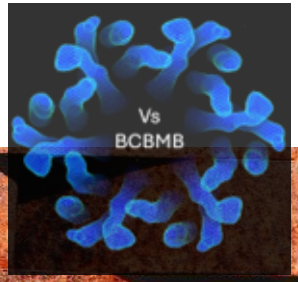


Biochemistry Fundamentals

PROTEINS

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

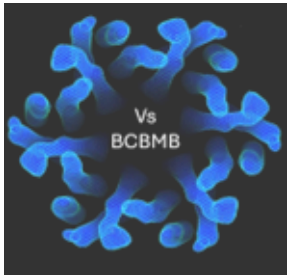
- **by thinking about/answering questions**,
(my answer is always provided on the next slide).
- **by completing a "short assignment"**
(it never will take more than a few minutes, if at all that long),
- by **watching a short video/clip**
(the embedded links will take you to my YouTube@VsBCBMB channel;
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 introductory primer on protein biochemistry can stand on its own, you will achieve a greater gain if you worked through the following silent lectures ahead of time

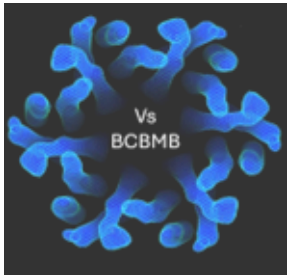
- How Do Molecules See Part 1 & 2
- Biochemistry Fundamentals – LIPIDS
- Biochemistry Fundamentals – CARBOHYDRATES
- Biochemistry Fundamentals – NUCLEIC ACIDS

the first two of the "Biochemistry Fundamentals" lectures listed above focused on how Nature managed to build **physical boundaries** that can 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.

the third of the "Biochemistry Fundamentals" lectures introduced the basic design principles of life's universal information repository, DNA, and how information flows from it to support cellular processes.

- specifically, you learned how the properties of H-bonds (linear, directional) are exploited to set up a pattern of H-bond donors and acceptors that due to the planarity of the aromatic nucleobases are displayed along a 1D-line.
  we saw why this patterning is essential/key for information storage: implementation of "molecular bits", which are easy to read because interaction partners do not need to scan a complex 3D surface to find what they are "looking for".
  moreover, the "stickiness" of the H-bond donor/accept pattern allows for very specific alignment, duplication and copying of the information.



Recap Cartoon

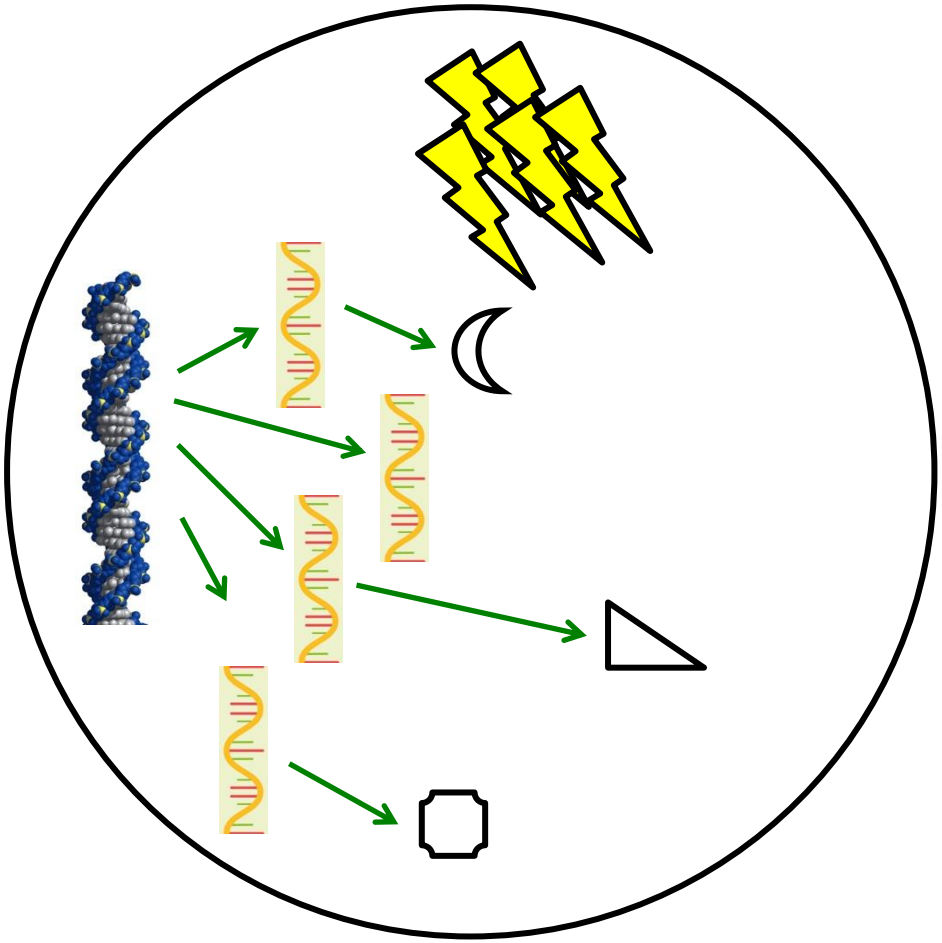
[that explains why becoming an artist was not really an option for me
….but this is good enough here]

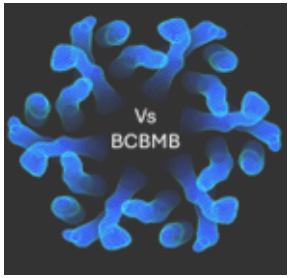
Learned
how to make boundaries (both flexible/responsive and hard → bilayers (**non-covalent)**; cell walls/exoskeletons **(covalent)**

Learned
how to store energy → starch **(covalent)**, emulsions/fat **(non-covalent)**

Learned
how to construct (long term) information storage → nucleic acids [DNA, RNA]

Learned
(1) how the "**Central Dogma**" summarizes the general idea of information flow from storage→system components
and
(2) that emergent feedback loops dynamically change information readout.

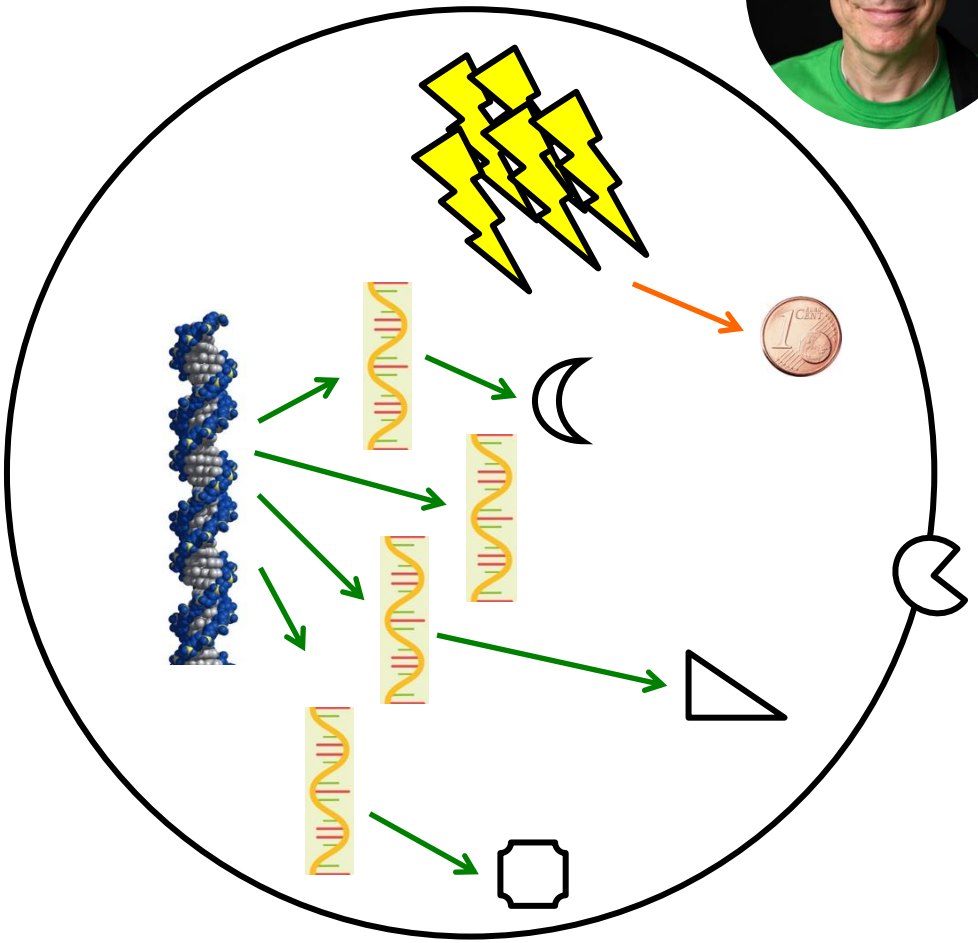




The Missing Link

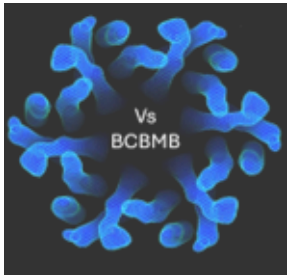
still a mystery:

- **how** any of the central dogma is implemented →
- **how** any and all of the cellular components are built, interconverted, destroyed (e.g. **how** energy storage is turned into currency), → →
- **how** things pass across the outer bilayer boundary
- ….**and on and on**...



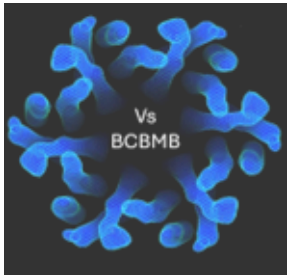
in other words: we only found molecular solutions to a tiny fraction of tasks that need to be accomplished within a biological unit.

what is missing is the "glue" that holds everything together and integrates the different macromolecular domains?

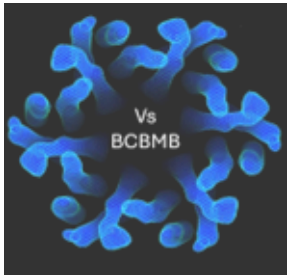


STRUCTURE DETERMINES…..?.....

...fill in the missing word....

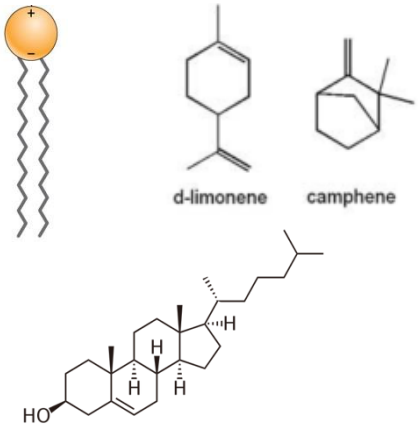


STRUCTURE DETERMINES *FUNCTION*



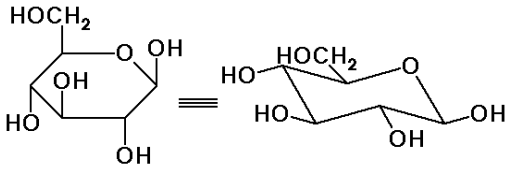
STRUCTURE DETERMINES FUNCTION

Lipids



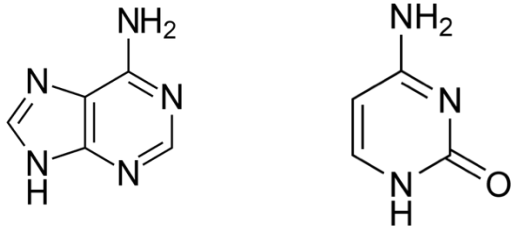
Boundaries;
Energy

Carbohydrates



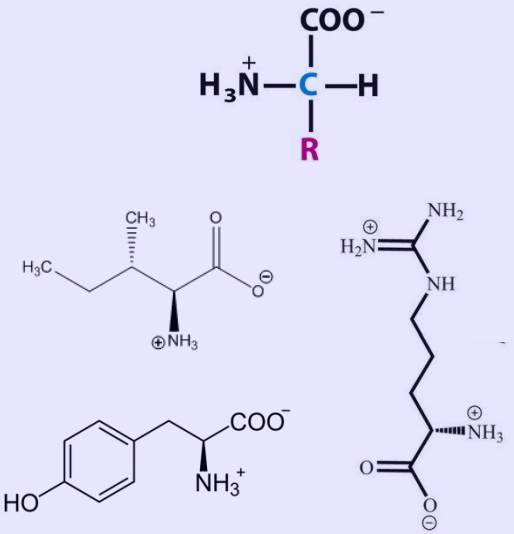
Boundaries;
Mechanical/Structural support;
Energy

Heterocyclic Aromatic "Nucleobases"



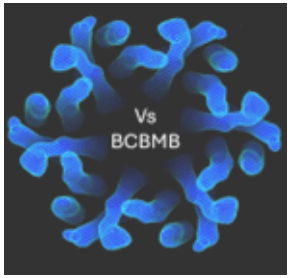
Information Storage +
Transmission

α–Amino Acids



?

Amino Acids are the only basic macromolecular inventory class left at this point …
→ are these what the **"glue"** is made from? If so, we should expect amino acids to be so versatile that they (or their oligo-/polymers) literally can do everything that still needs a mechanistic equivalent



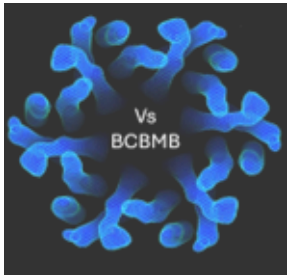
Goals Of This Lecture

the main goal of this closing lecture in the "Biochemistry Fundamentals" Collection is to help you understand why proteins (the polymers made from amino acids) are the “glue” that holds all of biology together

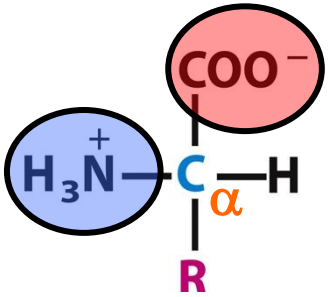
at the end of this handout, you should understand be able to explain:

1. what an amino acid is
2. what a peptide bond is and how it is formed in principle
3. the different levels of structural organization in polypeptides
4. key similarities and differences between polypeptides and other biological polymers
5. why polypeptides are functionally versatile, and
6. at a basic level, why polypeptides are uniquely suited for catalysis

even if you can explain all these points right now, you may want to still review the rest of this lecture (to refresh your memory) before working through the associated silent lecture in the "Advanced Biochemistry" Collection

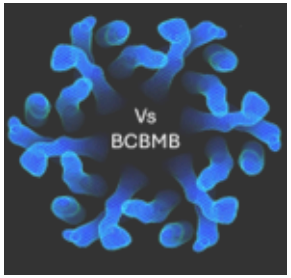


α–Amino Acids – Stepping Stone to Molecular Diversity



what do you see?

Answer: *…try to describe before looking at next slide…..*

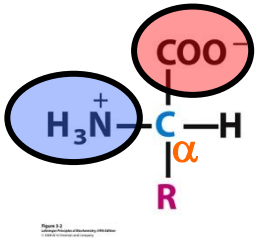


α–Amino Acids – Stepping Stone to Molecular Diversity

what do you see?

Answer: a generalized chemical template structure describing an organic molecule in which single bonds link a carbon atom to four groups, three of which are explicitly shown: carboxyl, amino, hydrogen. The fourth group, "R", denotes a variable entity that is not explicitly shown. […or something to that effect….]

