protein, RNA-ligand info)

HIV-1 TAR

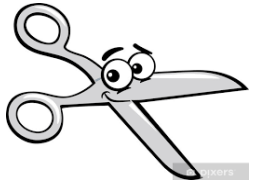




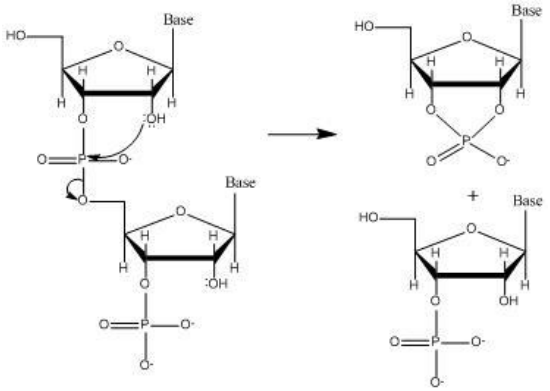
^{32}P-labelled or FAM-labelled oligo



RNAse digestions and alkaline hydrolysis as control



High-resolution PAGE



RNAses sequencing/footprinting + PAGE

Radioactive labelling
PAA 20% *7M UREA(denaturing)*
Sequencing gels

RNAse specificity: **ssG** **dsRNA** **ssC/U**

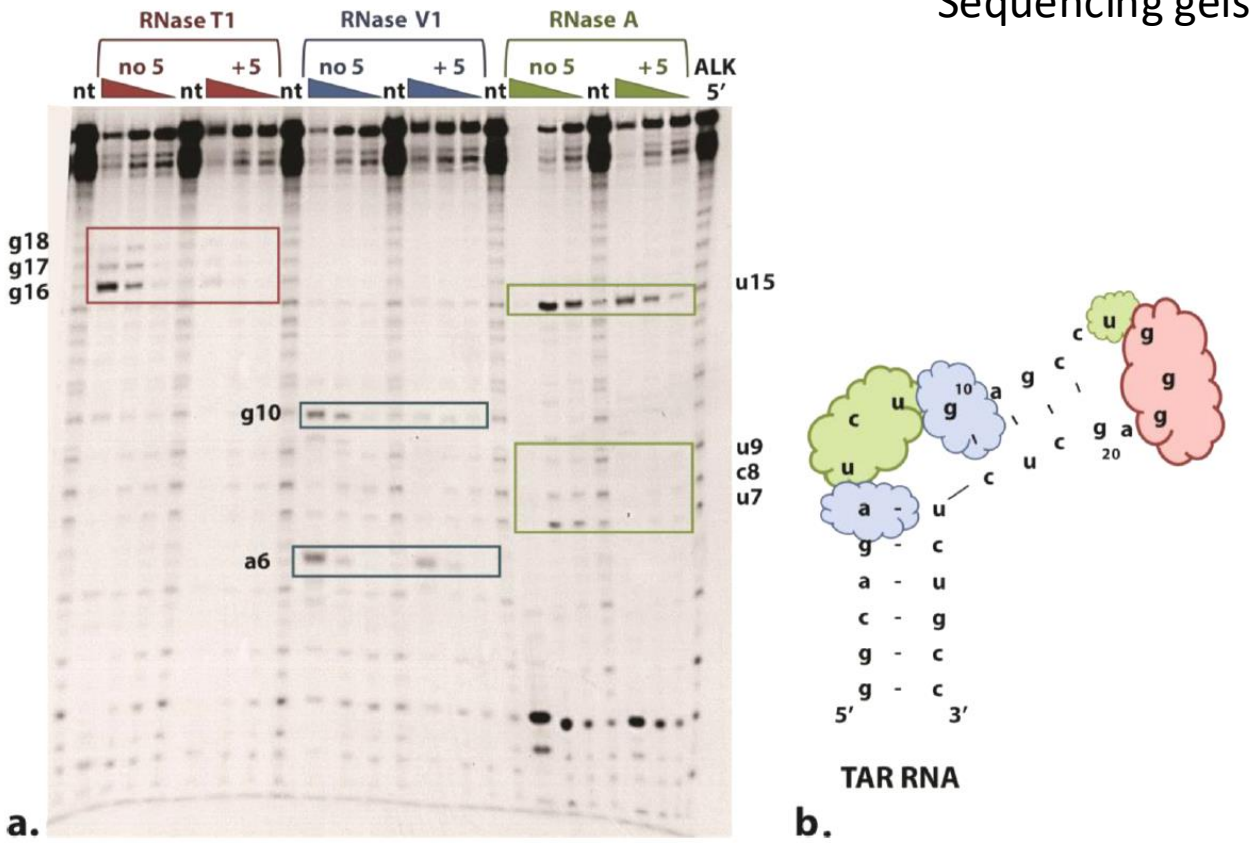


Figure 3. Nuclease footprinting analysis of TAR in the absence/presence of compound **5**. (a) Sequencing gel obtained from samples containing 0.6 μM of radiolabeled TAR without/with 300 μM of **5**, which were treated with RNase T1 (red), RNase V1 (blue), and RNase A (green). RNA was digested with three sequential 10-fold dilutions of each ribonuclease starting from RNase T1 1 U/μL, RNase V1 0.1 U/μL, and RNase A 1 μg/mL. The analysis was completed by denaturing gel electrophoresis (Polyacrylamide 20%, 7 M urea, TBE buffer) carried out at 95 W for ~1.5 h. Control samples that were not treated with RNases are indicated with "nt". A control submitted to alkaline hydrolysis by heating for 5 min is indicated with "ALK". (b) Summary of footprinting results. The cloudlike shapes identify nucleotides protected by the lysyl-peptidyl-anthraquinone **5**.

Alkaline hydrolysis

2'OH must be activated to be a nucleophile:

