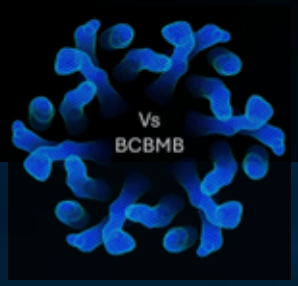


Advanced Biochemistry

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 shown in the slide-deck as still images, in case you don't want to watch the videos)

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

Why Give This a Go?

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

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

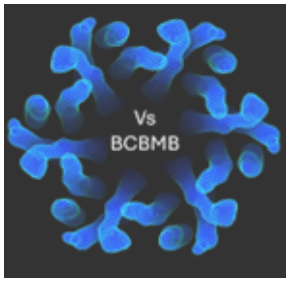
Advanced Biochemistry

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

while this tutorial can be used stand alone, you will benefit more from the "silent lectures" by

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

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



Overture

the discussion about chemical evolution has given us the “big picture” frame for all that is to come.

taking an engineering point of view – the first pit stop was to revisit the “system” aspect of our definition for life: *>>a …..system<<*
implies boundaries
(and the chapter on lipids * membranes told you a lot about those),

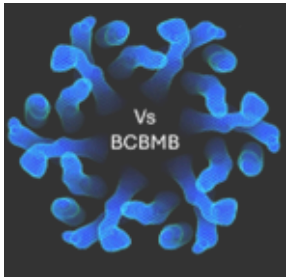
Returning to the minimalist definition of "life"
>>life is a process in which an open system maintains a homeostatically controlled directed flow of electrons<<

the next logical step to explore is
>> …open…. <<.

here you encounter another "stop – go" situation…..

the properties of lipids along with the hydrophobic effect predisposed some lipids to flexible boundaries and enclosed compartments. Although life could not have evolved without this, the appearance of cells and membrane bound compartments also created an enormous logistical challenge …

whatever the boundaries they need to be permeable to nutrients to enter and wastes to leave the cell. This is not trivial because ions, sugars, nucleotides and amino acids all are (quite) water soluble = plain lipid bilayer represent an impermeable barrier.



Compartmentalization Creates Challenges

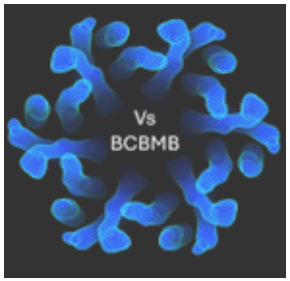
in other words: once nature figured out how to build molecular barriers that would allow the hallmark "non-equilibrium" conditions of life to be established, the very impermeability of bilayers to almost everything but very hydrophobic compounds became the singly most important problem to solve

further down the road, it also creates a need to coordinate and regulate what is happening where and when, and to learn how to communicate across the boundary. let's deal with these issues one step at the time.

first:

how do we deal with the impermeability?

…try to answer …. don't overthink it … what is your spontaneous thought?



Compartmentalization Creates Challenges

how do we deal with the impermeability?

Answer: need to create selective pores

yes – that is the right answer ….at least eventually ….

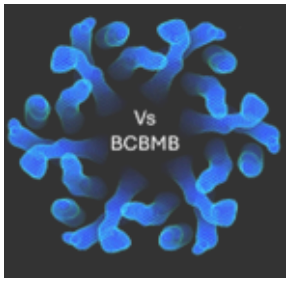
why "eventually"?

well … if you think about this carefully, and as we explore what this looks like later, you will realize that the "pores" are structurally quite complex
= they did not just pop up "out of nowhere"

meaning:

in the very beginning … bilayers were not as tight and impermeable as they are today … they were quite leaky = a lot of the small building blocks could still pass well enough to become "trapped" inside and allow for the processes that were needed to maintain the cell and develop new molecular inventory….

…including primitive "pores"…which in turn allowed the bilayer to improve its structure and tightness … and around you go until you arrive at "modern membranes" and modern "pores"



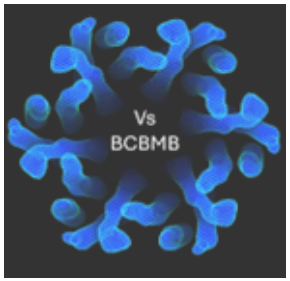
Compartmentalization Creates Challenges

Answer: need to create selective pores

let's not dwell on the exact timeline and the initial "messiness" of the system, but rather focus on

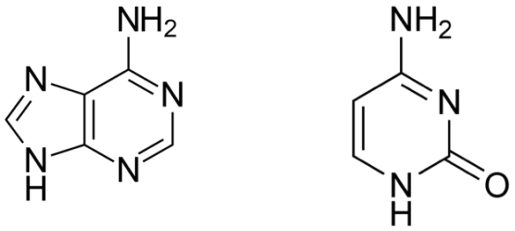
how do you create a selective pore??

Answer: the solution will come from examining our molecular inventory again – we have dealt with the lipids. What else is there?

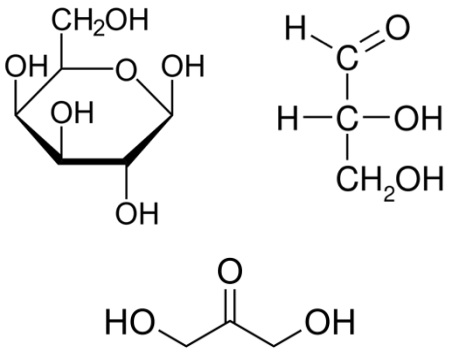


Recall – Macromolecular Inventory

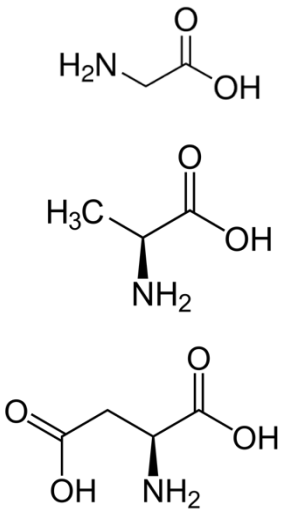
Heterocyclic Aromatic "Nucleobases"



Carbohydrates

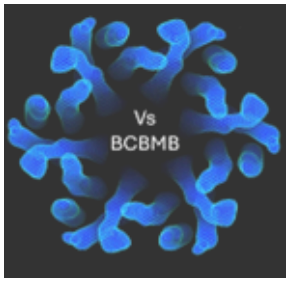


Amino Acids

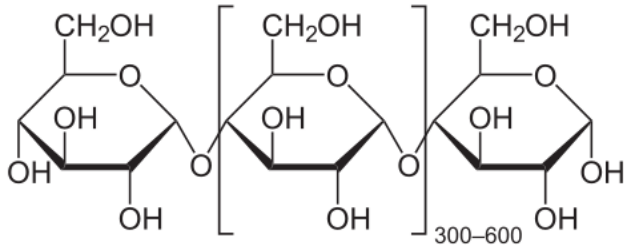


conclusion:

monomers of biologically relevant base compounds and their simple derivatives (such as ATP or other cofactors/coenzymes) are too small to form any type of pores → need to consider their oligo-/polymers for potential to form diversified pores.



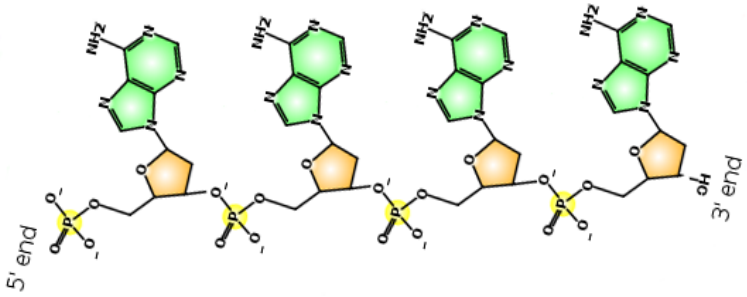
Assessing Biological Polymers for Pore Forming Potential

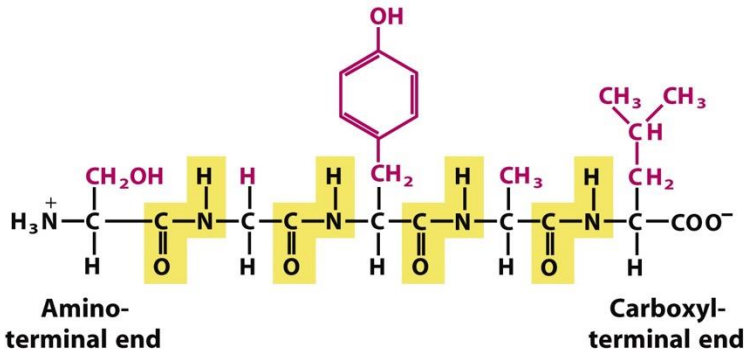


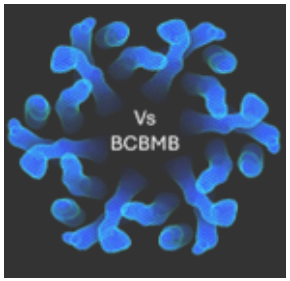
→ **no, too hydrophilic (as a single strand) to insert into hydrophobic membrane environment**

→ chemistry too limited to allow for transport of many different things

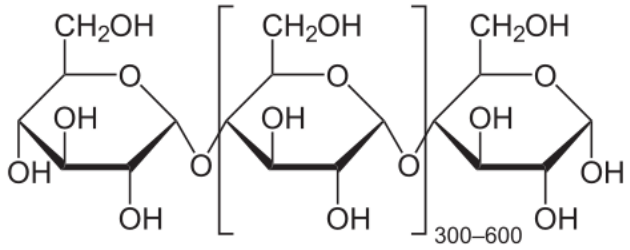
→ higher order structures too limited and rigid