α–Amino Acids

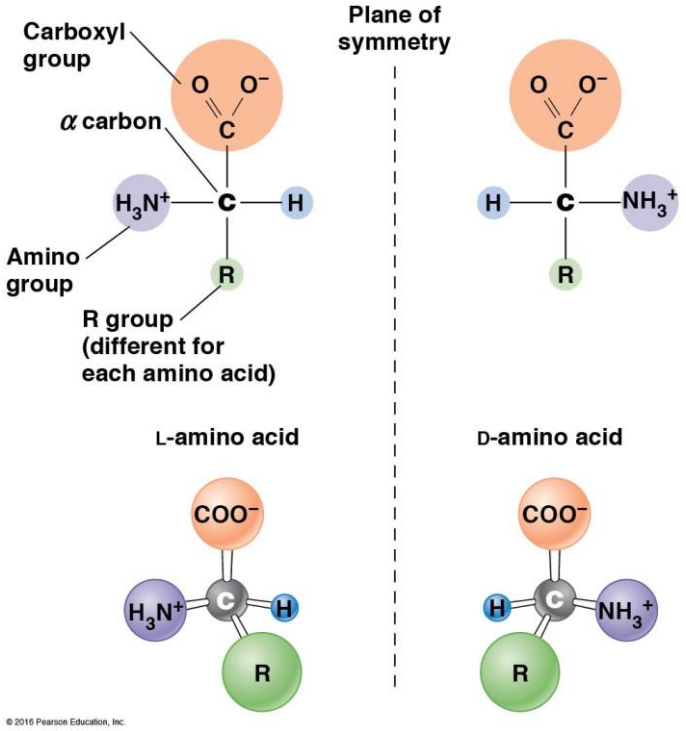


the lengthy description allows multiple conclusions:

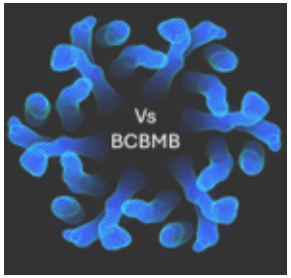
- explains name: α–**Amino Acids** [α denotes the C-atom adjacent to a "carbonyl function" (carboxyl, aldehyde, ketone)]
- multifunctional chemistry
- amino acids have a "constant" part (**scaffolding unit**) **and** a variable component (called **sidechain**)
- "R" likely has considerable structural/chemical variability that cannot be generalized to fit all

•for any R≠H, C_α is asymmetric (concept)

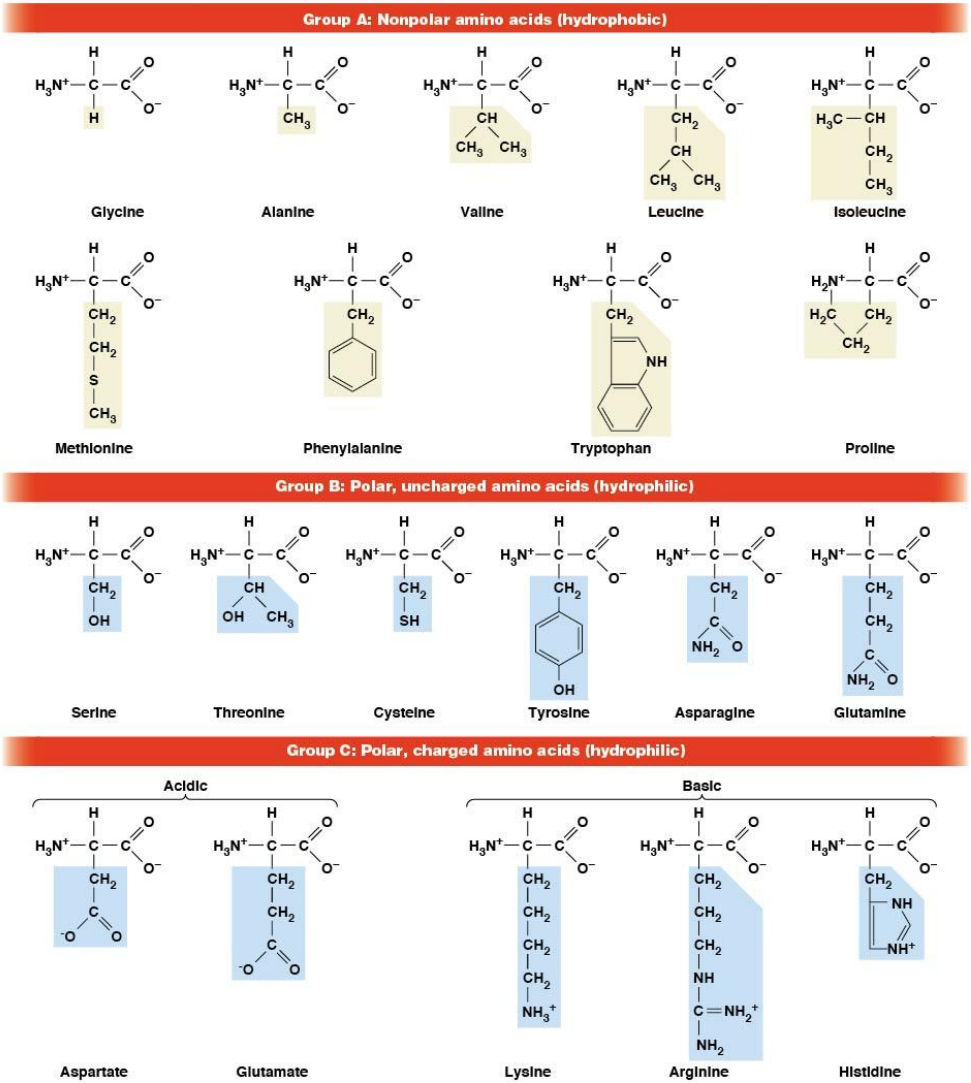
→ all but the simplest amino acid R=H exist as **stereoisomers** (remember carbohydrates…?)



L-Form prevailed in Biology/Life



α–Amino Acids – Stepping Stone to Molecular Diversity



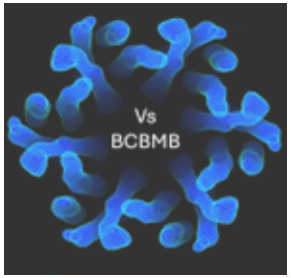
moving on to explicit structures what do you see?

Answer

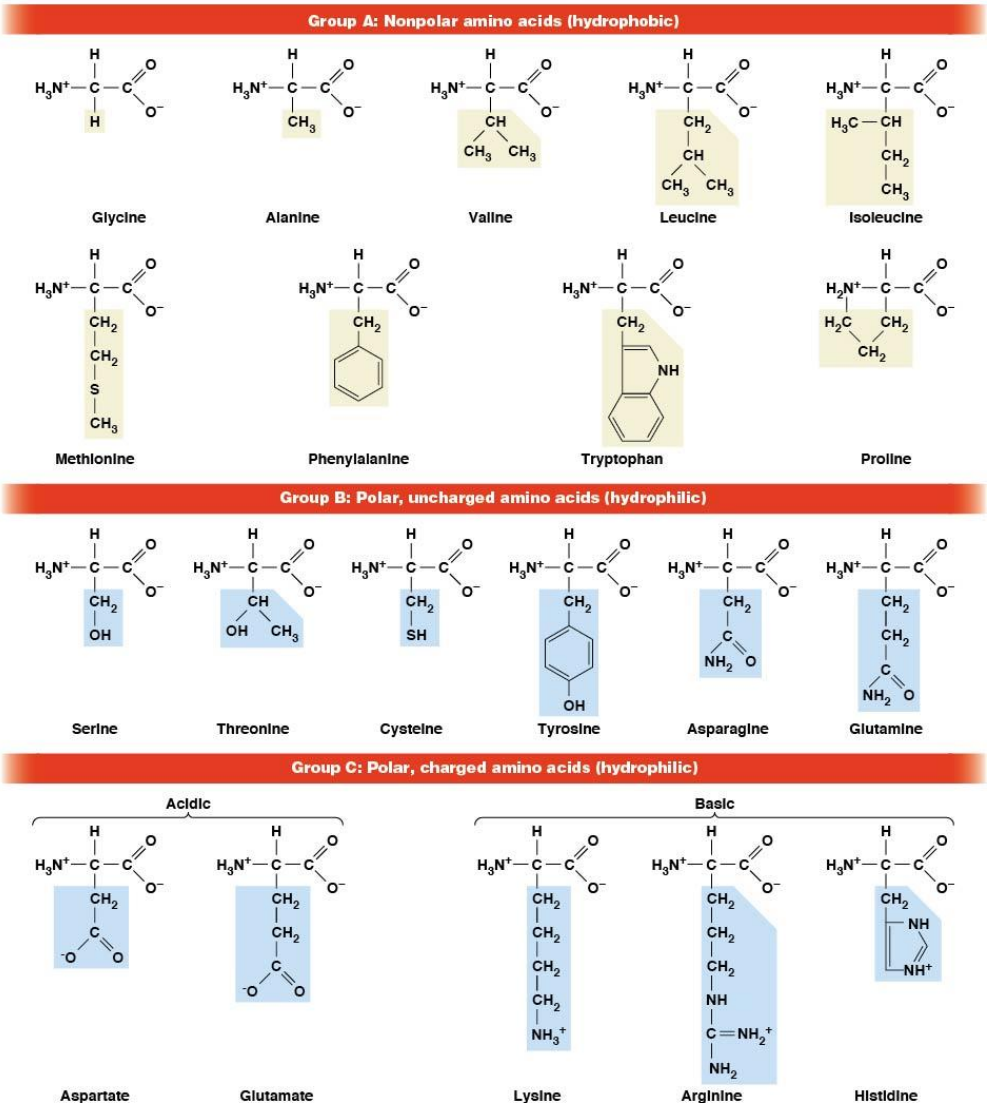
twenty different sidechains (“R”) that are chemically and structurally diverse, but seemingly fall into three categories: non-polar, polar/uncharged, polar/charged.

→ non-polar, polar/uncharged, polar/charged **should** trigger a specific memory….which one?

Answer: *…try to answer ….*



α–Amino Acids – Stepping Stone to Molecular Diversity



→ non-polar, polar/uncharged, polar/charged **should** trigger a specific memory….which one?

Answer:

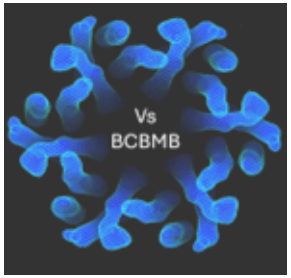
weak interactions (concept)

- non-polar: van der Waals
- polar/uncharged: H-bonds
- polar/charged: ionic bonds

→ sidechains form a group of related compounds that is not biased in its collective chemical properties

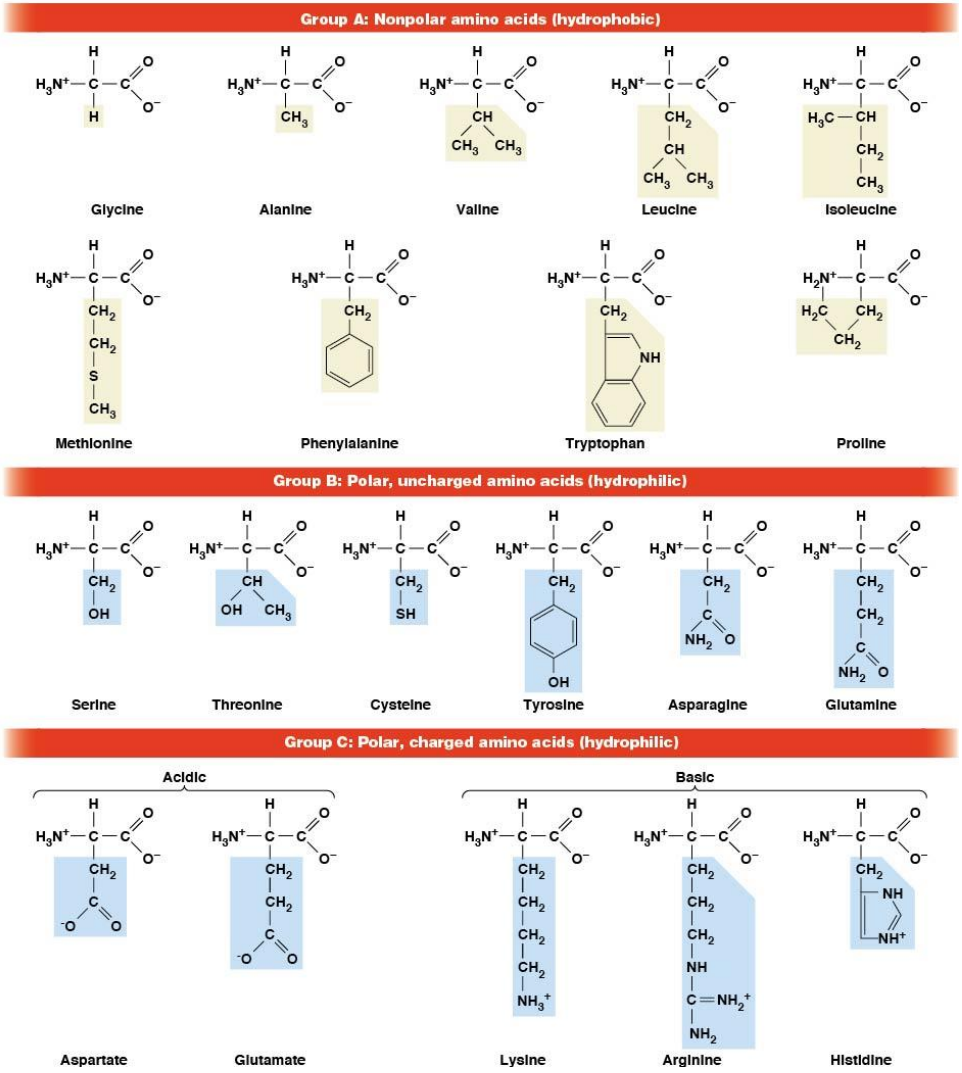
→ the balanced mix of chemical properties sets amino acids apart from lipids, carboyhdrates and nucleobases.

→ balance in chemistries suggests that a polymer from amino acids may (and does) behave quite differently from the other polymers we have encountered so far.

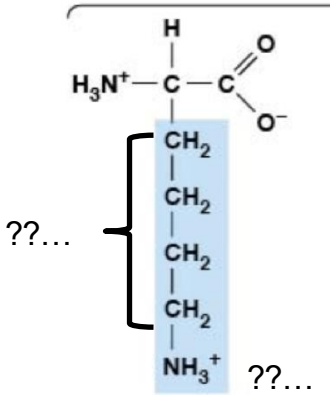


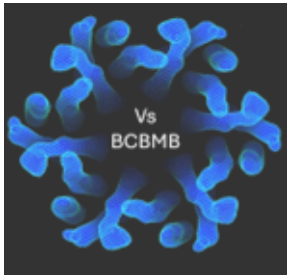
α–Amino Acids – Stepping Stone to Molecular Diversity

we are not done with this table …it's great in focussing your attention on the chemical balance and weak interaction potential of amino acids, **but** it also has (big) problems if you look more closely.....



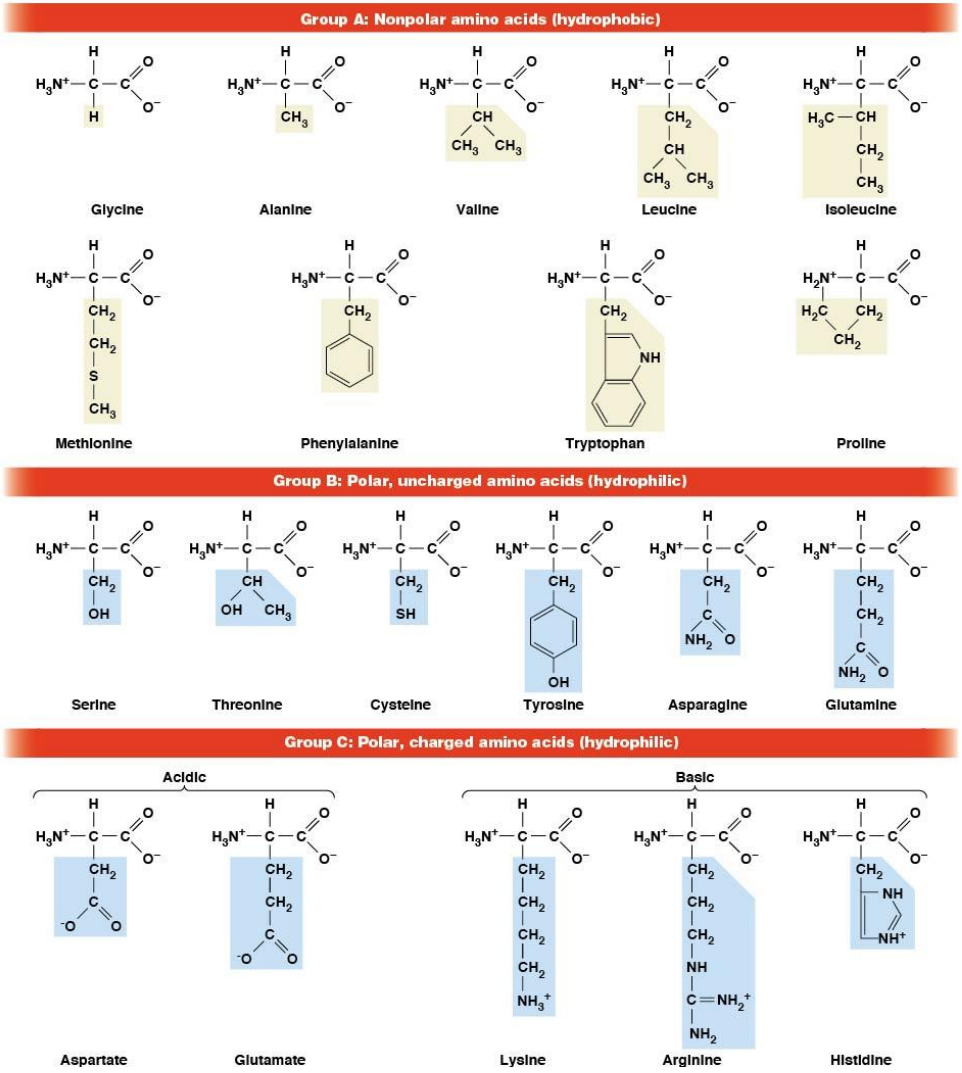
point in case: **what do you see?**



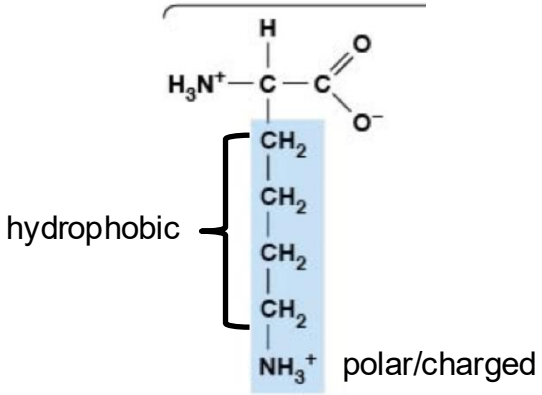


α–Amino Acids – Stepping Stone to Molecular Diversity

we are not done with this table …it's great in focussing your attention on the chemical balance and weak interaction potential of amino acids, **but** it also has (big) problems if you look more closely.....

