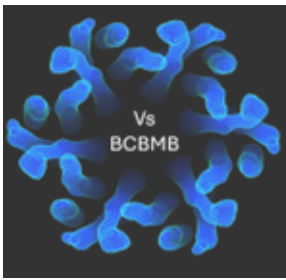


Advanced Biochemistry

NUCLEIC ACIDS



This is a Tutorial/Silent Lecture



What is a Tutorial/Silent Lecture?

a sequence of "slides" formatted to guide you through the exploration/study of the topic

you are the main actor in this active learning experience

think of it as working with a tutor without having to pay for it

as the slide sequence unfolds, you will get opportunities to engage with the material

➤ **by thinking about/answering questions,**

(my answer is always provided on the next slide).

➤ **by completing a "short assignment"**

(it never will take more than a few minutes, if at all that long),

➤ **by watching a short video/clip**

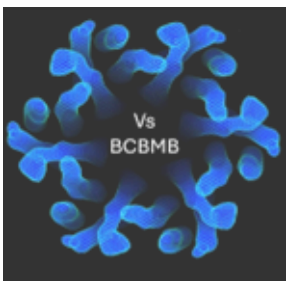
(the embedded links will take you to my YouTube@VsBCBMB channel or other creators;
key moments are shown in the slide-deck as still images, in case you don't want to watch the videos)

of course, you can skip the active learning aspect and look at the answers right away.

Why Give This a Go?

- **benefits: you set the pace** taking as much or as little time as you need.
- you **can turn tutorials/silent lectures into fully immersive experiences** (eg playing your favourite music while working through the content),
- **or invite friends to over the Q&A structured/guided materials together**, discussing the questions before looking at answers.

each of these features help you to hold on to the material.



Advanced Biochemistry



this collection of handouts builds on the "Biochemistry Fundamentals" collection
= the chapters assume that you know the basics covered in the "Fundamentals" tutorials (free downloads), and have some basic knowledge of molecular/cell biology

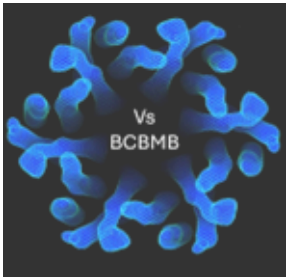
this tutorial can be used stand alone

but does reference other chapters

therefore, you will get the largest benefits if you

- work through other chapters in the order that they are posted in the "download gallery"
- **review the associated "Biochemistry Fundamentals" chapter** to refresh your memory
- spend 5-10 minutes to **summarize for yourself what you already know/remember** about the topics of the handout you are about to look at.
 - **take advantage of the "interactive" elements**

I welcome your thoughts and ideas for further improvements of the chapters. You can submit your comments by contacting me at pdf-comments@vsbcbmbstudy.com



Setting The Stage



following a narrative that builds **life from scratch**

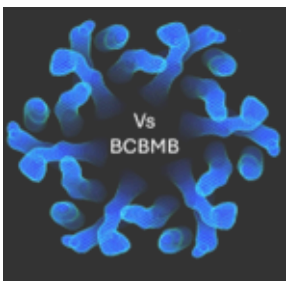
- we started off by **retracing the origins of simple biological macromolecules** and the emergence of polymers and cells
(chapter: Chemical Evolution)
- took reasonably detailed **look at how Nature succeeded in compartmentalizing space** by exploiting chemical properties of amphiphiles to build flexible membranes
(chapter Lipids and Membranes)
- caught up on and **extended what you know about proteins**, this time including membrane proteins that are needed to make membranes semi-permeable
(chapter: Proteins)
- and built a reasonably **detailed core knowledge of membrane associated transport processes**
(chapter: Membrane Transport)

continuing this narrative, we now

- **need a way to store information** about “cell identity, content and dynamics”
- **need a way to communicate with other units** and to interpret environment
 - **need a way to power the unit** and to manufacture the components

having been in that place before (**Biochemistry Fundamentals**), we spent a great deal of attention to **derive why nucleic acids are an ideal – in fact, the only – solution to large scale information storage and processing.**

back then, however, we very focused on understanding the molecular logic behind the design of nucleic acids - skipping core aspects related to the building blocks of nucleic acids, and the levels of structural organization in these important biological macromolecules

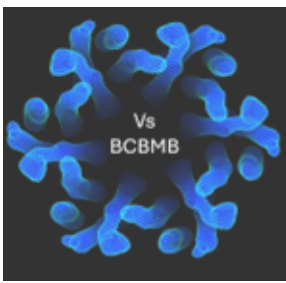


Goals of This Tutorial/Silent Lecture



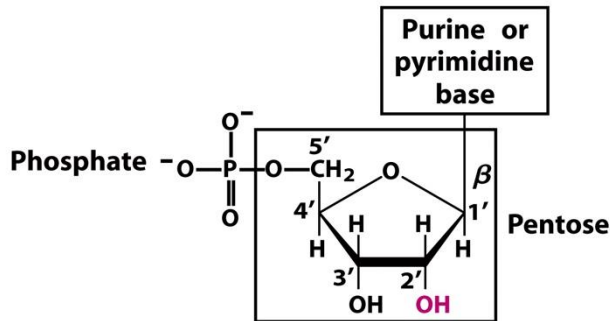
by the end of this tutorial

- ✓ you will have a **more complete understanding of the structure and chemical modifications of the molecular building blocks for nucleic acids**
- ✓ you will have learned about **subtle but important differences between DNA and RNA**
- ✓ you will have a **better understanding** of the hierarchy in **structural organization of nucleic acids** (1°, 2°, 3°, 4° structure)
- ✓ you will **understand how Nature overcame the challenges associated with selecting certain RNAs to play essential roles** that require RNA to adopt a defined structure.
- ✓ you will have **reviewed and compared the basic design principles of all biologically relevant polymer types**

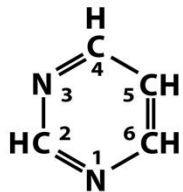


Nucleic Acids – Core Knowledge

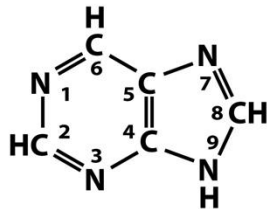
Nucleobases, Nucleosides, Nucleotides



- the building blocks of nucleic acids are composite molecules
heteroaromatic base + (deoxy)pentose = **nucleoside**
- addition of one or more phosphates creates the related family of **nucleotides**
- positions in base are numbered following IUPAC convention
- positions in pentose have a prime “'” added to distinguish them from positions in base. **anomeric carbon is 1'**



Pyrimidine



Purine

- linkage is a **N-glycosidic bond with N1 (pyrimidine) or N9 (purine)**

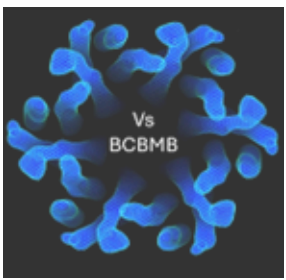
(if you are not familiar with the term "glycosidic bond", review CARBOHYDRATE Chapter, slides 20-23)

the **presence/absence of a hydroxyl at C2'** in the sugar divides two classes of nucleosides/nucleotides:

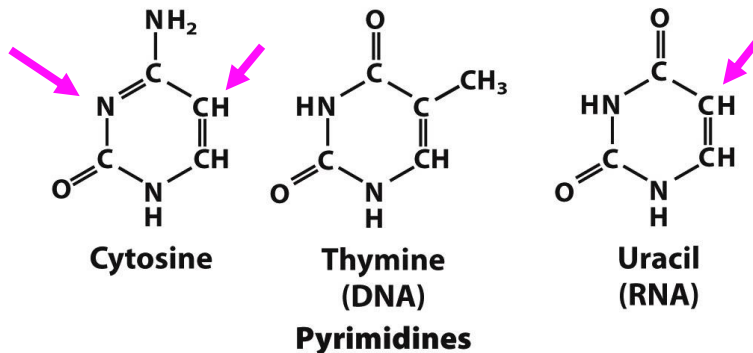
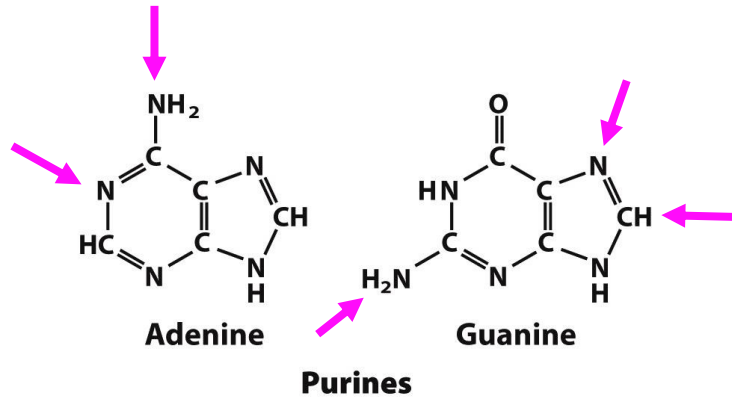
regular (ribose) → RNA

deoxy (C2' -deoxyribose) → DNA

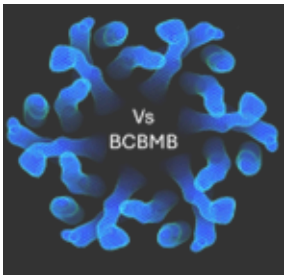
(recall: lack of the 2' -OH group increases the stability of the polymer backbone at neutral pH, "CHEMICAL EVOLUTION" Chapter, slide 69)



Chemical Modifications of Nucleobases - Overview

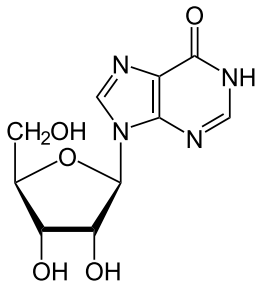
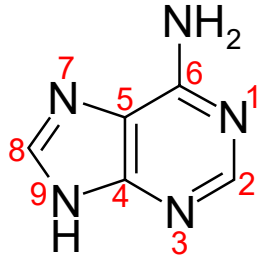


- arrows indicate positions that can be modified (most common: adding a methyl-group)
- modifications can occur during or after transcription/replication, and in cases like tRNA and rRNA are sequential
- modifications occur in all nucleic acids and are always functionally important
for example: facilitate recognition by DNA binding proteins, or protect DNA from degradation
- other minor bases exist; they have very specific functions.




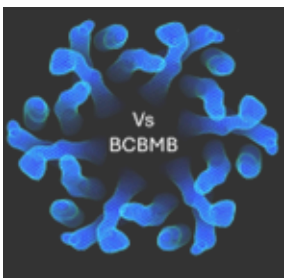
Chemical Modifications of Nucleobases

The Details: Adenine



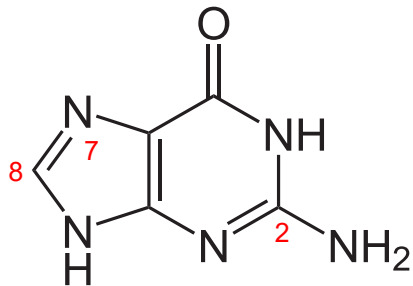
Inosine


Nucleobase 	Modification Name	Abbreviation	Common Roles & Significance
Adenine (A)	<i>N</i> ⁶ -methyladenosine	m6A	Most abundant mRNA modification; regulates splicing, transport, and stability.
	<i>N</i> ¹ -methyladenosine	m1A	Affects RNA decay and secondary structure; found in tRNA and mRNA 5'-UTRs.
	Inosine	I	Product of A-to-I editing; alters base-pairing (pairs with C) and protein coding.
	<i>N</i> ⁶ -isopentenyladenosine	i6A	Found in tRNA; enhances translation fidelity and efficiency.

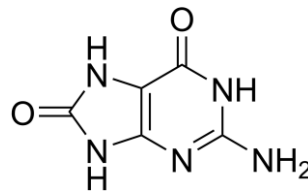


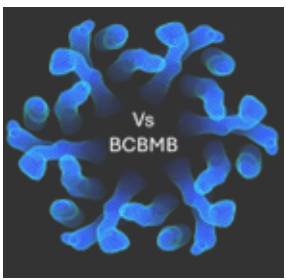
Chemical Modifications of Nucleobases

The Details: Guanine



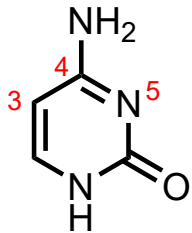
Nucleobase 	Modification Name	Abbreviation	Common Roles & Significance
Guanine (G)	7-methylguanosine	m7G	Essential part of the mRNA 5'-cap; also found internally in tRNA.
	<i>N</i> ² -methylguanosine	m2G	Contributes to proper tRNA folding and stability.
	8-oxoguanine	8-oxoG	Common product of oxidative DNA damage; can cause mispairing with Adenine.


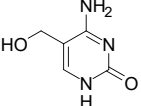


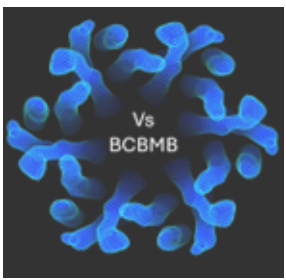


Chemical Modifications of Nucleobases

The Details: Cytosine

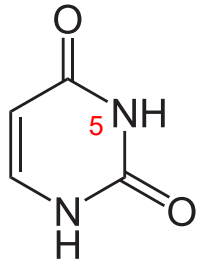



Nucleobase 	Modification Name	Abbreviation	Common Roles & Significance
Cytosine (C)	5-methylcytosine	5mC / m5C	Primary DNA epigenetic mark (gene silencing); also found in tRNA and rRNA.
	5-hydroxymethylcytosine	5hmC / hm5C	DNA demethylation intermediate; enriched in active brain genes.
	<i>N</i> ⁴ -acetylcytidine	ac4C	Found in tRNA and mRNA; enhances translation efficiency and stability.
	3-methylcytidine	m3C	Stabilises tRNA structure; found in the anticodon loop.

