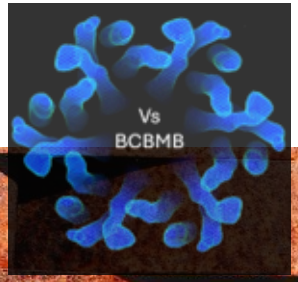


Biochemistry Fundamentals

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

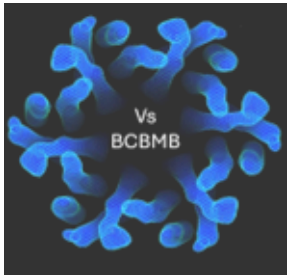
- **by thinking about/answering questions**,
(my answer is always provided on the next slide).
- **by completing a "short assignment"**
(it never will take more than a few minutes, if at all that long),
- by **watching a short video/clip**
(the embedded links will take you to my YouTube@VsBCBMB channel;
key moments are captured as still and are shown in the slide-deck, 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.



Setting the Stage

while this short primer on carbohydrate biochemistry can stand on its own, you will achieve a greater gain if you looked over the following silent lecturel (all free downloads):

- How Do Molecules See Part 1 & 2
- Biochemistry Fundamentals – LIPIDS
- Biochemistry Fundamentals - CARBOHYDRATES

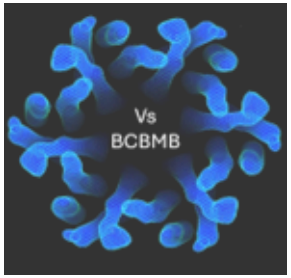
- last two silent lectures listed above focused on how Nature managed to build **physical boundaries** that serve to contain and to evolve life.

→ specifically, you learned how lipids – and strangely enough – carbohydrates accomplish formation of barriers through two fundamentally different strategies:
spontaneous self-assembly and directed chemical synthesis

building on what is covered in these prior lectures ("How do Molecules See?" - in particular) this handout continues to look at the chemical inventory of living systems from an “engineering“ point of view because that strategy will help you to understand **why** each class of biological macromolecules is uniquely suited for fulfilling certain roles.

more specifically, with the problem of "building/demarcating defined physical spaces" solved, the next question becomes what to put into these spaces, and how to coordinate the processes that interconnect the contents of these physical spaces.

ultimately this brings us to the challenge of how to efficiently store and process information



Goals Of this Lecture

by the end of working through this silent lecture

- you will **understand what "information" is** and how defining "information" impacts the design of biological systems

- you will **understand why nucleic acids are the only biological macromolecules that actually can be used for information storage in biologicl system**

- you will **know the design principles of nucleic acids** and why these principles are an ingenious solution to the information storage problem in biology

- you will **understand why to this day, DNA is unmatched in its ability to store information**

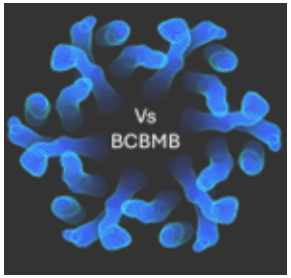
- you will **understand why DNA is a double strand** and how this makes heredity a beneficial side-effect, not the cause for the double stranded nature of DNA

- you will **learn about the general flow of information in cells** ("central dogma")

- you will **understand why the flow of information in cells in "unidirectional"** and how biology overcomes this challenge for duplicating the genetic information

first pit stop: **what does the word “information” mean?**

.....try to answer yourself first .....



Information

what does the word “information” mean?

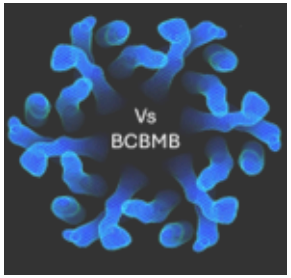
Answer

any non-random form of data that propagates the past through time and space.

(an informative, cross-referenced article can be found here: *https://en.wikipedia.org/wiki/Information*)

➔ what are the implications?

....try to come up with an answer.....



Information

what does the word "information" mean?

Answer

any non-random form of data that propagates the past through time and space.

(an informative, cross-referenced article can be found here: *https://en.wikipedia.org/wiki/Information*)

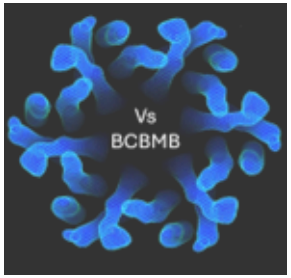
➔ what are the implications?

- information is linked to time = at a molecular level, **different time scales require different mechanisms to "store" and to process information**
  - ✓ short term memory (emergent; equivalent of RAM on computer), and
  - ✓ long-term term (hardwired; equivalent of hard disk on computer).

at the molecular level, these information related qualities must align with different sets of molecules:

✓**long-term memory** (genome) requires macromolecule that can reduce molecular complexity from high-dimensional chemical space into single dimension

✓**short-term** (decoding, integration of cell responses, and feedback loops) requires redundant multidimensional input (= multiple (types of) macromolecules whose ensemble averages determine functional state and recent history)



Information

what does the word "information" mean?

Answer

any non-random form of data that propagates the past through time and space.
(an informative, cross-referenced article can be found here: *https://en.wikipedia.org/wiki/Information*)

➔ **what are the implications?**

- information is linked to time = at a molecular level, **different time scales require different mechanisms to "store" and to process information**
  - ✓ short term memory (emergent; equivalent of RAM on computer), and
  - ✓ long-term term (hardwired; equivalent of hard disk on computer).

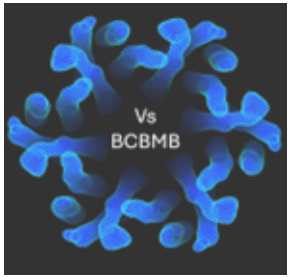
at the molecular level, these information related qualities must align with different sets of molecules:

✓**long-term memory** (genome) requires macromolecule that can reduce molecular complexity from high-dimensional chemical space into single dimension

✓**short-term** (decoding, integration of cell responses, and feedback loops) requires redundant multidimensional input (= multiple (types of) macromolecules whose ensemble averages determine functional state and recent history

➔ in this "Fundamentals" chapter, we will pursue

- why long-term **storage** can **only** be done by **nucleic acids**,
- **why transmission and processing involves all types of biological (macro)molecules**,
- and we will deal with some basic principles of short-term memory/information processing.



Long-term Term Information Storage
What Do You Know Already?

this slide is an open invitation to test your knowledge before we get started …
try to define/explain in your own words what each of the following terms means

likely heard about:

- Genes?
- Chromosomes?
- DNA (Deoxyribonucleic Acid)?
- RNA?
- Double Helix?
- Major Groove, Minor Groove??
- Deoxyribose-Phosphate Backbone?
- Nucleotides, Purines, Pyrimidines?
- A/T G/C – base pairs?
- Triplets/Codons?
- Complementarity?
- Replication?
- Transcription?
- Translation?

if you worked through "How Do Molecules See?" – you Definitely have heard about:

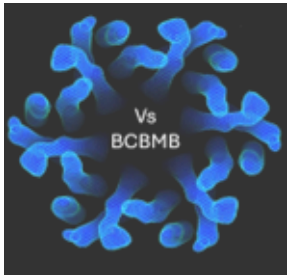
- Hydrogen Bonds
- Van der Waals Interaction
- Ionic Interactions

some of you may have heard about
(not a big deal if not … we'll cover that in later lectures)

- Base stacking?
- Promoters?
- Enhancers??
- Silencers
- Terminators?
- Histones?
- Replication Origins?

if you can explain all these terms – awesome!

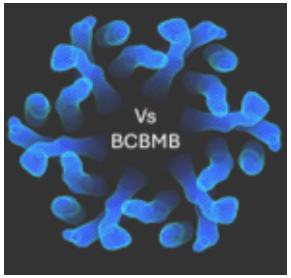
but do you know **why** DNA (on Earth) is the **only** biological macromolecule that **can** store large amounts of information?
and why that solution is also uniquely suited to be long-term term?



Long-Term Information Storage …. An Engineer's Perspective

what requirements drive the design of a long-term information carrier?
...try to answer … it really will make a difference ….

- .
- .
- . *…as many as you can think of*



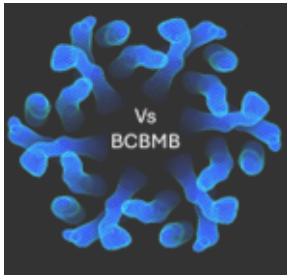
Long-Term Information Storage …. An Engineer's Perspective

what requirements drive the design of a long-term term information carrier?

the information should be

- stable
- easy to read and interpret
- unambiguous and easy to copy/synthesize.

→ **how** does that impact on design?



Long-Term Information Storage …. An Engineer's Perspective

what requirements drive the design of a long-term term information carrier?

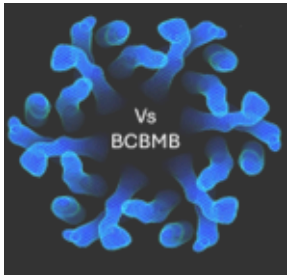
the information should be
- stable
- easy to read and interpret
- unambiguous and easy to copy/synthesize.

→ **how** does that impact on design?

Answer

ideally want the actual information to be encoded in **"bits"** (**b**inary dig**its**; 0 and 1 in computers [we will find that it is "2s" and "3s" in biology instead]).

why bits?



Long-Term Information Storage …. An Engineer's Perspective

what requirements drive the design of a long-term term information carrier?

the information should be
- stable
- easy to read and interpret
- unambiguous and easy to copy/synthesize.

→ **how** does that impact on design?

Answer

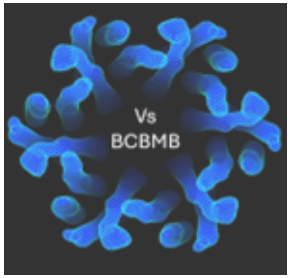
ideally want the actual information to be encoded in **"bits"** (**b**inary dig**its**; 0 and 1 in computers [we will find that it is "2s" and "3s" in biology instead]).

why bits?

Answer

"molecular bits" reduce complexity to a linear string
and - if designed properly –
allow for (easy) detection and correction of unintended changes (mutations).
["bit system" also has the highest probability of arising by chance + expand rapidly to increase/adapt information].

what is the implication of "linear string"?



Long-Term Information Storage …. An Engineer's Perspective

what requirements drive the design of a long-term term information carrier?

the information should be

- stable
- easy to read and interpret
- unambiguous and easy to copy/synthesize.

→ **how** does that impact on design?

Answer

ideally want the actual information to be encoded in **"bits"** (**b**inary dig**its**; 0 and 1 in computers [we will find that it is "2s" and "3s" in biology instead]).

why bits?

Answer

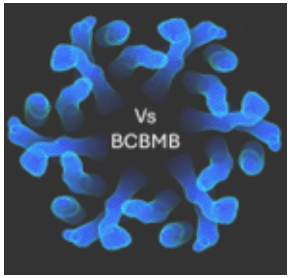
"molecular bits" reduce complexity to a linear string
and - if designed properly –
allow for (easy) detection and correction of unintended changes (mutations). ["bit system" also has the highest probability of arising by chance + expand rapidly to increase/adapt information].

what is the implication of "linear string"?