Deprotonation by OH^- at pH>8 in the alkaline hydrolysis

RNAse: specific nucleotide deprotonation by an OH^- ion coordinated to Mg^{2+}

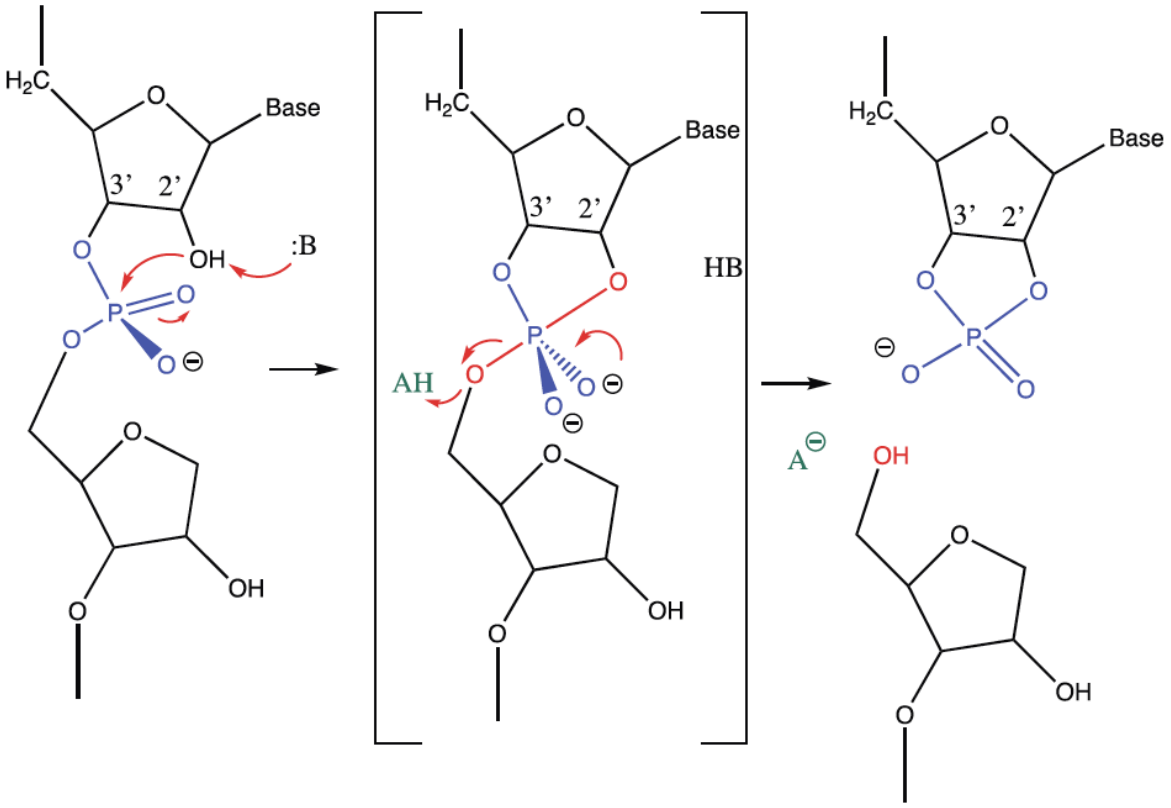


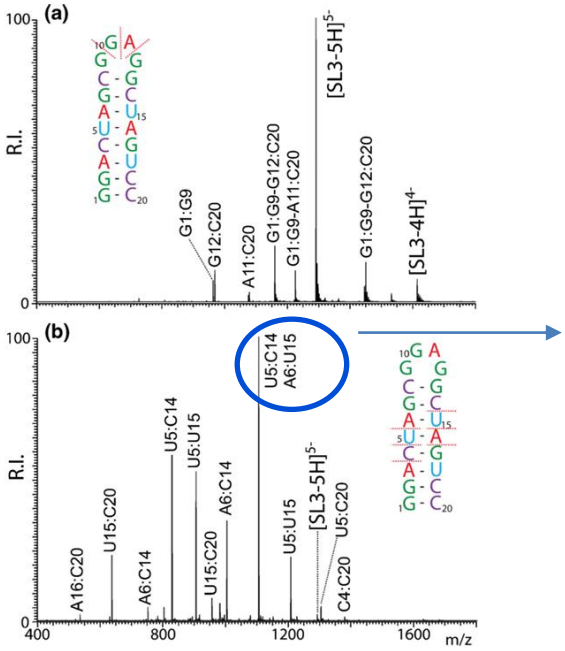
Fig. 3.2 A Mechanism of RNA Backbone Cleavage. The nucleophilic 2'-OH is activated by a general base and to act on the 3′ phosphate. In the transition state, the 2'-OH and the leaving phosphate have to be in-line and the phosphate has to adapt a trigonal bipyramidal configuration.

Structure specific RNAse + MS

- structure-specific RNAses are employed: **Mung bean** for ssRNA and **RNAse V1** for dsRNA