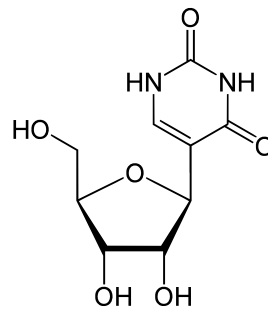


Chemical Modifications of Nucleobases

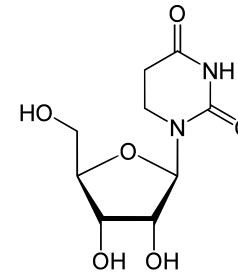
The Details: Uracil



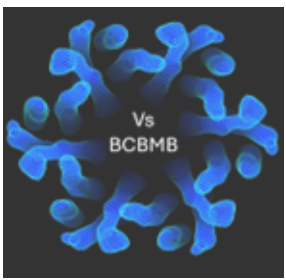
Nucleobase 	Modification Name	Abbreviation	Common Roles & Significance
Uracil (U)	Pseudouridine	Ψ	Most abundant RNA modification; affects mRNA translation and tRNA stability.
	Dihydrouridine	D	Increases RNA flexibility and destabilises base stacking; common in tRNA.
	5-methyluridine	m5U	Also known as ribothymidine; common in the T-loop of tRNAs.



pseudouridine

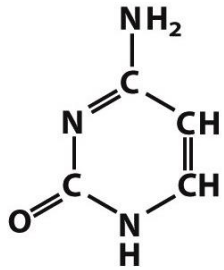


dihydrouridine

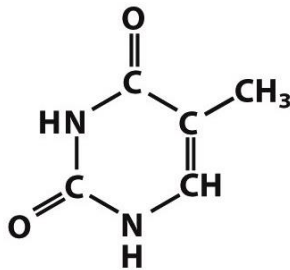


Nucleobase Differences

DNA vs RNA

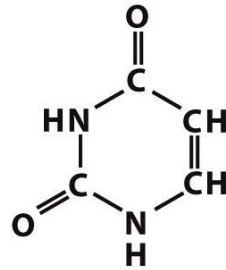


Cytosine



**Thymine
(DNA)**

Pyrimidines



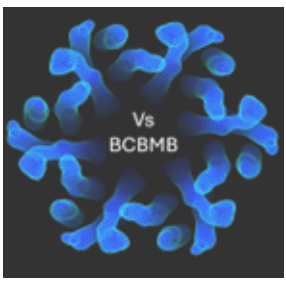
**Uracil
(RNA)**

**a curious question
why does DNA contain Thymine
instead of Uracil?**

...can you see the answer? ...

hint: water is everywhere

hint2: C pairs with U pairs with



Nucleobase Differences

DNA vs RNA

a curious question
why does DNA contain Thymine
instead of Uracil?

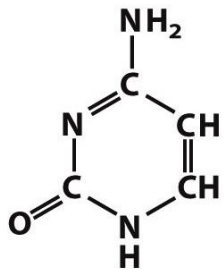
Answer:

➤ **spontaneous deamination of cytosine yields uracil and occurs in 1 out of 10^7 /day**
(~100/day in humans)

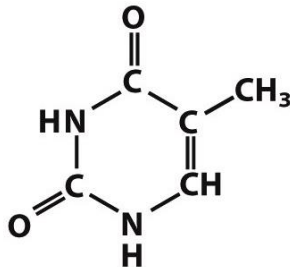
➔ **use of T instead of U in DNA unambiguously marks U as mutation**, which will activate the DNA repair machinery that excises U and restores C.

if not corrected, the C→U change will lead to a permanent transition

because C→U causes the C:::G pair to become a T:::A pair in half of the progeny
(that may not be too much of an issue in somatic cells, but if it hits germline cells, mutations will be inherited to offspring)

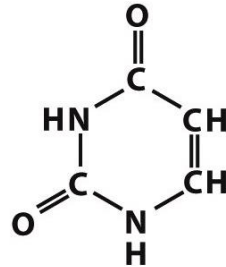


Cytosine



**Thymine
(DNA)**

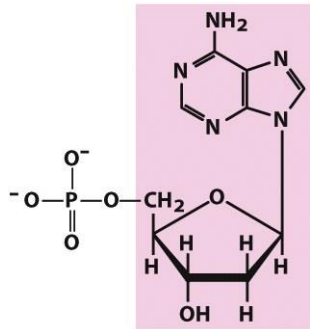
Pyrimidines



**Uracil
(RNA)**

closing out the chemistry of building blocks ... let's take a look at the full cast of standard building blocks that contribute to DNA and RNA synthesis....

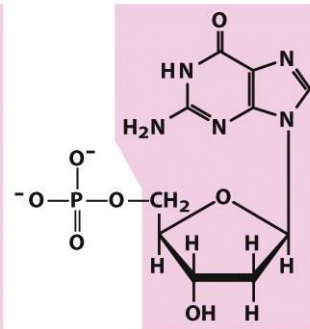
Nucleotides - The Full Cast of Characters



Nucleotide: Deoxyadenylate
(deoxyadenosine
5'-monophosphate)

Symbols: A, dA, dAMP

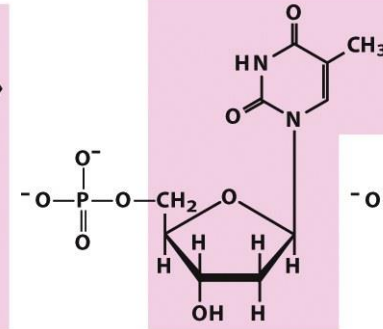
Nucleoside: Deoxyadenosine



Nucleotide: Deoxyguanylate
(deoxyguanosine
5'-monophosphate)

Symbols: G, dG, dGMP

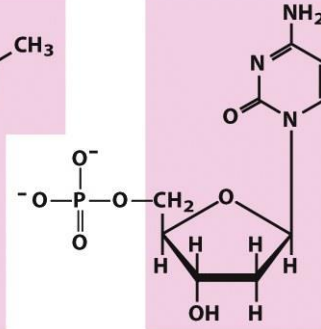
Nucleoside: Deoxyguanosine



Nucleotide: Deoxythymidylate
(deoxythymidine
5'-monophosphate)

Symbols: T, dT, dTMP

Nucleoside: Deoxythymidine

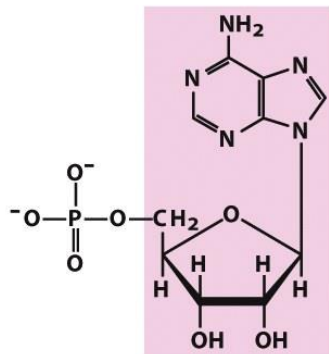


Nucleotide: Deoxycytidylate
(deoxycytidine
5'-monophosphate)

Symbols: C, dC, dCMP

Nucleoside: Deoxycytidine

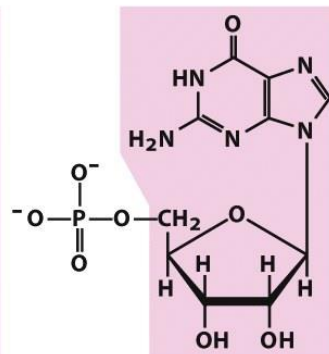
Deoxyribonucleotides



Nucleotide: Adenylate (adenosine
5'-monophosphate)

Symbols: A, AMP

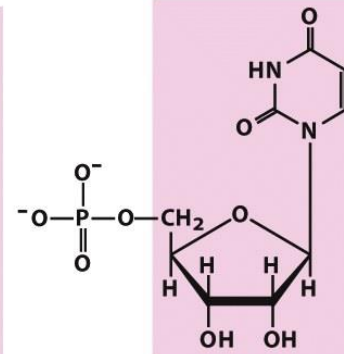
Nucleoside: Adenosine



Nucleotide: Guanylate (guanosine
5'-monophosphate)

Symbols: G, GMP

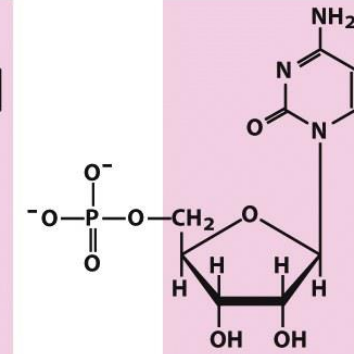
Nucleoside: Guanosine



Nucleotide: Uridylate (uridine
5'-monophosphate)

Symbols: U, UMP

Nucleoside: Uridine



Nucleotide: Cytidylate (cytidine
5'-monophosphate)

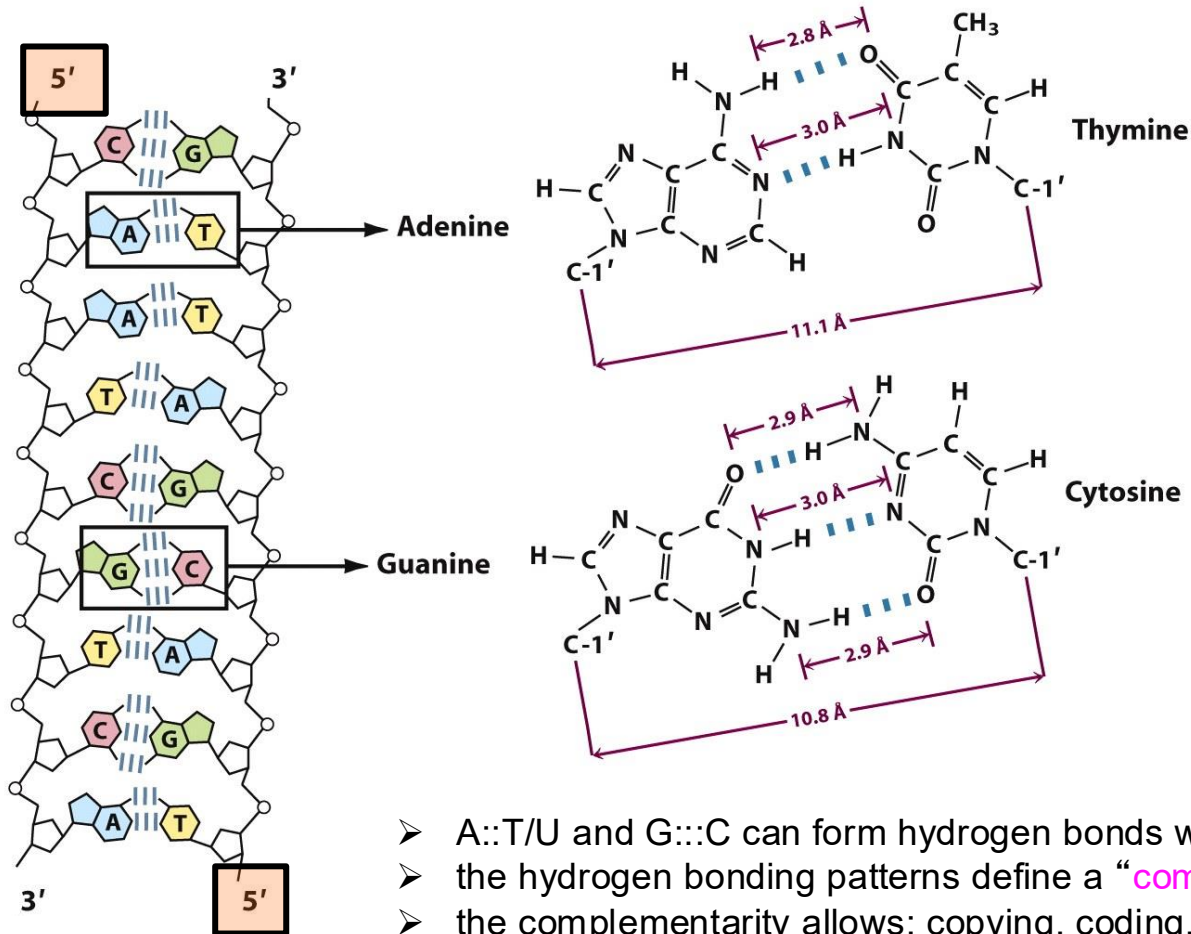
Symbols: C, CMP

Nucleoside: Cytidine

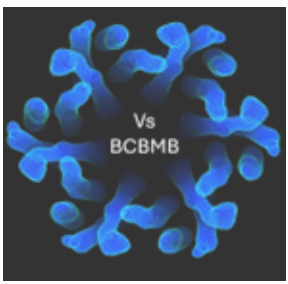
Ribonucleotides

The Molecular Basis of Heredity And Coding

(a more detailed derivation of the points can be found in the Biochemistry Fundamentals – NUCLEIC ACIDS chapter)



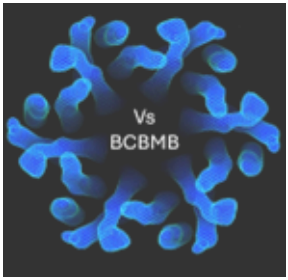
- A::T/U and G::C can form hydrogen bonds which each other
- the hydrogen bonding patterns define a “**complementarity**”
- the complementarity allows: copying, coding, preservation
- **note: the strands are antiparallel!**
- **new here:** Breaking with how 2^o/3^o/4^o structure were defined in proteins and carbohydrates, **we will refer to this double stranded arrangement as 2^o structure** (the reasons for that will become obvious later)



Guess What Comes Now?

*....how well have you caught on to
the way of thinking about
macromolecular structure?...*





Guess What Comes Now?

energetics ...because as you learned for proteins, the marginal stability of their native fold is key to proteins being able to function as catalysts....so, it is **curious to ask what stability can teach us about nucleic acid structure and function**

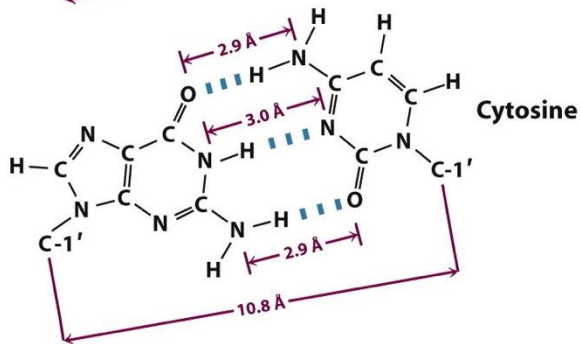
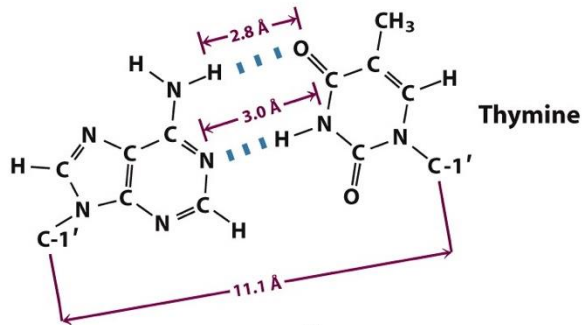


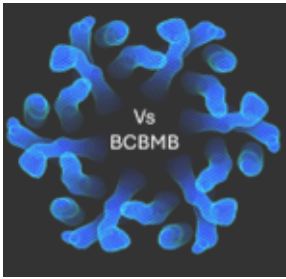
Stabilization of an AT pair?

...look below and tell ...!

Stabilization of an GC pair?

...look below and tell ...!





Guess What Comes Now?

energetics ... (of course) ...because as you learned in the case of proteins, their marginal stability of the native fold is key to proteins being able to function as catalysts....so, it is curious to ask what looking at stability can teach us about nucleic acid structure and function



Stabilization of an AT pair?