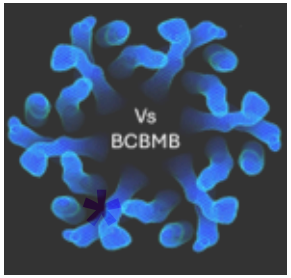


point in case: **what do you see?**

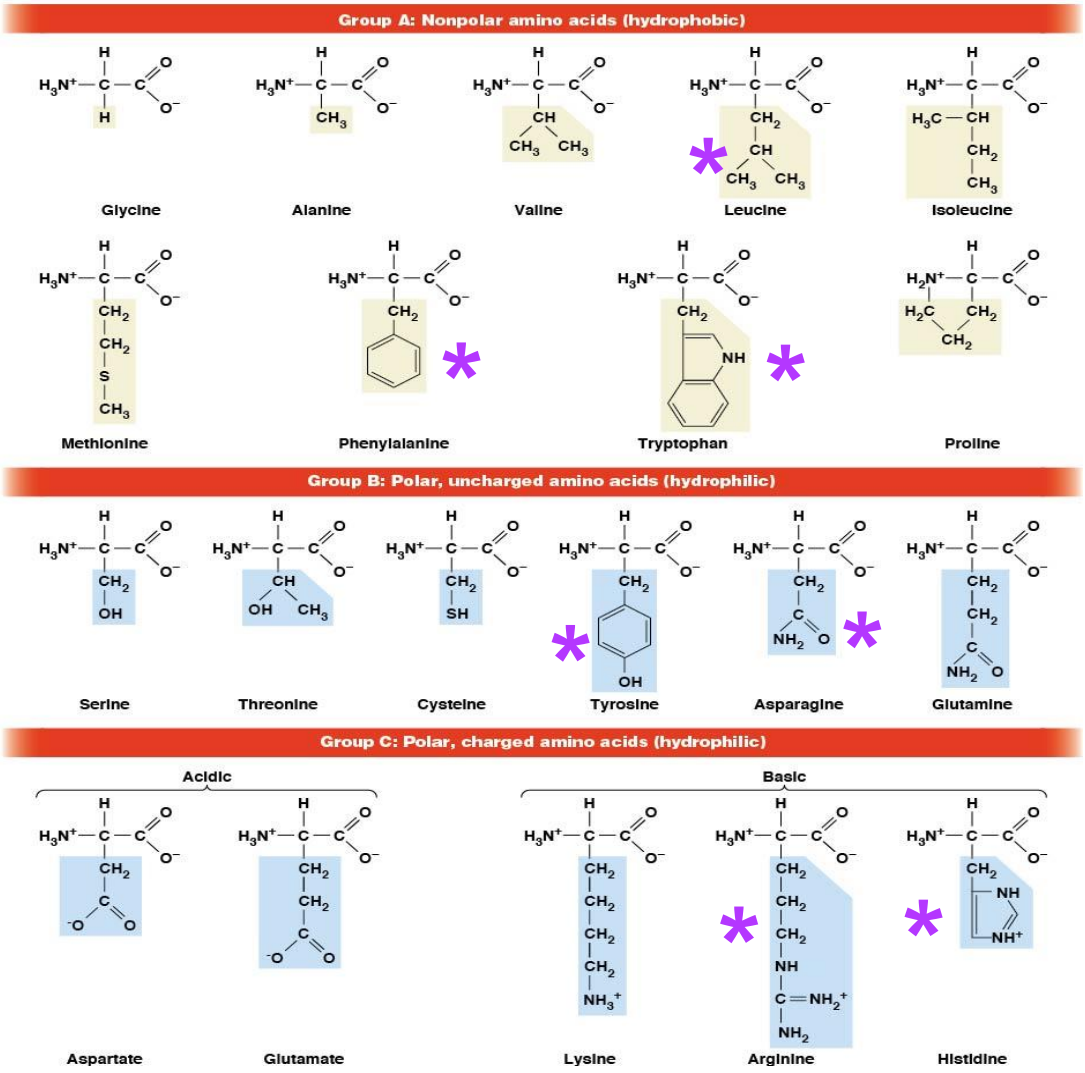


the Lysine sidechain combines opposing chemical characteristics…….
= if you look at just the sidechain, it is an amphiphile! (recall: phospholipids!)

→it's "OK" to list it with the charged amino acids **<u>BUT</u>** forgetting about the hydrophobic aspect is "fatal" because that aspect **<u>really</u>** matters in countless situations



α–Amino Acids – Stepping Stone to Molecular Diversity

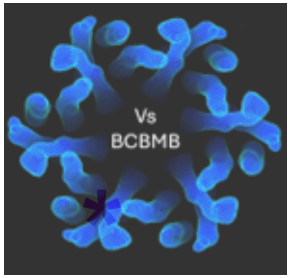


take a look at the marked amino acids and think about …..

- the sidechain itself
- the expected solubility in water
- the placing in the table

it's OK to use your phone/laptop to look up information you think would help.

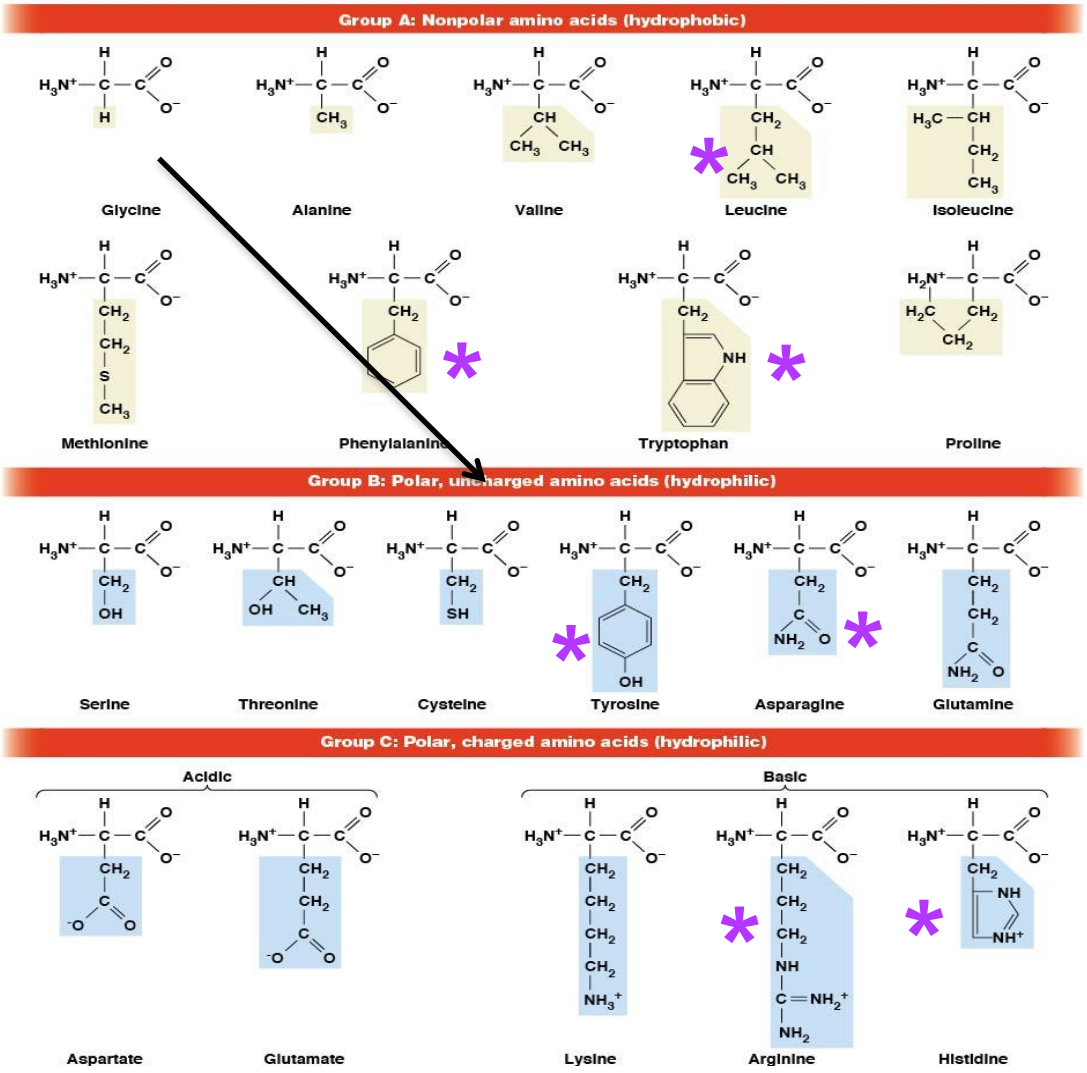
…engaging with this task will help you understand amino acids and, in turn, proteins….



α–Amino Acids – Stepping Stone to Molecular Diversity

take a look at the marked amino acids and think about …..

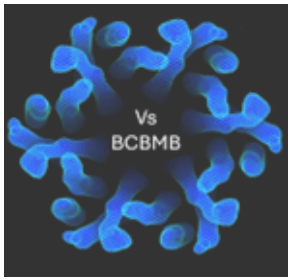
- the sidechain itself
- the expected solubility in water
- the placing in the table



here are the solubilities of the marked amino acids in water:

Glycine	**250.0 g/l**
Leucine	24.3 g/l
Phenylalanine	29.2 g/l
Tryptophane	1.4 g/l
Tyrosine	**0.4 g/l**
Asn (asparagine)	29.4 g/l
Arginine	148. g/l
Histidine	41.9 g/l

→ water solubility generally seems to be reflective of sidechain chemistry, but there are remarkable exceptions (e.g. Tyr)
and
one outright mistake:
Gly is **not** hydrophobic



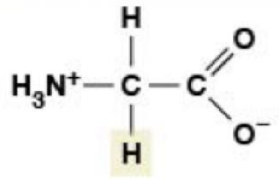
α–Amino Acids
Are Zwitterionic in Water (at pH7)

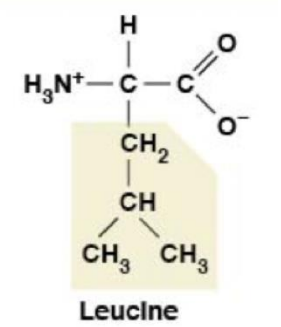
…contemplating water solubility

Gly 250 g/l

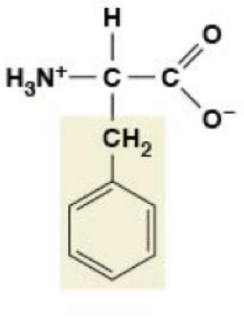
Leu 24.3 g/l

Phe 29.2 g/l





Leucine

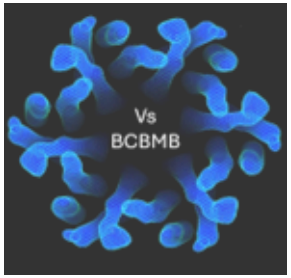


Phenylalanine

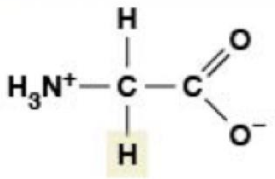
…Gly is extremely soluble and even aliphatic/aromatic Leu and Phe have significant water solubility.
(on their own, these sidechains are lipids!)

Why?

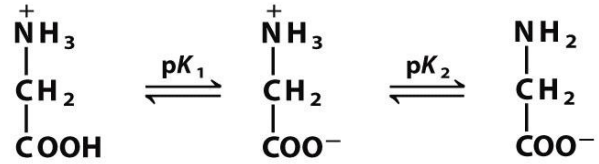
Answer: …*your guess?*



α–Amino Acids
Are Zwitterionic in Water (at pH7)

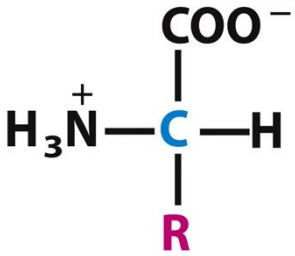


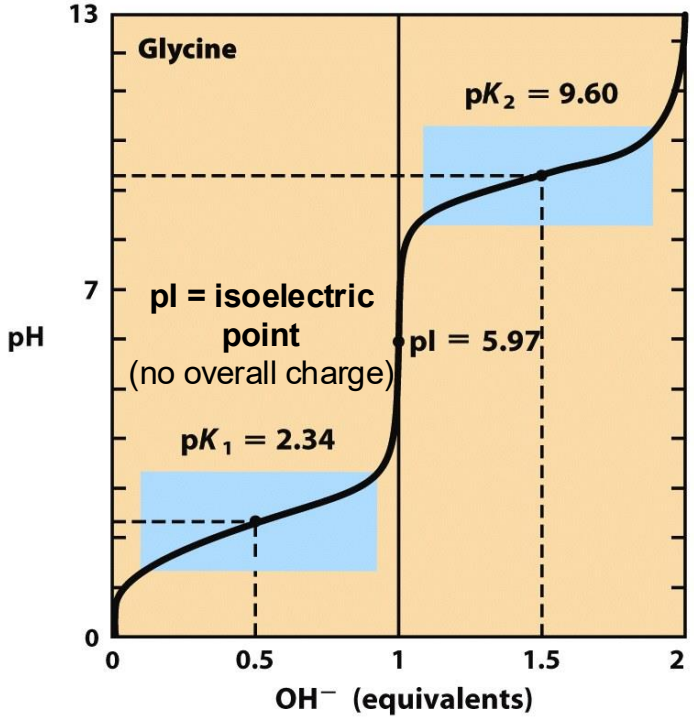
overall charge: **+1** **0** **-1**



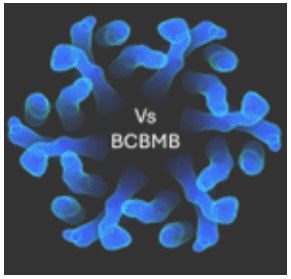
…Gly is extremely soluble and even aliphatic/aromatic Leu and Phe have significant water solubility.
why?

Answer
because all **monomeric** amino acids have titrable groups (*hello general chemistry (lab)*) and are **zwitterionic** (both positive and negative charge) **at neutral pH**



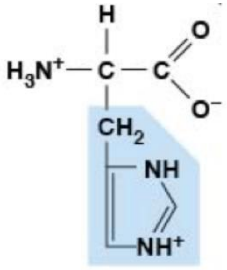


side note: "p" is short for "-lg" → pH = -lg [H^+], ….

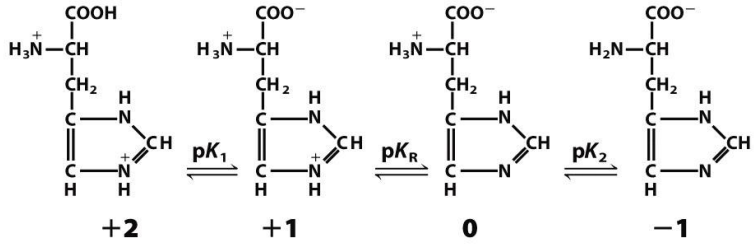


α–Amino Acids
Are Zwitterionic in Water (at pH7)

...**another pesky, but important detail** to look at....histidine....it's listed with the positively charged amino acids....why?



Answer
because the pK_a of the imidazole N-atom is within the physiological range.

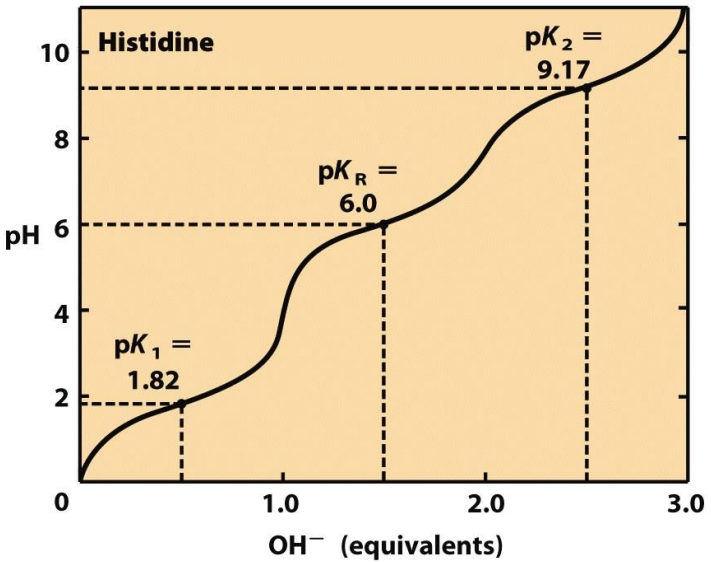


estimate of the pI: arithmetic mean between the two pK_a values that bracket the "zero charge" state.

$$pI = \frac{pK_R + pK_2}{2}$$

pK_R: acid constant (pK_a) of sidechain,
pK_2: pK_a of the $-NH_3$ group.

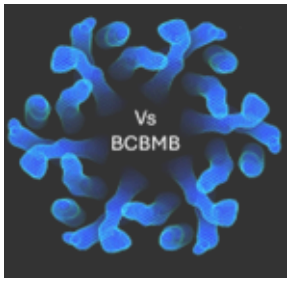
→~pI_{His}= (6+9.2)/2=7.6 [pH cytosol ~7.4]



→the charge/protonation state of histidine and any other sidechain with titrable groups depends on the environment!

this may seem "pointless and tedious" to take note of …. but the "environmental dependency of pKa" is EXTREMELY important in the case protein catalysts (enzymes)

we will look at that specifically in the "Advanced Biochemistry" lecture about catalysis



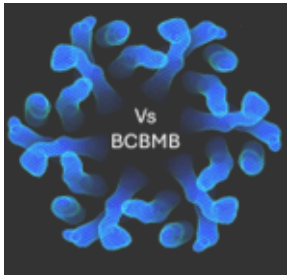
α–Amino Acids – From Monomer to Polymer → Peptide Bonds

in exploring molecular inventory, we focused on polymers because monomers are too small to register on the size scale of cellular structures. Here is the re-cap

- Lipids: **non covalent** emulsions and bilayers
- Carboyhdrates: **covalent** polymers that play various functions
- Nucleobases: require a covalent "(deoxy)ribose-phosphate" scaffold to form nucleic acids
  → **covalent, but composite polymer**

- **Amino Acids → do they form polymers? If so, will these polymers be covalent or non-covalent? Why?**

Answer: …try….



α–Amino Acids – From Monomer to Polymer → Peptide Bonds

- **Amino Acids → do they form polymers?**

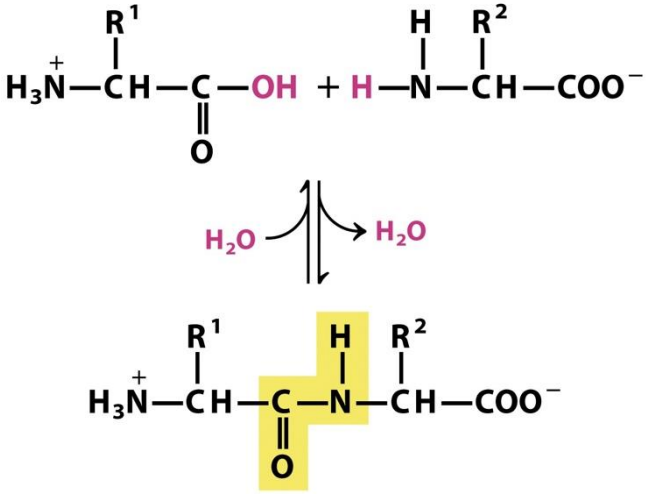
If so, will these polymers be covalent or non-covalent?