Answer

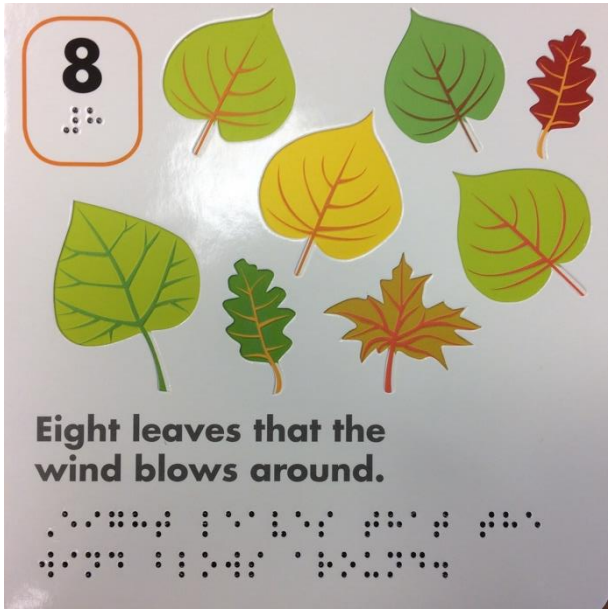
an unbranched covalent polymer (fun fact preview: when completely extended, the information coding for a human is ~2m long).

question now becomes: **what** does "easy to read" mean?
or in other words: **how** **is the information interpreted?**

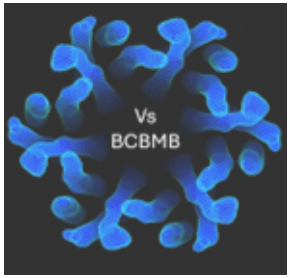


Interpreting Information at a Molecular Level

…..relies on **touch …**



review "How Do Molecules See?" Pt 1
slides 7-9
(to fully "get" references made along the way)



How do Conceptual Principles of Molecular Recognition Impact Molecular Design of Long-Term Information Storage?

flashback…..requirements

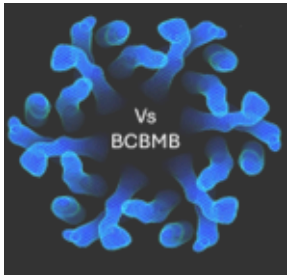
the information should be

- *stable*
- ➢ *easy to read and interpret*
- ➢ *unambiguous and easy to copy/synthesize.*

➢ ***"bits"** along an unbranched covalent polymer are desirable*

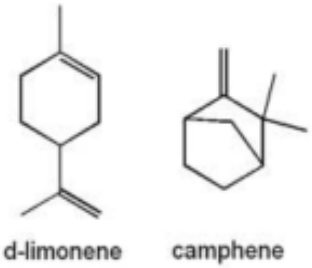
→ using conceptual principles (complementarity, weak interactions), allows different types of biological macromolecules to be evaluated for their ability to build what we want = a linear string of bits

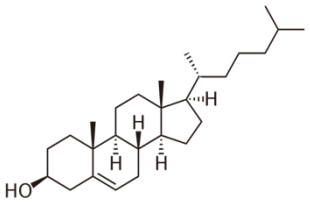
in previous lectures, we already covered …lipids and carbohydrates besides those, we have nucleobases and amino acids, which you already may know some things about …but for now, let's just add them to the inventory and take a look:



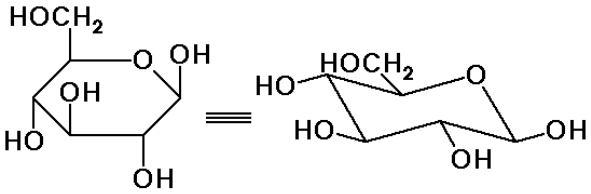
Life's Basic – Macromolecular Inventory

Lipids

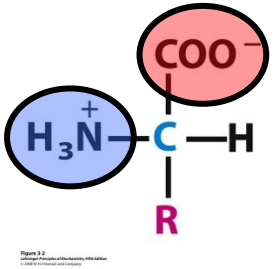




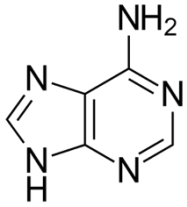
Carbohydrates

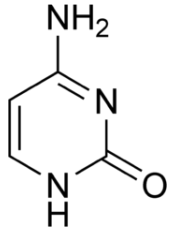


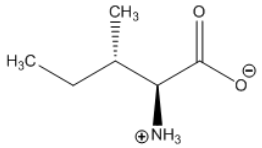
α–Amino Acids

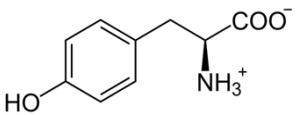


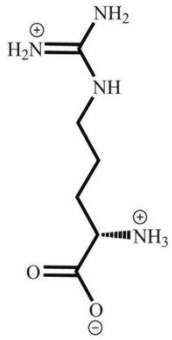
Heterocyclic Aromatic "Nucleobases"





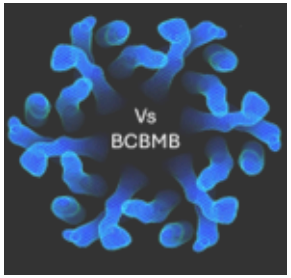






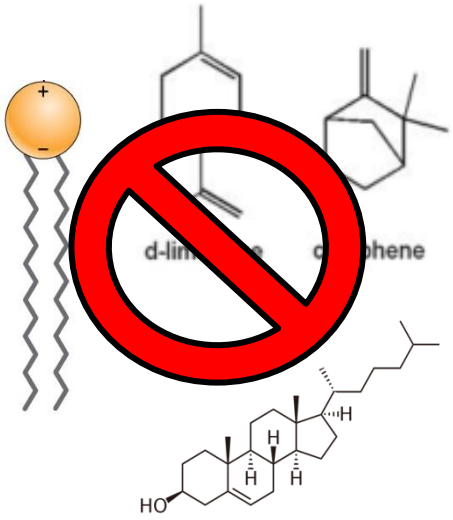
Lipids – Carbohydrates– Aromatic "Nucleobases" – Amino Acids

if information storage requires a covalent polymer –
then **which** of these **can be ruled out right away**?

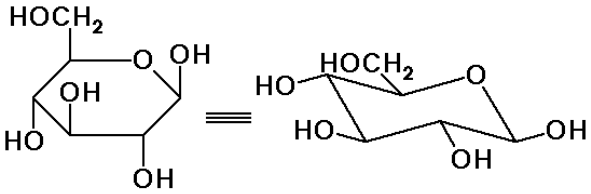


Life's Basic – Macromolecular Inventory

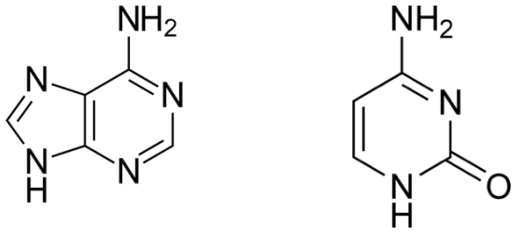
Lipids



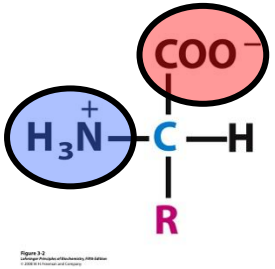
Carbohydrates

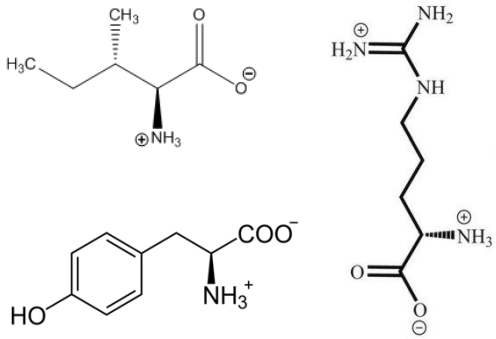


Heterocyclic Aromatic "Nucleobases"

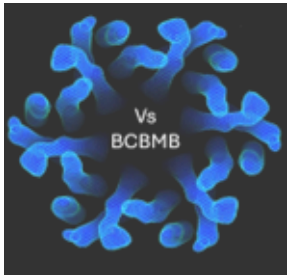


α–Amino Acids



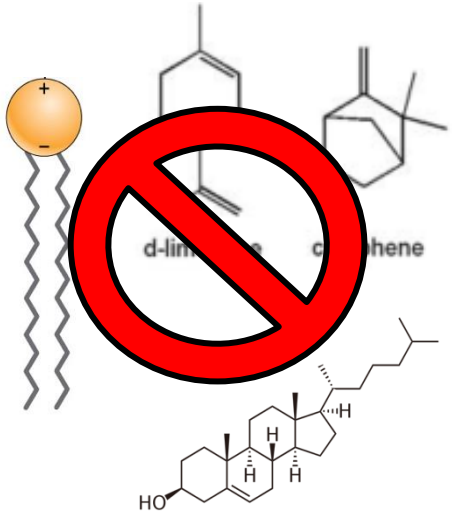


Answer: lipids – because they do not form covalent polymers → no defined structure that would stay unchanged over time.

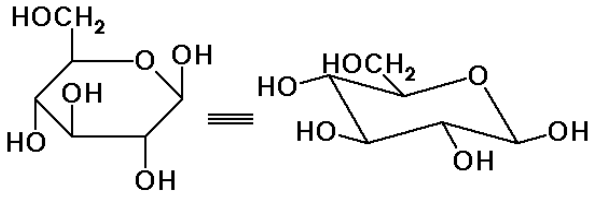


Life's Basic – Macromolecular Inventory

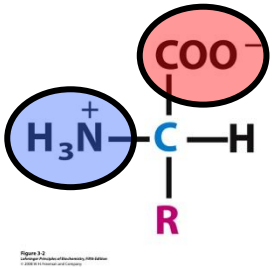
Lipids

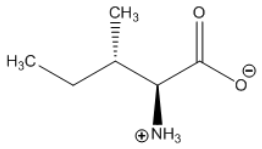


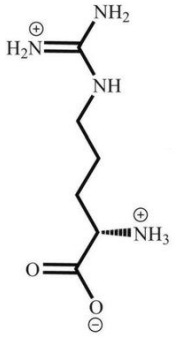
Carbohydrates



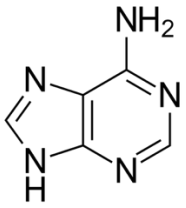
α–Amino Acids

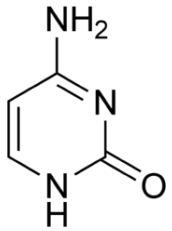


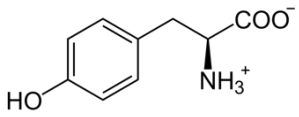




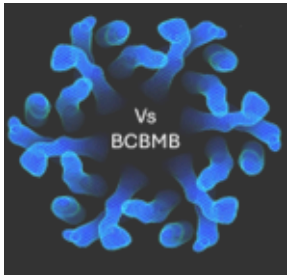
Heterocyclic Aromatic "Nucleobases"





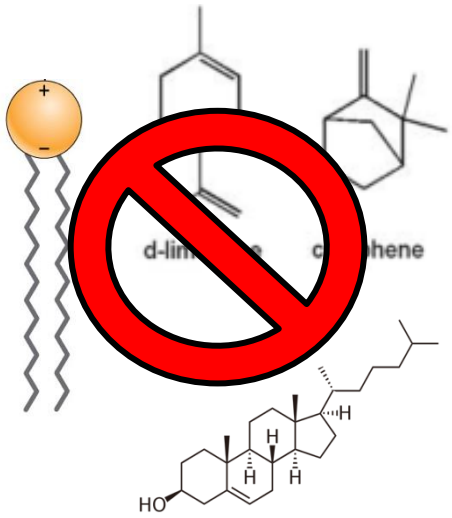


What About Carbohydrates?