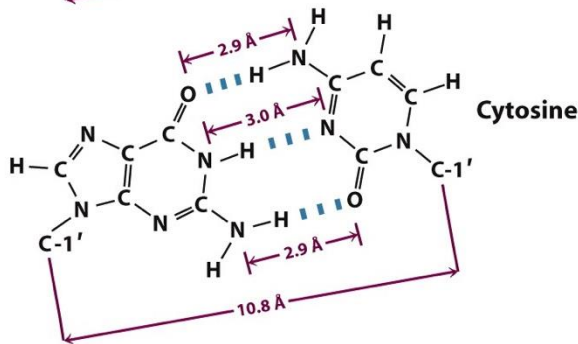
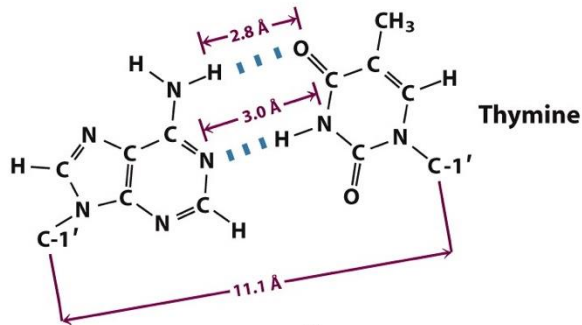
~5-8kcal/mol

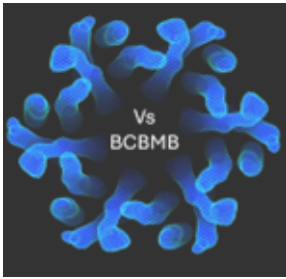
Stabilization of an GC pair?

~8-12kcal/mol

What do these numbers mean to you?

....put down a few thoughts....





Guess What Comes Now?

energetics ... (of course) ...because as you learned in the case of proteins, their marginal stability of the native fold is key to proteins being able to function as catalysts....so, it is curious to ask what looking at stability can teach us about nucleic acid structure and function



Stabilization of an AT pair?

~5-8kcal/mol

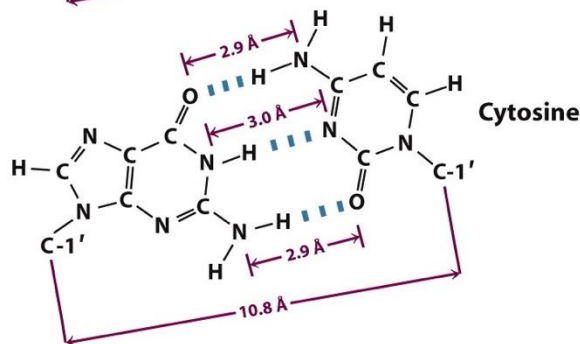
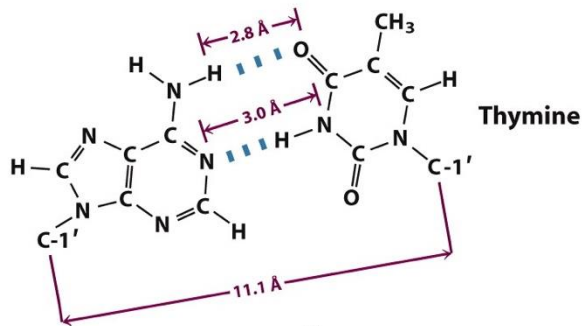
Stabilization of an GC pair?

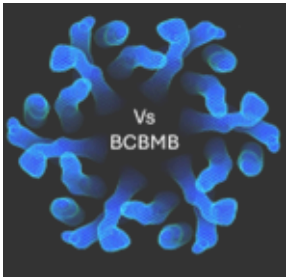
~8-12kcal/mol

comparison: protein fold is stabilized by ~10kcal/mol ... which is minute compared to the stability of a DNA double strand!!

→ to “melt”/denature DNA you need very high temperatures (as you will know if you ever have run a PCR)

but without going any further, there is more here ...
what can you deduce about the structure of DNA/RNA by just looking at the structure of a base pair?





Guess What Comes Now?



energetics ... (of course) ...because as you learned in the case of proteins, their marginal stability of the native fold is key to proteins being able to function as catalysts....so, it is curious to ask what looking at stability can teach us about nucleic acid structure and function

Stabilization of an AT pair?

~5-8kcal/mol

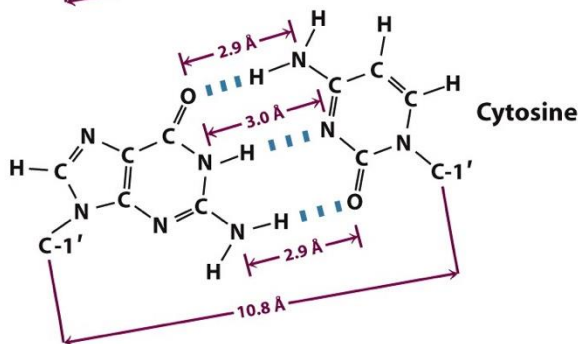
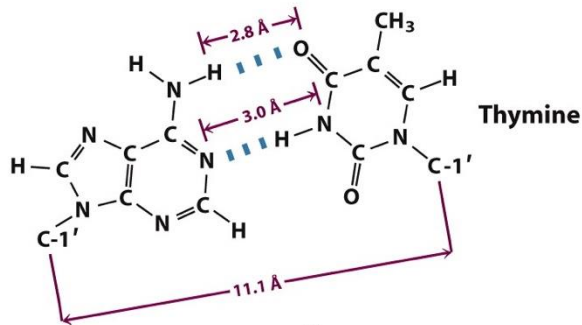
Stabilization of an GC pair?

~8-12kcal/mol

comparison: protein fold is stabilized by ~10kcal/mol ... which is minute compared to the stability of a DNA double strand!!

→ to “melt”/denature DNA you need very high temperatures (as will know if you ever have run a PCR)

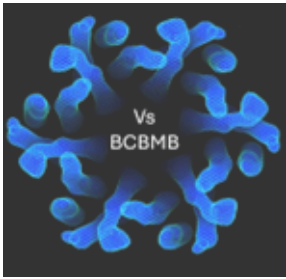
but without going any further, there is more here ...
what can you deduce about the structure of DNA/RNA by just looking at the structure of a base pair?



Answer

hydrogen bonds are co-axial → since the aromatic rings are planar, **the base pair is essentially flat**
 → in the 3D-structure, these “planes” **will want to stack because the stacking buries hydrophobic surfaces and compensates for low electron density in the pyrimidines.**

How does this tie in with concepts of 2°/3° structure formation in the other biopolymers?



Guess What Comes Now?

energetics ... (of course) ...because as you learned in the case of proteins, their marginal stability of the native fold is key to proteins being able to function as catalysts....so, it is curious to ask what looking at stability can teach us about nucleic acid structure and function

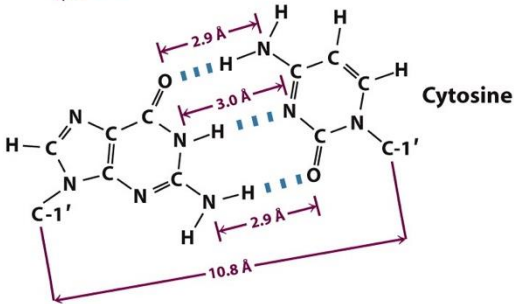
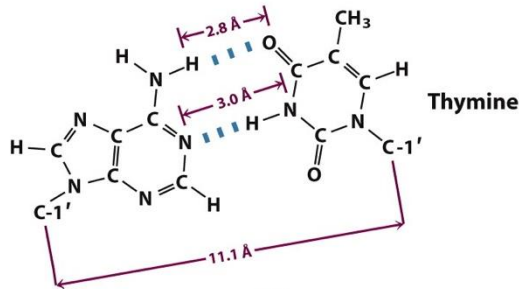


Stabilization of an AT pair?

~5-8kcal/mol

Stabilization of an GC pair?

~8-12kcal/mol



comparison: protein fold is stabilized by ~10kcal/mol ... which is minute compared to the stability of a DNA double strand!!

→ to “melt”/denature DNA you need very high temperatures (as all of you that have run a PCR will know)

but without going any further, there is more here ...

what can you deduce about the structure of DNA/RNA by just looking at the structure of a base pair?

Answer

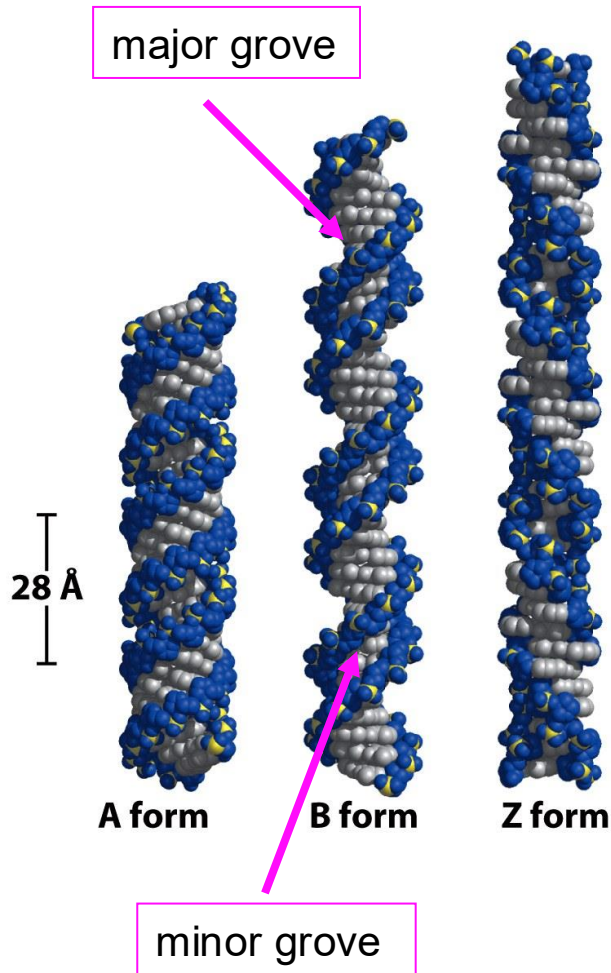
hydrogen bonds are co-axial → since the aromatic rings are planar, the base pair is essentially flat

→ in the 3D-structure, these “planes” will want to stack because the stacking buries hydrophobic surfaces and compensates for low electron density in the pyrimidines.

How does this tie in with concepts of 2°/3° structure formation in the other biopolymers?

Similar to proteins: hydrophobic effect and weak interactions drive 2°/3° structure formation
Less similar to carbohydrates: mostly weak interactions (hydrogen bonding)

DNA - Common 2° Structures - Stats



	A form	B form	Z form
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20 Å	~18 Å
Base pairs per helical turn	11	10.5	12
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Sugar pucker conformation	C-3' endo	C-2' endo	C-2' endo for pyrimidines; C-3' endo for purines
Glycosyl bond conformation	Anti	Anti	Anti for pyrimidines; syn for purines

protein α -helix

Right handed
~10Å
(including sidechains)
3.6 aa/turn

1.5Å/aa

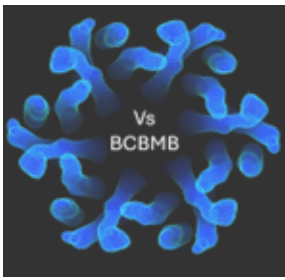
- of the three forms: **B-form is the most prevalent secondary structure of DNA**
- A and Z-form exist *in vivo* but are rare and temporary
- high negative charge from phosphates is shielded by **Mg²⁺**, polyamines and basic regions of proteins

(note: use of Mg²⁺ to bridge and screen negative charges from phosphate is a recurrent motif – we first encountered it when discussing the structure of phospholipid bilayers (LIPIDS and MEMBRANES, slide 78))

alternative secondary structures are possible: cruciform, triple stranded DNA, quadruplex and i-motif DNA

but we won't cover them here because their use in biology is very narrow.

If you interested in these structures, click on the links above which will take you to the corresponding "wiki page"



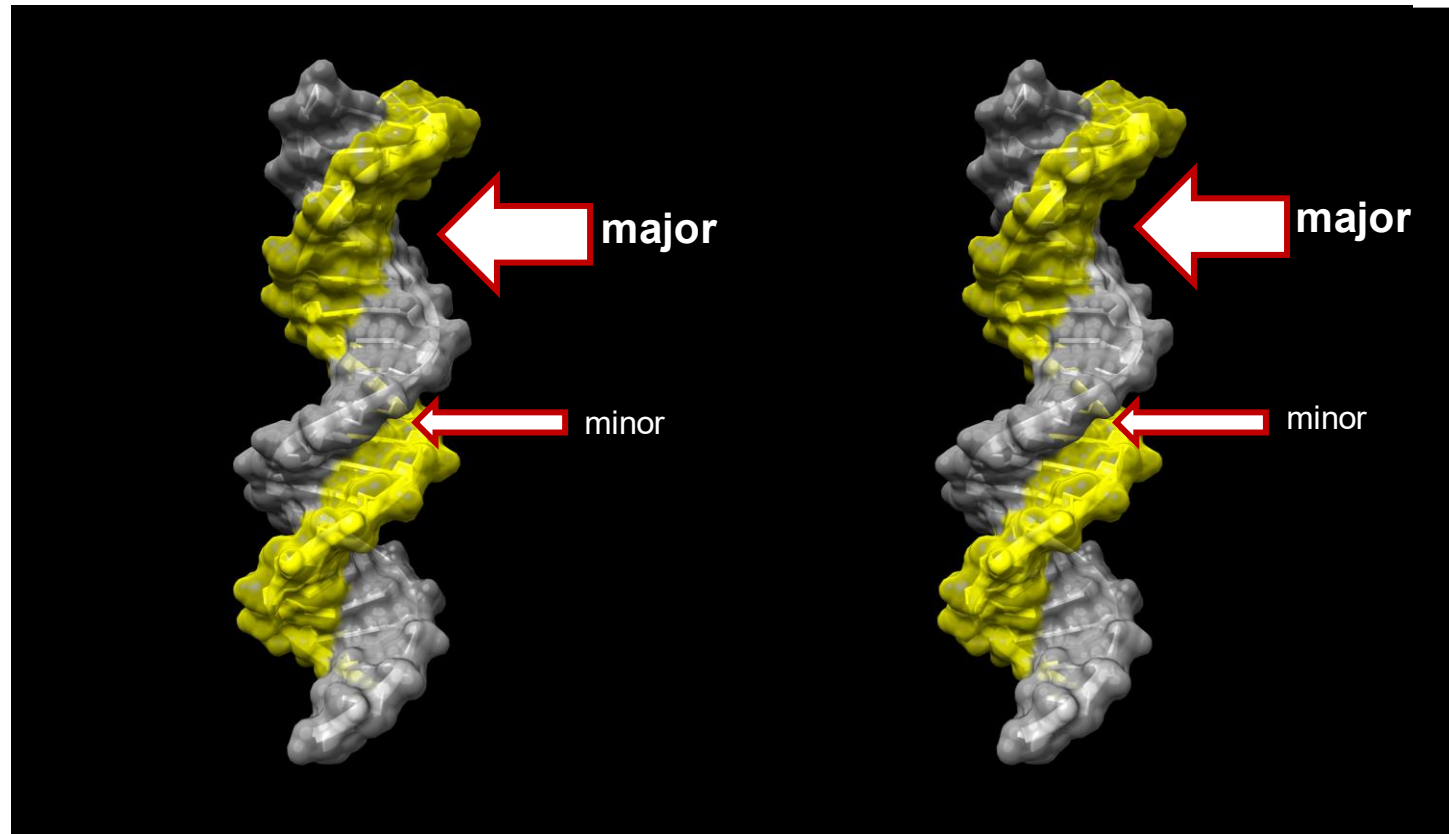
The Double Helix

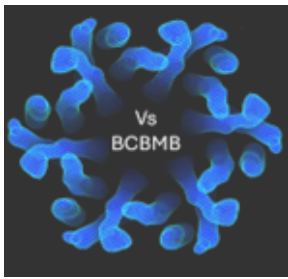
because B-form DNA is so central for life, let's look at its 3D structure again with a stereo pair

(taken from Biochemistry Fundamentals – NUCLEIC ACIDS, slide 40)



These are cross-eyed stereo images – hold ~30cm away from you, cross your eyes and adjust your eyes + head tilt until you achieve 3D view





DNA – Flashback Chromatin Structure ..New Pitch.....

in the Chromatin Chapter of the Molecular and Cell Biology Tutorial Series, you were introduced to the principle of "supercoiling" (slides 7-9), and how cells exploit this idiosyncratic elastic property of the DNA double helix to form complex compacted structures of the polymer.