Why?

Answer: yes – covalent

zwitterionic nature and chemistries of most sidechains rule out that they form emulsions or bilayers

how can a covalent linkage be created between amino acids?



Answer: condensation that eliminates water from the –COOH and H_2N- groups (ignoring their zwitterionic state here). The reaction takes place between the common cores and **does not** involve sidechain functional groups (same as in nucleic acids!)

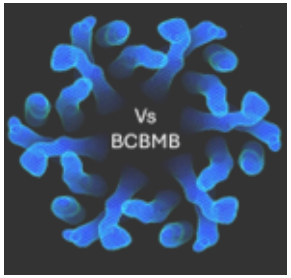
note: in solution, the **two protonation states shown in the reaction scheme cannot coexist at the same time. This is life saving because the natural, zwitterionic state of free amino acids is utterly unreactive** (= amino acids cannot ever polymerize spontaneously) → cells must use chemical tricks to deal with the big pK_a gap between the carboxyl and amino groups

(covered in "Molecular Biology and Cell Biology – TRANSLATION" Lecture)

the bond joining two amino acids is called **peptide bond**

and forms the basis of **covalent polymers** that are called **proteins (synonym: polypeptide)**

from the Greek word πρώτειος (*proteios*), meaning "primary".



Proteins vs ....

in contrast to nucleic acids,
formation of the repetitive (periodic) structure of the protein backbone **does not** require additional chemical components (sugar-phosphate) → this likens proteins to polysaccharides.

in contrast to polysaccharides,
the chemistry of the amino acid core unit **does not** allow for branching in the polypeptide main chain

in contrast to polysaccharides and nucleic acids
polypeptides are short (average: ~400 amino acids (also called "residues"), and very rarely exceed 1,000 residues; one notable exception: titin (a protein found in muscle) is ~27,000 residues long

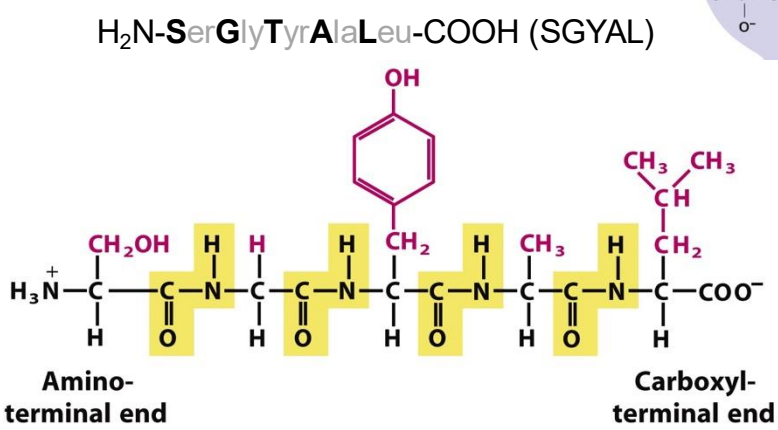
→compare cellulose: ~10,000 – 15,000 glucose molecules; average nucleic acid length varies depending on type (~100 nucleotides to 10s of millions)

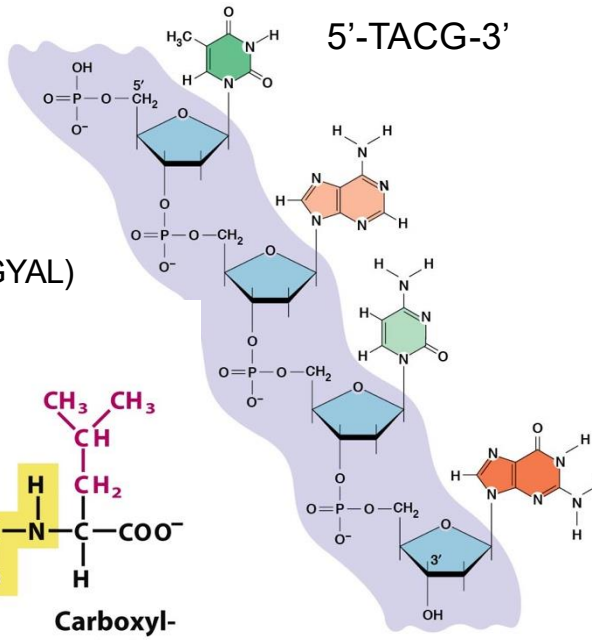
Like
nucleic acids and some polysaccharides**, (poly)peptide backbones are asymmetric!** (= ends are chemically different)

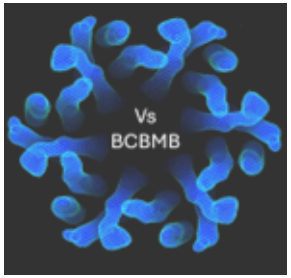
polypeptides: N terminus → C-terminus
nucleic Acids: 5' → 3'

notation: left to right (N→C; 5'→3')

the **explicit sequence** of covalent biological polymers is called
Primary Structure
eg: SGYAL
or: TACG



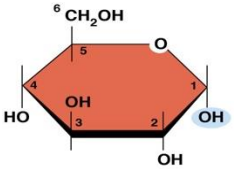


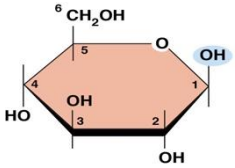


Proteins: Polymer Structure
.....Why 20 and Balance Matter....

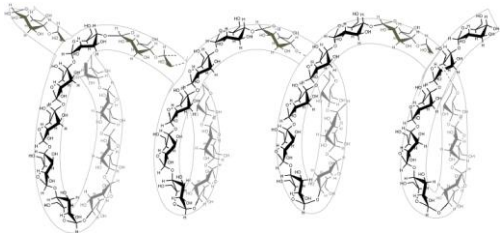
looking at **polysaccharides** – we found that the configuration at the **anomeric carbon atom is a key factor in determining the structural properties of the polymer**.....

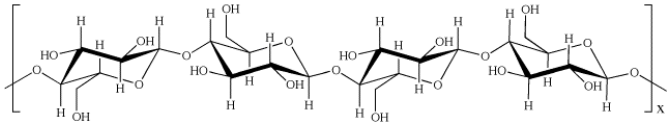
(Carbohydrate Lecture – Slides 30-32)





amylose





cellulose

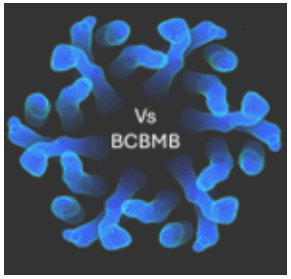
Gedankenexperiment 1

starting from a fully extended conformation,

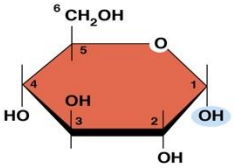
what will happen if the following nucleic acid is exposed to water?

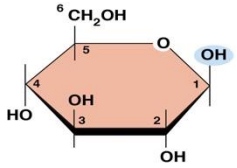
5’-GCGGACAAAGTTCCCGAGCAGGGAAACCGGCTGCTCCTTTATTAATTTGCGGGCCGAAATAATATAT-3’

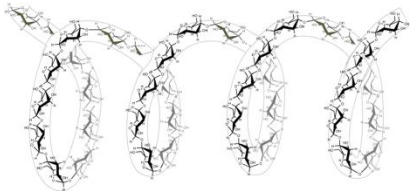
….think about it and try to verbalize your thoughts

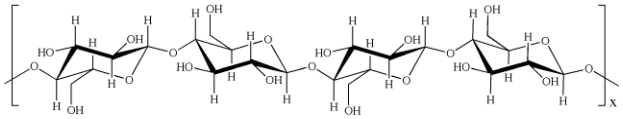


Proteins: Polymer Structure
...Why 20 and Balance Matter....









Gedankenexperiment 1:
starting from a fully extended conformation,
what will happen if the following nucleic acid is exposed to water?

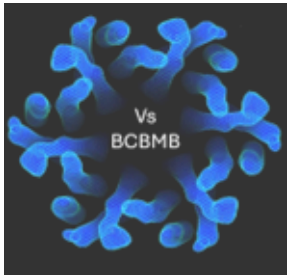
5’-GCGGACAAAGTTCCCGAGCAGGGAAACCGGCTGCTCCTTTATTAATTTGCGGGCCGAAATAATATAT-3’

Answer

the flexibility of the polymer will allow random encounters between different regions
(only takes 3 nucleotides to form a U-turn, redirecting the polymer to fold back on itself)
→any G hitting C, C hitting G, A hitting T, and T hitting A will cause a loose and transient entanglement through either 2 or 3 H-bonds ….if more than one interaction can be formed locally, the paired structures will quickly become **very** stable, and the process will get irreversibly stuck.

Just a few examples….

➔ multiple random outcomes that are stabilized mostly (and quite strongly) through H-bonds
(flashback: this was the reason why DNA is double stranded ….NUCLEIC ACIDS, slide 39).



Proteins: Polymer Structure
....Why 20 and Balance Matter....
Continued

Gedankenexperiment 2:
starting from a fully extended conformation,
what will happen if the following polypeptide (written in one letter code) is exposed to water?

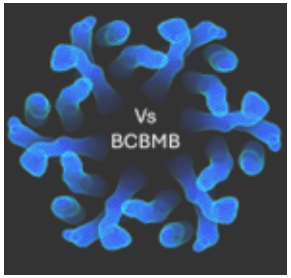
MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFILEKIKKTGKEVRSGKQL

…..this may be tricky for you to start thinking about if you do not know the one letter code for amino acids ....

….but **even without knowing what these letters stand for (exactly), there still is something you can predict …**

what it that?

…. *try ….*



Proteins: Polymer Structure ....Why 20 and Balance Matter.... Continued

Gedankenexperiment 2:
starting from a fully extended conformation,
what will happen if the following polypeptide (written in one letter code) is exposed to water?

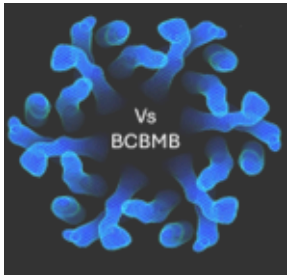
MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFILEKIKKTGKEVRSGKQL

…..this may be tricky for you to start thinking about if you do not know the one letter code for amino acids ....

….but **even without knowing what these letters stand for (exactly), there still is something you can predict …**

what it that?

Answer
just like for the nucleic acid, flexibility of the polymer will allow random encounters between different regions
but
since there is so many different sidechains whose chemistries are not all compatible with each other (eg hydrophobic vs charged)
the outcome will be very different from what happens with nucleic acids
let's take a closer look at why that is the right answer….



Proteins: Polymer Structure ....Why 20 and Balance Matter.... Continued

Answer

flexibility of the polymer will allow random encounters between different regions

but

the outcome will be very different from what happens with nucleic acids

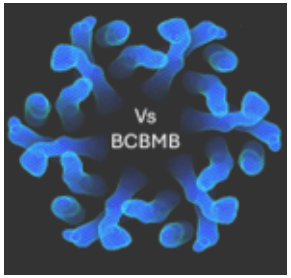
let's take a closer look at why that is the right answer….

to understand why the outcome will be very different from what happens for nucleic acids, we DO need to consider the "chemistries" of the amino acid sidechains…

→ to make this the least painful for you …I colored the letters based on the general chemical property of the sidechain

MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYD
FILEKIKKTGKEVRSGKQL

Yellow: hydrophobic; green: polar/uncharged; blue: polar/positive charge; red: polar/negative charge



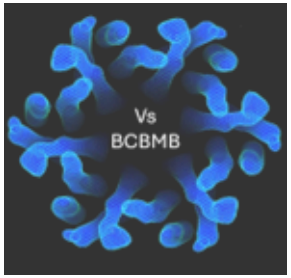
Proteins: Polymer Structure ....Why 20 and Balance Matter.... Continued

MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFIL
EKIKKTGKEVRSGKQL

Yellow: hydrophobic; green: polar/uncharged; blue: polar/positive charge; red: polar/negative charge

looking at this color pattern … **does this make ANY intuitive sense to you ….?**

Be honest……!!



Proteins: Polymer Structure ....Why 20 and Balance Matter.... Continued

MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFIL
EKIKKTGKEVRSGKQL

Yellow: hydrophobic; green: polar/uncharged; blue: polar/positive charge; red: polar/negative charge

looking at this color pattern … **does this make ANY intuitive sense to you ….?**

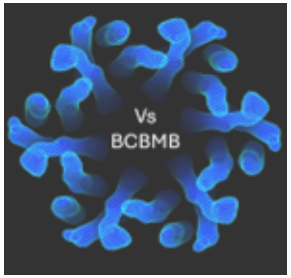
Be honest……!!

Answer

it would be surprising if it did make sense to you (unless you are already knowledgeable about protein structure). Assuming this were your first encounter, then likely this will look totally random … because you have

- all the chemistries intermingled,
- sometimes placing something hydrophobic right next to something charged…
- sometimes charges are bunched together,
- sometimes hydrophobics are grouped … but at any point in the chain it would be impossible to predict what the next amino acid is going to be, and what chemical properties it will add to the chain.

…. in sum: this is much more complicated than what we saw in nucleic acids where, in DNA, you only have ATGC & A::T, G:::C interactions ….. but if we "tidy up" these jumbled thoughts …you can summarize it like this:



Proteins ….Why 20 Matters....Continued

MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDI
SLEKQLVDVYTTLPYDFILEKIKKTGKEVRSGKQL

Yellow: hydrophobic; green: polar/uncharged; blue: polar/positive charge; red: polar/negative charge

Answer:

(1) no narrowly and preset defined matches akin to base pairs, and
(2) chemical properties vary in seemingly "random" fashion along the polymer….

(1) + (2) → when exposed to water, the **polypeptide** will utilize all types of weak interaction to **search for a more favorable conformation**. ..... in order to **minimize the Free Energy**...!

we already encountered the Gibbs Free Energy when we were looking at how molecules recognize/bind to each other ["How Do Molecules See? Pt2]
here we realize that we can think about intramolecular interactions (=interactions within a polymer))...the same way.....

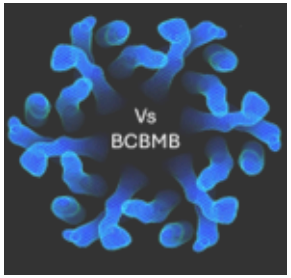
$\Delta G=\Delta H-T\Delta S$

what does "more favorable conformation" mean in case of a fully extended polypeptide?

Answer

non-polar sidechains do not want to be exposed to water
→ **hydrophobic effect** initiates conformational search
→ to what end? = **what is the chain searching for/trying to accomplish**?

…try to answer ….



Proteins ….Why 20 Matters....Continued

MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDI
SLEKQLVDVYTTLPYDFILEKIKKTGKEVRSGKQL

Yellow: hydrophobic; green: polar/uncharged; blue: polar/positive charge; red: polar/negative charge

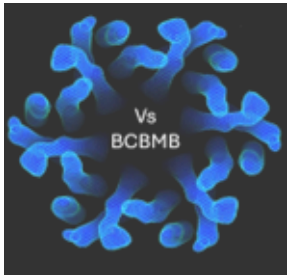
non-polar sidechains do not want to be exposed to water
→ **hydrophobic effect** initiates conformational search
→ to what end? = **what is the chain searching for/trying to accomplish?**

Answer
to "hide" the hydrophobic chains from solvent

formation of backbone abolishes the zwitterionic state of the amino acid core unit
➔ hydrophobic sidechains are now <u>much more</u> like lipids

…..now it may "click" if you worked through the LIPID lecture
→ In the case of **lipids, the hydrophobic effect caused aggregation** of molecules to form emulsions or bilayers
➔ reasonable to assume that the same phenomenon will drive the polypeptide chain to change its shape/conformation to bring together hydrophobic sidechains!