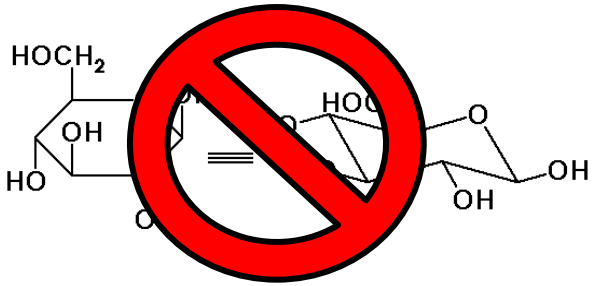


Life's Basic – Macromolecular Inventory

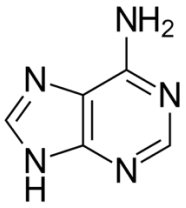
Lipids

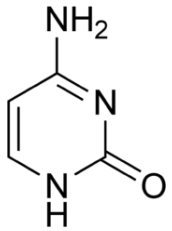


Carbohydrates

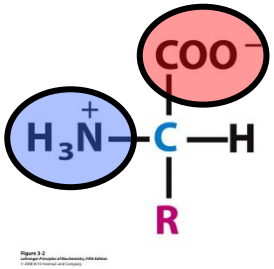


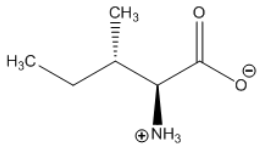
Heterocyclic Aromatic "Nucleobases"

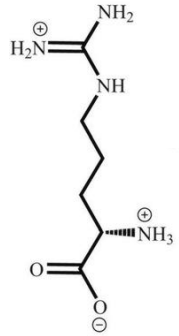


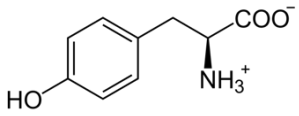


α–Amino Acids



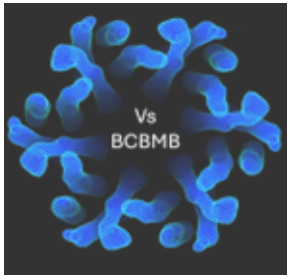






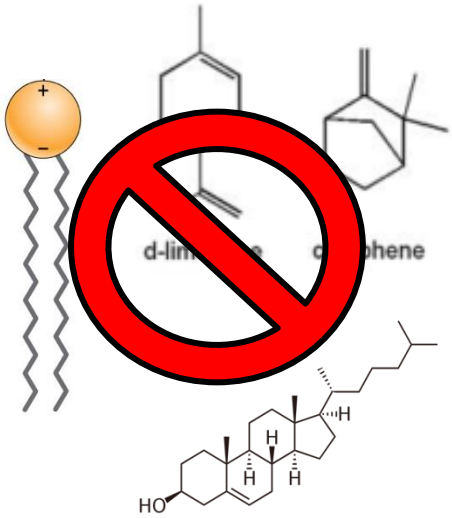
What About Carbohydrates?

cellulose, for instance, is linear …but if its **rigid properties** are any indication, then making even longer chains of something like cellulose is not going to help **+ having just –OH groups, or bulky modifications to build "bits" is not promising**

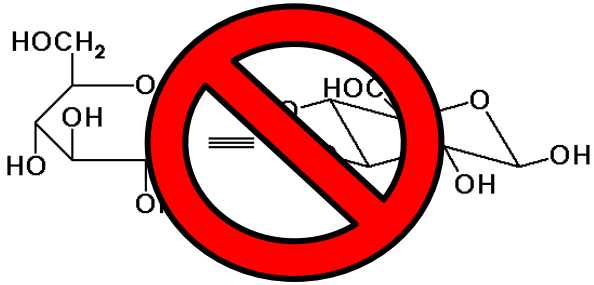


Life's Basic – Macromolecular Inventory

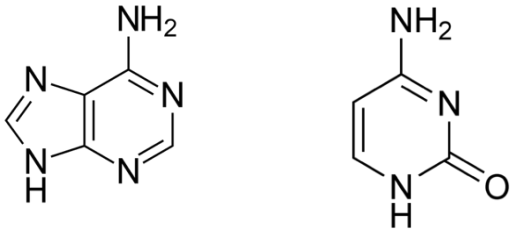
Lipids



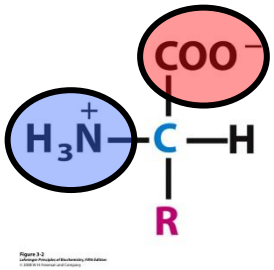
Carbohydrates

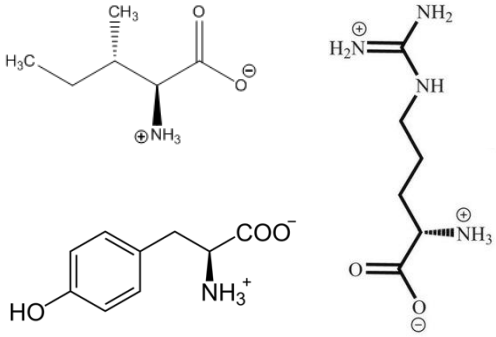


Heterocyclic Aromatic "Nucleobases"

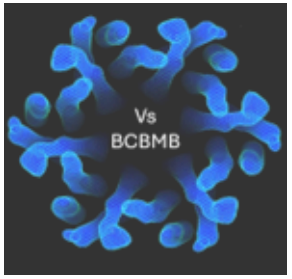


α–Amino Acids



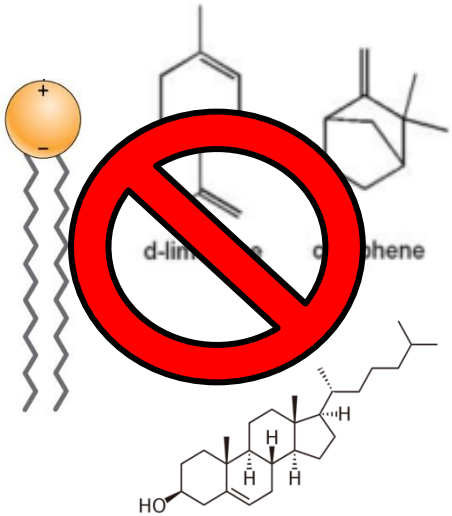


What About Amino Acids?

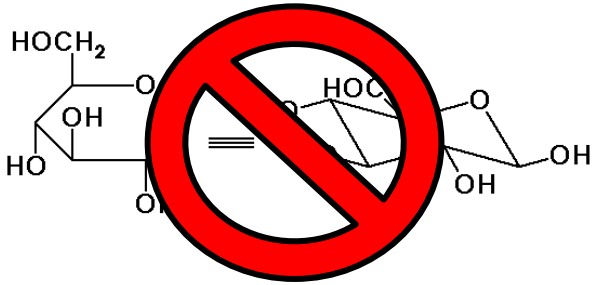


Life's Basic – Macromolecular Inventory

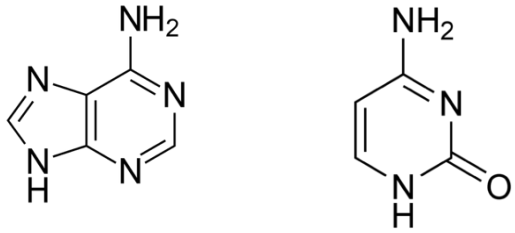
Lipids



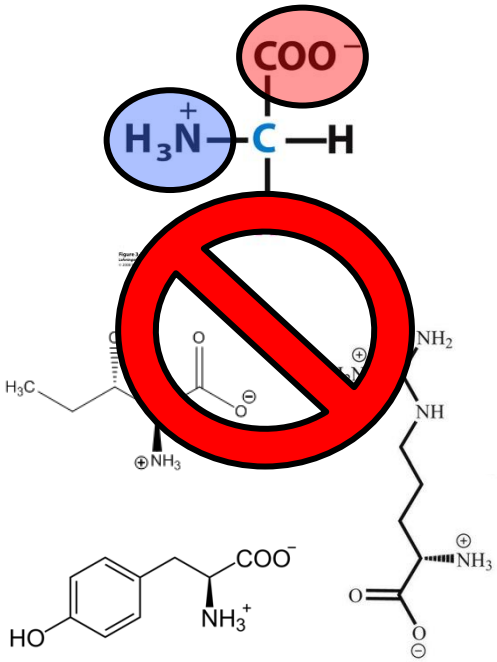
Carbohydrates



Heterocyclic Aromatic "Nucleobases"

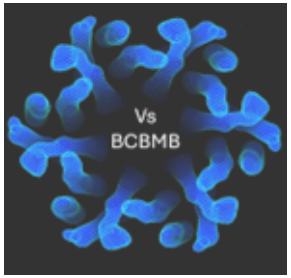


α–Amino Acids



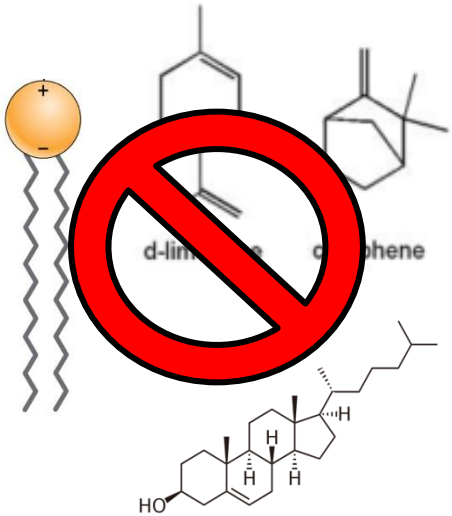
What About Amino Acids?

these can be polymerized for sure by linking amino acids through elimination of water to form a –CO-NH- linkage (we'll cover that in the next lecture), but the "R-groups" look like they combine shape and chemistry ….recognizing this by touch only …may not be straight forward…..not likely to be it...

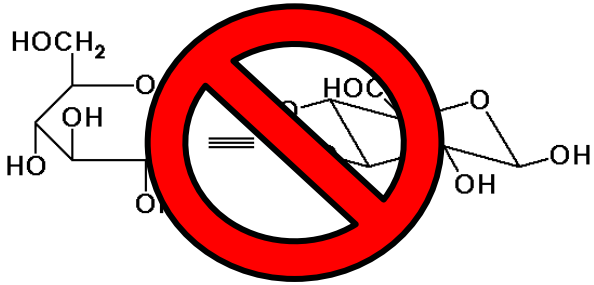


Life's Basic – Macromolecular Inventory

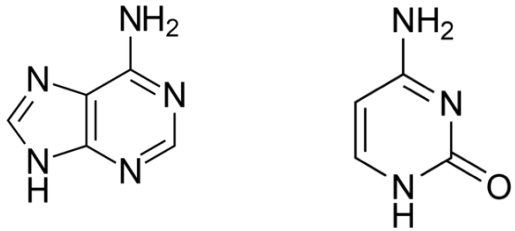
Lipids



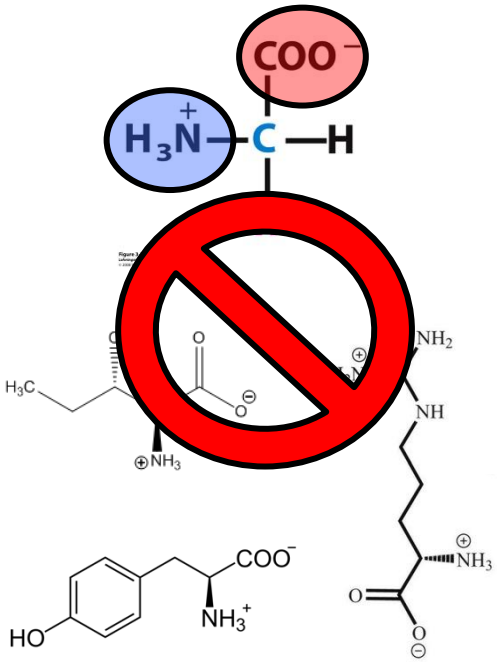
Carbohydrates



Heterocyclic Aromatic "Nucleobases"

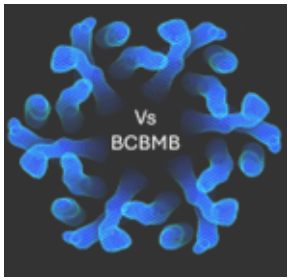


α–Amino Acids



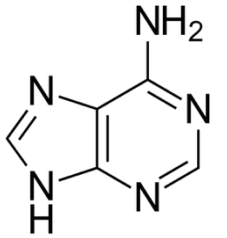
Nucleobases?