- structural probes to be employed **in combination with MS detection**. Both enzymes were found capable of operating in ammonium-based solutions that are preferred for high-resolution analysis by direct infusion electrospray ionization (ESI).

SL3



MS/MS for isobaric products with identical base composition and different sequences

tRNA

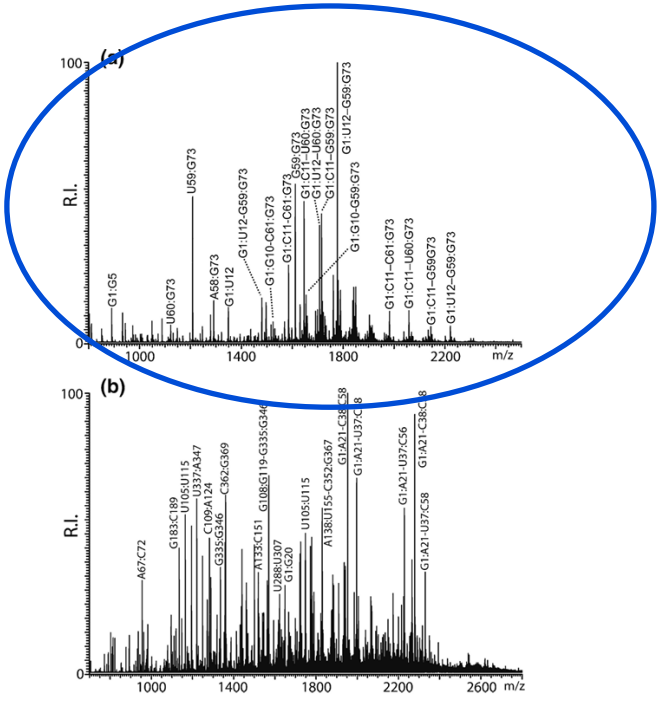


Figure 1.
ESI-MS spectra of digestion mixtures obtained by treating SL3 RNA with (**a**) Mung bean and (**b**) V1 nuclease. The insets illustrate the location of the respective cleavage sites. Dashes are used to indicate duplex products

Figure 4.
Representative ESI-MS spectra of samples obtained by Mb digestion of **(a)** human $tRNA^{Lys3}$ and **(b)** HIV-1 5′-UTR. Dashes are used to indicate duplex products