➔ and yes, that is what happens



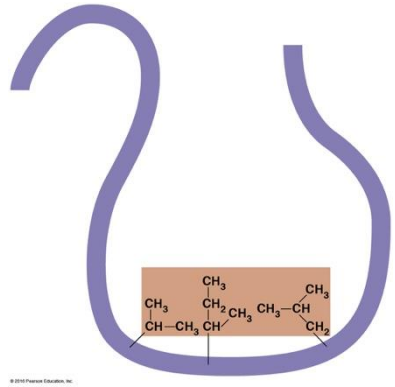
Proteins ….Why 20 Matters....Continued

$\Delta G = \Delta H - T\Delta S$

"aggregating" the hydrophobic sidechains constrains movement of the backbone

= loss of entropy (unfavorable)

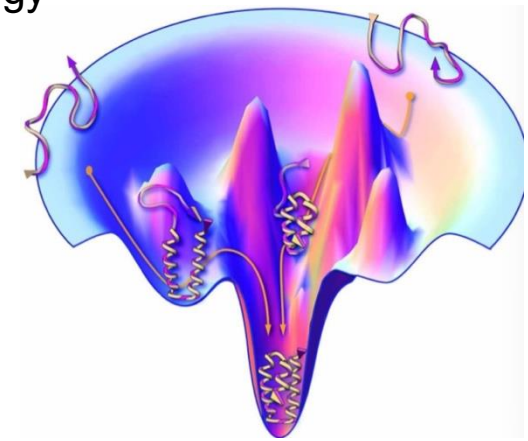
→ constraining the backbone needs "work" or some energy source to off-set the cost



➔ offset by gain of entropy in water molecules that are "tied up/confined" in clathrate cages around the hydrophobic sidechains but get liberated when the hydrophobic sidechains start interacting with each other

(recall "LIPIDS" lecture, slide 17)

Energy

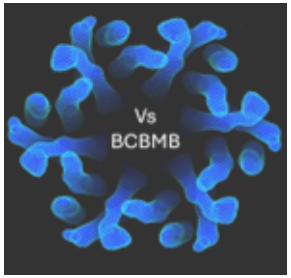


native structure

→ polypeptide "collapses" into a "molten globule" = a loosely structured state that is still very dynamic and keeps minimizing energy by trying to optimize weak interaction patterns until it finds a global minimum

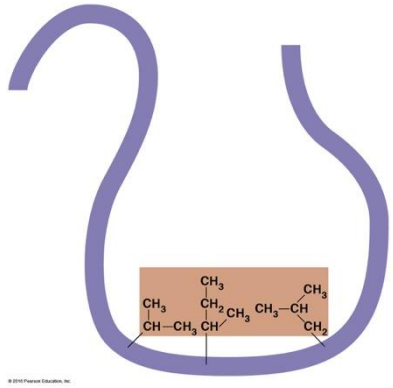
→ the **disorder to order transition** of the polypeptide is globally called: **protein folding**

$\Delta G_{folding} < 0$



Proteins ….Why 20 Matters....Continued

$\Delta G = \Delta H - T\Delta S$



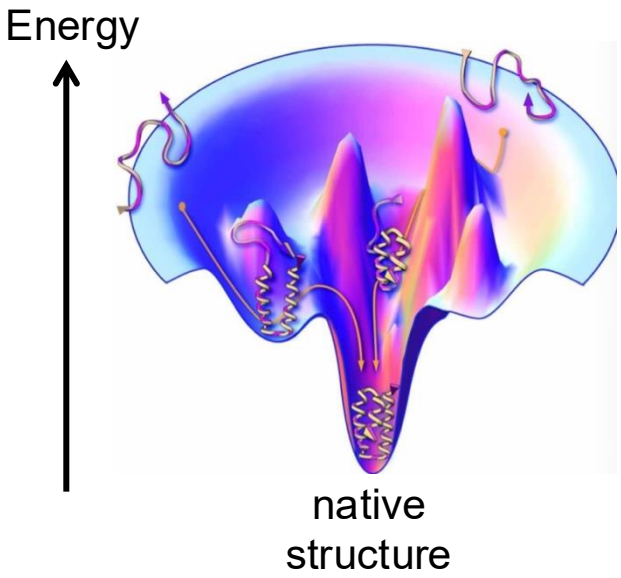
→ the **disorder to order transition** of the polypeptide is globally called: **protein folding**

$\Delta G_{folding} < 0$

→ ~90% of all polypeptides reach a global free energy minimum. The corresponding structure is called the "**native structure**".

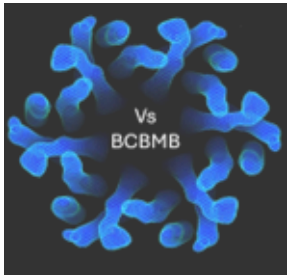
if a native structure exists, then the primary structure of the polypeptide carries all the information needed to attain it

= if you take 100 copies of the same polypeptide, extend it fully and let it fold, all chains will arrive at the same native structure if one exists ➔ this is fundamentally different from single stranded DNA … where you would end up with lots of different 3D structures, all very stable



→ in agreement with the paradigm that "structure determines function"

if a native structure exists, then only the native structure is functional.



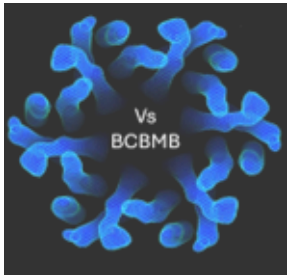
Up Close and Personal ….Meet a Protein

click on the embedded link to access a YouTube video tour of the protein whose sequence we used in the previous slides.

this video has sound
apologies for imperfections in the narration

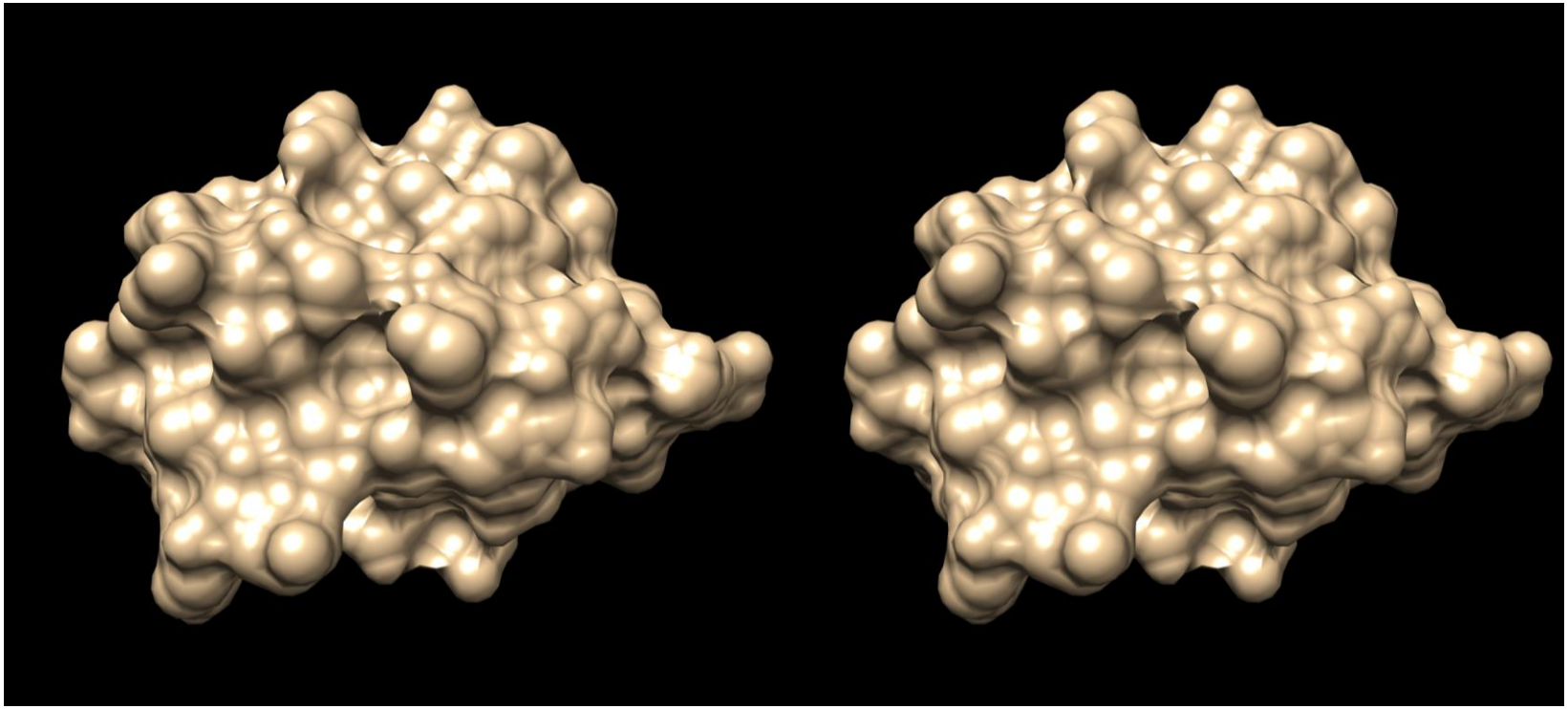
Virtual Tour of The Copper Chaperone ATX1

the following few slides summarize some of the key points from this video

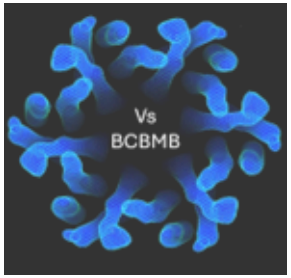


Still Recap Stereoviews of ATX1

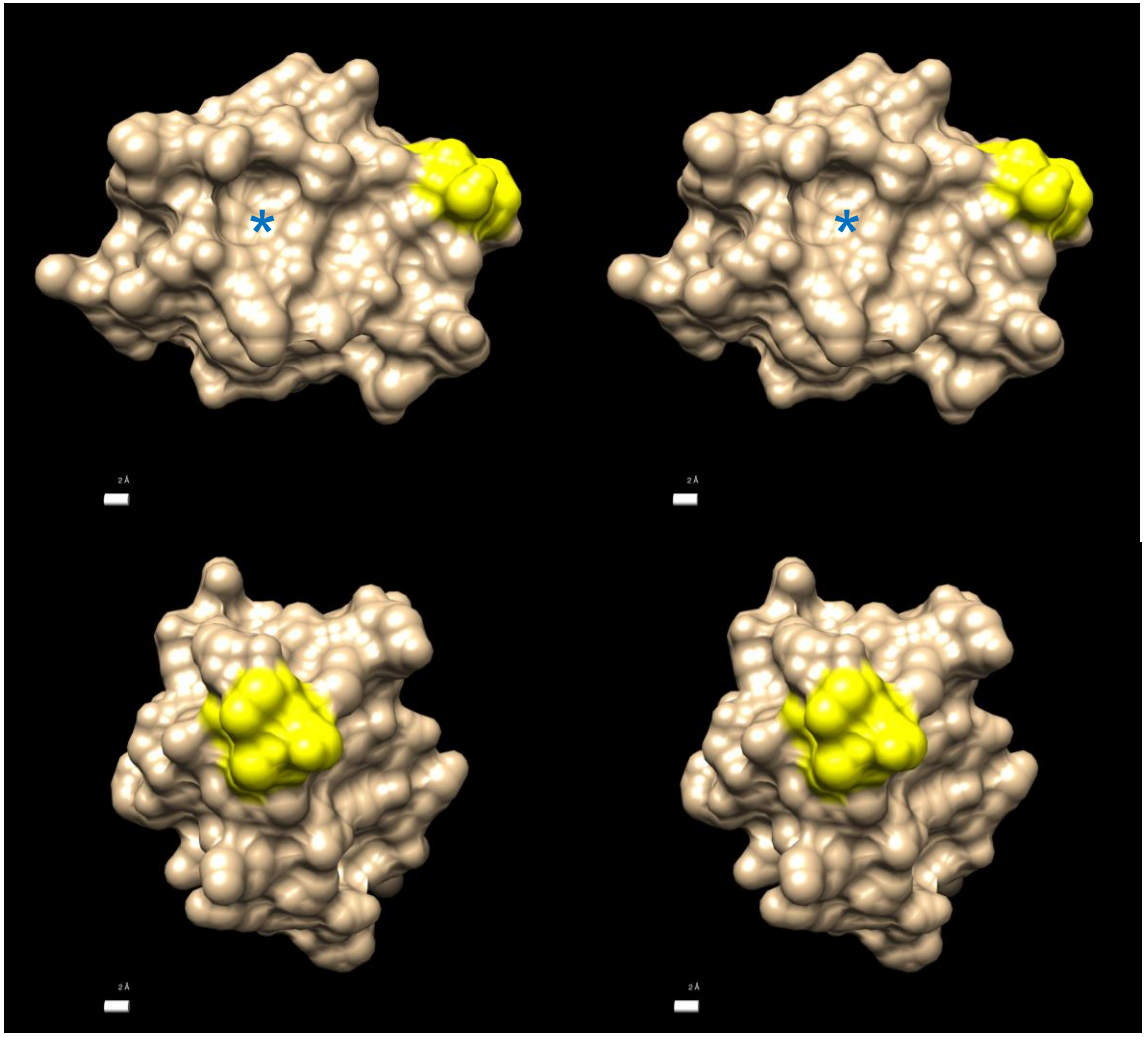
the images are cross-eyed stereoviews.
To achieve stereo vision:
hold the screen ~30cm away.
crossing your eyes and tilt your head a little left and right,
→ you should quickly see the depth perception come up.
If it takes a short while for the first one…don't give up, once you know how to do it, it is very easy and really helpful.



the Van der Waals surface of the protein appears irregular and rugged with lots of bumps and crevices



Still Recap Stereoviews of ATX1



two additional views.

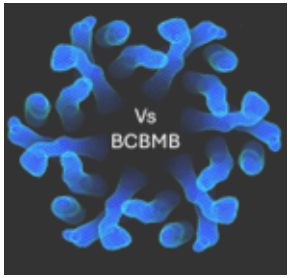
the 2Å scale bar indicates the approx. size of a copper ion

in the video you were "challenged" to identify the "binding site" for the copper ion.

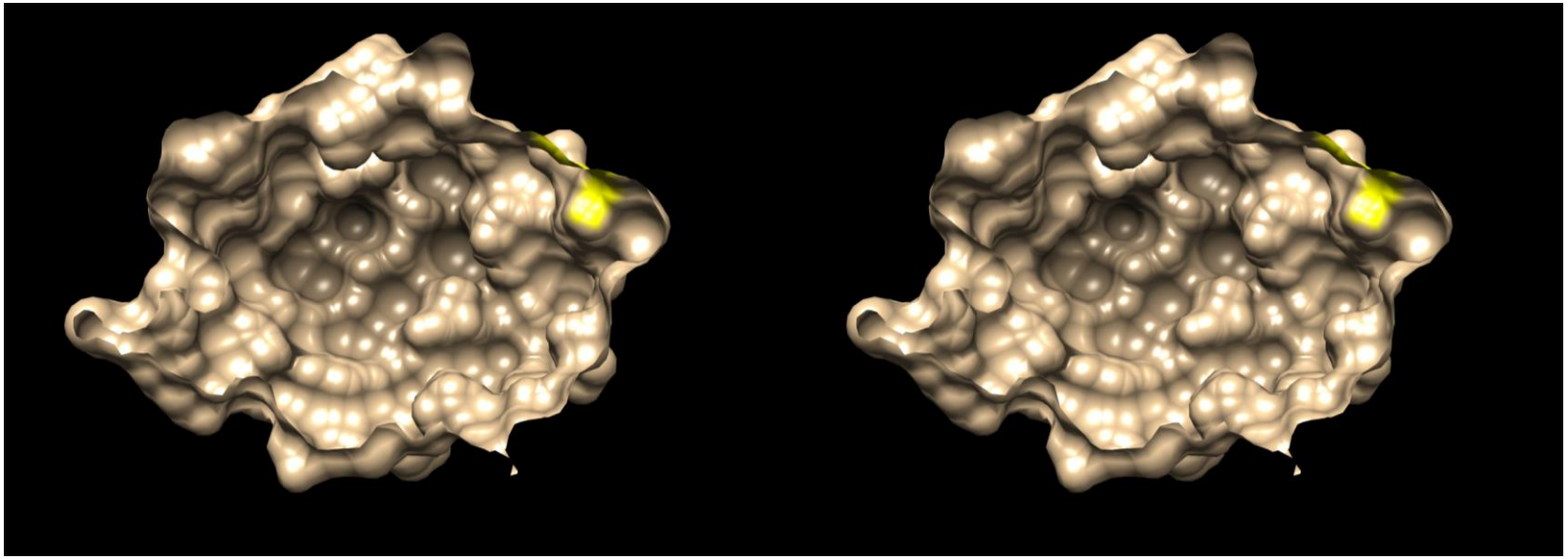
most intuitively:
one of the deep cavities on the surface (like the region marked with a blue "*")

…. but intuition can be deceiving like in this case, where the actual binding site is at an apex of the protein surface (highlighted in yellow).

in the video, you learned that this positioning makes more sense once you account for molecular motions of the protein and copper ions



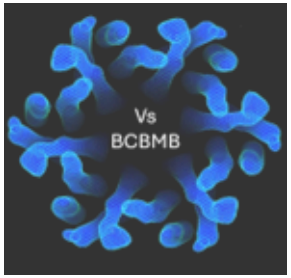
Still Recap Stereoviews of ATX1



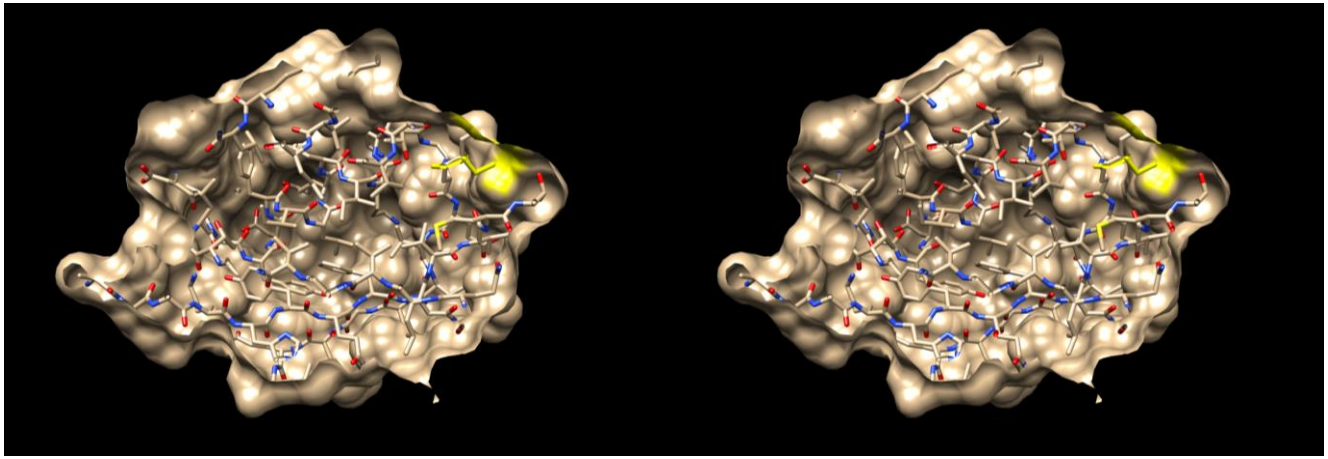
stepping through the surface and looking at the interior of the protein, it appears to be completely empty.