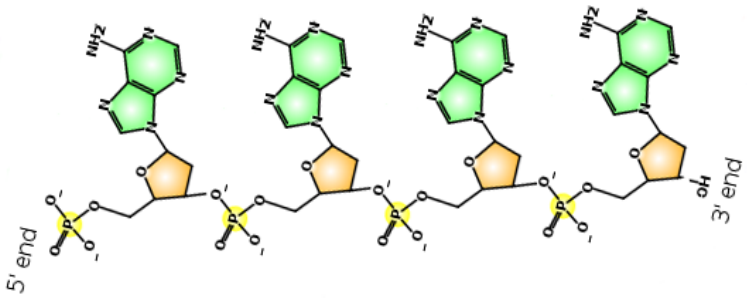
Assessing Biological Polymers for Pore Forming Potential



→ **no, too hydrophilic (as a single strand) to insert into hydrophobic membrane environment**

→ chemistry too limited to allow for transport of many different things

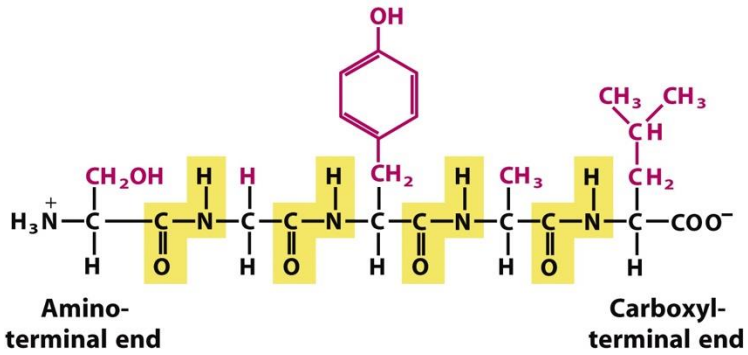
→ higher order structures too limited and rigid

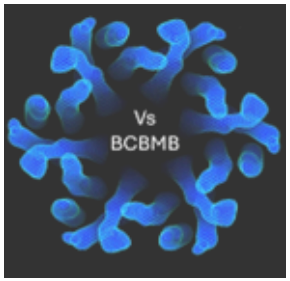


→ **some hydrophobicity, but the charged backbone is a killer if you think about how to accommodate it comfortably in the hydrophobic membrane environment**

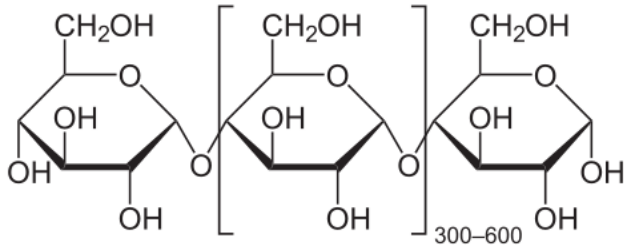
→ also: chemistry and ability to form a flexible shape is very limited = limits dynamics and what could be transported

---





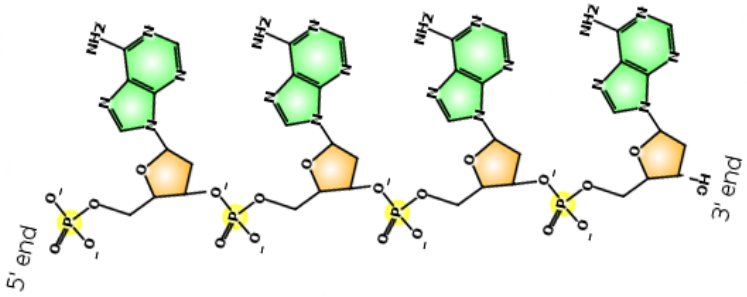
Assessing Biological Polymers for Pore Forming Potential



→ **no, too hydrophilic (as a single strand) to insert into hydrophobic membrane environment**

→ chemistry too limited to allow for transport of many different things

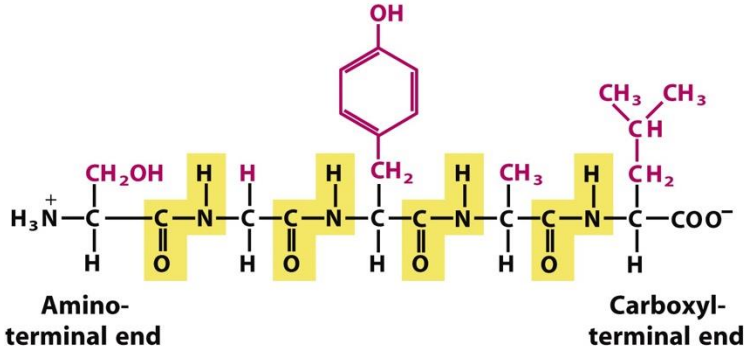
→ higher order structures too limited and rigid



→ **Some hydrophobicity, but the charged backbone is a killer if you think about how to accommodate it comfortably in the hydrophobic membrane environment**

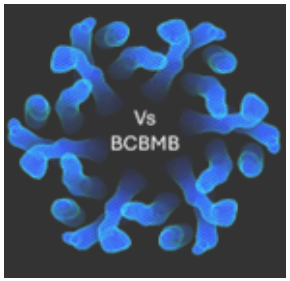
→ also: chemistry and ability to form a flexible shape is very limited = limits dynamics and what could be transported

---

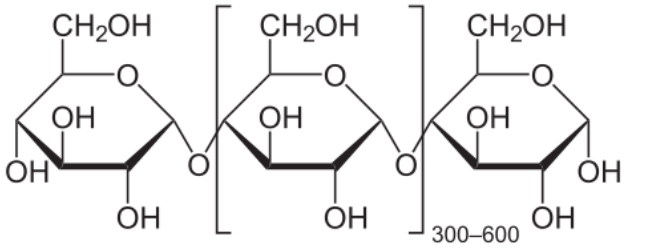


→ this shows promise – why?

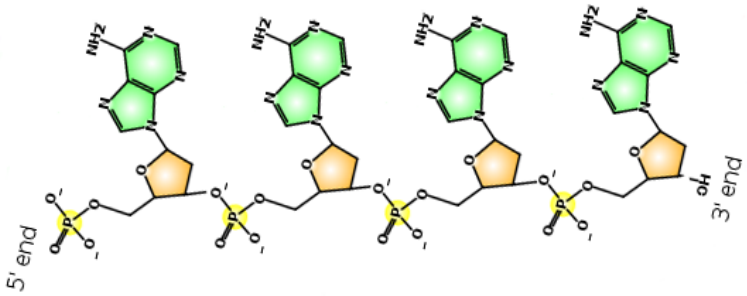
Answer: the backbone is merely polar, and the chemistry allows for hydrophobic surfaces
(that can be presented to the hydrophobic core)
+ structural diversity
(to create passageways for different compounds)



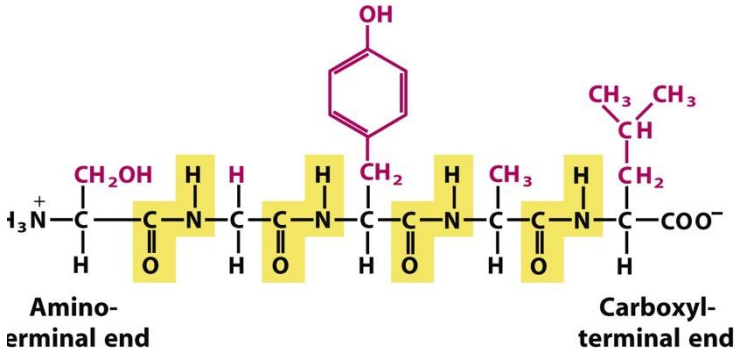
Assessing Biological Polymers for Pore Forming Potential



→ no, too hydrophilic (as a single strand)

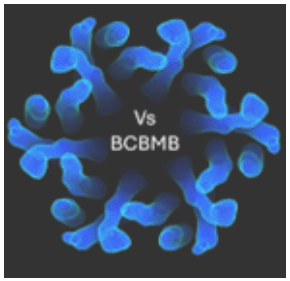


→ Some hydrophobicity, but the charged backbone is a killer



→ This shows promise – why?