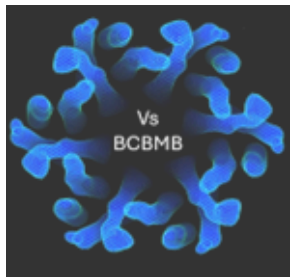


how does DNA topology (= state of supercoiling) fit into the scheme of organizational levels that are associated with biological polymers?



Answer:

DNA topology/supercoiling is the trigger for the formation of DNA 3° structures (beads on string, 30nm fiber, loops, rosettes, - next slide).



DNA – Flashback Chromatin Structure...A New Angle



in the Chromatin Chapter of the Molecular and Cell Biology Tutorial Series, you were introduced to the principle of "supercoiling" (slides 7-9), and how cells exploit this idiosyncratic elastic property of the DNA double helix to form complex compacted structures of the polymer.

how does DNA topology (= state of supercoiling) fit into the scheme of organizational levels that are associated with biological polymers?



Answer:

DNA topology/supercoiling is the trigger for the formation of DNA 3° structures (beads on string, 30nm fiber, loops, rosettes, - next slide).

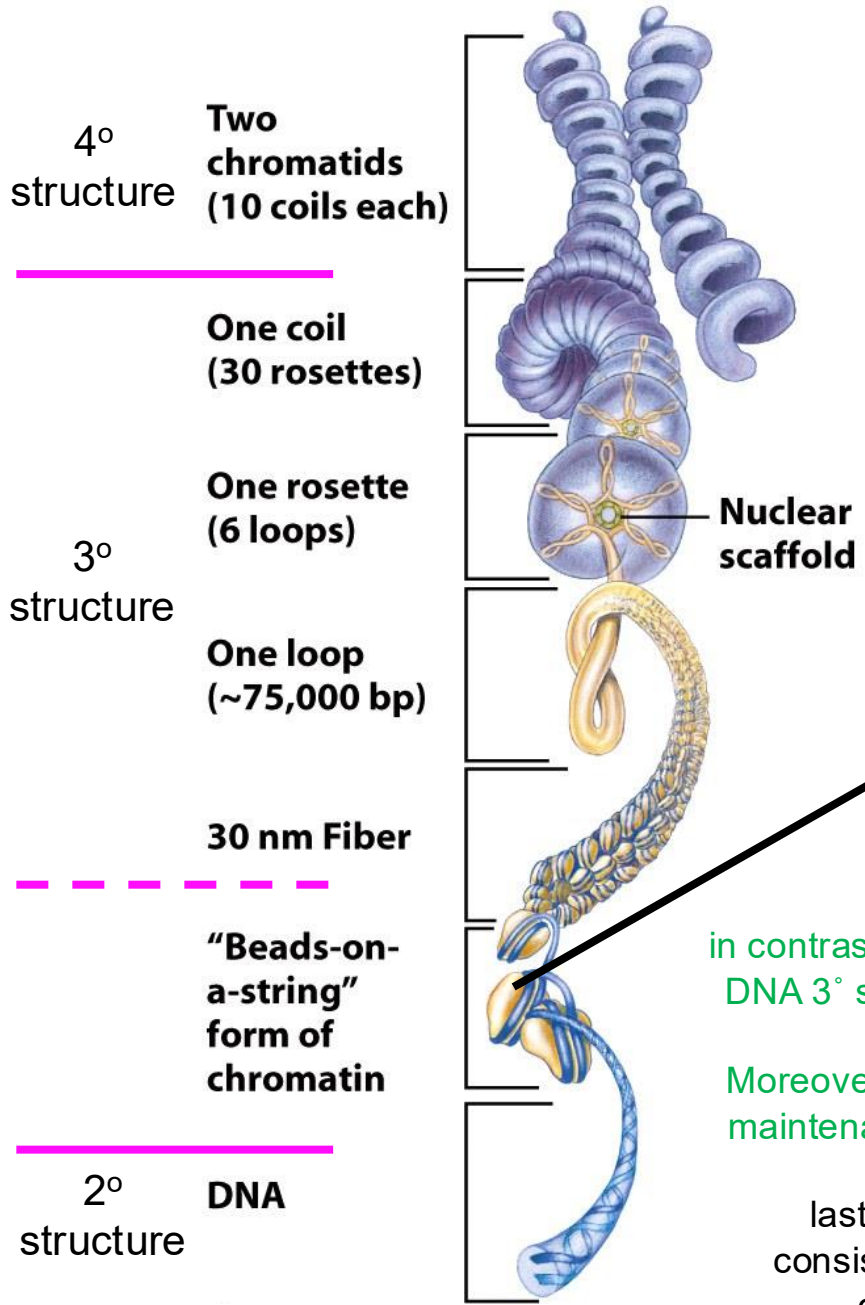
DNA's reliance on physical under-/overwinding for 3° structure formation is very different from 3° structure formation in proteins, polysaccharides, and even RNA

- **polysaccharides:** if formed at all, 3° structure is determined by branching
- **proteins:** 3° structure formation is entirely driven by hydrophobic effect & weak interactions that determine long range interactions
- **DNA:** 3° structure formation is driven by the mechanical response of the polymer backbone to the torsional stress that is imposed by supercoiling.
- **RNA:** 3° structure formation is similar to proteins = forms through long-range interactions between 2° elements; this may feel "weird" until you consider that RNAs are single stranded = single vs double stranded fundamentally changes the behavior of the polymer (and introduces a few complications as you will see shortly)

DNA Tertiary and Quaternary Structure

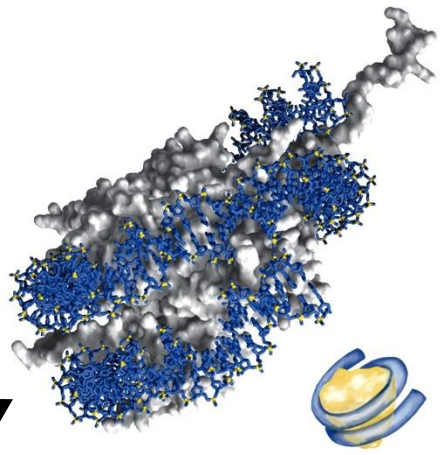
response to torsional stress from supercoiling causes various **levels of DNA compaction** that **represent DNA 3° structure**

(for a more detailed description see Molecular and Cell Biology – CHROMATIN Chapter, slides 15-22)



double stranded DNA wrapped around a histone octamer

note: the histone octamer itself and its complex with DNA (called nucleosome) represent 4° structure elements



→ **important aspect to keep in mind**

in contrast to proteins that have independently stable 3° structures, DNA 3° structures depend on "mixed 4° structure formation" with scaffolding proteins.

Moreover, beyond the stage of nucleosomes, establishment and maintenance of DNA 3° structures increasingly require enzymes

lastly: applying the hierarchy of structural organization consistently – a "DNA centric view" holds that paired sister chromatids represent a form of DNA 4° structure

(though strictly speaking, this 4° structure is a "4° structure of 4° structures")

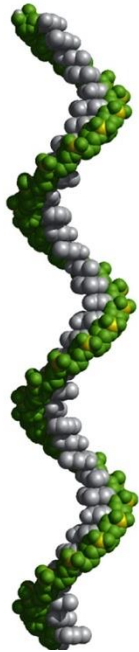
What about 2°, 3° and 4° structure in RNA?



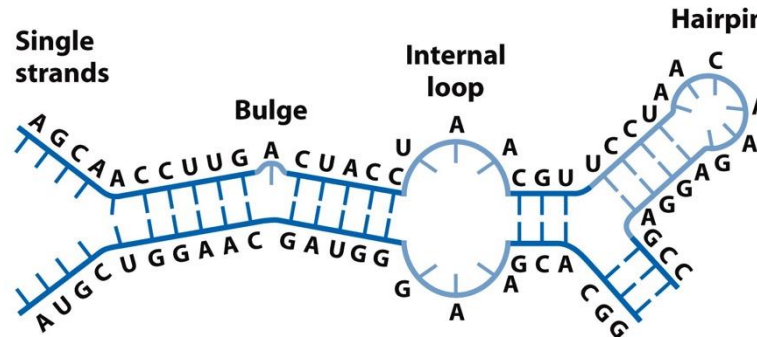
- RNA, typically single stranded
- if there were absolutely no regions of self-complementary sequence, RNA would adopt a right-handed helical structure

however, the degeneracy within the polymer (only four bases – AUCG - with only two pairings) **means that with almost no exceptions, RNA has self-complementary stretches**

→ recalling that 2° structure is defined as the local conformation of backbone atoms, this internal self-complementarity allows RNA to adopt complex secondary structures.



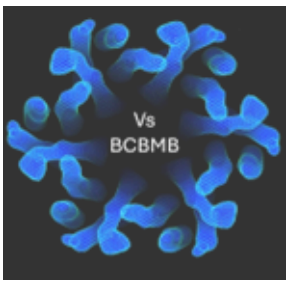
base stacking (grey)



note
paired regions
will adopt
double helical
conformations

some frequently encountered RNA 2° structures

here you may begin to understand why we wanted to define the "double helix" as a 2° element (even though in DNA it technically came about by pairing two independent DNA strands = 4° structure). doing so now allows us to define 2° structure elements that – as you will see next – can engage in long range interactions to form complex 3° structures.



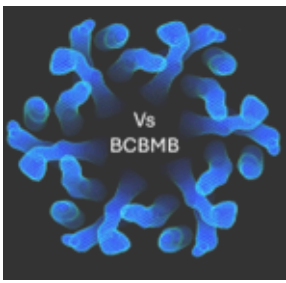
RNA 2°, 3°, 4° Structure – Continued



when first discussing nucleic acids in the Fundamentals Tutorial, we concluded that left to themselves, nucleic acids are prone to folding up on themselves, forming random 3D structures that by themselves would have little value for anything
(Fundamentals – NUCLEIC ACIDS slides 37-39).

with what you now learned about RNA 2° structure, you are ready for a more nuanced view

while the precise 3D (= 3°) structures of an ensemble of RNA molecules will show a distribution of outcomes – the outcomes will not be entirely random because locally the single strand will be able to adopt a limited number of 2° structures, which then will try to interact with each other to further minimize the Gibbs Free Energy of the whole chain.



RNA 2°, 3°, 4° Structure – Continued



when first discussing nucleic acids in the Fundamentals Tutorial, we concluded that left to themselves, nucleic acids are prone to folding up on themselves, forming random 3D structures that by themselves would have little value for anything
(Fundamentals – NUCLEIC ACIDS slides 37-39).

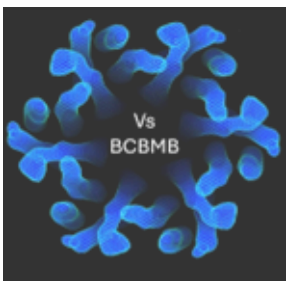
with what you now learned about RNA 2° structure, you are ready for a more nuanced view while the precise 3D (= 3°) structures of an ensemble of RNA molecules will show a distribution of outcomes – the outcomes will not be entirely random because locally the single strand will be able to adopt a limited number of 2° structures, which then will try to interact with each other to further minimize the Gibbs Free Energy of the whole chain

key difference to protein folding is that in proteins, the folding intermediates keep searching until the chain finds the global minimum (if it exists) – which is possible because most folding intermediates are not well stabilized (recall: the final fold is stabilized by only ~10kcal/mol)
= thermal energy (in most cases) can keep the search going, undoing interactions that are not optimal.

in stark contrast: RNA folding is prone to kinetic trapping
= the stabilization energy for even small 2° structure elements is very significant
→ the randomness of encounters during early folding stages results in a spread of final outcomes because the overall energy folding landscape is riddled with many steep local minima that arrest the conformational search, trapping a "suboptimal" overall conformation of the chain.

living by the paradigm "structure determines function"
how did Nature overcome this troubling folding behavior of single stranded RNA?

....what do you think?....



RNA 2°, 3°, 4° Structure – Continued

living by the paradigm "structure determines function"
how did Nature overcome this troubling folding behavior
of single stranded RNA?



surprising answer

it didn't

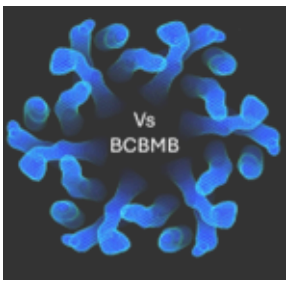
instead

it learned to live with it

what does that mean?

to answer this, you want to recall/consider what functions RNAs play in biology

....ready ... and Go!