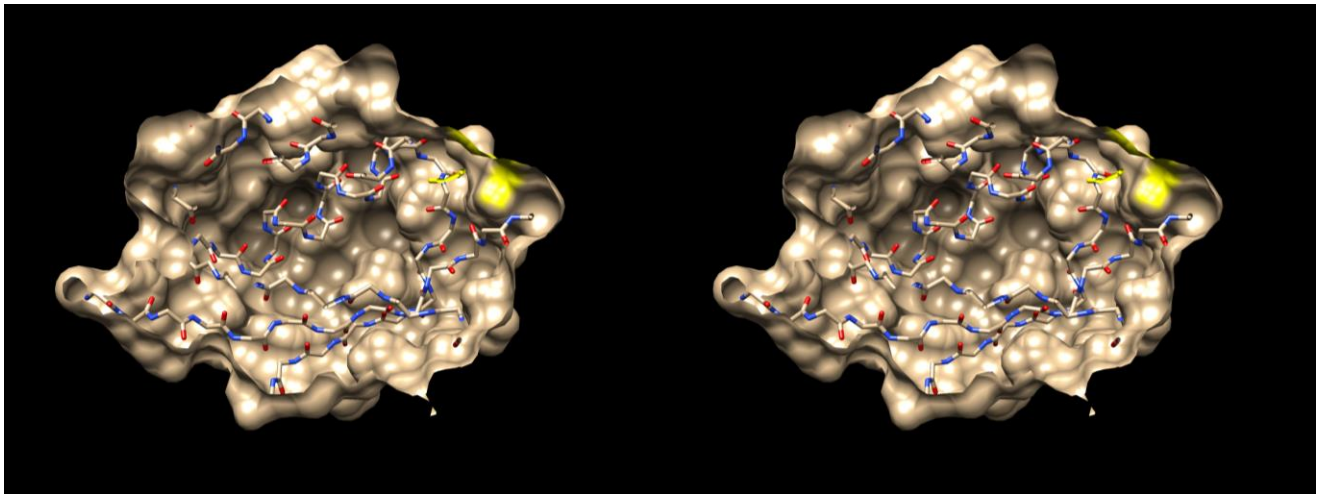
in the video, you learned that this "emptiness" means that atoms are so tightly packed against each other that their Van der Waals surfaces "merge" …giving you the sense of one continuous space that is bounded only by the Van der Waals surfaces of atoms that are located right at the boundary with the solvent/outside world.

the significance of this observation was that "globular", water soluble proteins like ATX1 do **not** have any water filled cavities in their interior.



Still Recap Stereoviews of ATX1





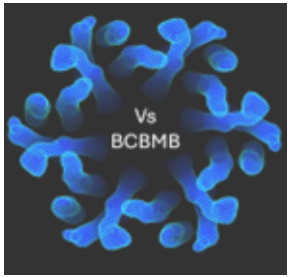
using a "skeleton" to visualize all the atom positions confirms that the interior of the protein is filled completely by amino acids.

while the skeleton explains how the bumps and crevices on the protein surface come about, it still appears that there is nothing to protein structure that seems regular or easy to appreciate …

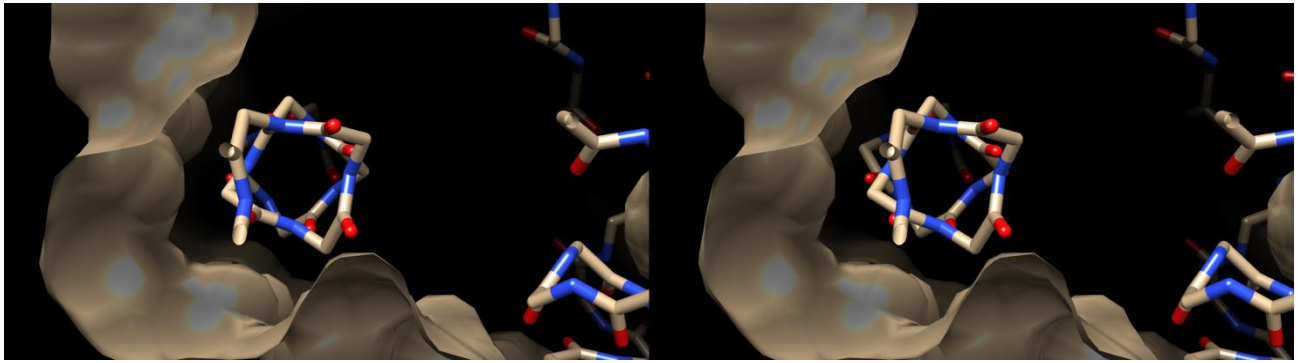
…until ….

we simplify further by only looking at the behavior of the protein backbone …

….looking at just that…it is less confusing and right away seems less random

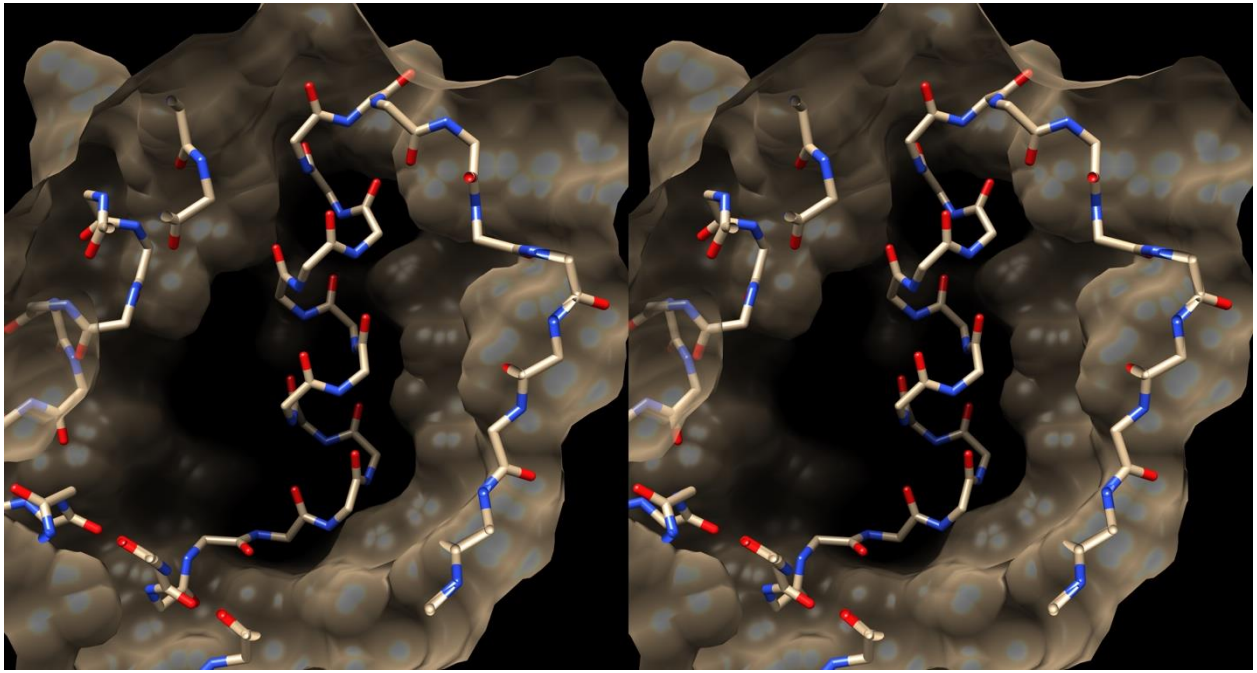


Still Recap Stereoviews of ATX1



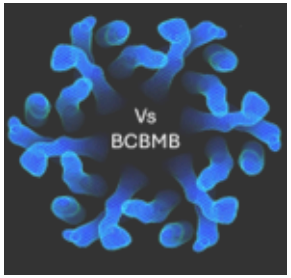
if properly positioned, some parts of the backbone appear to adopt a "spiral" = helical conformation.

seen along the helix axis (top). if they were displayed, then in this view, amino acid sidechains would point radially away from the center of the helix.

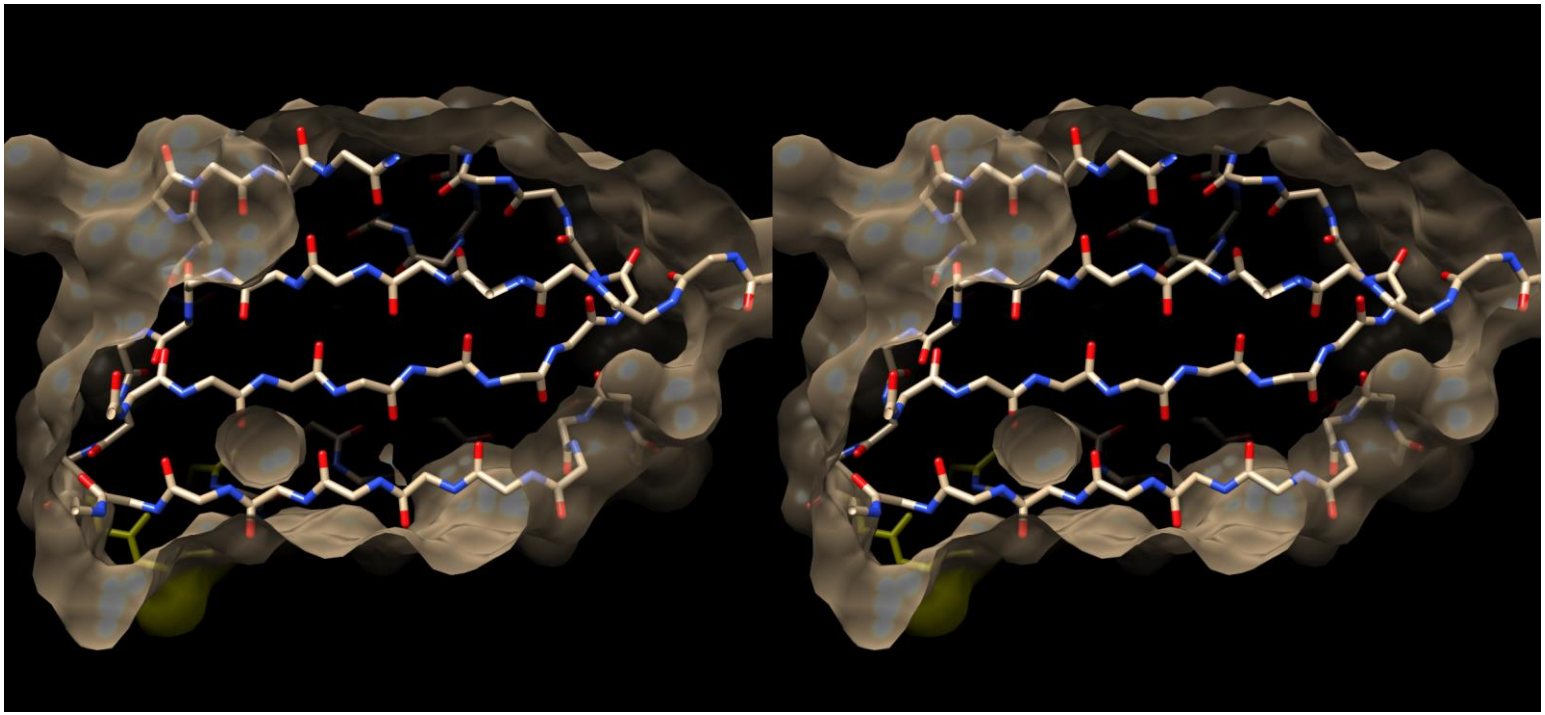


perpendicular to the helix axis (bottom)

note: along the right edge of the image to the left, you can also see a second very common type of protein backbone conformation. This is an "extended" conformation, lacking the "twist" of the helical segments.

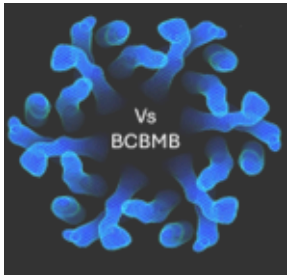


Still Recap Stereoviews of ATX1



a more complete view of this extended backbone conformation – called β–strand – is shown here. In protein structures, β–strands, associate laterally to form a structure called β–sheet. In this arrangement all the backbone atoms are within a plane, and sidechains are pointing up and down in a direction perpendicular to the plane defined by the backbone atoms (in the view shown here, the sidechains would alternate pointing at and away from you)

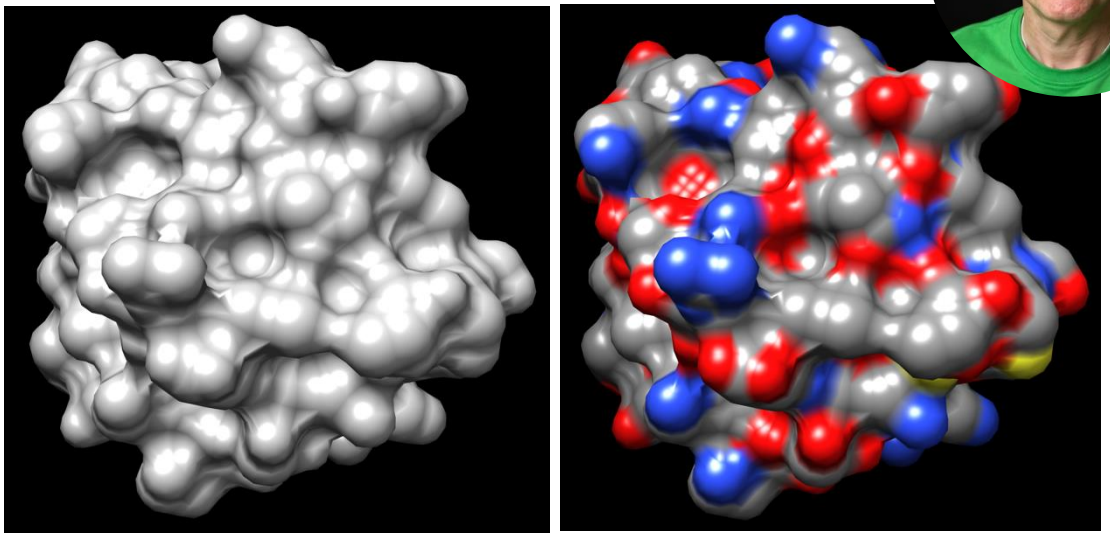
we will explore the details of these secondary structures in more detail in the "Advanced Biochemistry-PROTEINS" Lecture



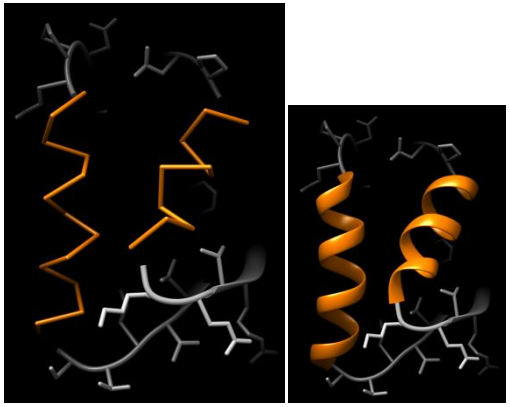
Folding of the Backbone Leads to Redundant Structural Elements Called Secondary Structure

the van der Waals surface that reflects backbone and sidechains looks irregular
left: just sterics; right adds chemical information.

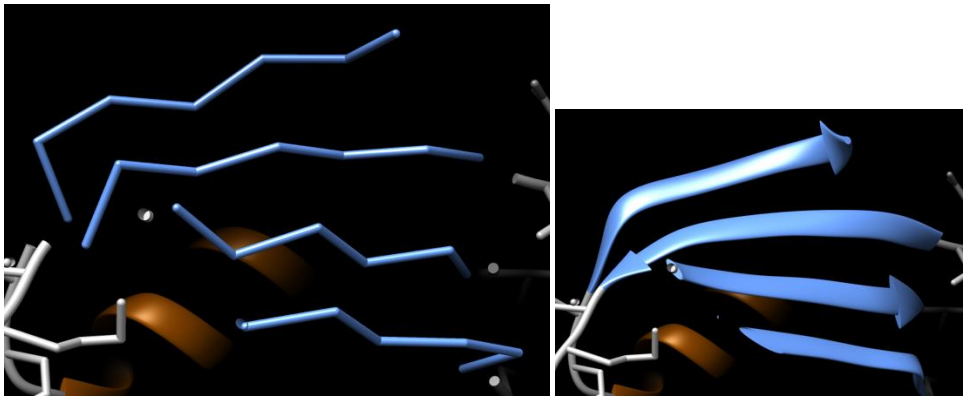
Color key (right panel)
Blue: nitrogen
Red: oxygen
Yellow: sulfur
Grey: carbon



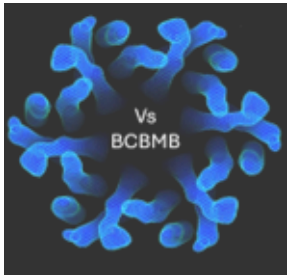
however**, the local conformation of just the backbone atoms – called secondary structure** - tends to be regular. The most common secondary structure elements in proteins are helical and extended conformations:



α–helix



β–strand/sheet

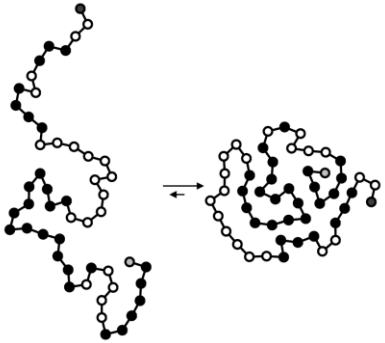


Interactions Between Sidechains Support the Core and Tertiary Structure of Water Soluble Proteins

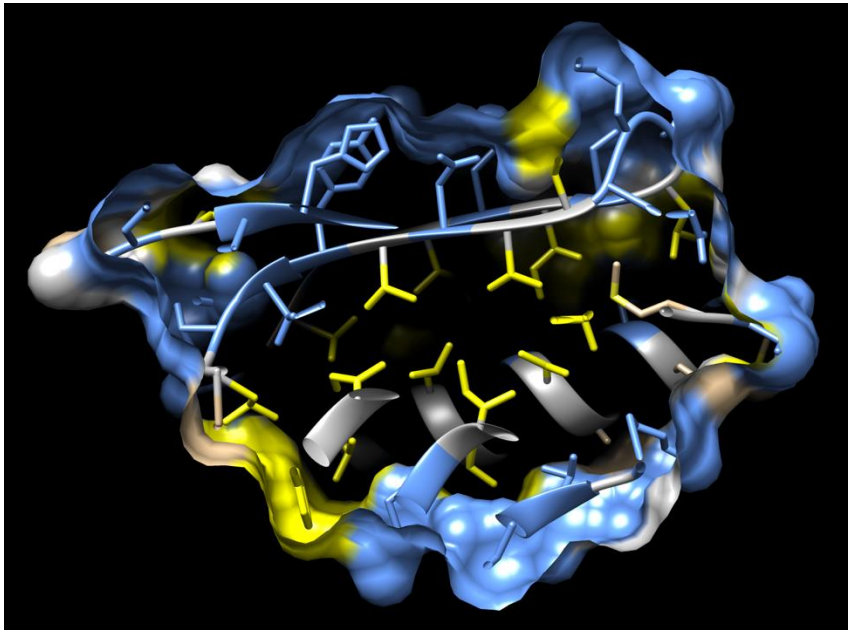
moving beyond secondary structure you want to pay attention to the type of protein – water soluble or membrane protein. The latter we deal with in the "Advanced Biochemistry - PROTEINS" lecture. Here, we want to focus on water soluble proteins, whose folding begins by the aggregation of hydrophobic sidechains (due to the hydrophobic effect).