Answer: the backbone is merely polar and the chemistry allows for hydrophobic surfaces
(that can be presented to the hydrophobic core)
+ structural diversity (to create passageways for different compounds)

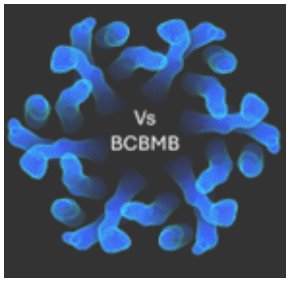


Goals of this Chapter

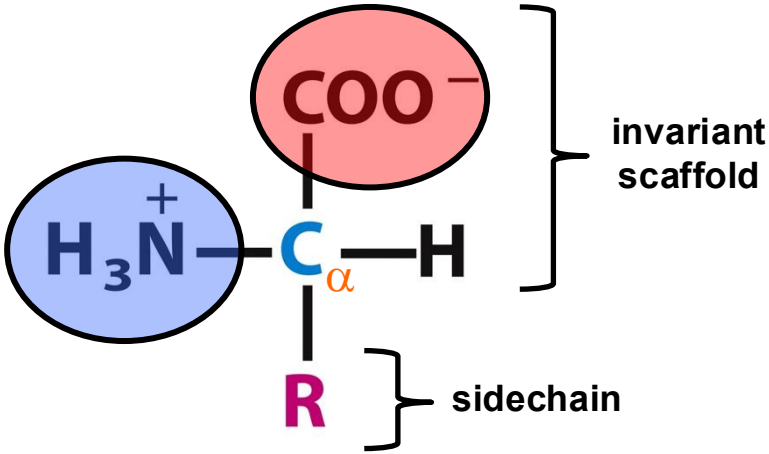
- want to refresh memory of protein basics
- what to deepen understanding of protein structure and folding
- want to explore the structure and function of membrane embedded proteins – a topic we completely ignored in the "Biochemistry Fundamentals – PROTEINS Chapter"

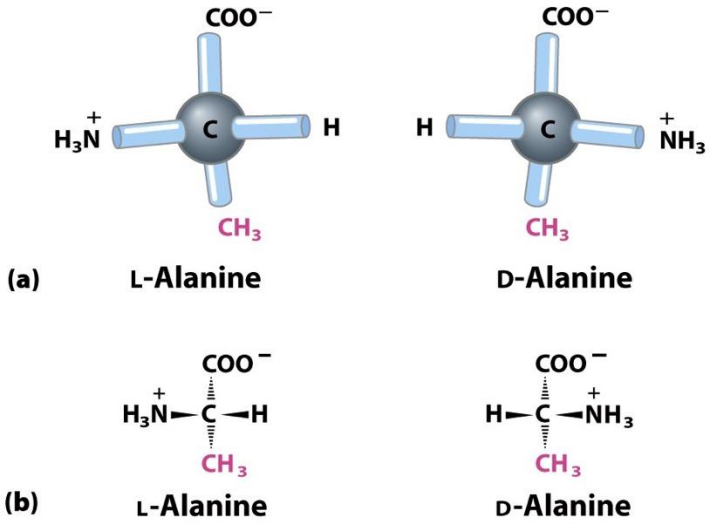
at the end of this chapter you should have a firm understanding of:

- principles that drive protein folding
- major structural elements in proteins
- correlation between protein structural organization and protein function
- the synergistic roles of polypeptide **and** bilayer properties that allow formation of membrane proteins
- the general functions that are carried out by membrane proteins
- general design principles in membrane proteins
- principles of membrane protein insertion into membranes



α–Amino Acids - The Building Blocks of Proteins (Recap)





L- and D- α-amino acids, except R=H are related like mirror images

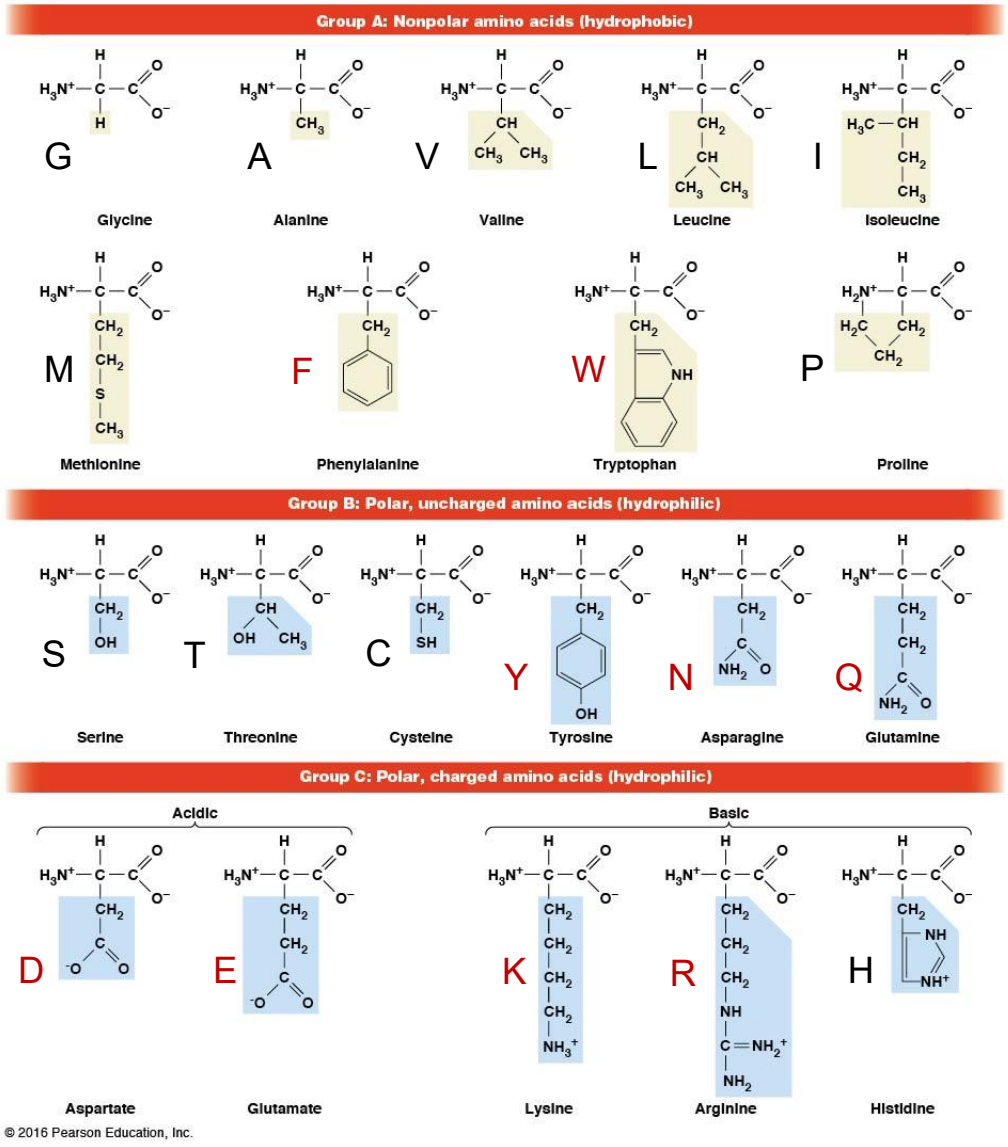
if you prefer the R,S nomenclature: L-Amino Acids = S (except Cys, which is the R-isomer)

This is another encounter with one of the important concepts: asymmetry

in proteins: only L-amino acids are used, and referred to as ”residues” (= a protein made from 400 amino acids is also said to have 400 residues)

Useful resource: https://en.wikipedia.org/wiki/Proteinogenic_amino_acid

α–Amino Acids – Recap and Extension



just like ATGC are used for the four bases when describing nucleic acids, a one letter notation is used to describe protein primary structure.

for convenience, the first letter of the amino acid name is used to abbreviate it with the following exceptions:

Phenylalanine → F
Tyrosine → Y
Tryptophan → W

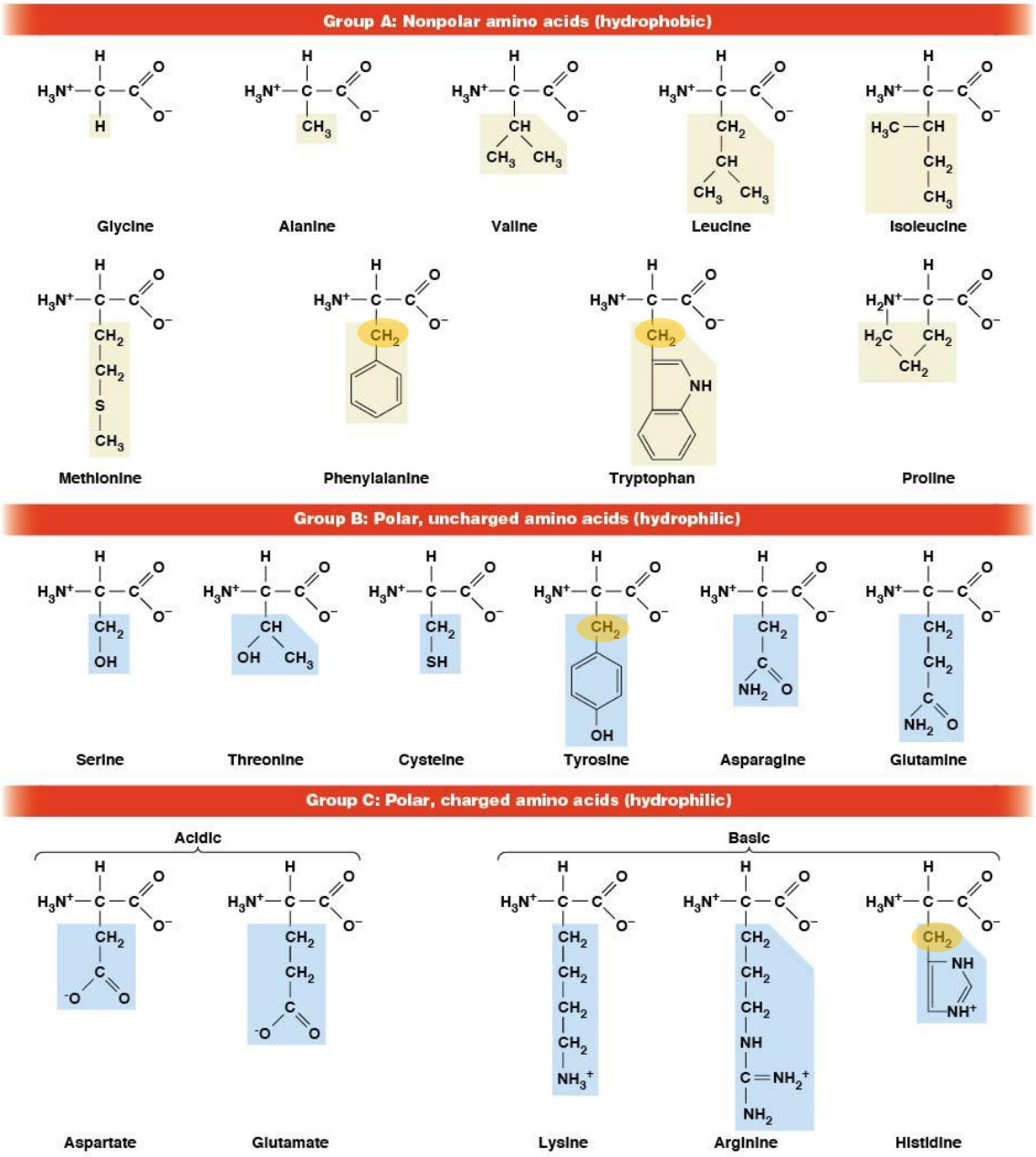
Aspartic Acid → D
Glutamic Acid → E

Asparagine → N
Glutamine → Q

Lysine → K
Arginine → R

Knowing the One Letter Code for the amino acids is REALLY helpful for following lectures, seminars, paper reading assigments etc …