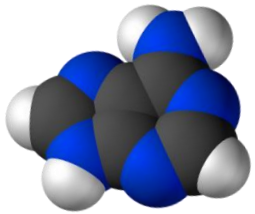
....you may get nervous because those are the only ones left in the race …. if they are not "perfect" for making "easy to read/recognize polymers" then "game over" but then…… since you **are** looking at this, you already know that they **have to be "it"…..but why?**



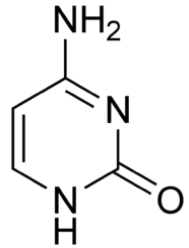
Long-Term Information Storage
Why are Heterocyclic Aromatic "Nucleobases" the Answer?

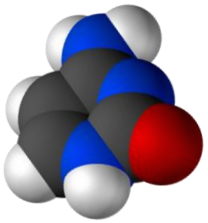
Adenine



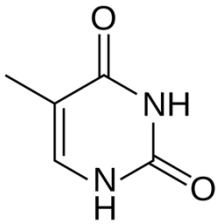


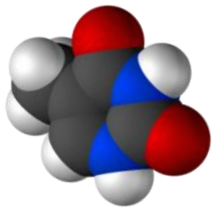
Cytosine



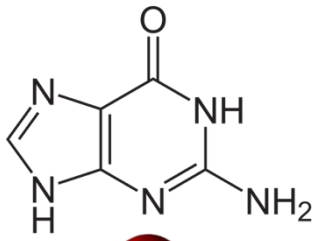


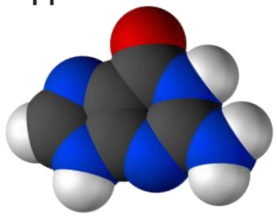
Thymine





Guanine

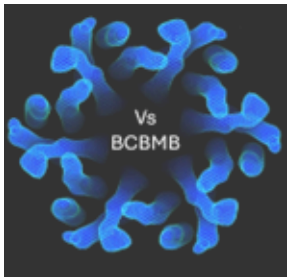




nucleobases are **aromatic** compounds
→ they are **planar (flat)** due to the sp^2 hybridization of the ring forming atoms.

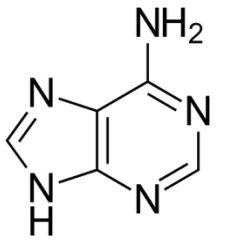
why is "planar" significant?

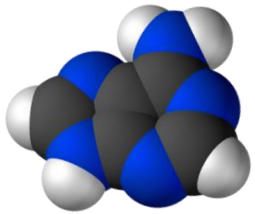
…can you see what the answer is here? … try!



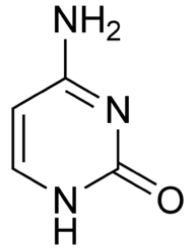
Long-Term Information Storage
Why are Heterocyclic Aromatic "Nucleobases" the Answer?

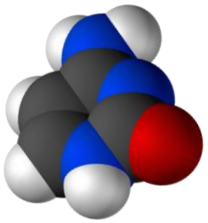
Adenine



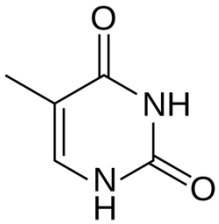


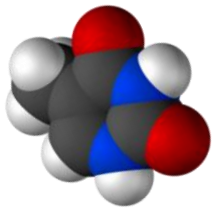
Cytosine



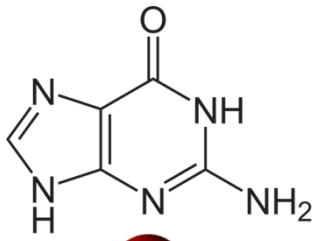


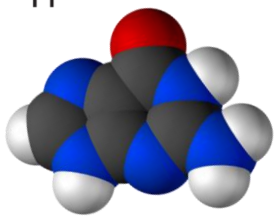
Thymine





Guanine





nucleobases are **aromatic** compounds
→ they are **planar (flat)** due to the sp^2 hybridization of the ring forming atoms.

why is "planar" significant?

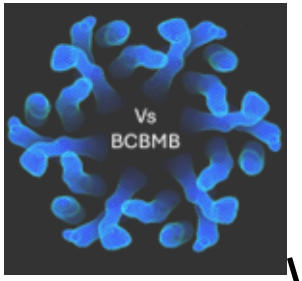
it's super significant because seen parallel to the ring it basically looks like this:

One nucleobase 

● functional group
H-bond donor or acceptor
"how do molecules see?" Pt 1
slides 15-17

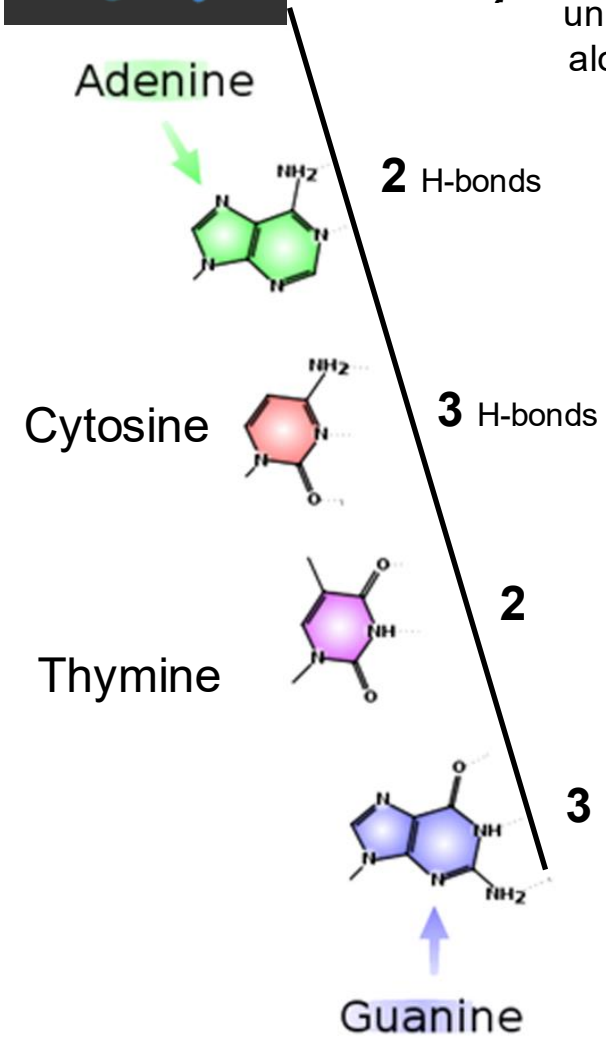
"Many" nucleobases

1D-line! ….**exactly** what is required.....



Long-Term Information Storage - Molecular Bits

combining the flat structure of the nucleobases with the very narrow geometric tolerance for H-bonds allows the creation of unambiguous pattern of H-bond donors and acceptors along a single line (= equivalent of a "bar code")

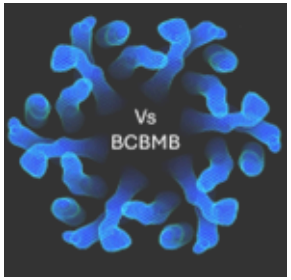


→no complicated shape to explore!
→"touch" just needs to distinguish between two possibilities: 2 or 3 H-bonds

If a string of computer bits look like this
1000001101010101010101010111110000000101010101000101010111110000….

Then its biological equivalent looks like
3222223232323232323232333332222223232323223222232323332222…..

→ while this design is easy, unambiguous, and efficient to create code, there is one big problem though – can you tell **what** it is?

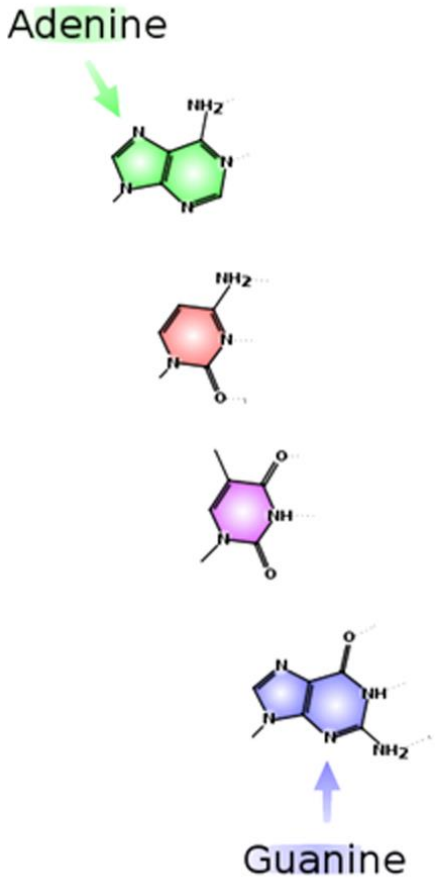


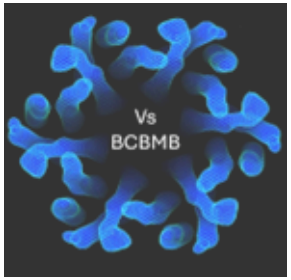
Designing the Long-Term Information Storage Polymer = Nucleic Acids

while this design is easy, unambiguous, and efficient to create code, there is one big problem though
can you tell **what** it is?

Answer: nucleobases on their own cannot readily polymerize ..and even if they could, the H-bond donors and acceptors would get so close to each other that the pattern would no longer be unambiguous
→ need a suitable scaffold to link and display bits.

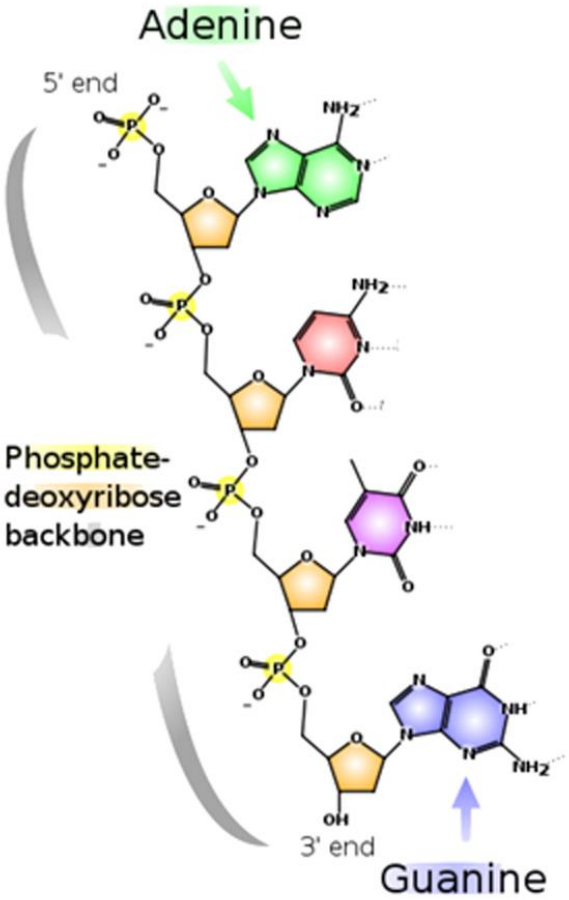
how do we fix this?





Designing the Long-Term Information Storage Polymer = Nucleic Acids

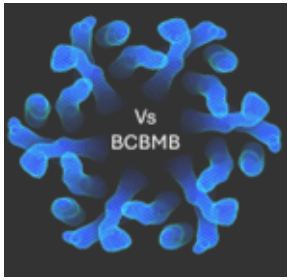
while this design is easy, unambiguous, and efficient to create code, there is one big problem though
can you tell **what** it is?



Answer: nucleobases on their own cannot readily polymerize ..and even if they could, the H-bond donors and acceptors would get so close to each other that the pattern would no longer be unambiguous
→ need a suitable scaffold to link and display bits.
how do we fix this?