Structure specific RNAse + MS

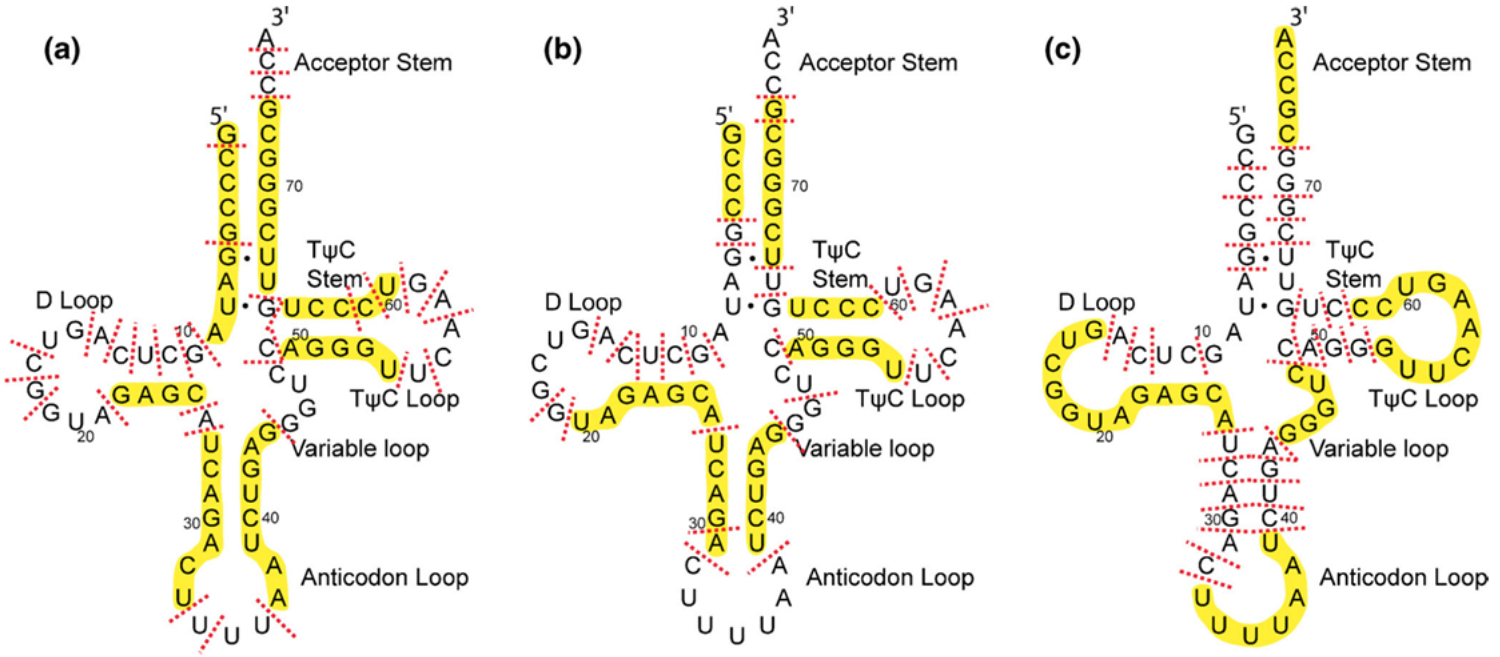
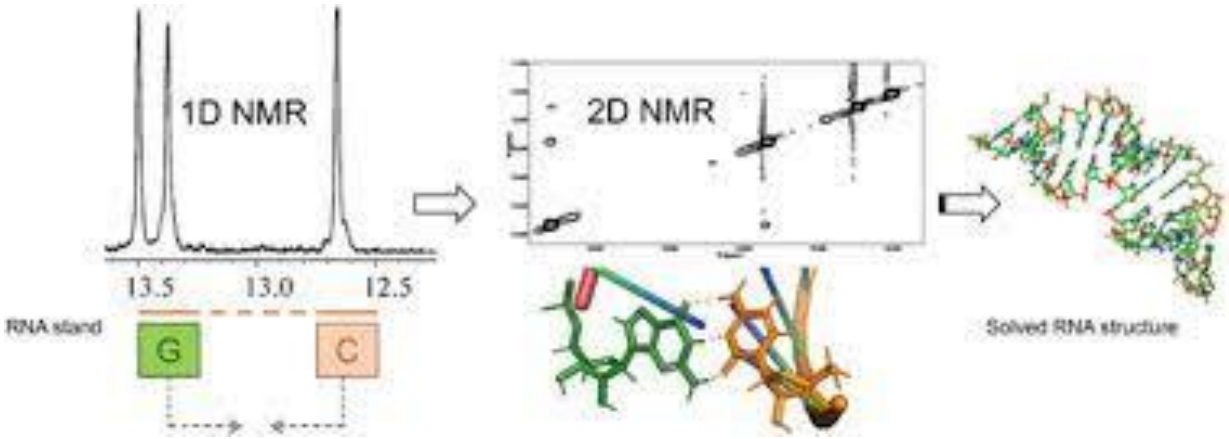


Figure 5.
Cleavage maps obtained by treating human $tRNA^{Lys3}$ with **(a)** Mung bean nuclease for 5 h, **(b)** overnight, and **(c)** V1 nuclease overnight (see Supplementary Table 2S). Dashed lines mark cleavage sites. Highlighted in yellow are regions marginally affected by hydrolysis

How to study 3D RNA structure ???