α–Amino Acids – Specialized Functional Roles



some amino acids have special functions beyond being a generic building block of proteins:

Methionine (M): all protein open reading frames start with M in the process of translation (but M is often removed post-translationally)

Proline (P): the unique structure of proline is exploited to form ”turns” in the polypeptide backbone.

Cysteine (C): the –SH thiol group can react with a –SH group of a second Cys residue to form a covalent –S-S- disulfide bond (cystine). Disulfide formation is an oxidation (loss of $2e^-$ and $2H^+$). This reaction can be undone if needed.

ALSO TAKE NOTE

- ring system of aromatic sidechains is separated from the scaffolding by a $–CH_2-$ spacer (= you can think of them as alanine sidechains where one –H atom is replaced with an aromatic ring(system)

- Sidechains of K and R have very significant aliphatic (hydrophobic) regions (→ are amphiphiles)

Protein Backbone Is Formed By Condensation (Recall)

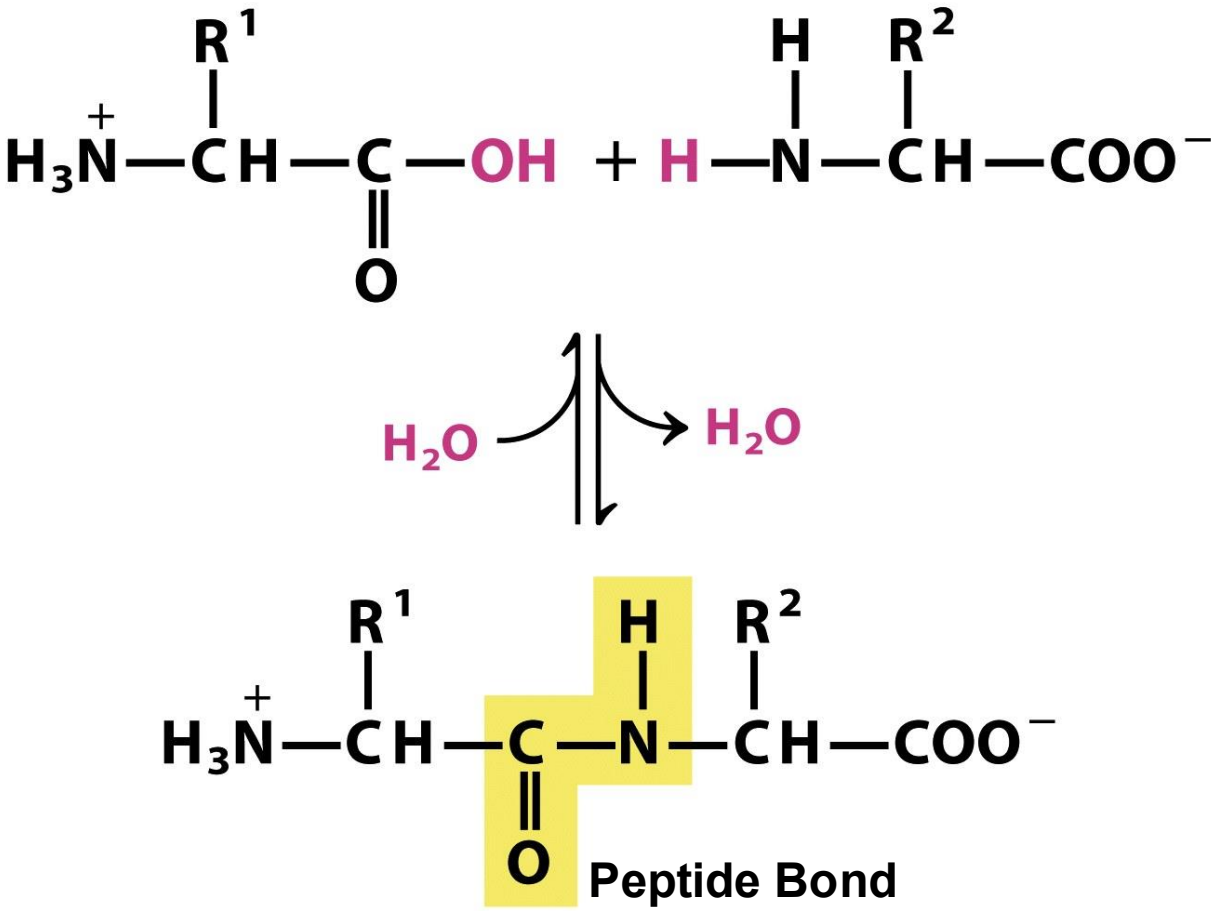


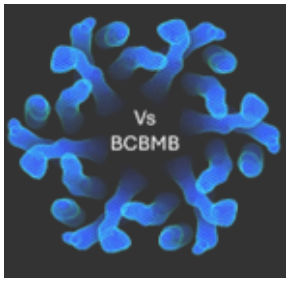
Figure 3-13
Lehninger Principles of Biochemistry, Fifth Edition

Also see:
Fundamentals – PROTEINS
Slide 22

Molecular and Cell Biology – TRANSLATION
(for details on the chemical issues with this bond formation and how peptide bond is synthesized)

definitions: oligopeptide = few amino acids (<50)

polypeptide = many amino acids



Before Moving on to More Details of Protein Structure let's Mix Things Up ….Take the QUIZ!

Protein Folding – What You SHOULD Remember/Know

Question 1: What is protein folding?

Question 2: What initiates protein folding?

Question 3: What types of intramolecular interactions contribute to folding of a polypeptide chain? Explicitly name them.

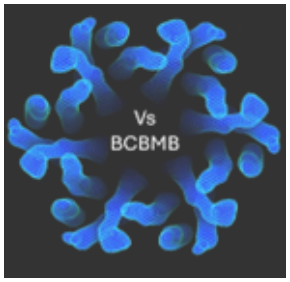
Question 4: How many different outcomes does the folding process have?

Question 5: Do all polypeptide chains fold? If so – why? If not – how does the paradigm "structure determines function" apply?

Question 6: Is quaternary structure formation part of folding? Why?

Question 7: What does the term "tertiary structure" refer to?

Question 8: What is a protein domain?



Before Moving on to More Details of Protein Structure let's Mix Things Up ….Take the QUIZ!

Protein Folding – What You SHOULD Remember

Question 1: What is protein folding? Disorder → Order transition of protein backbone

Question 2: What initiates protein folding? Hydrophobic Effect

Question 3: What types of intramolecular interactions contribute to folding of a polypeptide chain? Only looked at contributions of Van der Waals Interaction in Fundamentals Chapter

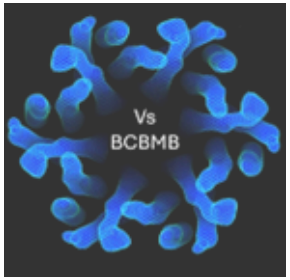
Question 4: How many different outcomes does the folding process have? Only one if the protein has a native structure

Question 5: Do all polypeptide chains fold? No. About 10% of protein chains are wholly unstructured (which is required for them to function properly)

Question 6: Is quaternary structure formation part of folding? Why? No. "Folding" refers to single chains of polypeptides. Quaternary structure is a higher order structure in which independently folded chains have associated (non) covalently to form a protein complex.

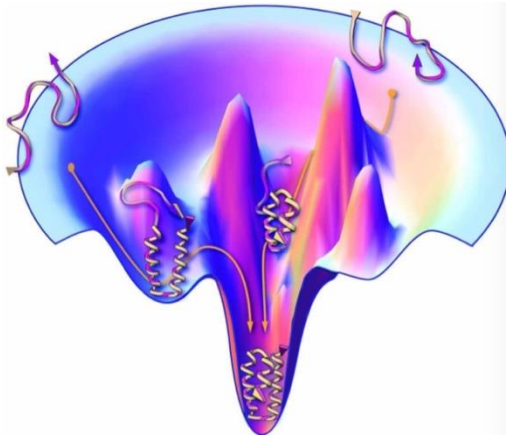
Question 7: What does the term “tertiary structure” refer to? The fully formed 3D structure of a polypeptide. Tertiary structure is formed through interactions of amino acid sidechains that are distant from each other in the primary structure (sequence).

Question 8: What is a protein domain? It's the smallest part of a polypeptide/protein that can independently fold into a functional region.