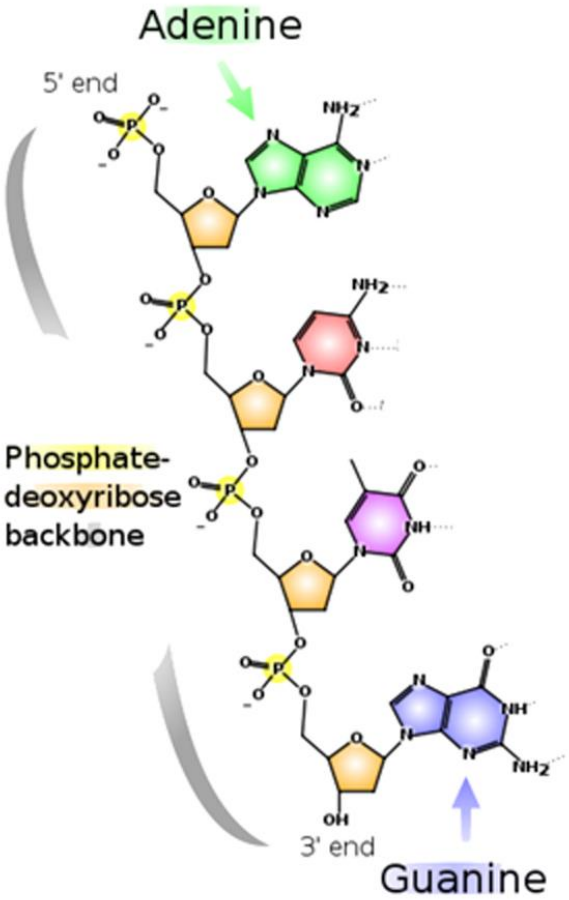
Answer: graft the “fingers” on a “hand/arm”.
ironically: carbohydrates, are just perfect for this purpose

why?



Designing the Long-Term Information Storage Polymer = Nucleic Acids

while this design is easy, unambiguous, and efficient to create code, there is one big problem though
can you tell **what** it is?

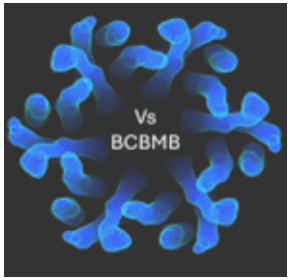


Answer: nucleobases on their own cannot readily polymerize ..and even if they could, the H-bond donors and acceptors would get so close to each other that the pattern would no longer be unambiguous
→ need a suitable scaffold to link and display bits.
how do we fix this?

Answer: graft the "fingers" on a "hand/arm".
Ironically: carbohydrates, are just perfect for this purpose
why?

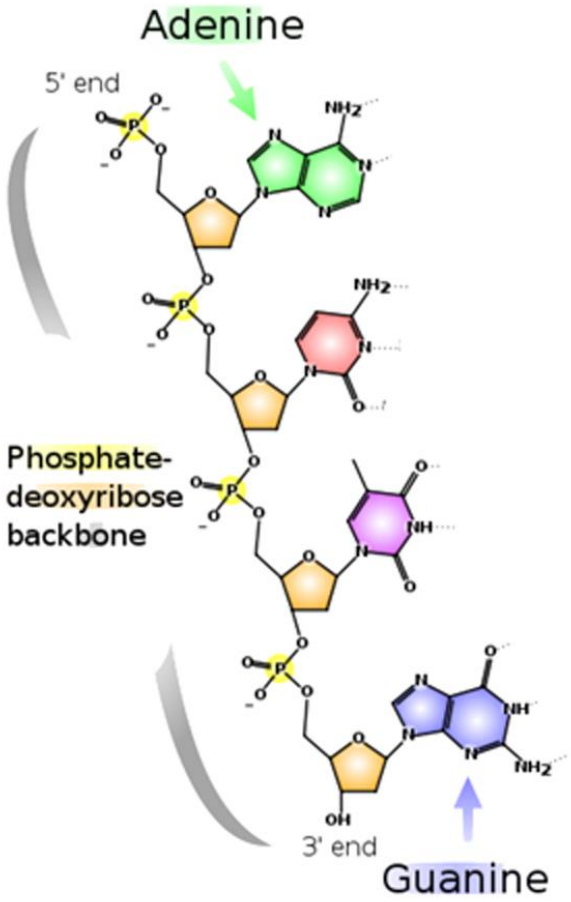
Answer: simple chemical functionalities (-OH groups) that allow multiple couplings on each sugar through a variety of chemical bonds (already saw "branching" in amylopectin as example, "Biochemistry Fundamentals – Carbohydrates" Slide 33).

→ but **aren't carbs a disaster in terms bulk and rigidity??** (think cellulose, chitin, amylose…)



Designing the Long-Term Information Storage Polymer = Nucleic Acids

while this design is easy, unambiguous, and efficient to create code, there is one big problem though
can you tell **what** it is?



Answer: nucleobases on their own cannot readily polymerize ..and even if they could, the H-bond donors and acceptors would get so close to each other that the pattern would no longer be unambiguous
→ need a suitable scaffold to link and display bits.
how do we fix this?

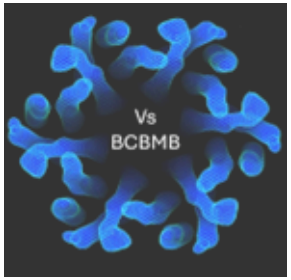
Answer: graft the "fingers" on a "hand/arm".
Ironically: carbohydrates, are just perfect for this purpose
why?

Answer: simple chemical functionalities (-OH groups) that allow multiple couplings on each sugar through a variety of chemical bonds (already saw "branching" in amylopectin as example, "Biochemistry Fundamentals – Carbohydrates" Slide 33).

→ but **aren't carbs a disaster in terms bulk and rigidity??** (think cellulose, chitin, amylose…)

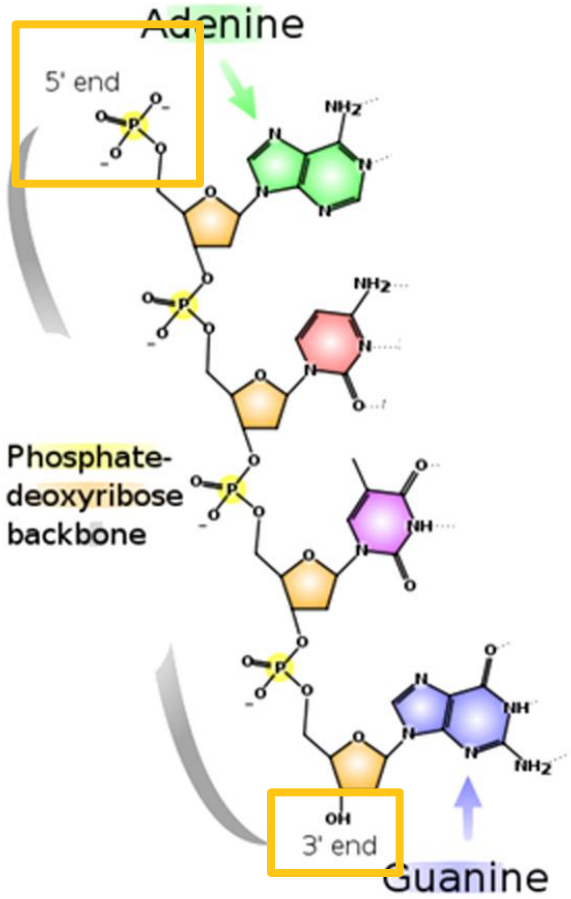
Answer
certainly would be - for exactly the reasons we encountered – if Nature had tried to use a pure polysaccharide template

→ how are nucleic acids different then?



Designing the Long-Term Information Storage Polymer = Nucleic Acids

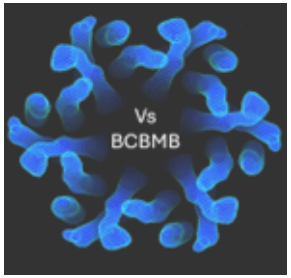
Answer
certainly would be - for exactly the reasons we encountered – if Nature had tried to use a pure polysaccharide template
→ **how are nucleic acids different then?**



note: this **polymer is asymmetric**: 5' (phosphate) → 3' (hydroxyl)

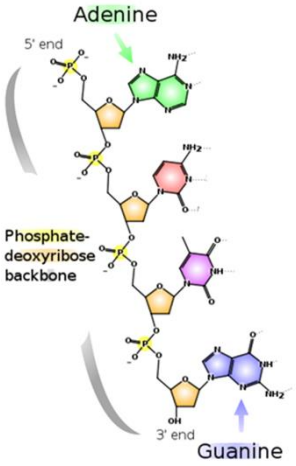
Answer
they are **composite polymers** in which the carbohydrate units are separated from each other through phosphate groups. This **solves a LOT of issues**
for instance:

- increases flexibility by spatially separating the sugar rings while maintaining free rotation on either side of the phosphate
- the added distance between the nucleobases now sets up a clear register of "2s and 3s"
- polyanionic backbone structural tendencies are independent of base pair that is coupled → can change "info=nucleobase" without impacting basic structure
- easy to synthesize (we'll cover that in the "Molecular Biology" Series of Lectures)
- makes polymer more resistant to chemical cleavage by water



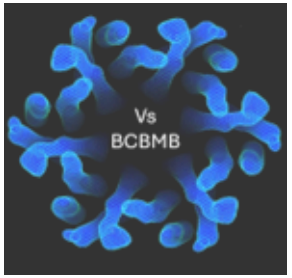
Interpreting Long-Term Information Storage

with a structural solution for information coding in hand and looking ahead to “transmission of information”, the question becomes:



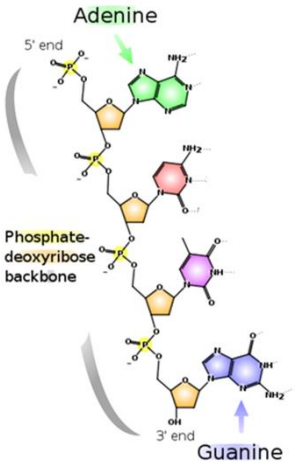
→how is information accessed and read out?
or in other words:
how do you best interpret a linear string of bits?

…try to find an answer here before looking at the next slide ….



Interpreting Long-Term Information Storage

how is information accessed and read out?
or in other words:
how do you best interpret a linear string of bits?