Studying the 3D structures of RNA poses greater challenges compared to proteins, primarily due to the inherent flexibility of RNA molecules.

- NMR



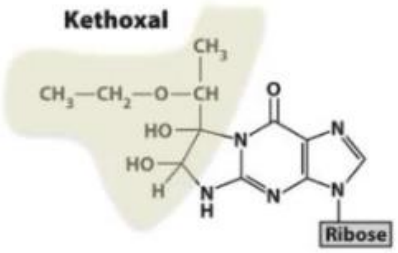
- X-ray crystallography

- cryo-electron microscopy (cryo-EM)
- small angle X-ray scattering (SAXS)

- structural probing reactions + analytical analysis (MS)

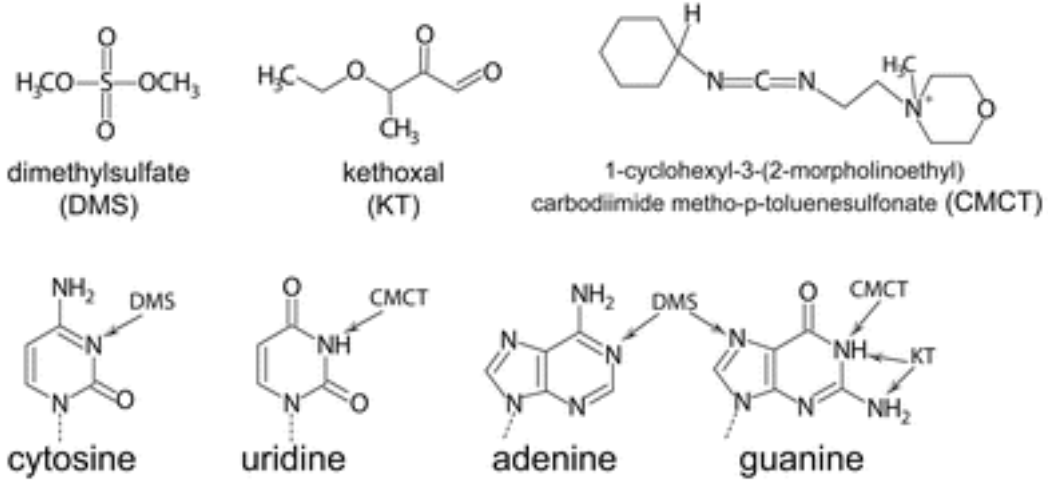
Monovalent probes

Kethoxal
(KT) to probe N1 and N2 of unpaired guanines



1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide (**CMCT**) used for N3-uridine

Dimethyl sulfate (**DMS**) for N1-adenine and N3-cytidine, and for N7-guanines in ssRNA

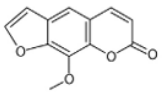


Divalent probes

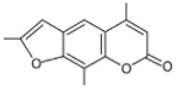
Table 1
Names and chemical structures of test compounds included in the study.

Psoralens

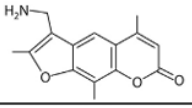
8-methoxypsoralen (8-MOP)



trioxsalen

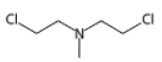


4'-aminomethyltrioxsalen (AMT)

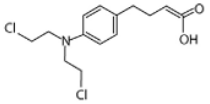


2-Haloethylamines

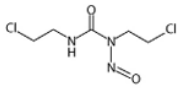
nitrogen mustard (NM)



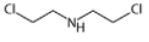
chlorambucil



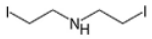
carmustin



bis(2-chloroethyl)amine

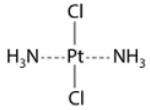


bis(2-iodoethyl)amine

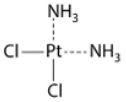


Platinated complexes

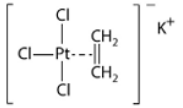
transplatin



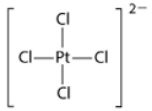
cisplatin



potassium trichloro(ethylene) platinate(II) hydrate



tetrachloroplatin

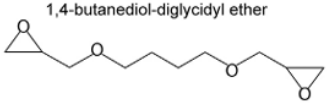


Epoxides

epibromohydrin

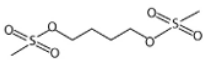


1,4-butanediol-diglycidyl ether



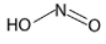
Alkyl sulfonate

busulfan



Others

nitrous acid

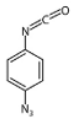


formaldehyde

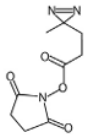


Photo-activatable

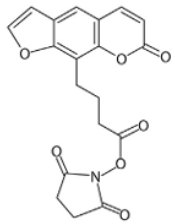
4-azidophenyl isothiocyanate



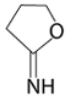
succinimidyl-4,4'-azidopentanoate (SDA)