Stages of Protein Folding (Recall) + Extension of Knowledge

→ initial burial of hydrophobic sidechains leads to an ensemble of poorly structured folding intermediates that are called “molten globule”



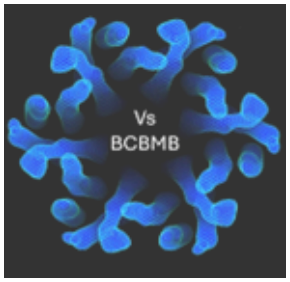
the “molten globule states” resolve by **locally** forming higher order structures

- helical
- extended, laterally packed or turning
- extended but unstructured

} **SECONDARY STRUCTURE**

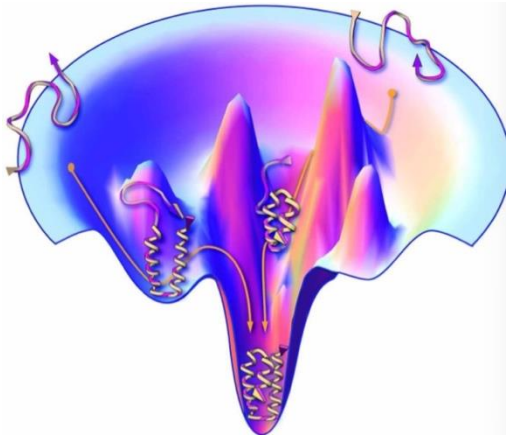
In the Fundamentals Chapter, we took some time to explore and "discover" the helical and extended secondary structure elements (slides 35, 39-42)

New here: the formation of secondary structure is still driven by the hydrophobic effect initially …..why?



Stages of Protein Folding (Recall) + Extension of Knowledge

→ Initial burial of hydrophobic sidechains leads to an ensemble of poorly structured folding intermediates that are called "molten globule"



the "molten globule states" resolve by **locally** forming higher order structures

- •helical
- •extended, laterally packed or turning
- •extended but unstructured

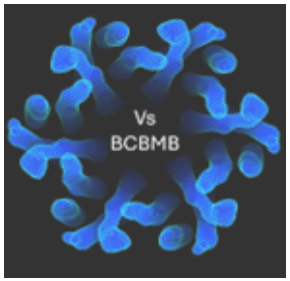
} **SECONDARY STRUCTURE**

In the Fundamentals Chapter, we took some time to explore and "discover" the helical and extended secondary structure elements (slides 34, 38-41)

new here: the formation of secondary structure is still driven by the hydrophobic effect …..why?

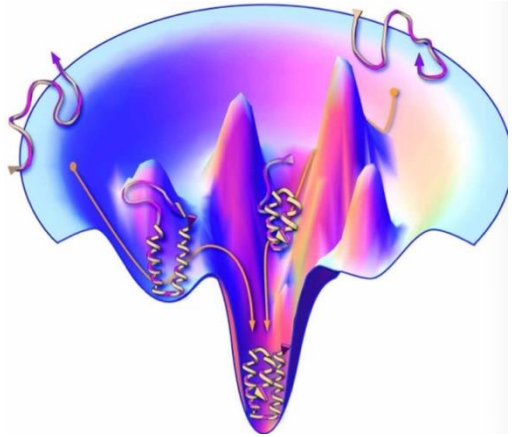
Answer: the peptide bond groups (-NH-CO-) do **NOT** like to be buried in the hydrophobic interior that begins to form ….. → polymer will attempt to find conformations that allow hydrogen bonding propensity of the –NH-CO- groups (donor – acceptor) to be fulfilled without opening the emerging hydrophobic core of the protein back up to the aqueous environment

→this conformational search leads to the formation of **secondary structure, which is defined as the local conformation of the protein's backbone atoms only** (slides 36-41)
(hydrophobic effect initiates secondary structure formation, then H-bonding takes over to drive it to completion)



Stages of Protein Folding (Recall) + Extension of Knowledge

→ Initial burial of hydrophobic sidechains leads to an ensemble of poorly structured folding intermediates that are called "molten globule"



the "molten globule states" resolve by **locally** forming higher order structures

- helical
- extended, laterally packed or turning
- extended but unstructured

} **SECONDARY STRUCTURE**

In the Fundamentals Chapter, we took some time to explore and "discover" the helical and extended secondary structure elements (slides 34, 38-41)

new here: the formation of secondary is still driven by the hydrophobic effect …..why?

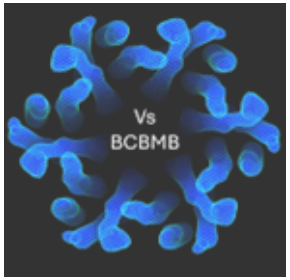
Answer: the peptide bond groups (-NH-CO-) do **NOT** like to be buried in the hydrophobic interior that begins to form ….. → polymer will attempt to find conformations that allow hydrogen bonding propensity of the –NH-CO- groups (donor – acceptor) to be fulfilled without opening the emerging hydrophobic core of the protein to open back up to the aqueous environment

this conformational search drives the formation of → secondary structure

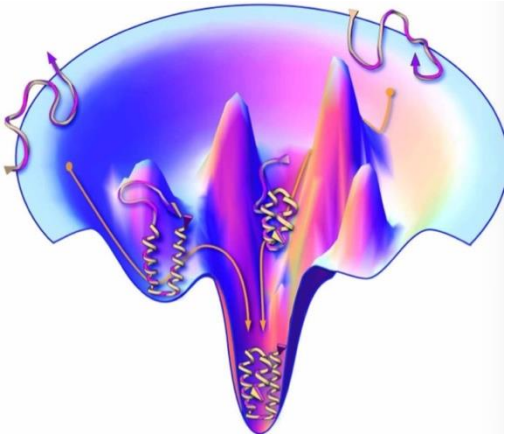
long-range interactions (H-bonds, charge, Van der Waals) between secondary structure elements - mostly through sidechains – then stabilize the fold further

→ **TERTIARY STRUCTURE**

REALLY IMPORTANT: this process is REVERSIBLE = use enough energy and protein chains will unfold!!



Stages of Protein Folding (Recall) + Extension of Knowledge

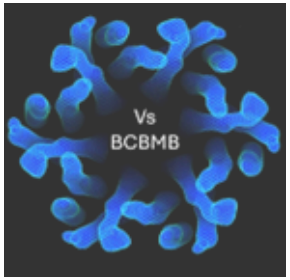


endresult is a compacted, densely packed polymer that typically does not have cavities on the inside and is stabilized by weak interactions (H-bonds, ionic, Van der Waals) & in some cases by disulfide bridges and/or divalent metal ions (mostly Zn^{2+}).

while clearly happening, formation of secondary/tertiary structure incurs a large entropic penalty (ΔS) on backbone atoms and many sidechains

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

→ while some of the penalty is offset by liberation of water molecules that are tied up in clathrate cages, **enthalpy** (ΔH) **becomes the dominant driver of folding** past the initial burial of hydrophobic sidechains because the cooperative formation of many H-bonds (as secondary/tertiary structure form) releases heat.

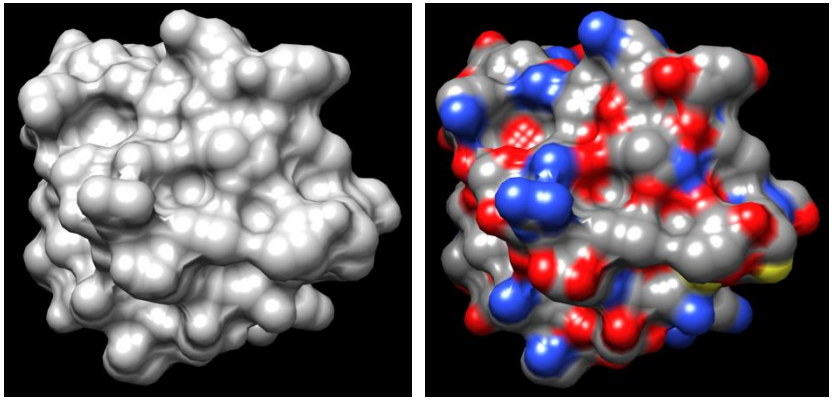


Folding of the Backbone Leads to Redundant Structural Elements Called Secondary Structure (Visual Recall).

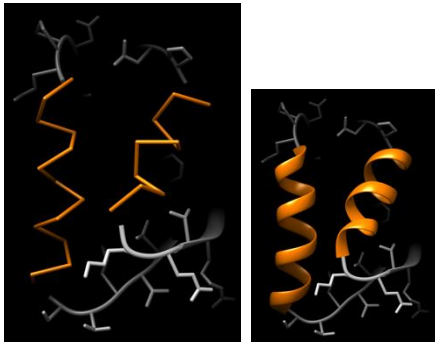
the protein's Van der Waals surface looks irregular

left: just steric surface topography
right: adds chemical information.