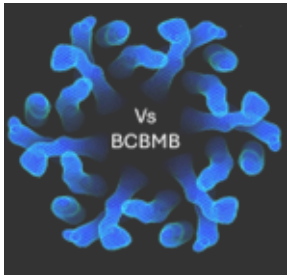
→at the end of folding **many (but not all) hydrophobic sidechains** have successfully been sequestered from the aqueous environment and are constrained to what is called **the "hydrophobic core"** (yellow color in figure below)

→similarly: polar/uncharged and polar/charged sidechains are facing the aqueous environment (blue color in figure below) and are rarely buried within the hydrophobic core.



the complex (long range) interactions between sidechains stabilize what is known as the proteins overall structure or **"tertiary structure"**



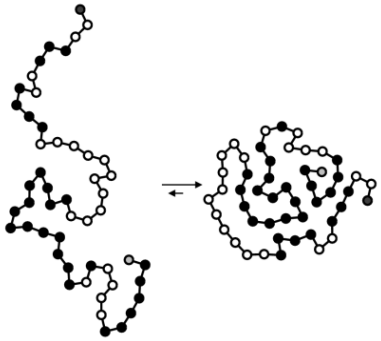


Tertiary Structure Explains the Apparent Randomness in the Sequence

if, for a moment, we step back and recall the "conundrum" we faced when looking at the sequence of Atx1

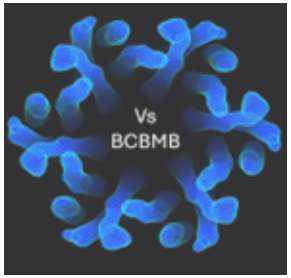
MAEIKHYQFNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDF
ILEKIKKTGKEVRSGKQL

Yellow: hydrophobic; green: polar/uncharged; blue: polar/positive charge; red: polar/negative charge

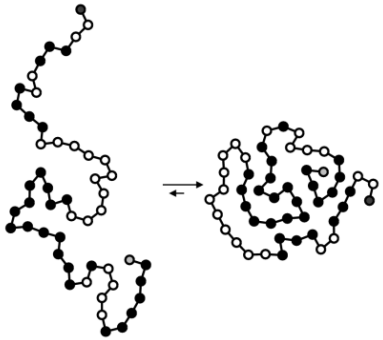


the seemingly random distribution of chemical properties along the sequence (primary structure), now comes into perspective as you realize that the **folding process** determines which sidechains will come close to each other in the 3D structure of the protein.

→ it's not surprising anymore that hydrophobic sidechains are scattered throughout the entire sequence, in fact – they must be to allow for formation of a folded protein that is stabilized by a hydrophobic core



Tertiary Structure Explains the Apparent Randomness in the Sequence



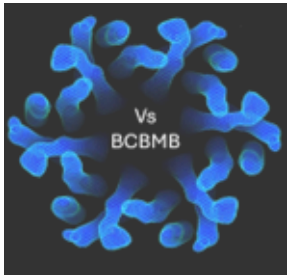
until 2018, the correlation between the exact primary structure (how many amino acids of what chemical nature and where) and the unique 3D native structure of a protein was a mystery, known as the "protein folding problem" = how does the sequence "know" what structure should form?

many tried to answer this question … but in the end "the machines won".

advances in computational methods brought about "Alpha Fold", a computer program that now allows accurate structure prediction based on amino acid sequence.

this advance is transformational; it bypasses the need to experimentally determine the structure (tedious, slow, sometimes impossible) to enable targeted functional studies that explain how exactly protein do their jobs. (….until AI will do that too….).

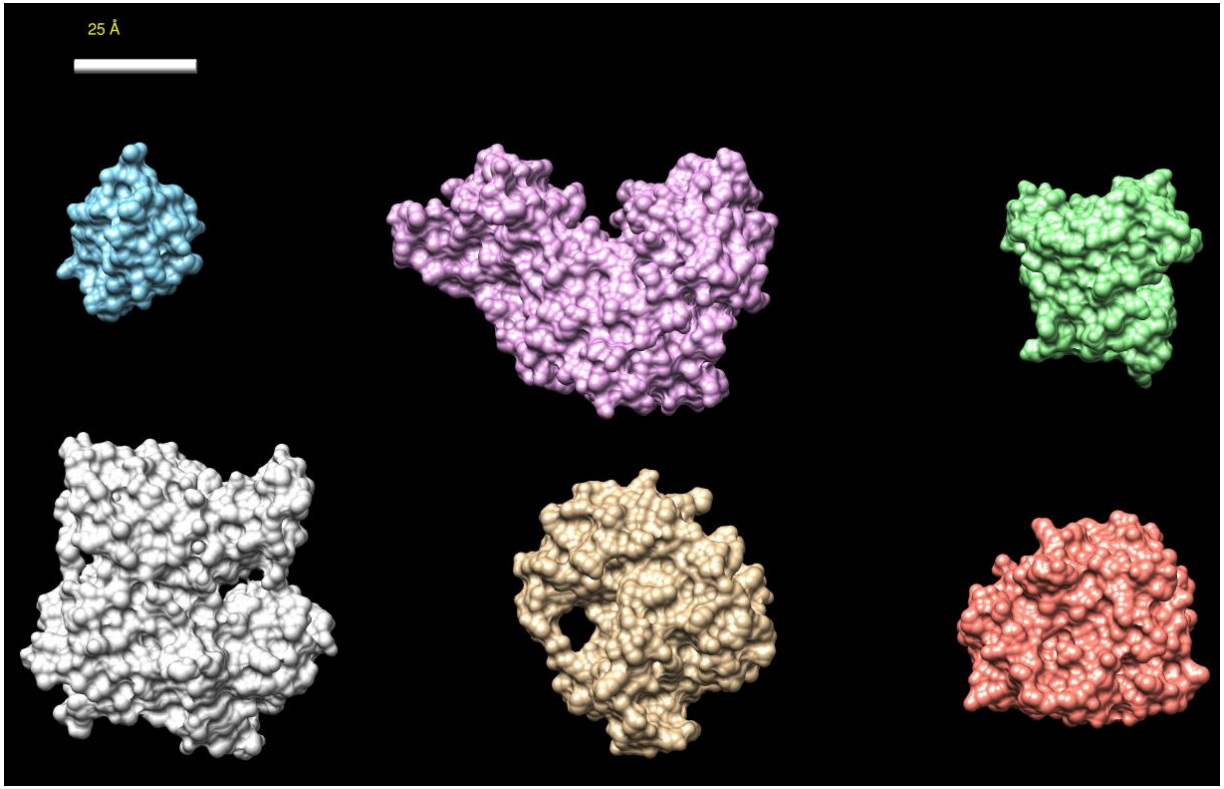
mind you: we still don't **understand** how, *in vivo,* protein folding gets it right every time – nor does the algorithm. It's a complete "black box" but it happens to get the right answer (most of the times). And with that … the crowds move on …. and so do we….



Proteins Come in Endless Shapes

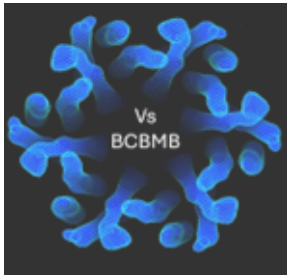
20-different sidechains + limited set of generic secondary structures + variable chain length

➔ **can build almost any shape with vast ability to tune chemical properties of the molecule.**



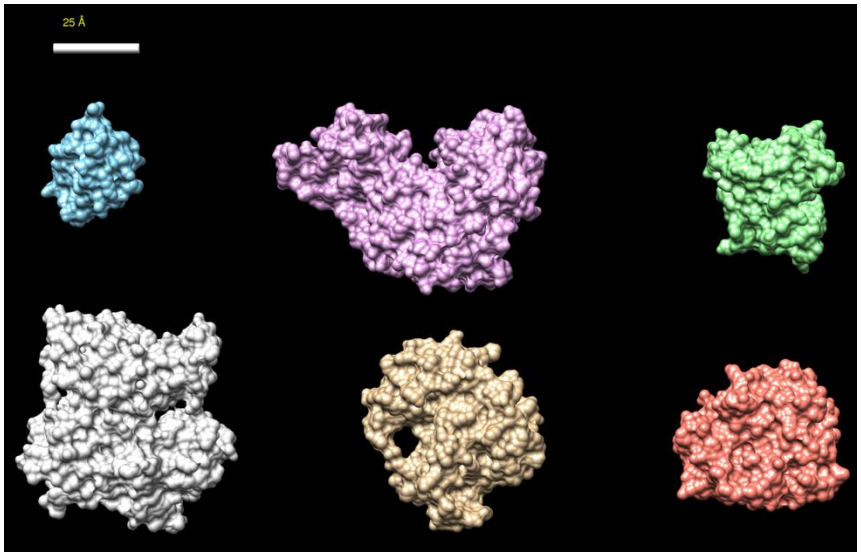
what is the significance of this pliability?

Answer: ….let's hear it....



Proteins Come in Endless Shapes

20-different sidechains + limited set of generic secondary structures + variable chain length
➔ **can build almost any shape with vast ability to tune chemical properties of the molecule.**



what is the significance of this pliability?

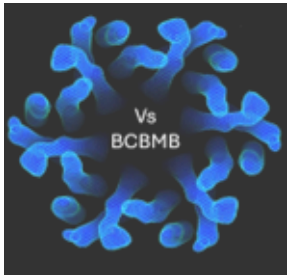
Answer

...Structure Determines Function.....

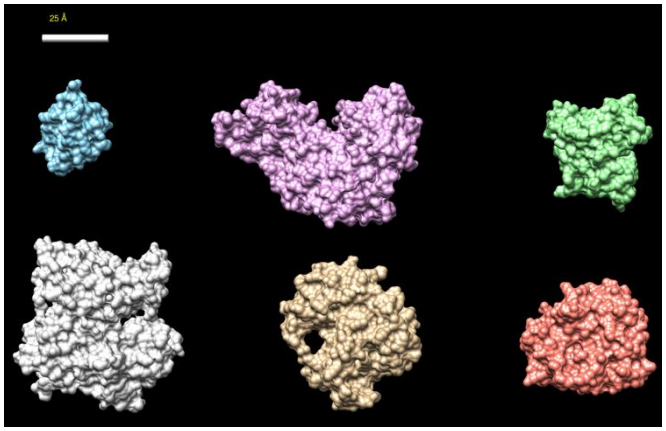
➔large shape repertoire = vast potential to fill many functional needs

let's emphasize this <u>really</u> important point once more
when we thought about DNA, we learned that DNA/nucleic acids were perfect for information storage because the composite sugar-phosphate backbone adopts a helical structure independent of what nucleobases are bound and in what order
➔ you can code through the oder of nucleobases on a "structurally invariant" (most of the time) template.

we find the same basic idea here in proteins: the polypeptide backbone prefers a small number of secondary structures (most common: helical, β-strands, turns). Meaning: these universal backbone templates can be decorated with an infinite number of sidechain combinations, allowing for a large number of different shapes/functions = this is, indeed, the "glue" we have been looking for ....!



Proteins Come in Endless Shapes



.... these universal backbone templates **can be decorated with an infinite number** of sidechain combinations, allowing for a **large number** of different shapes/functions = this is, indeed, the "glue" we have been looking for ....!

.... huhhhh....??!!? *Infinite* **is not the same as** *large* .... is that me being "sloppy" with words ... or is there something else here?

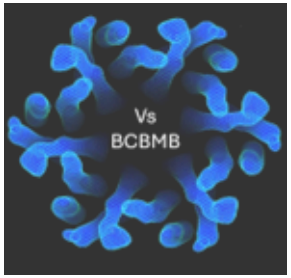
Answer

there is something else and without it, proteins would not have been too useful. We will cover this in detail in the "Molecular Biology and Advanced Biochemistry Lectures", but the preview we already want to look at here is this:

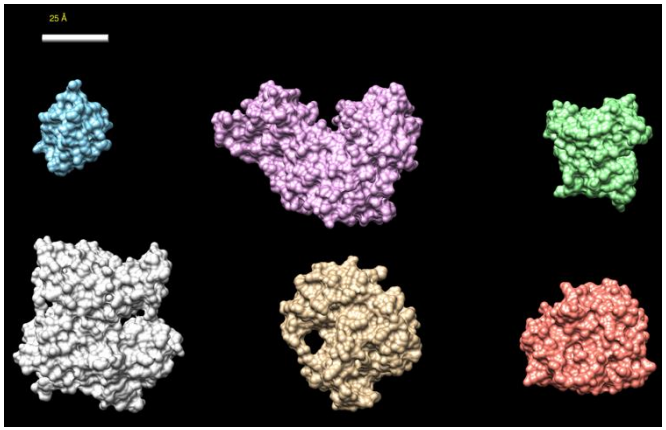
two polypeptide chains will adopt the same overall fold if they share at least 30% of sequence identity along their primary structure. (that is what causes **infinite combinations** to become **large number** ...).

how is this significant??

....flex your brainmuscle .... formulate hypotheses!



Proteins Come in Endless Shapes



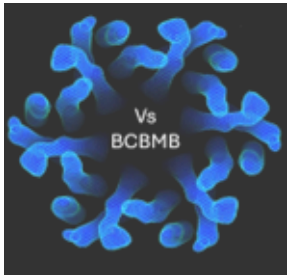
two polypeptide chains will adopt the same overall fold if they share at least 30% of sequence identity along their primary structure. (that is what causes **infinite combinations** to become **large number** ...).

how is this significant??

Answers

(1) it protects against loss of function if mutation occurs because the mutation will not change the overall structure and spatial alignment of functionally important amino acid sidechains (unless one of those gets changed … and then ... game over....)

(2) being able to make certain changes without destroying the overall fold allows "tweaking" a protein's functionality (eg: adding phosphate to glucose can become adding the phosphate to a structurally very similar, but functionally distinctively different sugar like galactose).



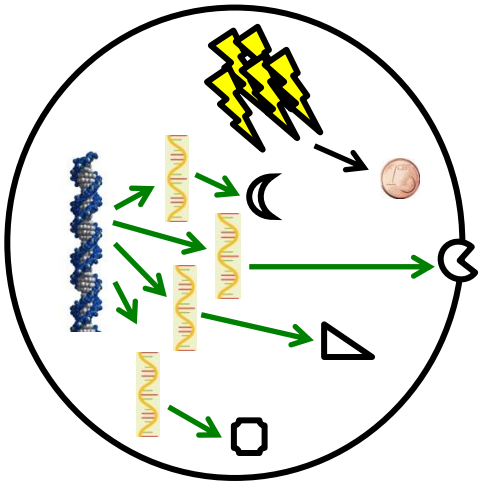
.

Lots of Information ... All Interesting (hopefully)

But What is Missing Still is the Perspective on What Proteins Actually Do

=

Time to Wrap This Up



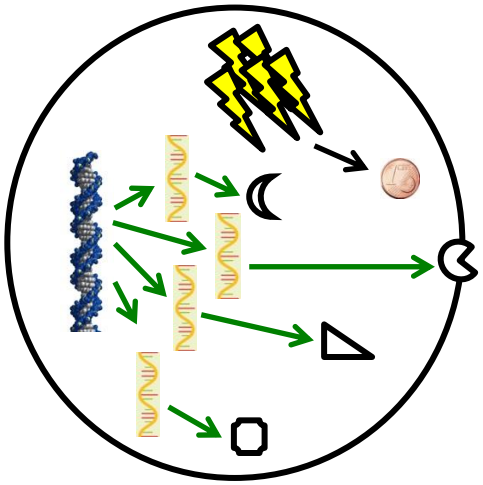
From Structure to Function

revisiting our starting point - **what is the most pressing need that remains unmet to make this "cell" functional at a basic level?**

Answer: catalysis **(another concept) –** i.e. the increase of the rate of chemical reactions or processes

why is that necessary?

...try to answer ....



From Structure to Function

revisiting our starting point - **what is the most pressing need that remains unmet to make this "cell" functional at a basic level?**

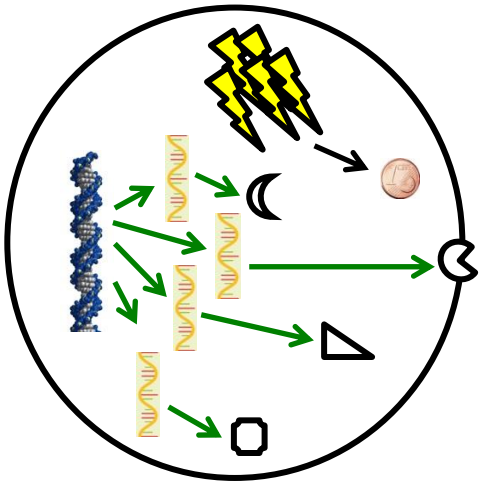
Answer: catalysis (another concept) – i.e. the increase of the rate of chemical reactions or processes

why is that necessary?