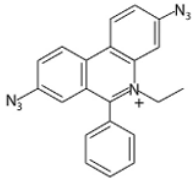
succinimidyl-[4-(psoralen-8-yloxy)]butyrate (SPB)



2-imino-thiolane

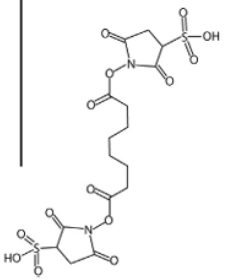


ethidium diazide



Others

bis(succinimidyl)-suberate (BS3)



Divalent probes

Table 1. Characteristics of candidate bifunctional probes

Name	Base specificity[a]	Δm (Da)[b]	Adduct stability[c]	Other products[d]	Crosslinking distance (Å)[e]	Probe/ substrate[f]	Other reaction conditions
BKT	G	230.079	+	L	4.4–6.4	5:1	20mM ammonium borate (pH 8.0), 37°C, 2h
PDG	G	190.027	+	L	7.0–8.0	5:1	20mM ammonium borate (pH 8.0), 37°C, 2h
CPT	G>A>C	225.995	+++	B	3.2	2:1	10mM ammonium cacodylate pH6, 37°C, 24h
NM	G>A>C	83.073	++	H, D	8.0–11.0	9:1	10mM ammonium acetate/cacodylate (pH 8.0), 37°C, 2h
CHB	G>A>C	231.126	++	H, D	8.0–11.0	9:1	10mM ammonium acetate/cacodylate (pH 8.0), 37°C, 8h
sTT	G	110.962	++	H, D	2.4	60:1	50mM ammonium borate, (pH 6.0), 37°C, 2h

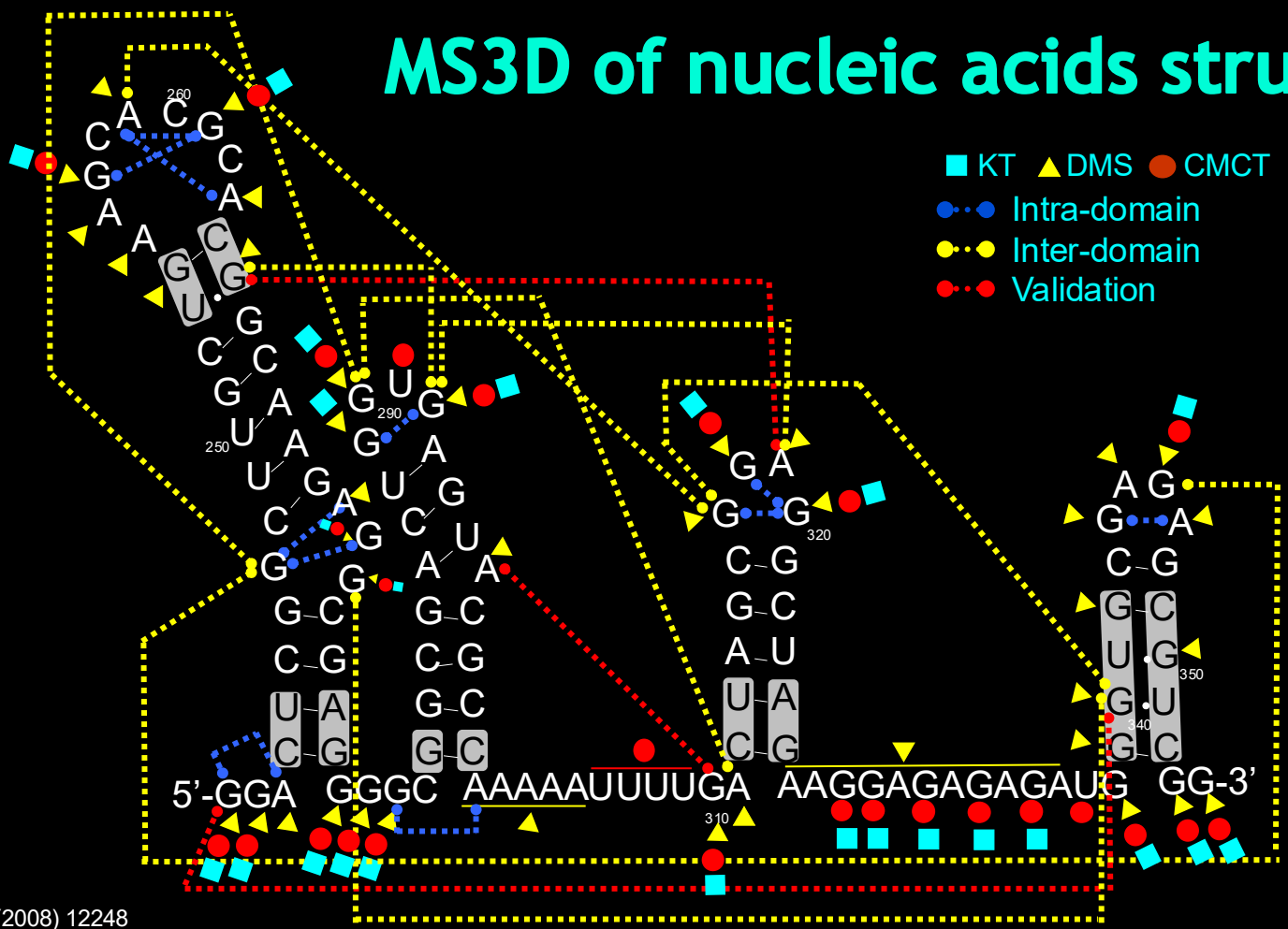
[a]Preferential targets determined under the described conditions (see text for the position of alkylation).
[b]Incremental mass of bifunctional adduct.
[c]Stability estimated at the conditions provided: +++ very stable; ++ stable; + somewhat reversible.
[d]Additional products: B buffer interaction (e.g., acetate coordination); D depurination; H hydrolysis of reagent or mono-functional adduct; L water loss from adduct.
[e]Distance between crosslinked functions in final adduct. A range is provided for the more flexible reagents.
[f]The probe/substrate ratio was calculated in moles of reactive functions in the probe per moles of susceptible sites in the substrate.