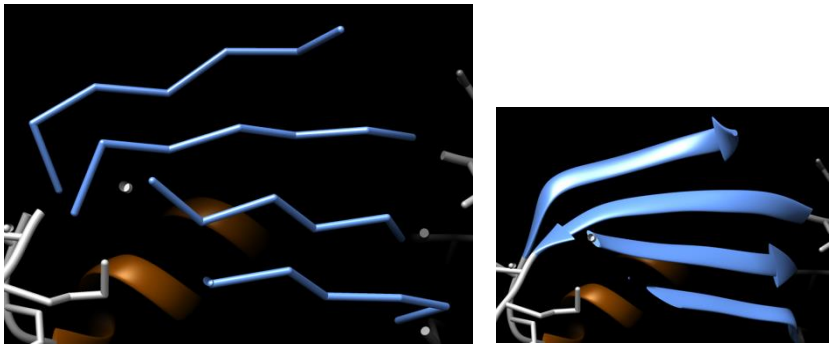
Color key (right panel)
Blue: nitrogen
Red: oxygen
Yellow: sulfur
Grey: carbon



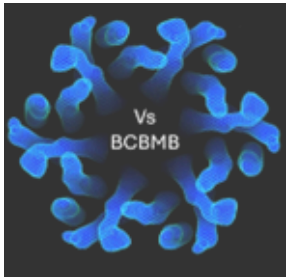
local conformation of just the backbone atoms (= secondary structure) tends to be regular most common



α–helix



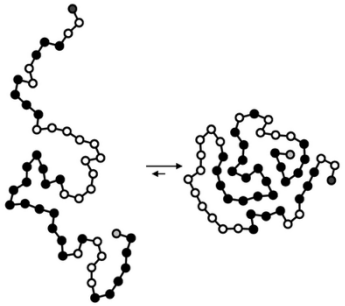
β–strand/sheet



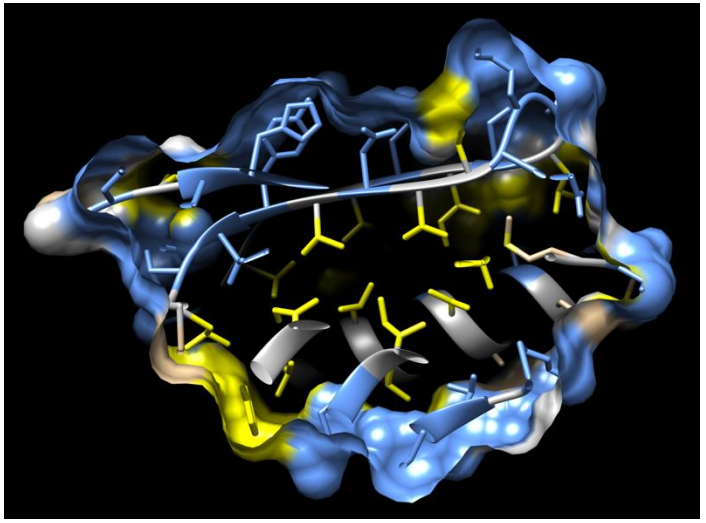
Tertiary Structure – Visual Recall

spontaneous folding of a polypeptide in the cytosol results in the formation of a **"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.

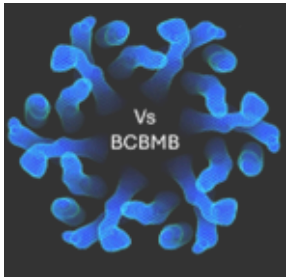


the complex (long range) interactions between sidechains stabilize the protein's overall structure = "tertiary structure"



reaching this point in the discussion/recall of protein folding, you **should** start feeling uneasy ….

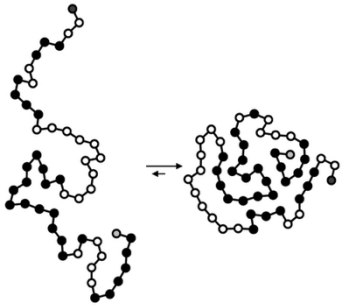
why?



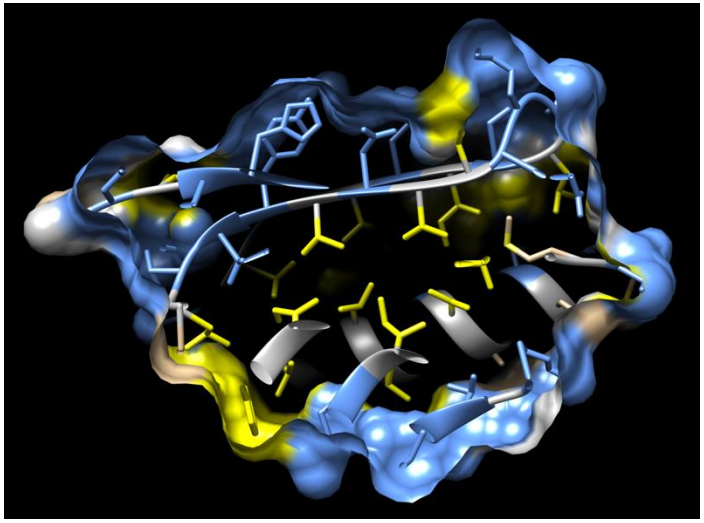
Tertiary Structure – Recall

spontaneous folding of a polypeptide in the cytosol results in the formation of a **"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.

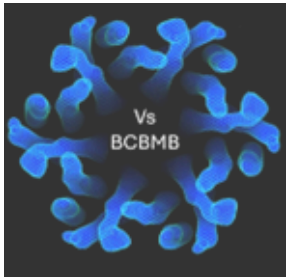


the complex (long range) interactions between sidechains stabilize the proteins overall structure or **"tertiary structure"**



reaching this point in the discussion/recall of protein folding, you **should** start feeling uneasy …. **why?**

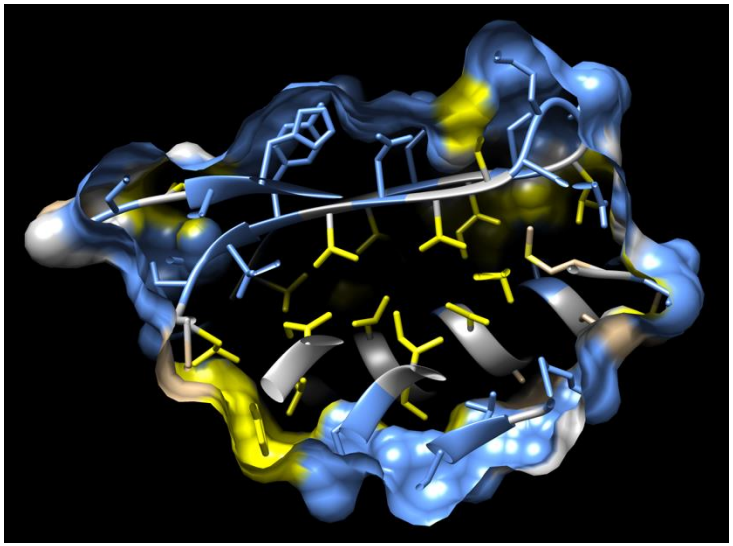
Answer: proteins are supposed help us to build pores that make membranes selectively permeable…..yet…if anything …what we have so far doesn't help at all …..why?



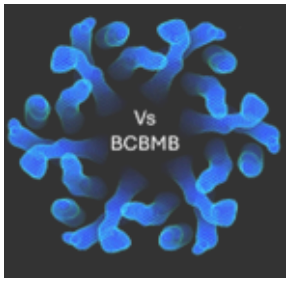
When Depth of Understanding Helps ….

Answer:

membrane interior is hydrophobic …if folding hides hydrophobicity in the core of the protein and aims to expose polar, water loving groups on the surface, then how could that ever be accommodated in a membrane?? Something is missing here…….

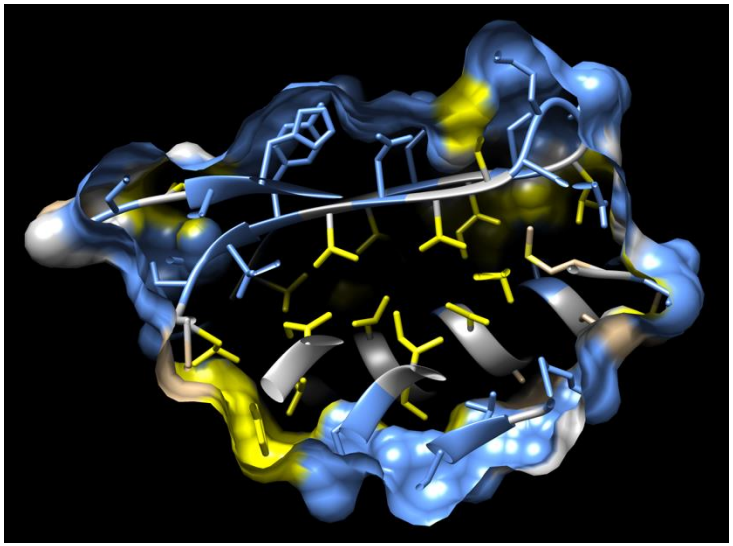


first thing to settle: if spontaneous folding buries hydrophobicity then formation of membrane embedded proteins ……. (*complete the sentence…be bold and/or creative!*)



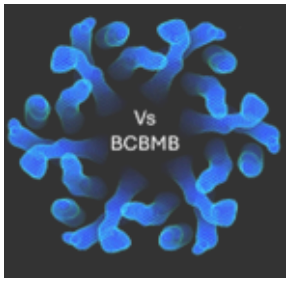
When Depth of Understanding Helps ….

Answer:
membrane interior is hydrophobic …if folding hides hydrophobicity in the core of the protein and aims to expose polar, water loving groups on the surface, then how could that ever be accommodated in a membrane?? Something is missing here…….