RNA 2°, 3°, 4° Structure – Continued



living by the paradigm "structure determines function"
how did Nature overcome this troubling folding behavior
of single stranded RNA?

surprising answer

it didn't

instead

it learned to live with it

what exactly does that mean?

to answer this, you want to recall/consider what functions RNAs play in biology

messenger

mRNA

adapter

tRNA

RNA - Functions

catalytic

rRNA

ribozymes

snRNA (spliceosome)

RnaseP

regulatory

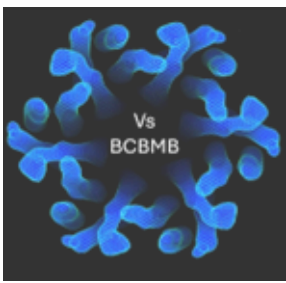
miRNA (micro)

siRNA (small interfering)

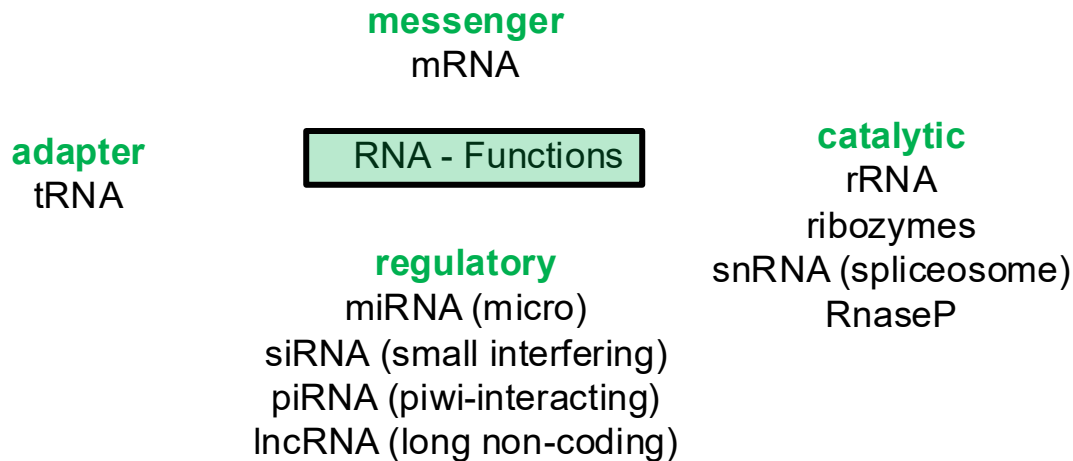
piRNA (piwi-interacting)

lncRNA (long non-coding)

how does knowing the functions help understanding what "live with it" means/implies?



RNA 2°, 3°, 4° Structure – Continued



how does knowing the function help understanding what "live with it" means/implies?

it helps because it allows you to evaluate how critical it is for the RNA to reproducibly adopt a specific 3° structure, and hence what – if anything – you must do to make things work

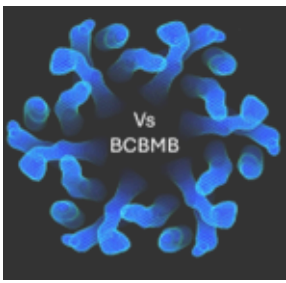
mRNA: the value of it is in the linear 1° structure → while mRNAs do have certain conserved structural features, the **exact overall 3° structure is irrelevant** because ribosomes act on the linear, unfolded mRNA

tRNA, all catalytic RNAs: 3° structure is critical = learning to "live with it" means you **MUST develop dedicated machinery to supervise the folding** of these molecules

regulatory RNAs are somewhere in between

regulatory RNAs: 3° structure is critical ... BUT miRNA, siRNA & piRNA are very short (20-30nt) = fewer kinetic traps, and fewer deep but false minima = learning to "live with it" is not a big deal

for lncRNAs (can be 100dreds of nt long)... **only the structure of small defined regions matters** → supervising folding of those subregions is **manageable**



RNA 2°, 3°, 4° Structure – Continued



mRNA: exact 3° structure is irrelevant

tRNA, all catalytic RNAs: 3° structure is critical = MUST develop dedicated machinery to supervise the folding

regulatory RNAs: 3° structure is critical ... BUT miRNA, siRNA & piRNA because they are very short
for lncRNAs (can be 100dreds of nt long)... only the structure of small defined regions matters

what you learn here is that there is a molecular equivalent to the saying
"you can't have your cake and eat it too"

= where it was necessary, Nature DID work out solutions to overcome the temperamental folding behavior of RNAs

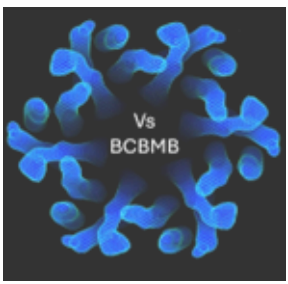
if you had to design such solutions – how would you approach this?

...come up with a step-by-step reasoning...

"what is the issue that needs attention...?"

why does it exist?

→ how could you avoid the issue = change the game?



RNA 2°, 3°, 4° Structure – Continued

= where it was necessary, Nature DID work out solutions to overcome the temperamental folding behavior of RNAs



if you had to design such solutions – how would you approach this?

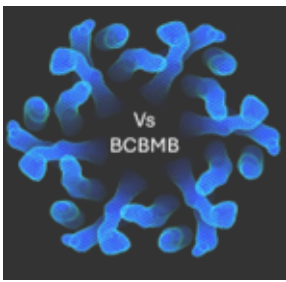
what is the issue?

an abundance of local Free Energy minima traps folding intermediates that - because of the stability of even small 2° elements - cannot spontaneously resolve

→ the issue is: **kinetic trapping** → **need to prevent or minimize this.**

why does the issue exist?

how can you avoid this?



RNA 2°, 3°, 4° Structure – Continued



= where it was necessary, Nature DID work out solutions to overcome the temperamental folding behavior of RNAs

if you had to design such solutions – how would you approach this?

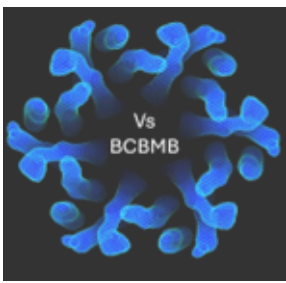
what is the issue?

an abundance of local Free Energy minima traps folding intermediates that - because of the stability of even small 2° elements - cannot spontaneously resolve
→ the issue is: **kinetic trapping** → need to prevent or minimize this.

why does the issue exist?

flexibility of the single chain polymer + degeneracy
(only 4 bases, forming 2 pairs)
= **self-complementarity creates many opportunities for kinetic trapping**

how can you avoid this?



RNA 2°, 3°, 4° Structure – Continued



how can you avoid/minimize kinetic trapping?

if degeneracy and stability of base pairs are behind all of this you could change this by

✓ keeping the RNA length as small as necessary to serve its function

✓ minimize self-complementarity

(efficient but severely limits what sequences you can use + since some self-complementarity is necessary to form the 2° structures you **do want** ...you would have to rely on the emergence of spatially distant self-complementarities and lock them in once they show up = lot of dead-end trials)

✓ chemically modify bases in strategic positions to interfere with formation of inappropriate intermediates by avoiding false local Free Energy Minima

✓ cover up parts of the RNA and then sequentially uncover parts to allow folding

✓ could also cleave (and perhaps trim) primary transcripts at strategic positions to resolve false minima or allow re-arrangement of fragments to form the correct, more extensive 2° structure elements.

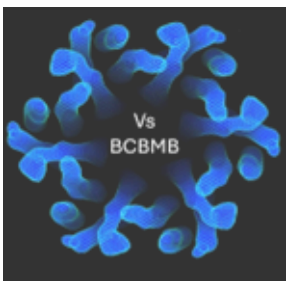
and there you have it it not only sounds logical ... as all these strategies are used

let's look at three examples that illustrate "learning to live with it"
starting with the least complex

RNaseP

a catalytic RNA, ~300-400nt long (dependent on species) that is essential for trimming the 5'-end of every tRNA molecule to generate the mature tRNA from its precursor.

RNaseP is found in every branch of life ...



RNA 2°, 3°, 4° Structure – Continued

RNaseP

a catalytic RNA, ~300-400nt long (dependent on species) that is essential for trimming the 5'-end of every tRNA molecule to generate the mature tRNA from its precursor.

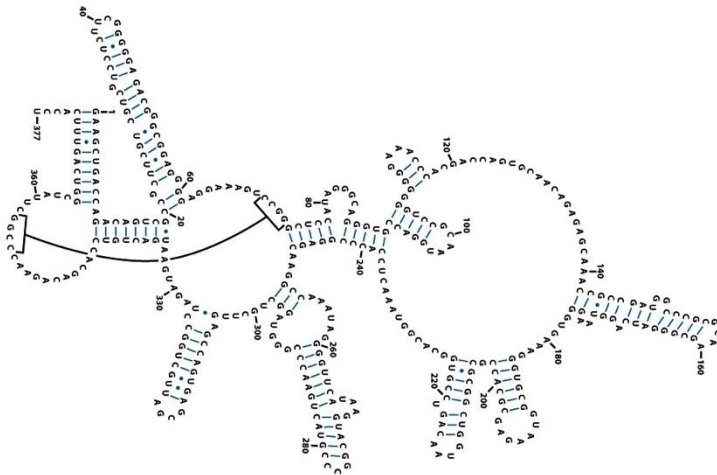
RNaseP is found in every branch of life ...

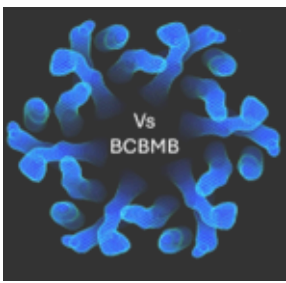


- secondary structure diagram of the M1 RNA component of E coli RNaseP shows that the secondary structure can be quite complex, featuring multiple hairpins, bulges and internal loops

looking at the diagram – does that remind you of anything you (may) have encountered in the real world?

...try to think of something





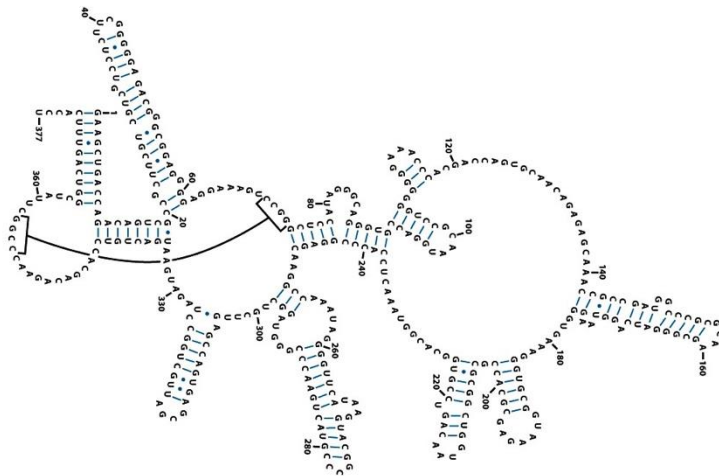
RNA 2°, 3°, 4° Structure – Continued



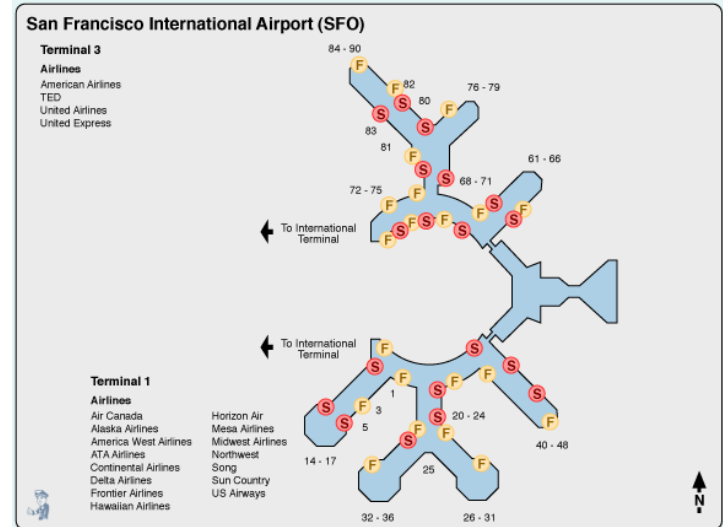
RNaseP

a catalytic RNA, ~300-400nt long (dependent on species) that is essential for trimming the 5'-end of every tRNA molecule to generate the mature tRNA from its precursor.
RNaseP is found in every branch of life ...

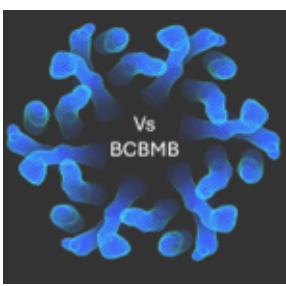
- secondary structure diagram of the M1 RNA component of E coli RNaseP shows that the secondary structure can be quite complex, featuring multiple hairpins, bulges and internal loops



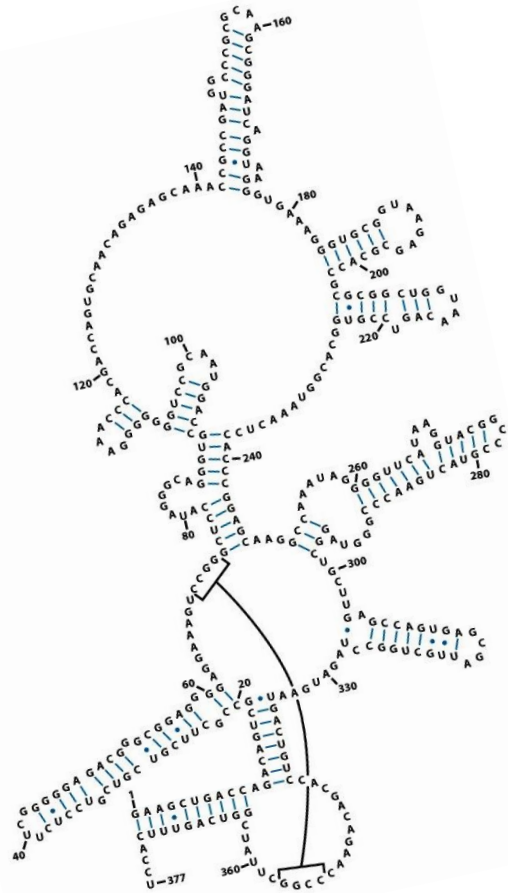
looking at the diagram – does that remind you of anything you (may) have encountered in the real world?