Divalent probes

Table 2
Evaluation of putative cross-linking reagents by using selected nucleic acid and protein-nucleic acid substrates (see *Experimental*). The production of inter-molecular crosslinks was assessed by using both denaturing PAGE and MS analysis. The observed performance was rated as excellent (✓✓✓), acceptable (✓✓), limited (✓), or negligible (×). The "n.d." abbreviation indicates reagents that did not display detectable reactivity under the explored conditions.

Name	Class	DNA type 0	RNA type 0	Protein type 0	DNA-DNA crosslink	RNA-RNA crosslink	Prot.-RNA crosslink
8-Methoxypsoralen	Psoralens	✓✓✓	✓	n.d.	✓✓✓	×	×
Trioxsalen	"	✓✓	✓	n.d.	n.d.	×	n.d.
4′-Aminomethyltrioxsalen hydrochloride	"	✓✓✓	✓✓✓	n.d.	×	×	×
Transplatin	Platinum complexes	✓✓✓	✓✓✓	✓✓	✓✓✓	✓✓✓	✓✓✓
Cisplatin	"	✓✓✓	✓✓✓	✓✓	✓✓✓	✓✓✓	✓
Trichloro(ethylene) platinate(II)	"	n.d.	✓✓✓	n.d.	n.d.	✓✓✓	✓✓✓
Tetrachloro platin	"	n.d.	✓✓✓	✓✓	n.d.	✓✓✓	✓✓✓
Nitrogen mustard	Mustard	n.d.	✓✓✓	✓✓	n.d.	✓✓✓	✓✓✓
Chlorambucil	"	n.d.	✓✓✓	n.d.	n.d.	✓✓✓	✓✓
Carmustine	"	n.d.	✓✓	n.d.	n.d.	✓✓	✓✓✓
Bis(2-chloroethyl)amine	Alkylamine	n.d.	✓✓	✓✓	n.d.	n.d.	✓✓
Bis(2-iodoethyl)amine	"	n.d.	✓✓	✓✓	n.d.	n.d.	✓✓
Epibromo hydrin	Epoxides	n.d.	×	×	n.d.	×	×
1,4 Butanediol diglycidyl ether	"	n.d.	×	×	n.d.	×	×
Busulphan	Alkyl sulphonate	n.d.	×	×	n.d.	n.d.	×
Nitrous acid	Nitrous acid	n.d.	×	×	n.d.	n.d.	×
Formaldehyde	Aldehyde	n.d.	✓✓	✓✓	n.d.	n.d.	✓✓✓
Ethidium diazide	Photo-cross-linker	n.d.	n.d	n.d	n.d.	×	×
p-Azidophenyl isothiocyanate	"	n.d.	n.d	✓✓✓	n.d.	n.d.	×
SDA (succinimidyl 4,4-azipentanoate)	"	n.d.	×	✓✓✓	n.d.	×	×
Succinimidyl-[4-(psoralen-8-yloxy)]-butyrate (SPB)	"	n.d.	n.d	✓✓✓	n.d.	×	×
2-Iminothiolane	"	n.d.	×	✓✓	n.d.	n.d.	×
Bis(sulfosuccinimidyl) suberate	Activated carboxylic acid	n.d.	×	✓✓✓	n.d.	×	×