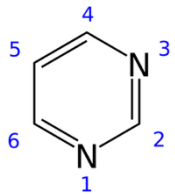


Answer

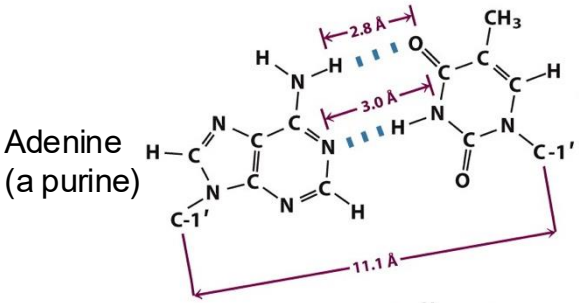
through another 1D-string that is **complementary** to the template
= able to match 2:2 and 3:3

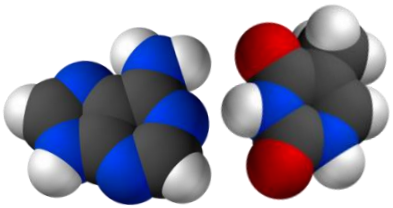
→ "only" challenge for Nature was to find these "base pairs" (**bp**)

Purine (parent)

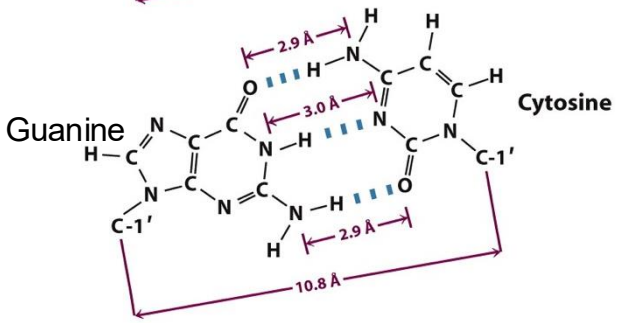


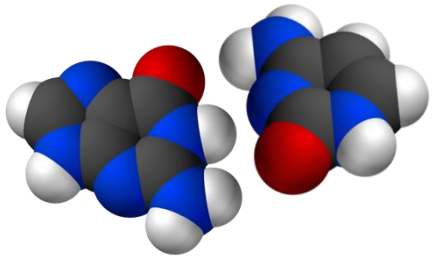
Pyrimidine(parent)



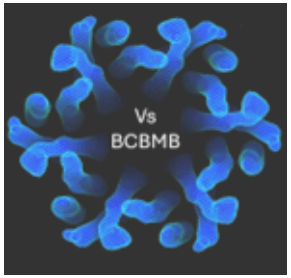


A **T**



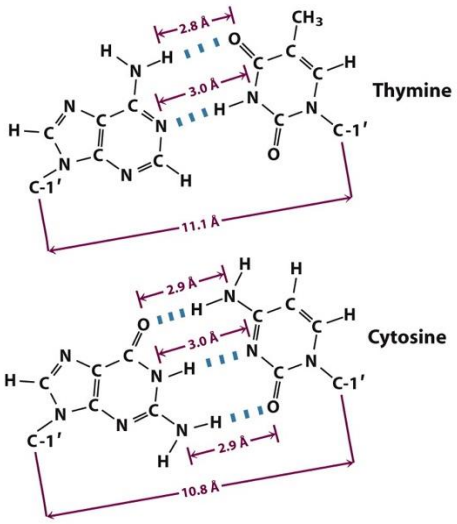


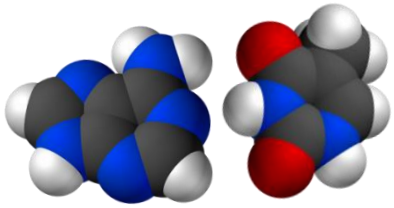
G **C**



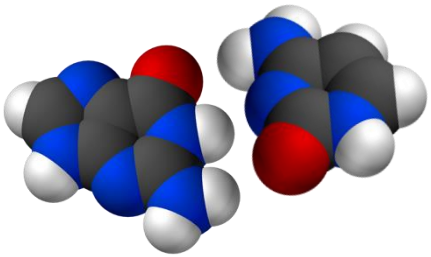
Interpreting Long-Term Information Storage

there is one last thing to take note of here





A **T**



G **C**

the space filling model visually drives home why Nature chose H-bonds as the principal weak interaction to define information coding….

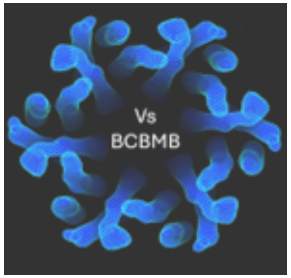
…because charge:charge or hydrophobic Van der Waals interactions have spherical geometry = they can form in any direction ….

H-bonds, on the other hand, are **directional** (donor → acceptor) and have a **very narrow tolerance for deviation from linear geometry** (= the donor and acceptor have to "directly look at each other" with no or only very little rotational offset from a straight line)

→**these interactions will be confined to a plane**!!

their energy is also …large enough to make these contacts "sticky" = things can lock into register without slipping

(we had seen this also in how cellulose forms very strong "cables"; Biochemistry Fundamentals – Carbohydrates, Slides 21&31)

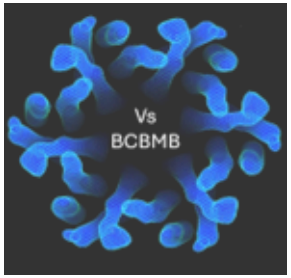


The Power Of Biological Information Storage

the ability to match **A with T** and **G with C** allows for an unthinkable degree of **data compression** ….because this molecular solution reduces the **<u>entire</u>** molecular complexity into instructions that are coded by just 4 letters

(compare: shortest alphabet in use is 12 letters (language: Rotokas; 4300 speakers; longest is 74 letters (language: Khmer; 13 million speakers)).

...try to grasp what that actually means, it's mindblowing ...



The Power Of Biological Information Storage

the ability to match **A with T** and **G with C** allows for an unthinkable degree of **data compression** ….because this molecular solution reduces the **entire** molecular complexity into instructions that are coded by just 4 letters

perspective

average human body has ~37 trillion cells, each of which is made of 10s of thousands of different chemicals at any one point in time...number of letters for the **instruction manual that tells "what" "when" "where" "how long for" and "how":**

4 letters

total number of letters: ~$6x10^9$ of which ~$6x10^7$ bp are read out to make other molecules

➔ the degree of **lossless** compression is easily >10^{10}

(human engineering accomplishes ~100-fold),

not even accounting for the fact that you start it all from 1 egg and 1 sperm....

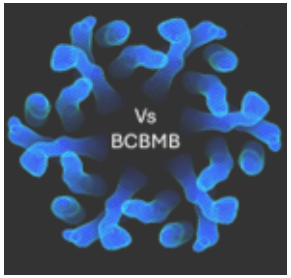
....that (to me) is mindblowing.

the three images illustrate what compressing information does to an image

up to a point, you get away with it, but there will be a point where the image becomes uninterpretable

….

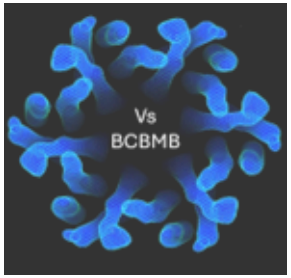
life cannot tolerate loss because loosing information quickly becomes deadly



When Simplicity Becomes a Problem

while necessary to allow life to evolve and succeed, **the conceptual simplicity of information storage also causes a significant problem**
do you see what that could be?

…try to think of an answer here ….



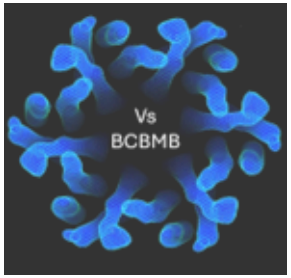
When Simplicity Becomes a Problem

while necessary to allow life to evolve and succeed, **the conceptual simplicity of information storage also causes a significant problem**
do you see what that could be?

Answer
degeneracy
meaning: 4 “sticky” letters that match in 2 complementary pairs (A/T; G/C), forming a flexible polymer with ~$6x10^9$ characters (in a somatic human cell)

➔ **if left to itself, this will curl up on itself**
➔ **that would be a HUGE problem**

why?



When Simplicity Becomes a Problem

while necessary to allow life to evolve and succeed, **the conceptual simplicity of information storage also causes a significant problem**
do you see what that could be?

Answer
degeneracy
meaning: 4 “sticky” letters that match in 2 complementary pairs (A/T; G/C), forming a flexible polymer with ~$6x10^9$ characters (in a somatic human cell)

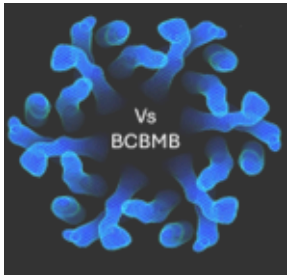
➔ **if left to itself, this will curl up on itself**
➔ **that would be a HUGE problem**

why?

Answer
because there will be many possible ways how this long, very flexible polymer can fold on itself
(to turn the chain around by 180˚only needs 3 base pairs)
whatever structure forms will be random
(=n copies of the string will form n different 3D structures)
and very stable because it is held together by many H-bonds (2 to 3 **per** base pair)

→ if that were to happen, spatial and temporal control over the use of the information is lost
= you would have the **entire** instruction manual, but the instruction is permanently locked and cannot be executed (a molecular form of "locked in Syndrome")

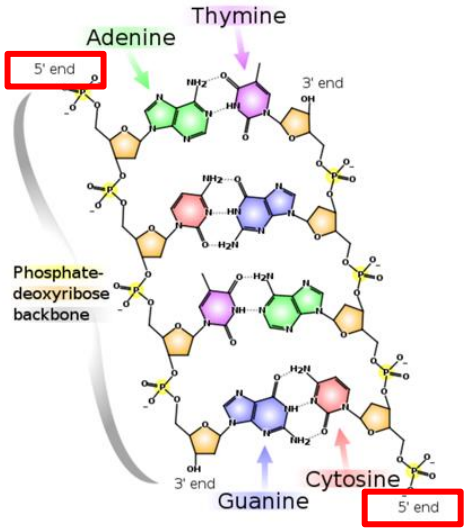
→ how can you avoid this disaster from happening?

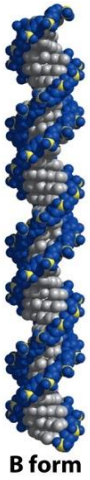


The Double Helix

Answer

surprisingly simple…just **make a second copy of the polymer that complements the first copy over the entire length**, and store them as pair.





Note: the two strands are antiparallel

still leaves the task/challenge to spatially organize it/compact it, but that is doable and can be controlled

(covered in "Molecular Biology and Cell Biology – GENOMES" and "CHROMATIN" Lecture)

incidentally, the resulting double helical structure is extraordinarily stable.

in fact, it's the most stable biopolymer with nothing else coming close

→ only biopolymer that can be retrived from specimen like mummies and amber preserved insects long after the organism perished.

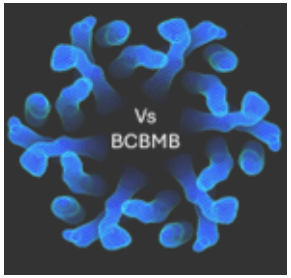
→ **and there it is: the molecular answer to the problem of long-term storage is a Double Helix**

the particular type of nucleic acid used for long-term information storage is called

DNA = deoxyribonucleic acid

deoxyribo: refers to the sugar used in the backbone (deoxyribose)
nucleic: DNA was first discovered in the nucleus of cells
acid: the large number of phosphate groups make DNA very acidic