I doubt the architects were aware of borrowing design principles from the nucleic acid world



RNA 2°, 3°, 4° Structure – Continued

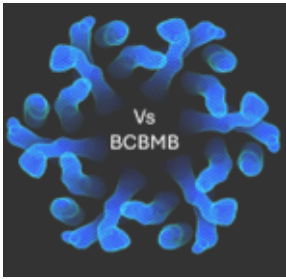


but seriously and for real now

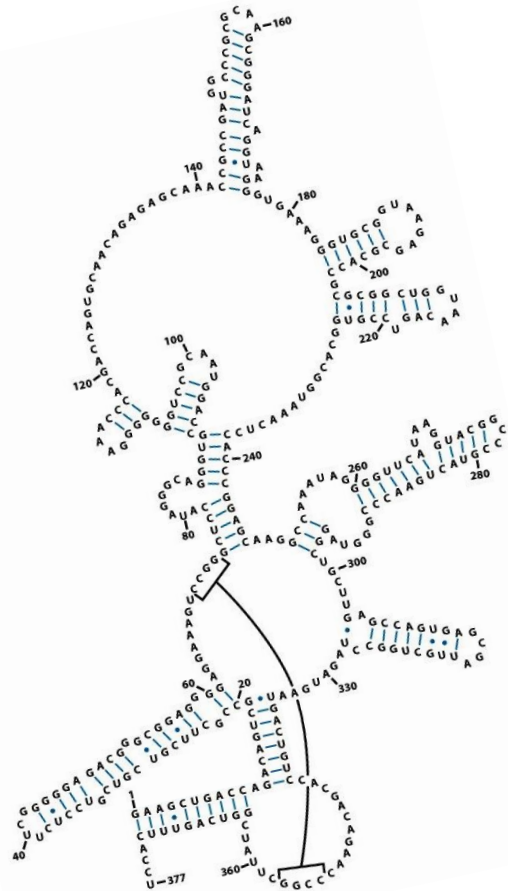
looking at the 2° diagram with your sight set at 3° structure

how can you build a complex 3D structure from this?

hint ... if you can control it, then 3° structure formation is similar to how it forms in proteins ... which is through



RNA 2°, 3°, 4° Structure – Continued



how can you build a complex 3D structure from this?

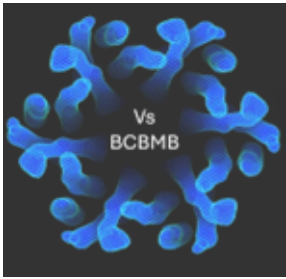
...if you can control it, then 3° structure formation is similar to how it forms in proteins ... which is through

long-range weak interactions between sidechains

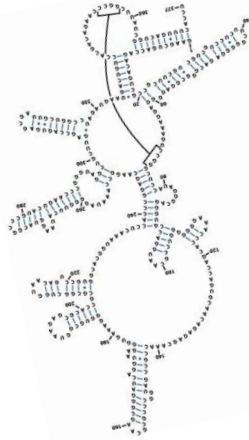
→ looking at the 2° diagram you realize that **in RNAs the unpaired nucleobases in the single stranded regions, bulges and internal loops are the "sidechain" equivalents**

= these unpaired bases **can engage with other unpaired bases in spatially distant 2° structure elements** or interact with the helical stems of hairpins

→ **elements interacting through space = fold = complex, compacted 3° structure**



RNA 2°, 3°, 4° Structure – Continued

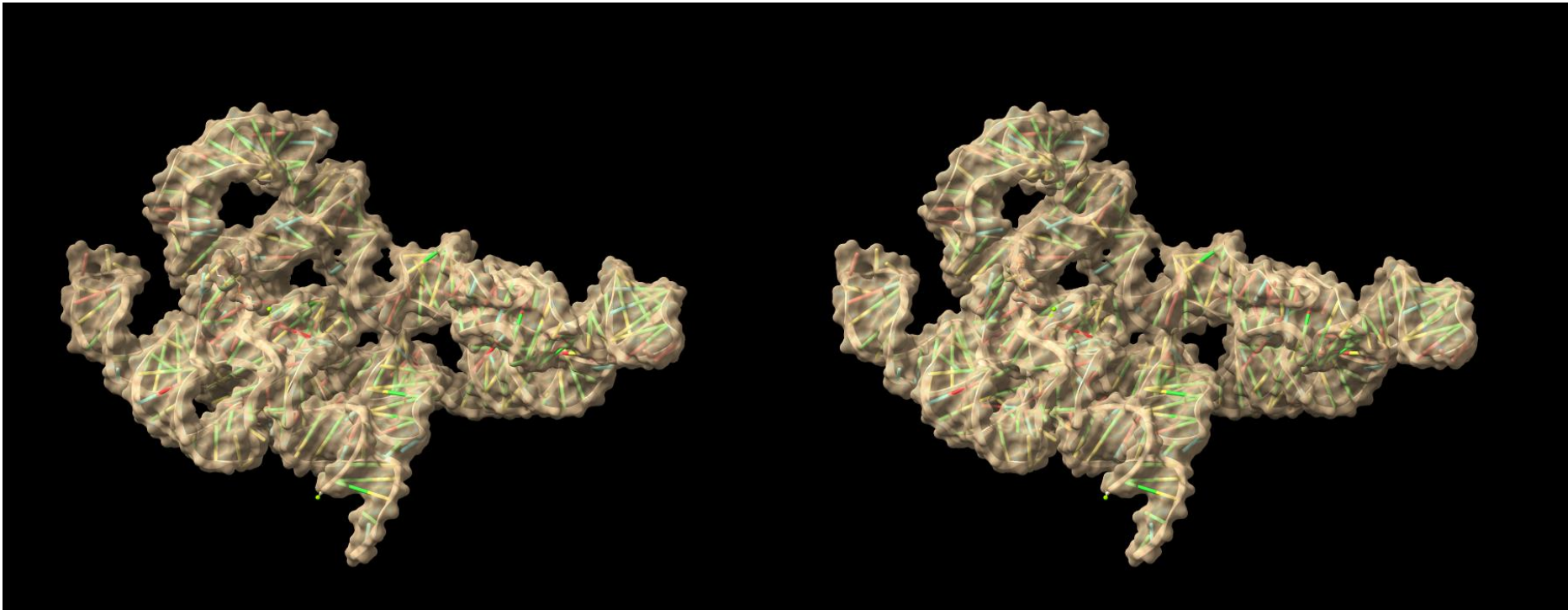


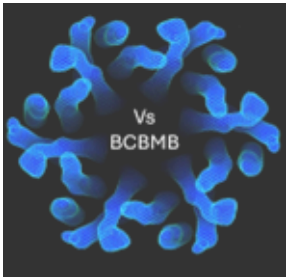
long-range interactions between 2° elements allow RNaseP to adopt a compact fold

= turning the diagram on the left into the 3D structure shown below

cross-eye stereoview of RNaseP RNA 3° structure

hold at ~30cm distance, cross your eyes, look at the image, adjust eyes and head tilt (left-right) to gain 3D depth perception





RNA 2°, 3°, 4° Structure – Continued

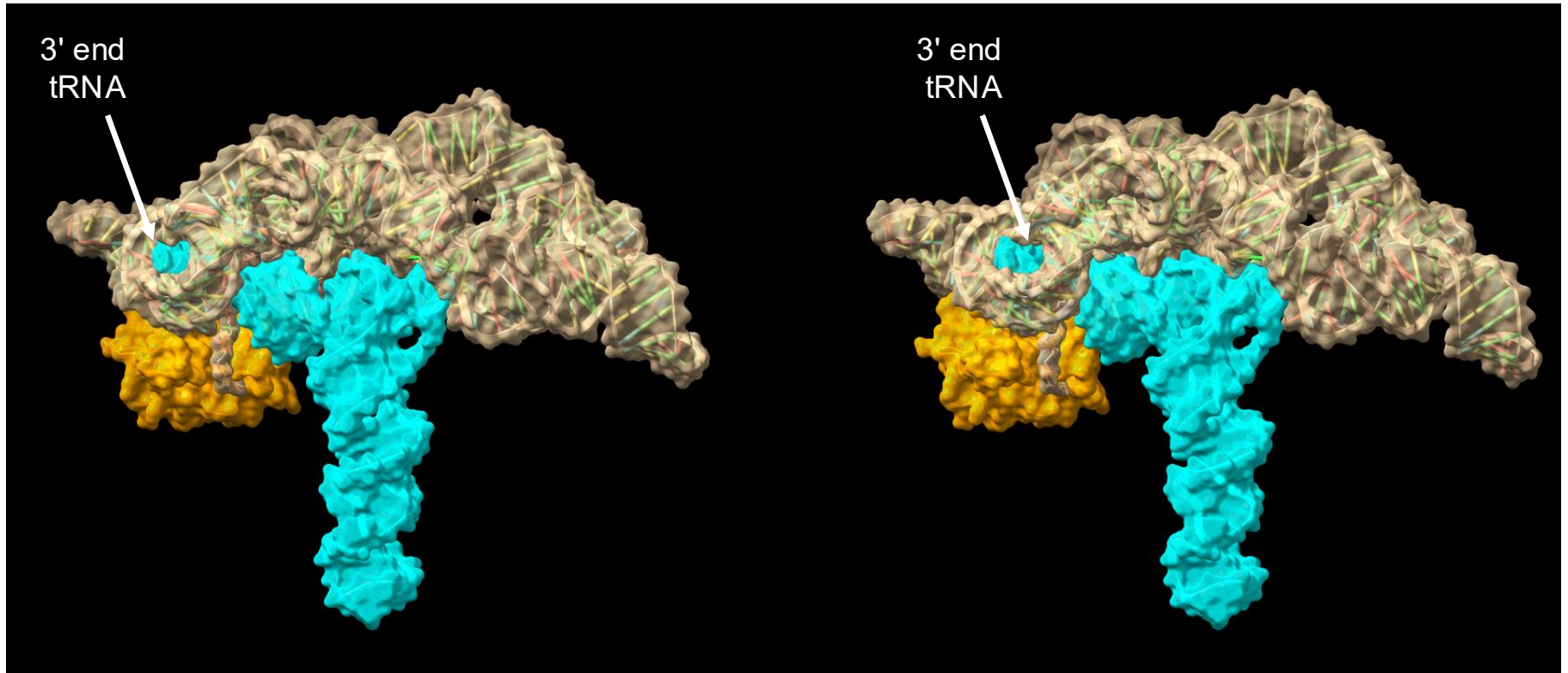


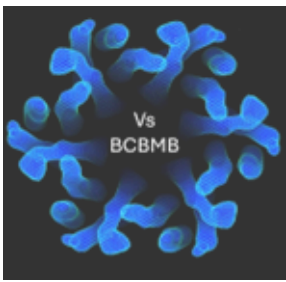
as complicated as the 2° structure diagram and 3° structure may look like, **the bacterial (=evolutionary early) version of RNaseP RNA** - called the M1 RNA – is unique in that it **can fold unassisted in a test tube under carefully controlled conditions** = **THAT is a GREAT starting point because it means the 1° does not set up too many kinetic traps that are deep enough to scramble the 3D structure.**

In vivo, though, intracellular conditions require the presence of a small RNA chaperone-like protein called C5. This protein stabilizes the near native fold and facilitates binding of the tRNA substrates.

cross-eye stereoview of holoRNaseP complex (4° structure)

C5 protein: orange; t-RNA: cyan. **Note** how the M1 RNA cradles the single stranded 3' end of the tRNA, perfectly defining where cleavage of the precursor should occur





RNA 2°, 3°, 4° Structure – Continued



"learning to live with it"so farso good

... but

there is a "sad twist"

eukaryotic RNaseP has lost the ability to fold almost independently.

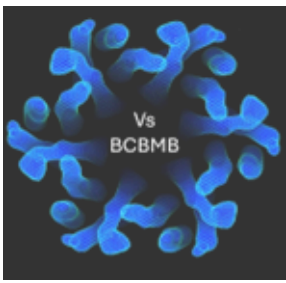
point in case: **in humans a total of 10 "core proteins" are involved** in supervising the folding process and continued function of the equivalent RNaseP RNA – called H1.

without these proteins, RNaseP will neither fold nor function = the proteins stay associated with the folded ribozyme

→ this "twist" is a **great first example for "learning to live with"** because it illustrates how Nature "grudgingly" developed a more complex machinery once an essential process - protein synthesis – had been built around an initially "innocent looking" ribozyme.

and to no surprise:

.....what do you think come next ?



RNA 2°, 3°, 4° Structure – Continued



there is a "sad twist"

eukaryotic RNaseP has lost the ability to fold almost independently.

point in case: in humans a total of 10 "core proteins" are involved in supervising the folding process and continued function of the equivalent RNaseP RNA – called H1.

without these proteins, RNaseP will neither fold nor function = the proteins stay associated with the folded ribozyme

→ this "twist" is a **great first example for "learning to live with"**

because it illustrates how Nature "grudgingly" developed a more complex machinery once an essential process - protein synthesis – had been built around an initially "innocent looking" ribozyme.

and to no surprise:

once supervising some aspects of RNA folding became possible, the stage was set to evolutionary optimize the new tools to tackle far more difficult cases

...strangely...

we do not have to go far to find another example because RNaseP's substrate – tRNA - is another example, and it is far more perplexing than RNaseP itself.

Why?