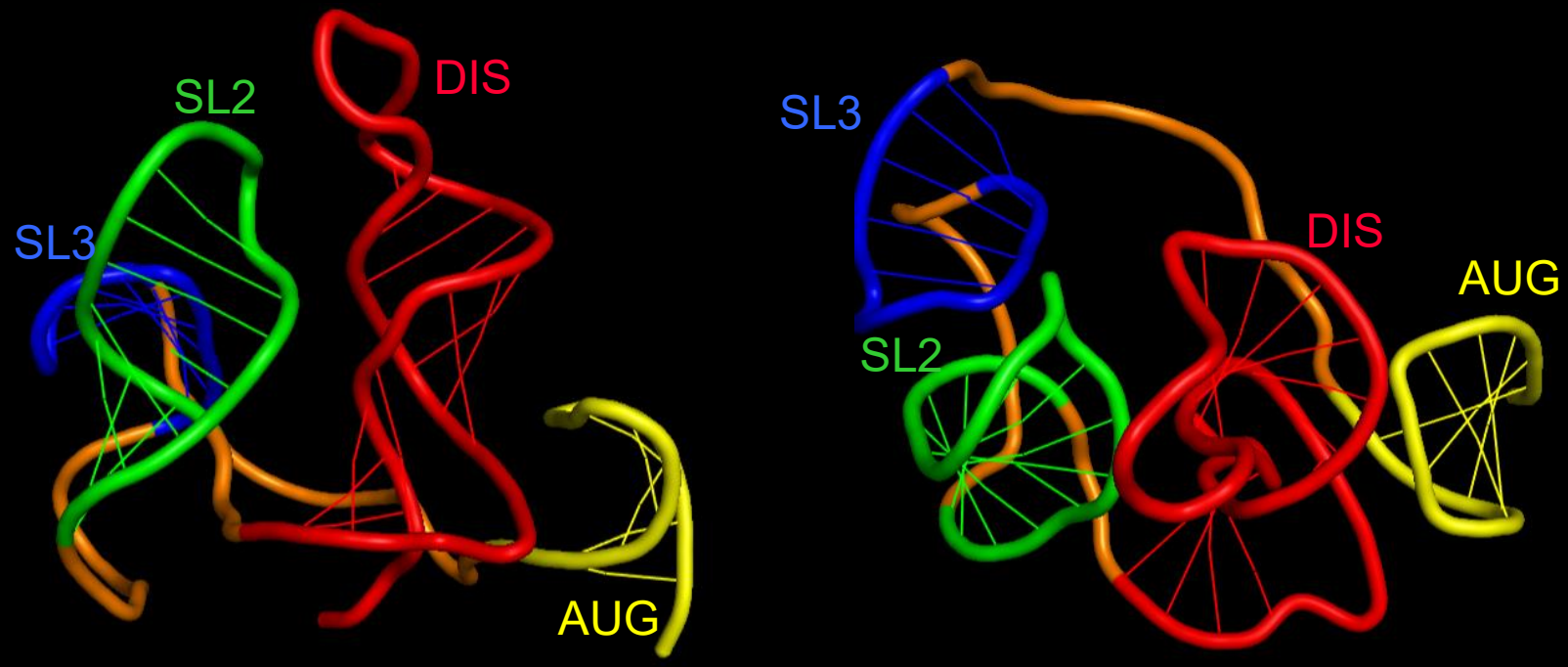
MS3D of nucleic acids structures



Yu *et al.*,***PNAS*** *105* (2008) 12248

Fabris D. et al. J Mass Spectrom 45 (2010) 841-860

MS3D of nucleic acids structures



Yu *et al.*,***PNAS*** *105* (2008) 12248

Fabris D. et al. J Mass Spectrom 45 (2010) 841-860