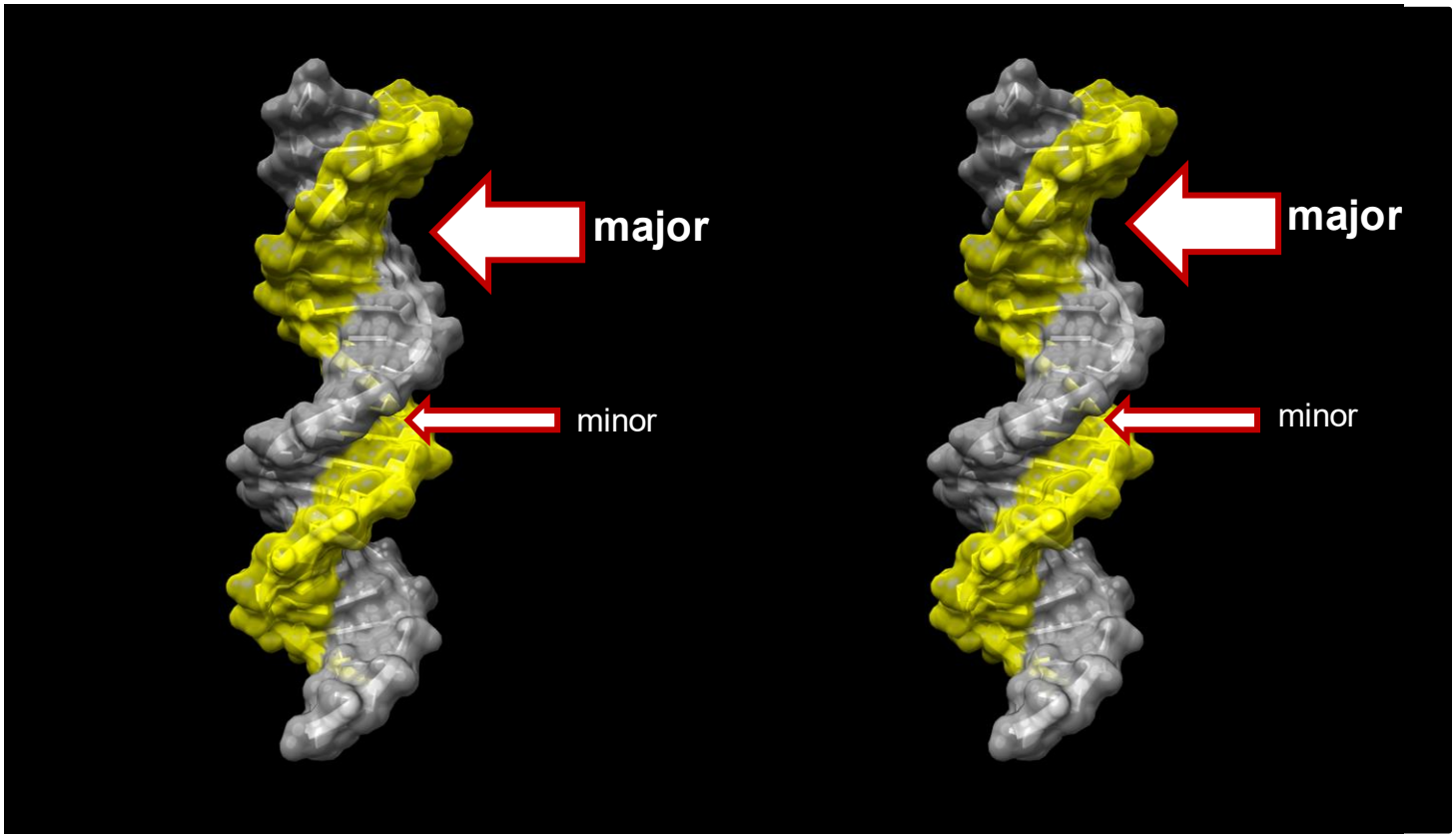
a related polymer, ribonucleic acid (RNA) uses ribose as sugar and the nucleobase U instead of T; covered in the "Advanced Biochemistry" Collection of Handouts

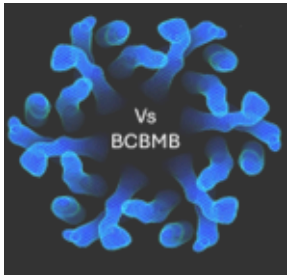


The Double Helix

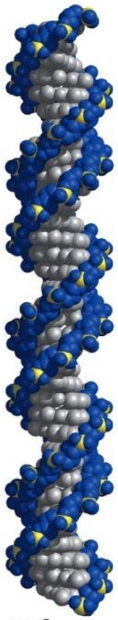
let's look at DNA structure in 3D because it helps you understand its defining structural features.…
the antiparallel orientation of the two strands causes the formation of two distinct regions – called "grooves" - along the surface of the double strand.
the narrower groove is called "minor groove", the wider one "major groove".

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





Designing the Long-Term Information Storage Polymer = Why Double Helix?



B form

the determination of DNA structure by Rosalind Franklin (collected the data), and Watson & Crick (“stole” data and built the model from it) was transformative for understanding of life because it suggested simple mechanisms for heredity and usage of the genetic information in a cell.

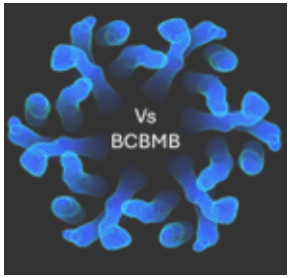
if you are interested in the controversy around who "discovered the structure of DNA" and why the determination of DNA structure counts as one of the "darkest hours" in biophysics, then here is a link for you to get started.

Rosalind Frankin

as exciting as the emerging field of molecular biology was, even back then there were some that were not afraid of voicing concerning implications of these discoveries. If you are interested, read the ½ page commentary that I included after the last slide.

this commentary was written in 1967…. and just think where we are today, able to do gene therapy, editing babies (sooner than later) and such …

…also consider the similarities with the emergence of AI and its implications ….



Double Helix and Heredity

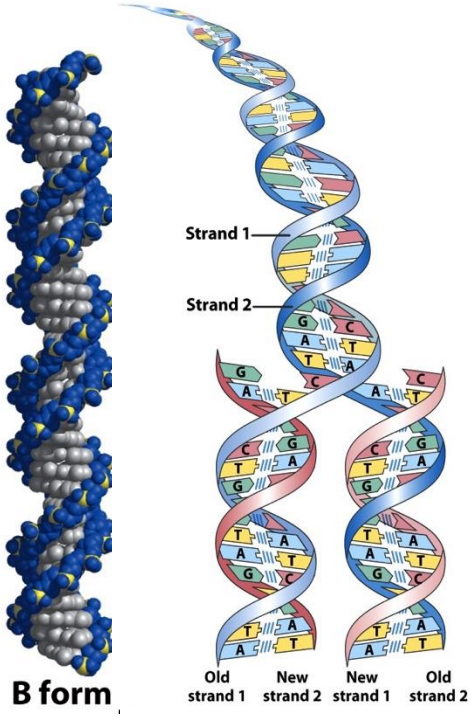
the double-stranded structure of DNA suggests a very intuitive way how the instruction manual can be passed on over time.

similarly

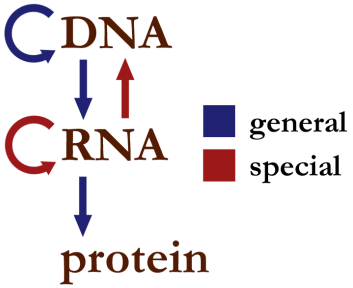
- the basic principle how to copy DNA predicts that the most efficient mechanism to execute the stored information involves single stranded intermediates that serve as “messages”

- base pair complementarity also predicts the existence of adaptor molecules that translate the nucleic acid level code into a different molecular form.

taken together – these ideas summarize the principal flow of information in biological systems, and are known as “**Central Dogma**”

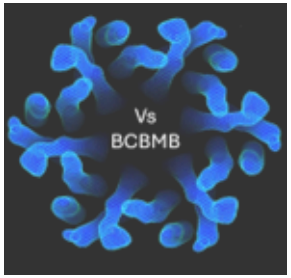


RNA: ribonucleic acid (nucleic acid that uses ribose instead of deoxyribose)
Protein: polymer made from amino acids ("Biochemistry Fundamentals" - Proteins)



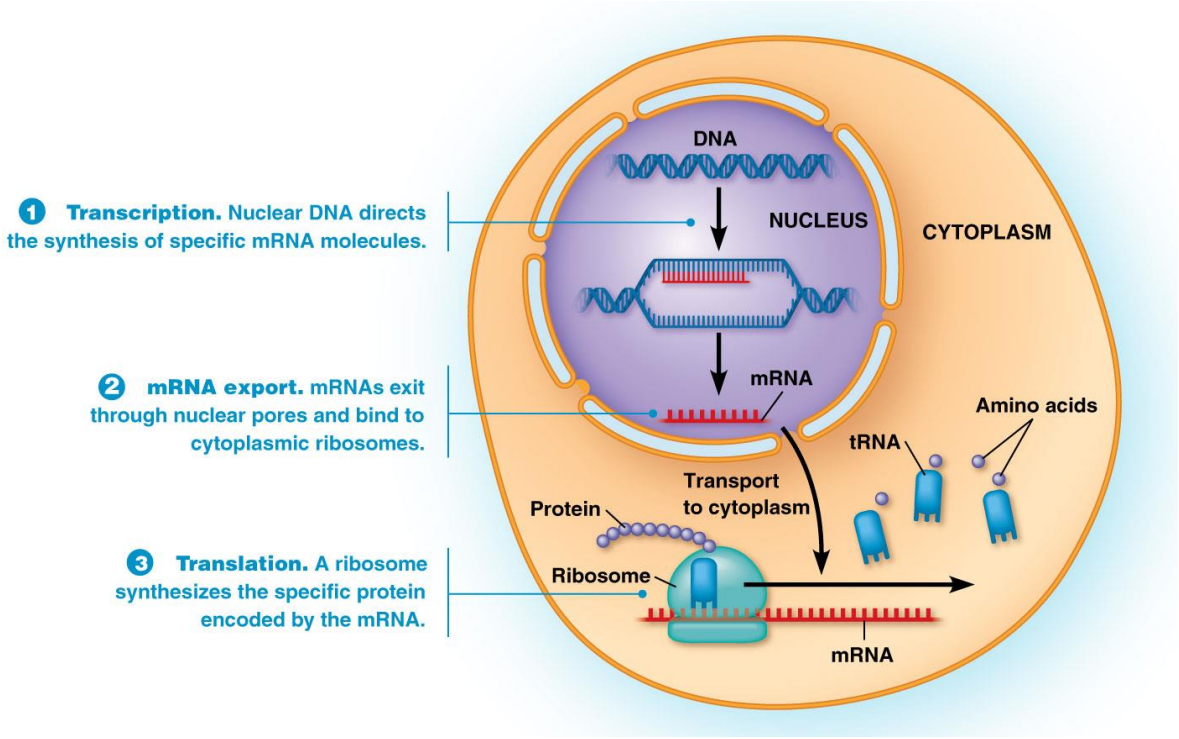
https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology

important note to take though…..while ALL these things (replication/heredity, transcription, tRNA, …) are intuitive and obvious once you see the DNA structure – **NONE** of them are the **cause/reason** for why DNA is double stranded. To just repeat this one more time: the **reason** for DNA being double stranded is to create a structural form of the polymer that remains "readable" and manageable.



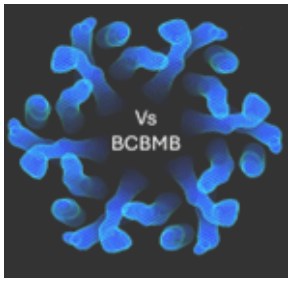
Central Dogma of Information Flow
Words → Visual

put into context of a eukaryotic cell, where DNA is confined to a membrane bound organelle, this cartoon certainly captures the general ideas of information flow from instruction → product.



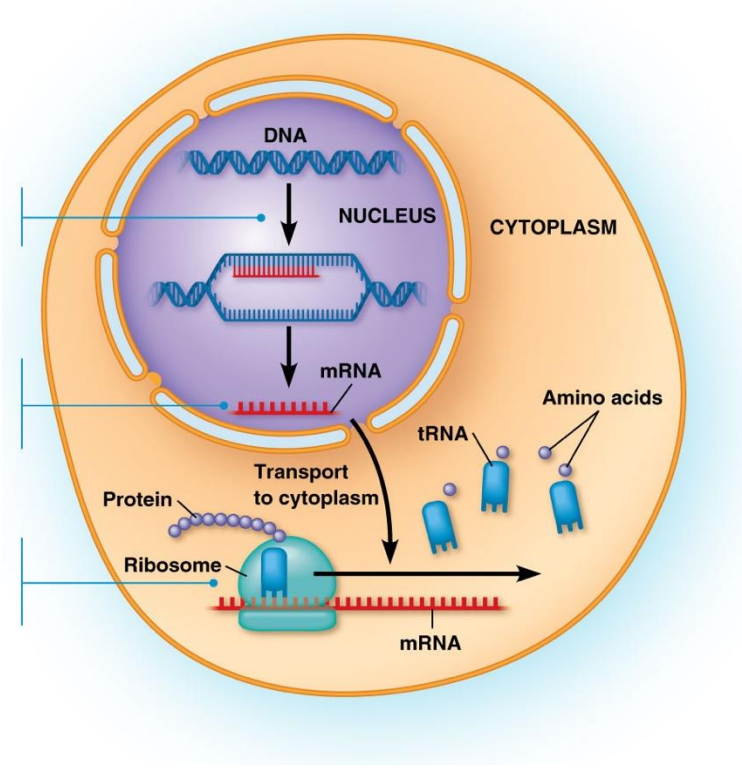
…but: when thinking about “transmission/processing”...there is something “odd” about this picture that might (should) bother you based on what we covered in previous lectures (especially: "How Do Molecules See? Pt2" slide 18)

→ what is puzzling here?