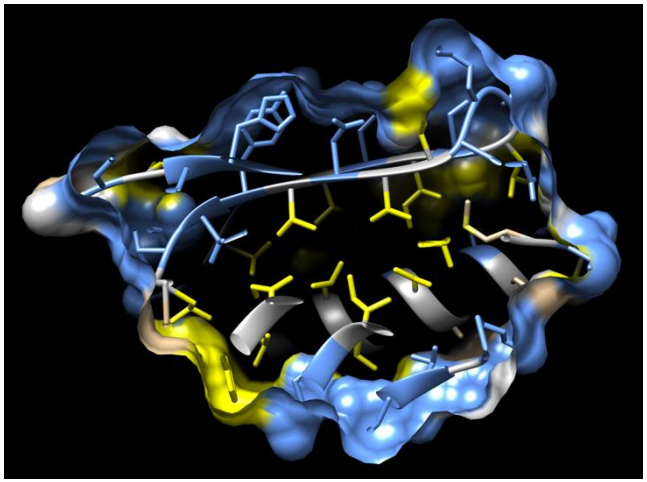
first thing to settle: if spontaneous folding buries hydrophobicity then formation of membrane embedded proteins ……. (*complete the sentence*)

…must be the opposite …. like an "umbrella…. flipping in strong wind"
= **ideally,** folding is manipulated in a way that somehow prevents sequestration of hydrophobicity, effectively turning protein structure **"inside-out"** = creating a **hydrophilic interior and hydrophobic surface……..**
….just thinking this should make you smile – WHY?



When Depth of Understanding Helps ….

Answer: membrane interior is hydrophobic …if folding hides hydrophobicity in the core of the protein and aims to expose polar, water loving groups on the surface, then how could that ever be accommodated in a membrane?? Something is missing here…….

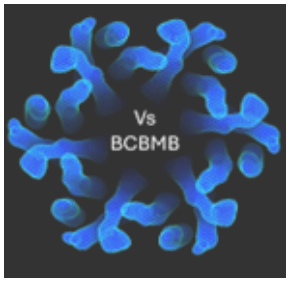


first thing to settle: if spontaneous folding buries hydrophobicity then formation of membrane embedded proteins ……. (complete the sentence)

mist be the opposite … like an "umbrella…. flipping in strong wind"
= **ideally,** folding is manipulated in a way that somehow prevents sequestration of hydrophobicity, effectively turning protein structure **"inside-out"** = creating a **hydrophilic interior and hydrophobic surface……..**
….just thinking this should make you smile – WHY?

because: this simple idea/conclusion/hypothesis/wish is **absolutely correct <u>AND</u>** would accomplish **EXACTLY what we need** … a structure that can help with moving a water-soluble molecule across the hydrophobic bilayer.

at a mechanistic level, this "inversion" is accomplished by co-translational membrane insertion (in most cases), and we will briefly look into it at the end of the chapter.
beyond the mechanistic differences in the synthesis of water soluble and membrane proteins, further understanding comes from taking a more in depth look at structural idiosyncrasies of polypeptides ….starting with …..



with the brief review of protein folding and
some new insights about protein folding out of
the way it's time to move on to

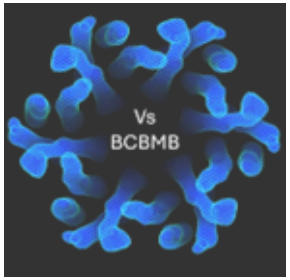
A Very Dense, and Terse Summary About Peptide Bonds and Secondary Structure Elements

before looking at some additional aspects of protein structure and the principles of membrane protein design (which we completely ignored in the Fundamentals Chapter)

as much as I love to break down things into easy to swallow pieces and to explain the "whys" …. this would make for a super long chapter that goes WAY beyond even an advanced biochemistry course.
if the particulars interest you - you want to take a specialized course on protein structure.

that said – I tried to locate and link helpful YouTube videos and to give you some rationale for why some amino acids prefer to be in helices over being in beta sheets.

some of the properties and stats summarized over the next few slides will come back into play when we think about how to make functional membrane proteins … so hang in there … you will understand some of the relevance of these facts a little later in the chapter.

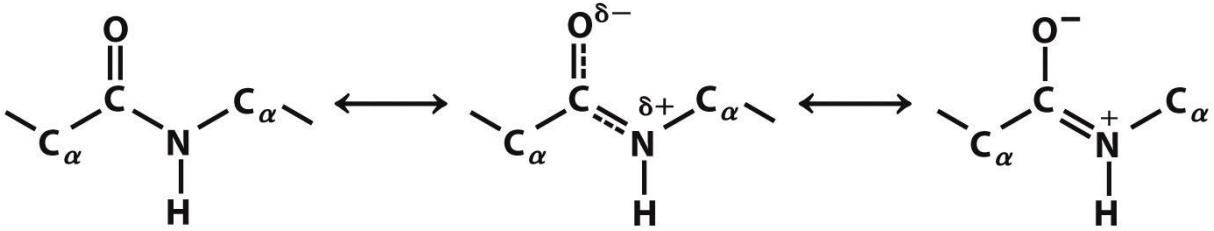


Peptide Bonds Are Very Stable Due to Partial Delocalization of the Nitrogen Free Electron Pair

Curious Trivia: what does "very stable" mean?

means: to chemically cleave a polypeptide you place it into a sealed vial containing at least 1M HCl and heat to 110˚C for at least 12 hours … that, by all means … is very stable and resilient to chemical attack by water ….

where does this stability come from??

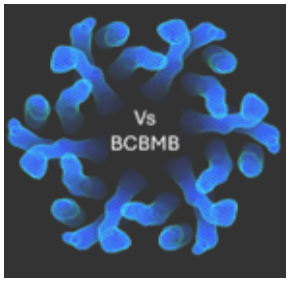


the carbonyl oxygen (CO) has a partial negative charge and the amide nitrogen (NH) a partial positive charge, setting up a **small electric dipole**.

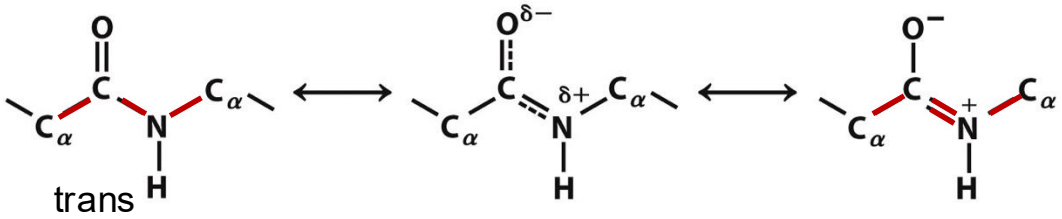
the small dipole **facilitates delocalization** of the free electron pair on the nitrogen atom over the entire peptide bond, giving it **partial double bond character**.

accordingly, all atoms participating in a peptide bond (C_α to C_α, O and N) localize to a single plane and are subject to "cis-trans" stereochemistry.

virtually all peptide bonds in proteins occur in this "trans configuration".

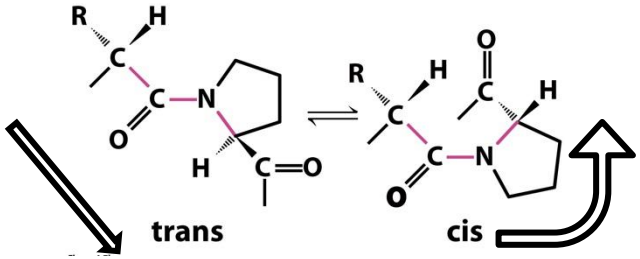


Peptide Bonds Are Very Stable Due to Partial Delocalization of the Nitrogen Free Electron Pair



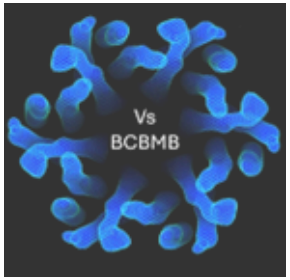
an exception involves Proline (Pro, P)
in this amino acid, the sidechain bonds back onto the invariant scaffolding unit that contains the backbone atoms.

in the context of the polypeptide chain rotation about the peptide bond allows for the "cis" conformation
→ **take note** how this "turns" the direction of the backbone around



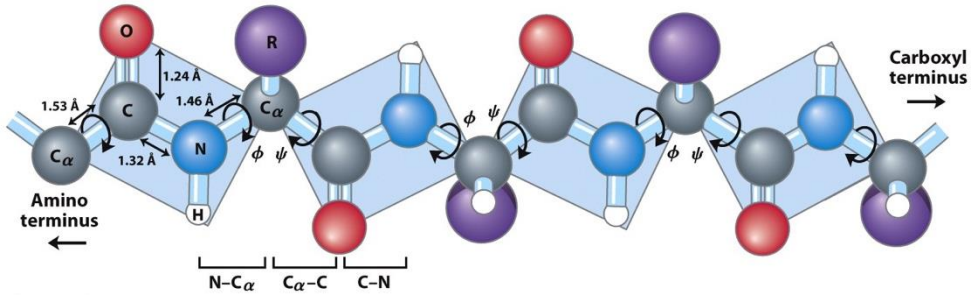
prevalent in turns

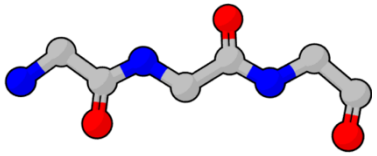
curious fact: this rotation does **NOT** occur spontaneously. It is catalyzed by an enzyme. This makes sense because changing proline geometry from trans to cis completely alters the direction of the polypeptide backbone ….which can wreck the proteins 3D structure (unless it is specifically designed to exploit this structural change….but even then, this change needs to be regulated to be useful)



Geometry of Peptide Bonds

yes .. this is painful …..
a really thorough step by step explanation/animation of this can be found here
https://www.youtube.com/watch?v=JyUMLSsbecl