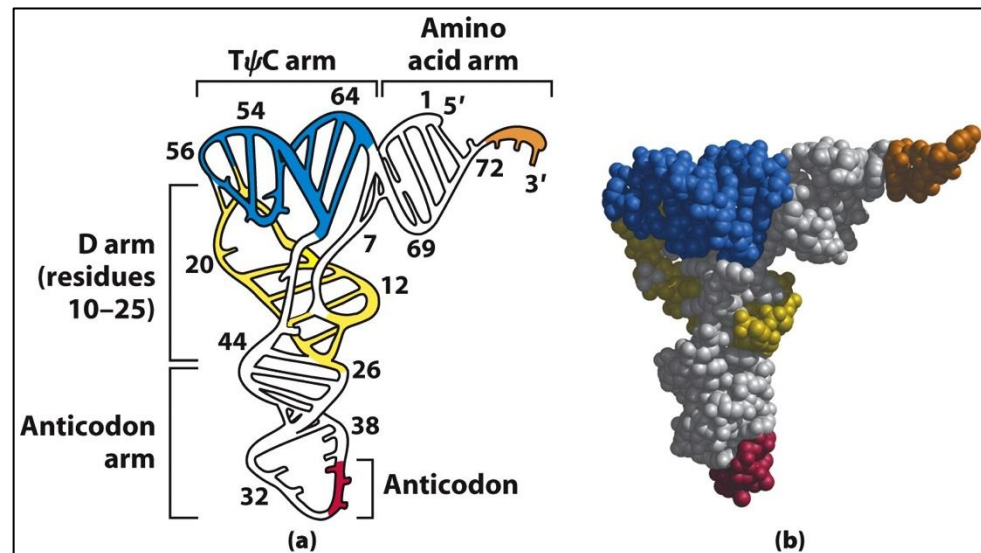
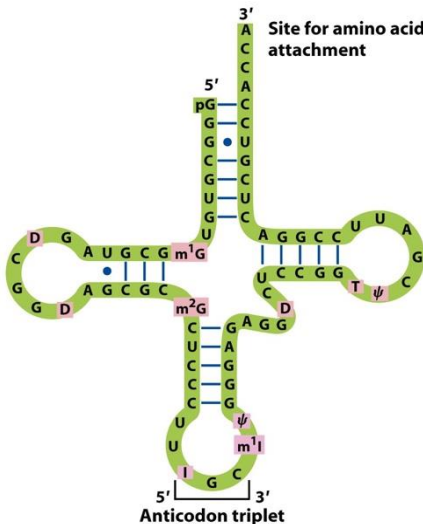
tRNAs – The Devil is in the Details



Answer

because tRNAs are a lot shorter than RNaseP RNA
70-100 nucleotides vs 400-500 nucleotides

→ since bacterial M1-RNA was able to fold almost autonomously, one might easily assume that the shorter tRNAs should have even fewer issues finding the right 2° and 3° ... shown to the left and right respectively



assuming that tRNA folding should be easy is a big mistake because it really is not.

in fact, if you were to make a tRNA transcript in the test tube – it would not fold at all and any conformations that resemble the characteristic L-shaped native structure would not be able to function.

if, instead, you **isolate the tRNA from the cell + unfold it, it will refold easily even in the absence of any protein factors. How can that be explained?**

...try to think of reasons (hint: look at the 2° structure diagram ... what is "weird" about it?) ...

tRNAs – The Devil is in the Details



Answer

looking at the 2° diagram more closely... you notice that certain positions are marked "pink"

→ these are the sites where the primary transcript has been edited by modifying the original base pairs

each tRNA molecule contains 8-15 essential modifications

(out of a repertoire of >100 modifications that have been identified)

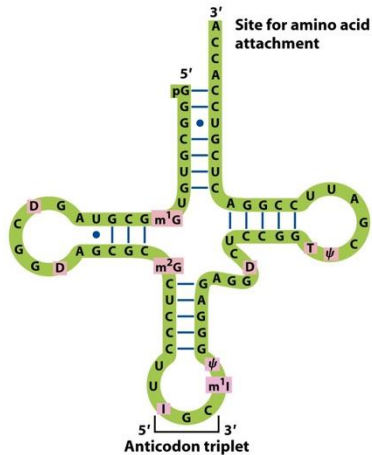
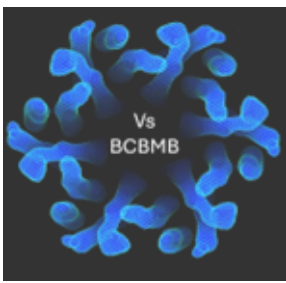
most common: inosine (slide 8), methylated guanine (slide 9) or pseudo- and dihydrouridine (slide 11)

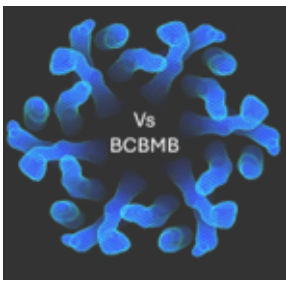
tRNA maturation and base modifications occur during a strictly staged and sequential post-transcriptional process that involves 12-15 processing enzymes, 2-5 chaperones, and final proofreading of the 3° structure by the matching Aminoacyl-tRNA-synthetase that eventually will add the correct amino acid

this amount of **processing is extreme** given the short length of the tRNA – **but so are the number of constraints that determine what can and cannot function**
specifically: **the small tRNA molecule needs to be**

- ✓ able to interact with the correct aminoacyl-tRNA synthetase (determinants: anticodon, acceptor stem, and "elbow" between T&D arm)
 - ✓ The ribosome (determinants: various regions depending on ribosomal binding site)
 - ✓ The mRNA (determinant: anticodon)

while tRNA processing is impressive ... making functional ribosomes is even more complex





rRNAs Have Exactly Defined Folds



ribosomes – the cellular protein factories – are highly conserved quaternary structures, made from ribosomal RNAs, and proteins that stabilize ribosome structure and modulate ribosome function

characteristically: ribosomes are made from two subunits

small

large

Feature	<i>E. coli</i> (30S Subunit)	Human (40S Subunit)
Main Functions	mRNA decoding; tRNA selection	mRNA decoding; scanning; initiation
Total Mass	~0.8 – 0.9 MDa	~1.3 – 1.5 MDa
rRNA Type	16S rRNA	18S rRNA
rRNA Length	~1,542 nucleotides	~1,870 nucleotides
Number of Proteins	21 proteins (S1–S21)	~33 proteins (eS1–eS31 + others)
Protein Naming	S-prefix (e.g., S5, S12)	eS or uS (Universal/Eukaryotic)
Protein Mass	~35% of subunit mass	~50% of subunit mass

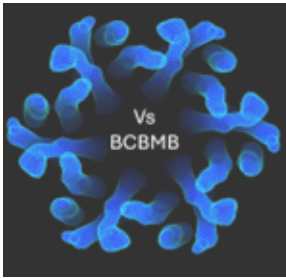
Feature	<i>E. coli</i> (50S Subunit)	Human (60S Subunit)
Main Functions	Peptidyl transferase activity; Peptide exit tunnel	Peptidyl transferase activity; Peptide exit tunnel
Total Mass	~1.6 MDa	~2.8 MDa
rRNA Types	23S and 5S rRNA	28S, 5.8S, and 5S rRNA
rRNA Length	~2,900 (23S) & 120 (5S) nts	~4,700 (28S), 160 (5.8S), & 120 (5S) nts
Number of Proteins	33-34 proteins (L1–L36*)	~47 proteins (eL1–eL43 + others)
Protein Naming	L-prefix (e.g., L2, L3)	eL or uL (Universal/Eukaryotic)
Protein Mass	~30% of subunit mass	~40% of subunit mass

note: the size of subunits and the fully assembled ribosome is historically given in "Svedberg" [S] – which measures the sedimentation rate of a molecule under high gravitational forces.

Larger S values represent faster sedimentation.

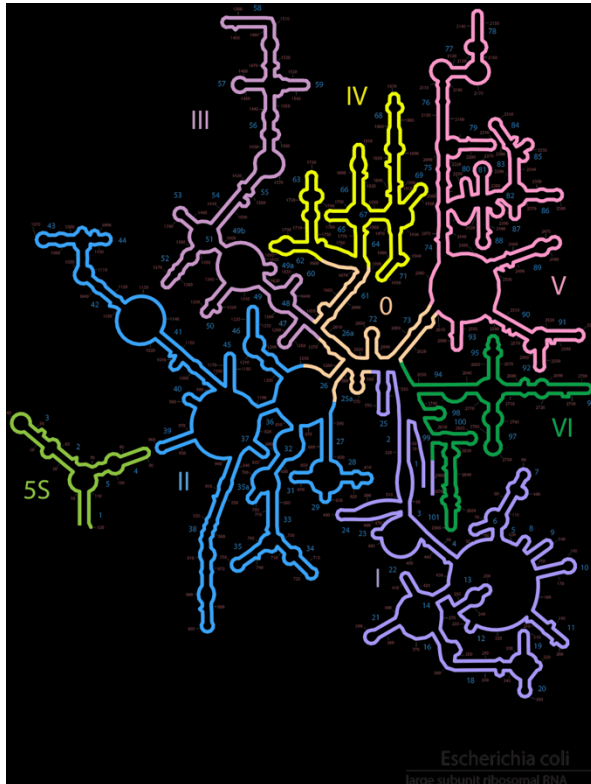
perspective: an 80S ribosome (see below) will sediment at a rate of ~16 μ m/s @ 200,000 x g (a typical gravitational force applied in a preparative ultracentrifuge) → at that force would take ~45minutes to pellet in a tube that is 4cm long

important: the values are not additive = the human ribosomal subunits are 40S and 60S respectively, but the whole ribosome sediments as an 80S particle



rRNAs Have Exactly Defined Folds

since nucleic acids are the focus of this tutorial and continuing our quest to answer the question how Nature learned to live with the idiosyncrasies of RNA folding, we want to focus on the two ribosomal rRNAs of the E coli large subunit.



on the left:

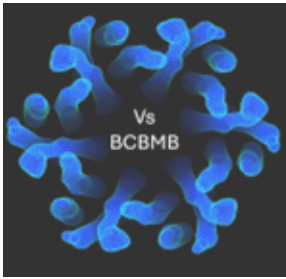
2° structure diagrams of the E. coli 23S & 5S rRNAs that form the 50S large subunit

even without scrutinizing the details – just considering the length of the 23S rRNA (2,900 nt) you would not expect this rRNA to spontaneously adopt the exact same 2° and 3° structure each time.

yet – failure to reach the correct answer every time is deadly since everything depends on having functional proteins.

so ... "learning to live with it" in this case meant to evolve

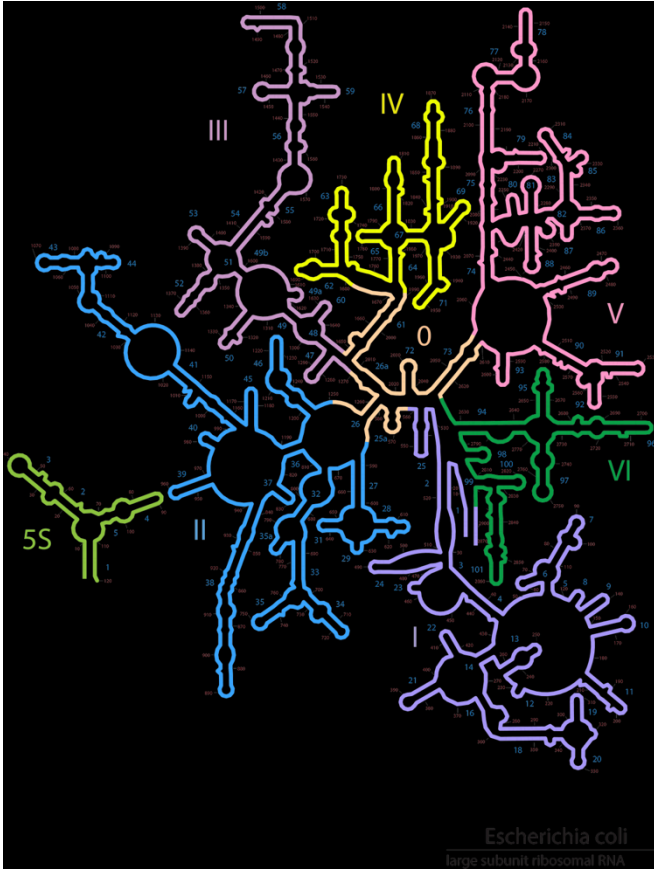
....any guess...now is the time to say



rRNAs Have Exactly Defined Folds



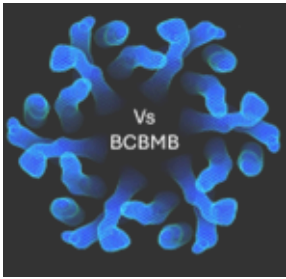
so ... "learning to live with it" in this case meant to evolve a process that includes the following players and processes



rRNA Species	Precursor Intermediate	Key Primary Enzyme	Secondary Maturation Enzymes (5' End)	Secondary Maturation Enzymes (3' End)
16S rRNA	17S (115 nt leader)	RNase III	RNase E , RNase G, RNase AM	Unknown endonuclease (possibly YbeY)
23S rRNA	25S (~7 nt 5', 8-9 nt 3')	RNase III	RNase AM, RNase G	RNase PH, RNase T, RNase R
5S rRNA	Pre-5S (~3 nt ends)	RNase III	RNase E, RNase AM	RNase E, RNase T

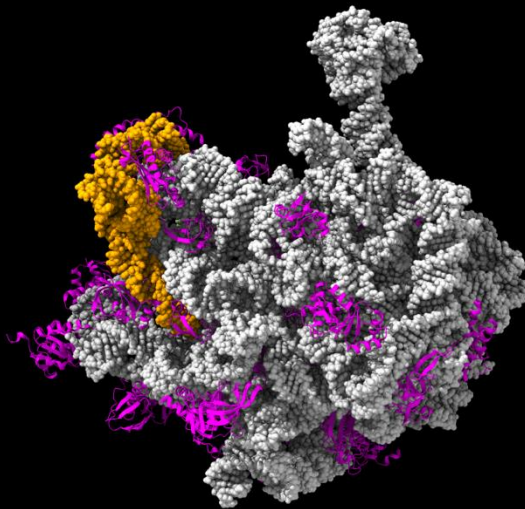
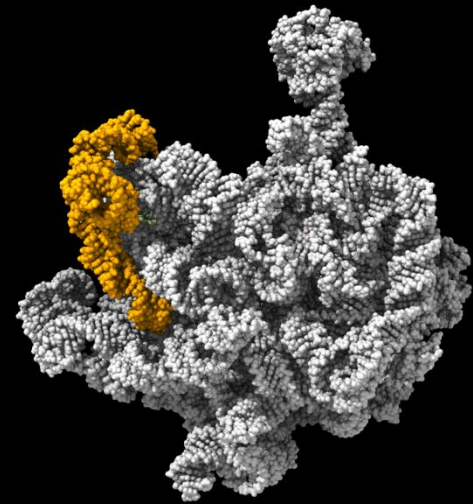
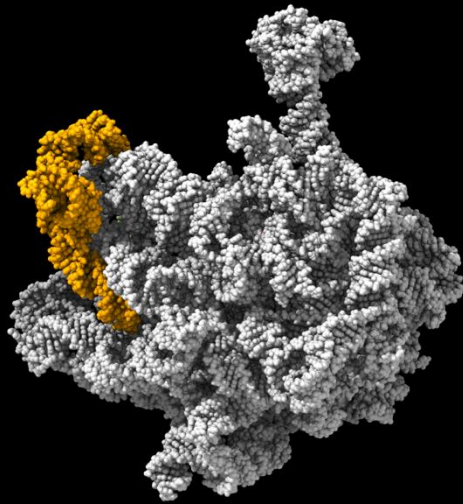
rRNA Species	Total Mods	Main Modification Types	Key Enzymes	Primary Function/Location
16S rRNA	11	10 Methylations, 1 Pseudouridine (Ψ)	RsmA-J , RsuA	Clustered in the Decoding Center ; vital for translation fidelity.
23S rRNA	25	14 Methylations, 9 Ψ , 1 Methylated Ψ , 1 Unknown	RlmA-N , RluA-F, RlhA	Clustered in the Peptidyl Transferase Center (PTC) ; stabilize catalytic structure.
5S rRNA	0	None	N/A	Primarily structural.