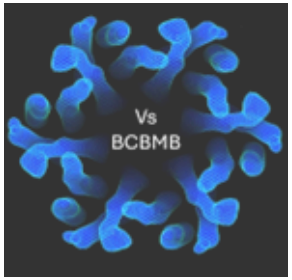


Information Flow/Processing: Coming Full Circle

the flow of information seems unidirectional
→ that is odd because everything is based on equilibria…!
→ **why is there no way back?**

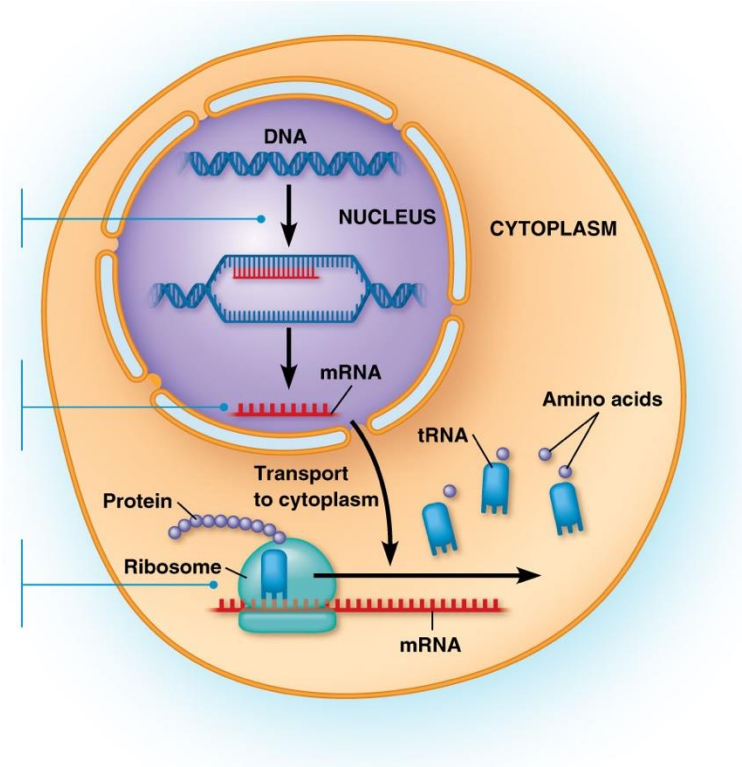


Answer: *…try to answer first before looking at the next slide….*



Information Flow/Processing: Coming Full Circle

The flow of information seems unidirectional
→ that is odd because everything is based on equilibria…!
→ **why is there no way back?**



Answer

- the chemistry of the processes that are involved is irreversible under physiological conditions

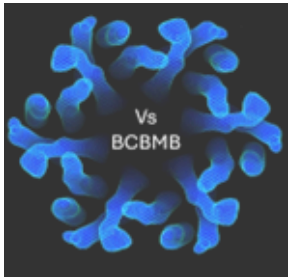
- **more important though**

crossing over into "protein space" is the point where the "decompression" of information content occurs – that step is irreversible

(think of putting a foam mattress back into its tightly packed roll without special tools after it expanded; or stuffing an airbag back into its compartment after it deployed)

➔ **how can the central dogma be reconciled with life?**

….try to think of an answer…..



Information Flow/Processing: Coming Full Circle

Answer:

- the chemistry of the processes that are involved is irreversible under physiological conditions.
- **more important though**: crossing over into "protein space" is the point where the "decompression" of information content occurs – that step is irreversible

➔ **how can the central dogma be reconciled with life?**

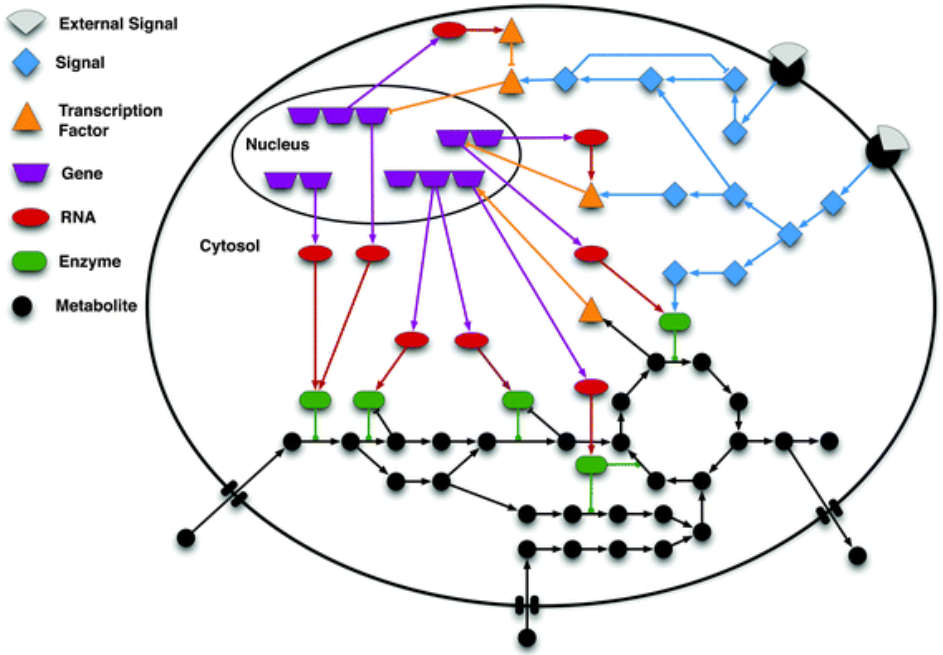
Answer:

"responsiveness"

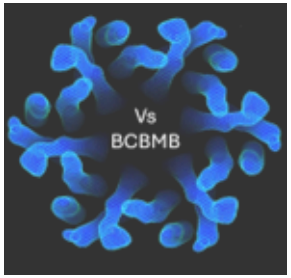
at the outset we distinguished between long-term and short-term memory.

→the cell-level cartoon of the central dogma ignores that cells respond to changes in (internal) environment through use of emergent, non-permanent information processing.

→changes in boundary conditions trigger feedback loops that dynamically change the behavior of the source code, including duplication, which allows closing the circle.



so in some sense: the cell is constantly "falling forward" … which at some point triggers creation of a new copy of the highly compacted information data base to pass to the next generation. This is possible only by constantly expending chemical energy …. (yes...mostly in the form of ATP hydrolysis (for those that already know a little more)).



SUMMARY

- **Nucleic Acids** are Nature's answer to the challenge of having to store large amounts of information in a compact, yet accessible way
- **Nucleic Acids** are **polymers,** named nucleic acids based on the acidic nature of the optically dense material that dominates the cell nucleus

- Nucleic Acids are built from small building blocks called nucleotides that themselves are made of three components: a "nucleobase" (heteroaromatic compounds), a sugar molecule with five carbon atoms (called ribose → RNA or deoxyribose → DNA), and inorganic phosphate

- Polymerization of these building blocks results in an asymmetric (two distinguishable and chemically different ends) **composite polymer** in which adjacent bases are separated by an inorganic phosphate bridge.

- This design is the **only** biological polymer that **can** store information because the flat structure of the nucleobases allows the creation of a polymer in which **all** information carrying functional groups (H-bond donors and acceptors) can be presented along a **1D-line**

- This "**1D-line"** design is **essential** for information storage because it is the **only** design that allows rapid and unambiguous molecular recognition/reading by a **complementary** nucleic acid, exploiting the linear geometry and directionality (donor → acceptor) of H-bonds

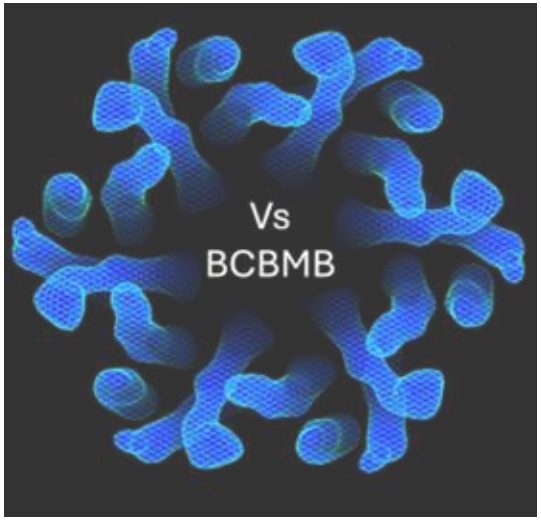
- Incidentally, DNA only requires **four nucleobases (**2 purines, 2 pyrimidines) to encode the basic information
  (additional layers of information imprint on DNA is covered in the "Molecular Biology" and "Advanced Biochemistry" Lectures)

- To create a stable, readable and manageable data repository, DNA **must be** double stranded

- **Heredity, and transcribability** are advantageous **side effects of,** but not the reason for, **the double stranded nature of DNA.**

- In a cellular context, only information that is needed is accessible for making the shorter copies that serve as templates for protein synthesis and regulatory processes

- ➔ the core tenet of the central dogma of cell biology holds that information flows: DNA→RNA→Protein



Thank You for Working Through This
if you found this Silent Lecture useful – recommend it to friends and consider watching other free silent lectures on YouTube @VsBCBMB

or download other the free pdfs (with clickable links) from this site

Will Society Be Prepared?

New information is being obtained in the field of biochemical genetics at an extremely rapid rate. Thus far, this knowledge has had relatively little effect upon man. More information must be obtained before practical application will be possible, and the technical problems that must be overcome are formidable. However, when these obstacles have been removed this knowledge will greatly influence man's future, for man then will have the power to shape his own biologic destiny. Such power can be used wisely or unwisely, for the betterment or detriment of mankind.

Salvador Luria has said: "the progress of science is so rapid that it creates an imbalance between the power it places in the hands of man and the social conditions in which this power is exerted. Then neither warnings of scientists, nor breadth of public information, nor wisdom of citizens may compensate for inadequacies of the institutional framework to cope with the new situations."

The public understands to some extent the recent developments in biochemical genetics, but has only a vague notion of what may be expected in the future, in spite of the efforts of many scientists to inform the public about probable future developments.

Where do we stand today? The genetic language now is known, and it seems clear that most, if not all, forms of life on this planet use the same language, with minor variations. Simple genetic messages now can be synthesized chemically. Genetic surgery, applied to microorganisms, is a reality. Genes can be prepared from one strain of bacteria and inserted into another, which is then changed genetically. Such changes are inheritable. Thus far, it has not been possible to program mammalian cells in this way.

What may be expected in the future? Short but meaningful genetic messages will be synthesized chemically. Since the instructions will be written in the language which cells understand, the messages will be used to program cells. Cells will carry out the instructions, and the program may even be inherited. I don't know how long it will take before it will be possible to program cells with chemically synthesized messages. Certainly the experimental obstacles are formidable. However, I have little doubt that the obstacles eventually will be overcome. The only question is when. My guess is that cells will be programmed with synthetic messages within 25 years. If efforts along those lines were intensified, bacteria might be programmed within 5 years.

The point which deserves special emphasis is that man may be able to program his own cells with synthetic information long before he will be able to assess adequately the long-term consequences of such alterations, long before he will be able to formulate goals, and long before he can resolve the ethical and moral problems which will be raised. When man becomes capable of instructing his own cells, he must refrain from doing so until he has sufficient wisdom to use this knowledge for the benefit of mankind. I state this problem well in advance of the need to resolve it, because decisions concerning the application of this knowledge must ultimately be made by society, and only an informed society can make such decisions wisely.—MARSHALL W. NIRENBERG, *National Heart Institute*

This editorial is adapted from remarks made in accepting the Research Corporation's 1966 award.

Appendix

Thoughts of An Independent Thinker