Answer

necessary because an energy barrier known as "activation energy" prevents stable organic compounds to react with other molecules at rates that are large enough to sustain life

proteins are uniquely suited to carry out catalysis – **why?**



From Structure to Function

revisiting our starting point - **what is the most pressing need that remains unmet to make this "cell" functional at a basic level?**

Answer: catalysis (another concept) **–** i.e. the increase of the rate of chemical reactions or processes

why is that necessary?

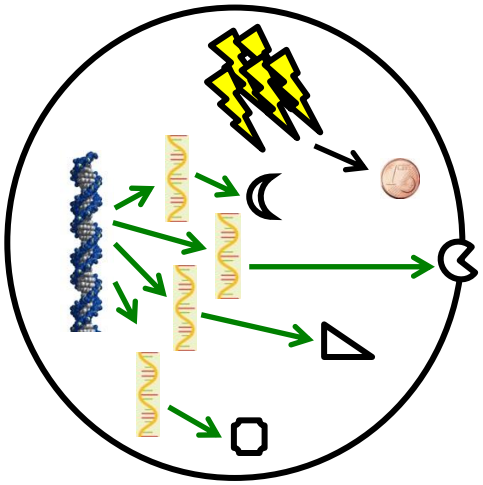
Answer: necessary because an energy barrier known as "activation energy" prevents stable organic compounds to react with other molecules at rates that are large enough to sustain life

proteins are uniquely suited to carry out catalysis – **why?**

Answer: practically unlimited structural diversity of **proteins** allows them to **form customized reaction chambers** that are capable of specifically recognizing reactants, to lower activation energies, and to increase the rate of bond rearrangments

(we'll cover that in the "Advanced Biochemistry CATALYSIS" lecture)

is catalysis the only function proteins can have?



From Structure to Function

revisiting our starting point - **what is the most pressing need that remains unmet to make this "cell" functional at a basic level?**

Answer: catalysis (another concept) **–** i.e. the increase of the rate of chemical reactions or processes

why is that necessary?

Answer: necessary because an energy barrier known as "activation energy" prevents stable organic compounds to react with other molecules at rates that are large enough to sustain life

proteins are uniquely suited to carry out catalysis – **why?**

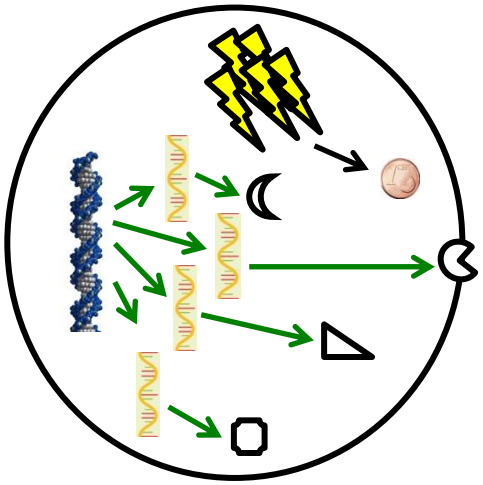
Answer: practically unlimited structural diversity of **proteins** allows them to **form customized reaction chambers** that are capable of specifically recognizing reactants, to lower activation energies, and to increase the rate of bond rearrangments

is catalysis the only function proteins can have?

Answer: no – the ability to very specifically engage other biological macromolecules also makes proteins very useful as scaffolding/structural templates, signaling receptors, and membrane embedded transporters for movement of polar substances across bilayers

(all covered in other Silent Lectures)

are there any functions that proteins cannot fulfill?



From Structure to Function

revisiting our starting point - **what is the most pressing need that remains unmet to make this "cell" functional at a basic level?**

Answer: catalysis (another concept) **–** i.e. the increase of the rate of chemical reactions or processes

why is that necessary?

Answer: necessary because an energy barrier known as "activation energy" prevents stable organic compounds to react with other molecules at rates that are large enough to sustain life

proteins are uniquely suited to carry out catalysis – **why?**

Answer: "unlimited" structural diversity of **proteins** allows them to **form customized reaction chambers** that are capable of specifically recognizing reactants, to lower activation energies, and to increase the rate of bond rearrangments

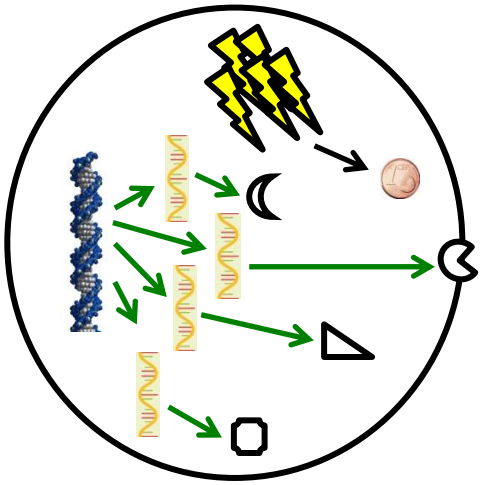
Is catalysis the only function proteins can have?

Answer: no – the ability to very specifically engage other biological macromolecules also makes proteins very useful as scaffolding/structural templates, signaling receptors, and membrane embedded transporters for movement of polar substances across bilayers.

are there any functions that proteins cannot fulfill?

Answer: yes – proteins cannot be used for information **storage**. **why?**

From Structure to Function



Answer: no – the ability to very specifically engage other biological macromolecules also makes proteins very useful as scaffolding/structural templates, signaling receptors, and membrane embedded transporters for movement of polar substances across bilayers.

are there any functions that proteins cannot fulfill?

Answer: yes – proteins cannot be used for information **storage**. **why?**

Answer

proteins violate two key design principles for an information storing protein

(1) protein surfaces are too complicated to be easily and rapidly interpreted and

(2) the relationship between primary structure and shape is degenerate

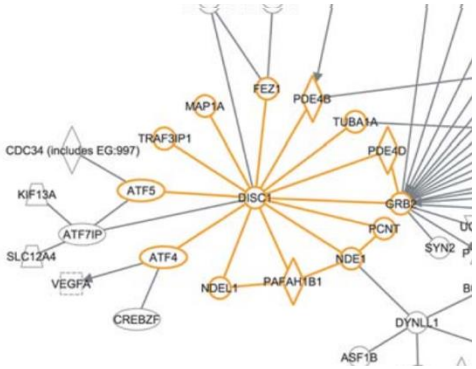
= as pointed out earlier: multiple different sequences can have same overall folding patterns of their backbone

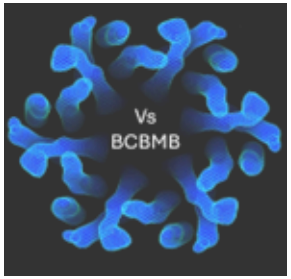
(plausible if you consider: polypeptide of 400 amino acids could give 20^{400} primary structures...! = no way that all of these form unique overall structures....)

while this is great for creating and diversifying (a given) function, it **kills** information storage

unless you can live with *GENERIC* = *SPECIFIC* = *TASTY* = any other number of things ....

Diversifying Function Through Structural Complexity

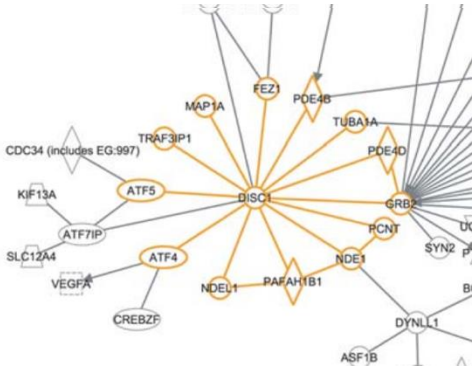


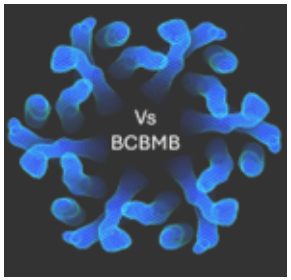


contemplating the nature of biological systems in the introduction lecture to cells – protein interactoms served as an example to illustrate the complexity of molecular interconnectedness ("Introduction to Cells", Slide 23)

→ a puzzling observation was that **some proteins are quite promiscuous = can recognize and interact with multiple other proteins**
how is polyvalency possible?

Diversifying Function Through Structural Complexity





contemplating the nature of biological systems in the introduction lecture to cells – protein interactoms served as an example to illustrate the complexity of molecular interconnectedness ("Introduction to Cells", Slide 23)

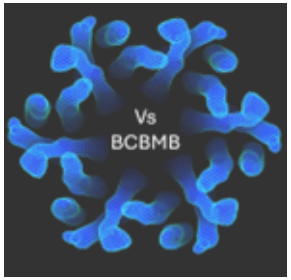
→ a puzzling observation was that **some proteins are quite promiscuous = can recognize and interact with multiple other proteins**
how is polyvalency possible?

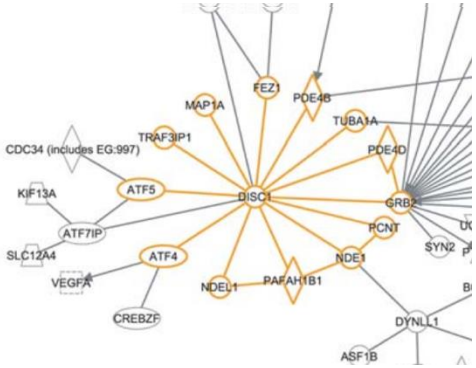
Answer
surface display of multiple specific patterns that allow different proteins to bind

how to implement this?

....brainstorm .... formulate hypotheses!

Diversifying Function Through Structural Complexity



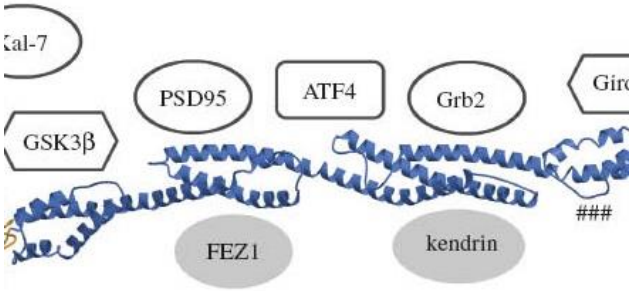


Answer: surface display of multiple specific patterns that allow different proteins to bind
how to implement this?

1 - simplest solution: (overlapping) regions on an extended [or spherical in some cases] polypeptide

analogy: row of seats in a theater

example: "disconnected in schizophrenia 1" protein from the "introduction to cells" ("central hub" in figure above; part of its predicted structure is shown to the right



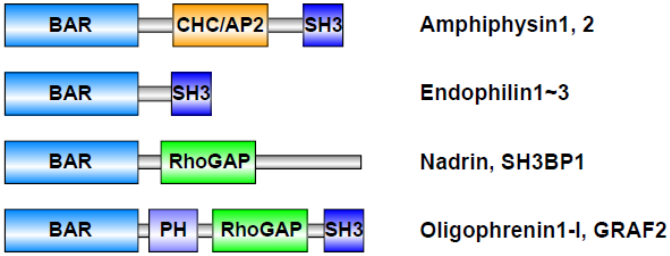
2 – second solution: domains

defintion: a domain is the smallest part of a protein that can independently fold into a stable tertiary structure.

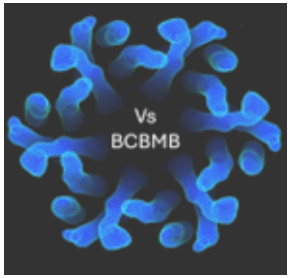
analogy: storefront with several different businesses

➔Domain sizes range from ~50-300 amino acids (average: ~100-120 residues)

➔typically: 1 domain = 1 function → emergence of new proteins can follow a "copy, cut, paste & tweak" approach (analogy: eg different types of Fast Food chains)



BAR: membrane bending activity; SH3: recognizes peptides enriched in proline residues, CHC/AP2: interacts with adapter proteins in endocytosis, Rho/GAP has a role in molecular signaling



Diversifying Function Through Structural Complexity

so far: design solutions to functional polyvalency in proteins are constrained to a single polypeptide.

are there additional alternatives to achieve functional complexity in proteins?

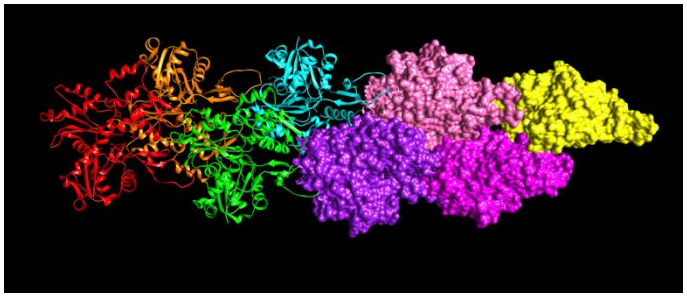
Answer: yes

formation of (mostly) **non-covalent aggregates** from independently folded polypeptides. Each of the independent polypeptides in a complex is called a **"subunit"**

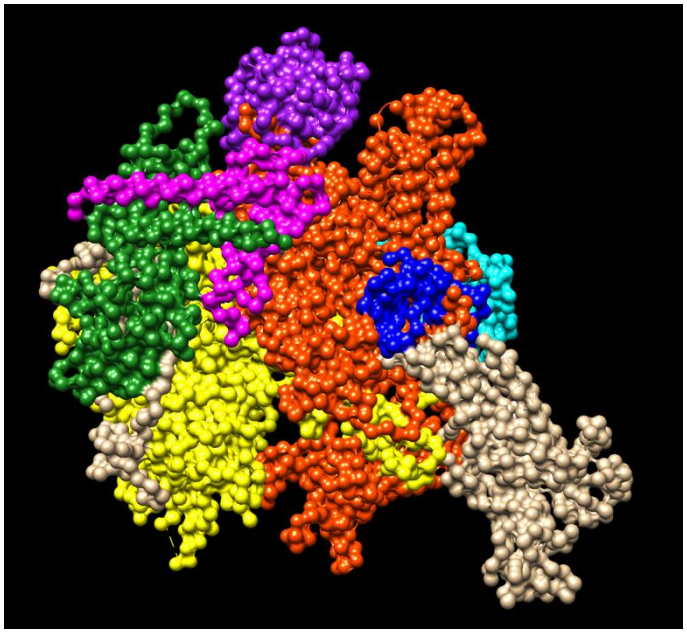
in more general terms: the (mostly) non-covalent aggregation of ≥2 independently folded biological polymers gives rise to a level of structural organization that is called **quaternary structure**

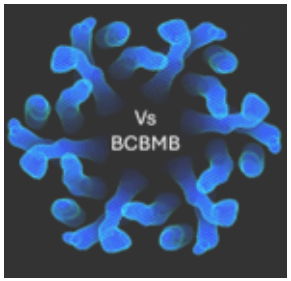
- ✓ quaternary structure can involve the aggregation of completely different classes of polymers (most common: proteins + nucleic acids; proteins + lipids)
- ✓ complexes made from **multiple copies of a single subunit** are called **homo oligo-/polymers** (e.g. actin filaments, top image)
- ✓ complexes made from **subunits with different primary structures** are called **hetero oligo-/polymers** (e.g RNA-polymerase/transcription factor complex)

Actin Filament



RNA-Polymerase/TFII complex





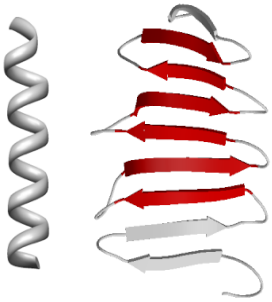
Summary of Protein Structural Organization

Primary Structure: sequence of amino acids ….for instance:

PEPTIDESMAKE**MACHINES**THAT
ACCELERATELIFES**CHEMISTRY**….

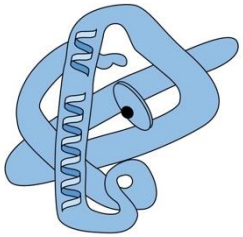
⇩

Secondary Structure: local conformation of polypeptide backbone atoms (= amino acid constant core unit)



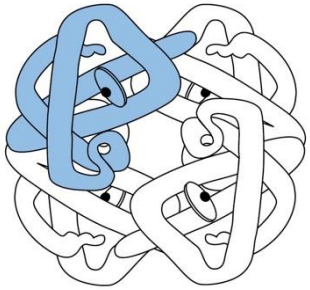
⇩

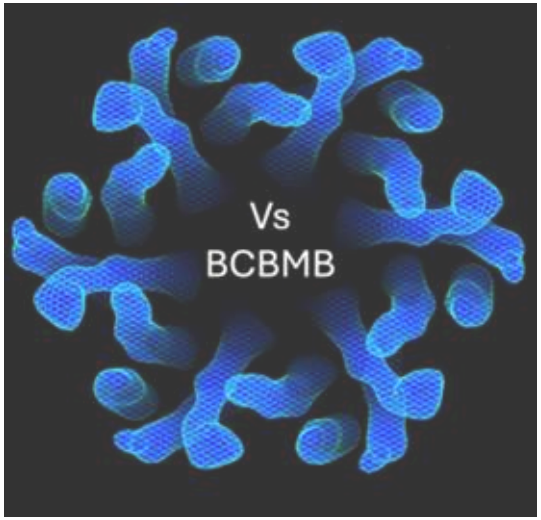
Tertiary Structure: overall spatial arrangement of all of a polypeptide's atoms that is caused and stabilized by long range interactions (mostly) between amino acid sidechains



⇩

Quaternary Structure: aggregate of ≥2 independently folded (identical) polypeptide chains.





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