add to that the 33-34 ribosomal proteins that bind and stabilize the mature rRNA fold of the large subunit



rRNAs Have Exactly Defined Folds

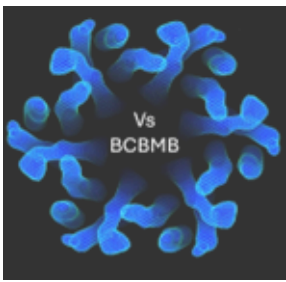
the complex maturation and supervised folding of the rRNAs assures that rapidly growing E coli cells produce 2,500– 3,000 large ribosomal subunits every minute that in their final folded state look like shown in the cross-eye stereo views



stereo-views

Top
rRNA only,
5S rRNA in orange

Bottom
ribosomal proteins
added
magenta color



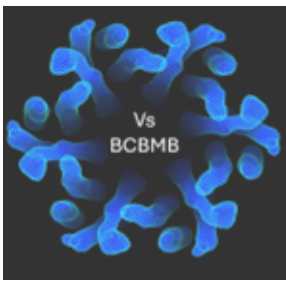
rRNAs have exactly defined folds



so ...how do you feel about this amount of supervision?
did it match your expectations ... was it less than you thought
....or more?

regardless of your answer
...you definitely will have have the right answer if asked
**how what you just learned about the large ribosomal subunit
of E coli compares to ribosome biogenesis in humans....**

....and that answer is?



rRNAs have exactly defined folds

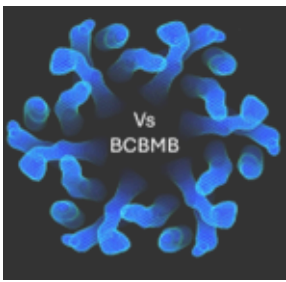
....and your answer is?

correct!

in our own cells, the process is significantly more complicated



Feature	<i>E. coli</i>	Human (Rapidly Dividing Cell)
Subunits per Minute	~2,500 – 3,000	~1,200 – 2,000
Assembly Time	~2 minutes	~30 – 60 minutes
Location	Cytoplasm (Coupled)	Nucleolus, Nucleoplasm, & Cytoplasm
Total Ribosomes/Cell	~70,000	~5 to 10 million
Feature	<i>E. coli</i> (23S & 5S)	Human (28S, 5.8S, & 5S)
Total Base/Ribose Modifications	~25	~130 – 140 mostly: methylation of 2' –OH on ribose! methylation base very rare
Modification Mechanism	Site-specific protein enzymes	snoRNA-guided (mostly)
Processing/Maturation Factors	~10 – 15	~200 – 250
Primary Nucleases	RNase III, E, G, T, PH	RNASE13, XRN2, Exosome complex
Assembly Chaperones	Few (e.g., SBD, ObgE)	Extensive (Helicases, GTPases, AAA-ATPases)



RNA – Folding and Processing



Final Thoughts (finally....!!)

the past slides most certainly threw a lot of facts at you ...
as in other places throughout the tutorials, this "overloading" is not aiming to
get all those facts into your head

...

putting it up here, is an invitation for you to
slow down,
to not just look for the bottomline,

and to let yourself marvel about the incredible feats your body accomplishes
every second, 24/7, 365 days a year without you being aware of any of it ...

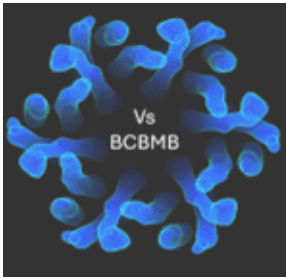
...until something, somewhere throws a little wrench into the whole
system...making you sick

.

if you stuck the course and took time to think along as we went – thank you!
I hope you enjoyed it
(even if only a little)....

to close this out before turning our attention to the Grand Finale
(a comparison between the properties of biological polymers)

here is one last piece of perspective....



RNA – Folding and Processing



Final Thoughts (finally....!!)

we started the deep dive by taking note of the fact that Nature had to "learn to live" with the temperamental folding behavior of RNA....

the comment/question that may linger on your mind:

***if Nature went through so much trouble to make this all work
(instead of trying to find a different solution to making proteins)
... then it better be worth it***

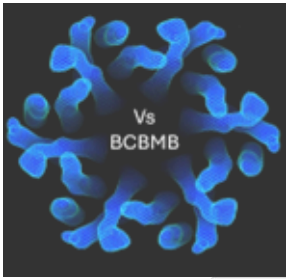
what percentage of RNAs actually need all this attention to "get in shape"?

1. **rRNA (Ribosomal RNA):** ~80% of total RNA. It is the most abundant because ribosomes are needed in massive quantities for protein synthesis.
2. **tRNA (Transfer RNA):** ~15% of total RNA. While small, there are many molecules required to shuttle amino acids to the ribosomes.
3. **mRNA (Messenger RNA):** ~1–5% of total RNA. Although it gets the most attention, individual mRNA strands are highly unstable and have a high turnover rate.
4. **Other RNAs (snoRNA, snRNA, miRNA, etc.):** <1% of total RNA. These regulatory and processing molecules exist in much smaller concentrations.

>95% ... so yes making an effort to work around the "folding problems" of single stranded RNAs **really** mattered

in fact, it matters so much that proliferating human cells expend ~30% of the total energy budget on making and maintaining just the rRNA components of ribosomes (= **not including** the cost for making the required protein factors).

....which may make you curious just exactly how a cell's energy budget looks like....

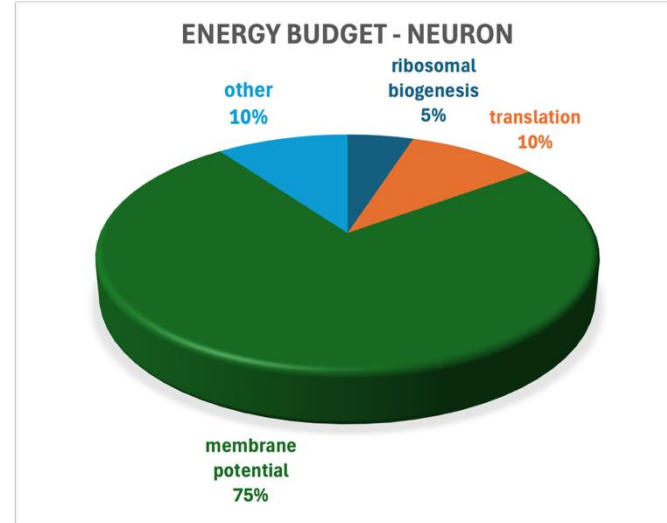
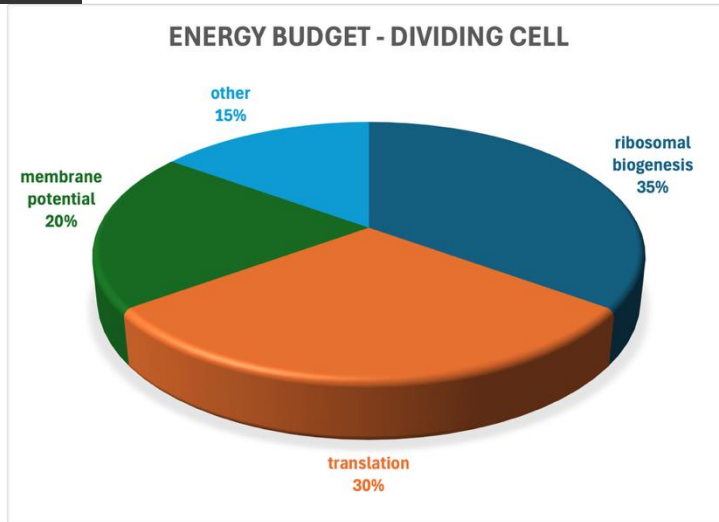


RNA – Folding and Processing



Final Thoughts (finally....!!)

like this



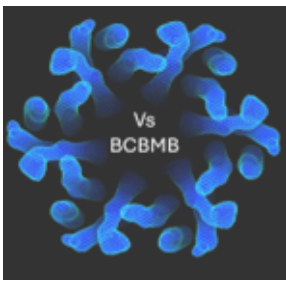
percentages will vary depending on how exactly processes are classified.
however, the overall take home message stays the same
different cells allocate their energy budget differently

these differences are not super surprising – but I wanted to bring them up here because the differences explain beautifully why at the time of death the brain dies first

in neurons almost ALL of the energy budget is spent on maintaining the membrane potential to interpret and generate action potentials

→ requires Na^+/K^+ -ATPase = no oxygen, no ATP is made = cell dies fast
in somatic cells, processes simply can slow down to preserve energy for a while
though eventually that will stop too.

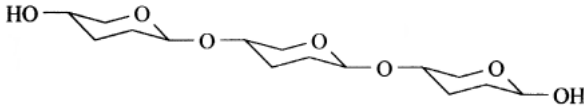
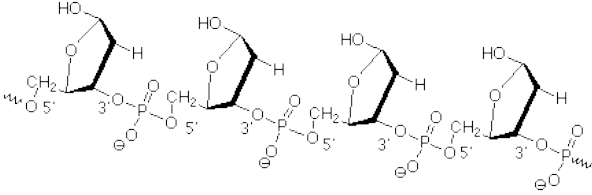
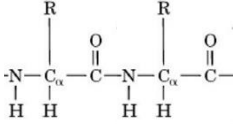
and with that ...let's close this by taking one last comparative look at the different biological polymers....



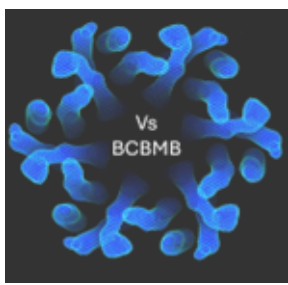
Polymer Design Principles

The Grand Picture



Polymer	=	Backbone	+	Decoration
Emulsions		N/A		N/A
Membranes		N/A		N/A
Polysaccharides		 <p>uniform; linear or branched</p>		-H, -CH ₂ OH, -COOH, -OH, -NH ₂ , -NHCOCH ₃ , -SO ₃ ⁻ , -NHSO ₃ ⁻ , -PO ₃ ²⁻
Nucleic Acids		 <p>composite; linear, deoxyribose or ribose</p>		A,T,C,U,G chemically modified bases (mostly methylation)
Proteins		 <p>uniform; linear</p>		21 sidechains, large diversity of chemical modifications (post- translational)

Polymer Design Principles The Grand Picture Continued



Emulsions

Membranes

1°

N/A

N/A

2°

N/A

N/A

3°

N/A (liquid)

bilayer

4°

N/A

“yes”

eg Golgi, nuclear envelope

Polysaccharides

Yes

backbone polarity if one end of chain is reducing
Spirals, linear, branched
(stabilized by H-bonds from -OH "decoration")

no complex folds;
limited to interactions of spatially close sugars

yes, but limited
eg: cellulose, starch

Nucleic Acids

Yes

5' → 3'

DNA: A,B,Z double helices, cruciform, triple, quadruple helices

RNA: extended/linear, hairpins, bulges, internal loops

DNA: beads on string, 30nm fiber, loops, rosettes, coils, chromatid

RNA: defined only for tRNA, rRNA, ribozymes, regulatory RNAs

considering just the nucleic acid

DNA: yes but very limited to metaphase chromosomes

RNA: yes
e.g. spliceosomes, ribosomes

Proteins

Yes

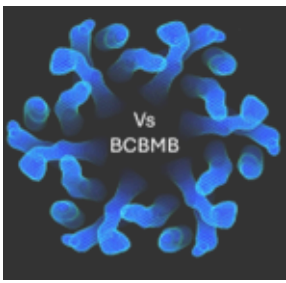
N → C

α-helices, β-strands/sheets,
β-turns, others

(stabilized mostly through H-bonds between the invariant scaffolding units backbone atoms)

large diversity in folds

yes
large diversity

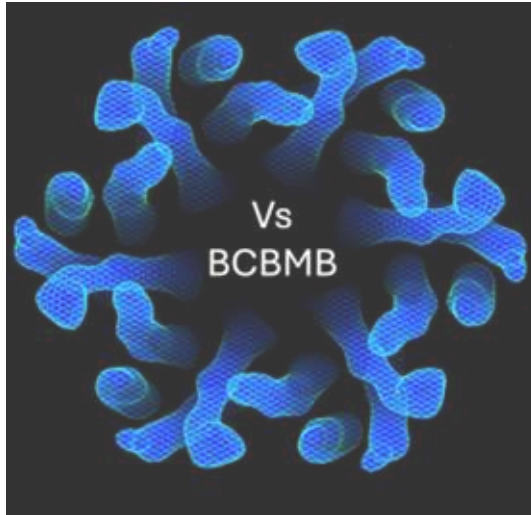


Nucleosides, Nucleotides and Nucleic Acids

Selected Facts To Know



- Unlike for carbohydrates, the name “nucleotide” does not refer to chemical composition, it refers to the cellular location (nucleus) - which originally was identified as an optically dense region in the cell
- Chemically, **nucleic acids are composite polymers** made from nitrogenous nucleobases that are grafted onto a backbone made from alternating sugar and phosphate moieties.
- Similar to many lipids, **nucleotides (the monomeric building blocks of nucleic acids)** are fairly complex substances consisting of a “base”, a sugar, and phosphate
- **Nucleotides** fulfill four functions in a cell: **energy currency, signaling, phosphate donors, building blocks for nucleic acids and certain co-factors**
- Two different types of polymers: **DNA, RNA** which differ in the type of sugar used in the nucleotide; both polymers are linear
- **DNA** solely serves for **storage** of the genetic information; except for some viruses, DNA contains **two complementary strands** of the polymer
- RNAs are further subdivided into different types: **messenger RNA** (mRNA is a portable instruction for a polypeptide), **transfer RNAs** (tRNAs are adaptors used for decoding), **ribosomal RNAs** (rRNAs are components of ribosomes, which synthesize proteins), and **micro RNAs** (micro RNAs play regulatory roles); some viruses use RNA instead of DNA as genetic material (e.g. HIV, Hepatitis C)
- **RNAs are typically single stranded**, which causes them to adopt **complex 3D-folds**; where applicable: RNA bases undergo significant post-transcriptional modifications (most common: methylation) to stabilize the 3° and 4° structures that are formed through highly supervised folding processes. Curiously, the majority of RNAs that depend on specific 3D structures are involved in mRNA splicing (eukaryotes) and translation.



Thank You for Working Through This

if you found this Tutorial/Silent Lecture useful – recommend it to friends and consider downloading other the free tutorials (with clickable links) from this site