The Take Home Messages Are:

exploiting the planar geometry of the peptide bond (indicated by blue rectangles), you can describe the geometry/progression of the polypeptide by just two angles that are called: **dihedral angles ϕ (phi) ψ (psi)**

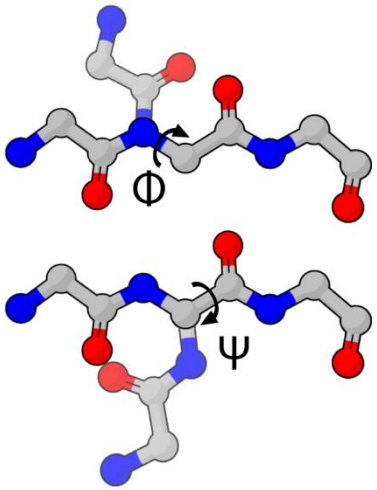
they are determined by evaluating the angle between the CO or NH groups of adjacent amino acids as it appears if you look down the C_α–NH bond (ϕ), or the C_α–CO bond (ψ) □ watching the video linked above REALLY helps here.

a third angle, omega (ω), is the angle of the CO-NH bond itself. It is fixed at 180˚ because of the trans conformation of the peptide bond (exception for Pro if it is cis)

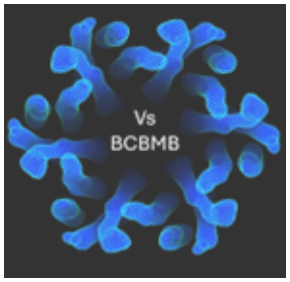
ϕ,ψ in principle can adopt any value between -180˚ and + 180˚ , but in reality are more constrained due to steric hindrance (e.g. ϕ,ψ both 0˚ is forbidden)

why do we introduce this? → because it is the basis of understanding secondary structures = the fact that backbone conformation are not random

… (as we saw during the ATX1 tour in the Fundamentals Chapter)



These panels give you another illustration of what a 180˚ change means for each of the **ϕ (phi) ψ (psi)** dihedral angles ..

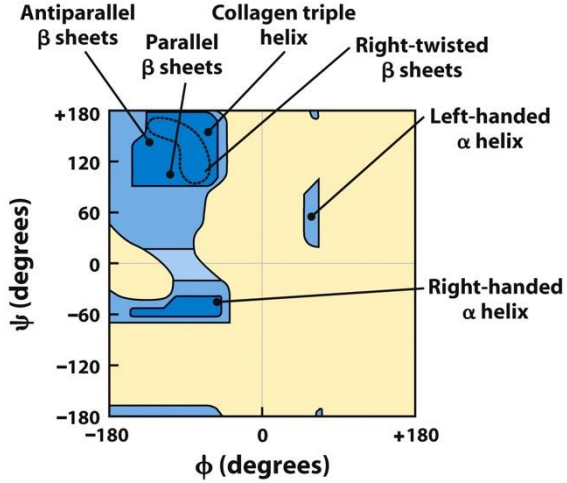


The Ramachandran Plot Describes the ϕ,ψ-Space Explored By Polypetide Chains

G.N. Ramachandran (Indian Biophysicist, 1922-2001)

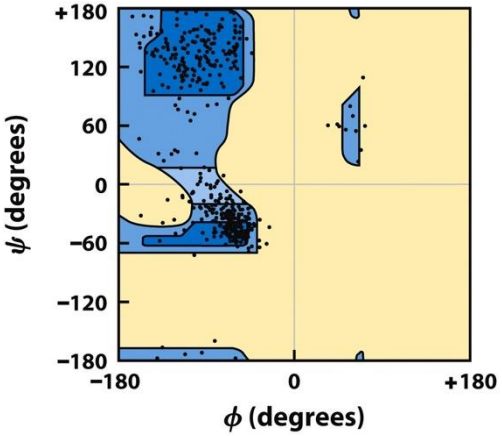
along with acknowledging the existence of dihedral angles … the question became:
are all angular combinations possible?
driven by curiosity, Ramachandran - for each possible combination of angles - plotted if they were possible without causing steric clashes between sidechains or mainchain atoms. Doing this, he found that there are distinct regions of favorable angle combinations and some regions that just cannot happen.

Theory



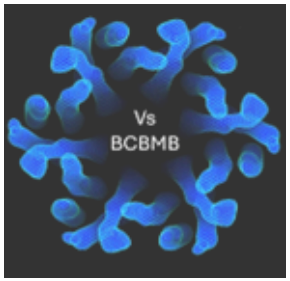
blue regions are "allowed" = no steric hindrance
yellow regions are "forbidden" = steric hindrance

Practice
(pyruvate kinase)



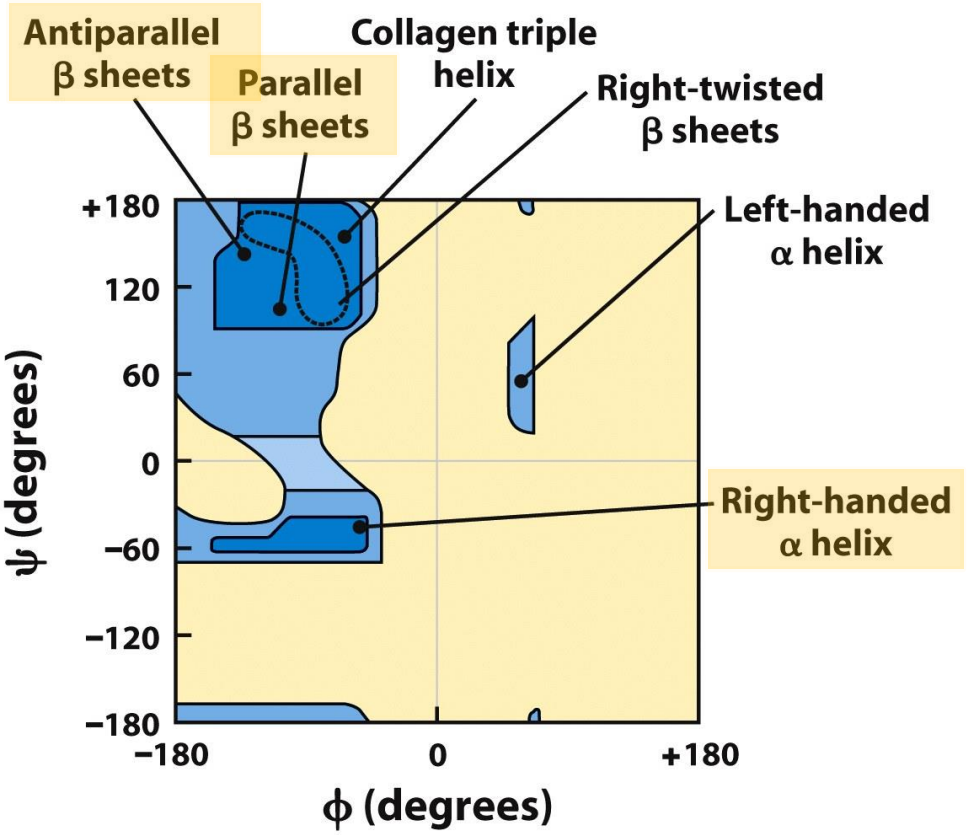
unless … the sidechains found in disallowed regions are G (sidechain = H-atom = Gly can contort itself into impossible conformations)

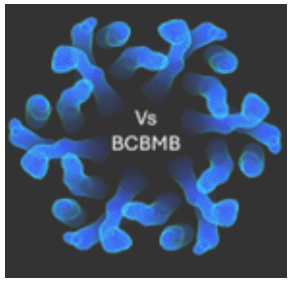
to this day – experimental structures have to pass this "geometry check" during model validation …. if the model contains too many angular value combinations outside the allowed regions, the model is deemed invalid



The Ramachandran Plot Describes the ϕ,ψ-Space Explored By Polypetide Chains

let's take a closer look at the three most common secondary structure elements (orange underlay)….

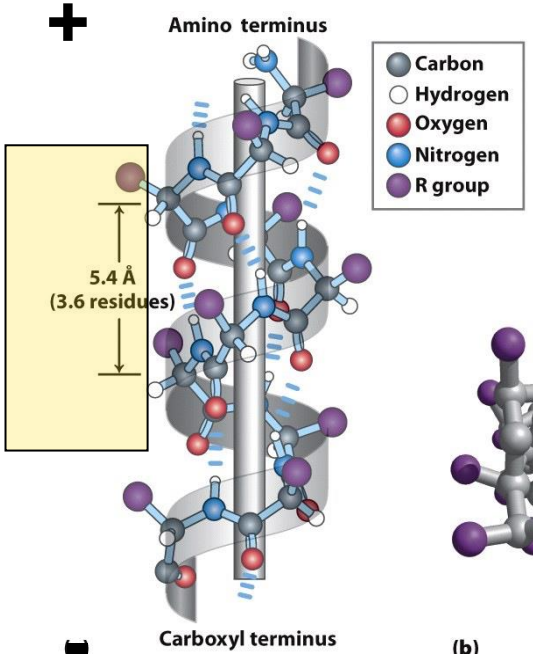


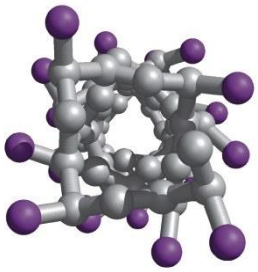


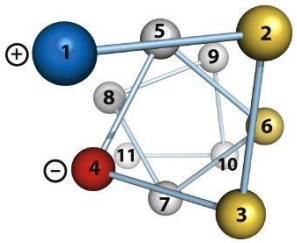
Alpha-Helix – The Stats….

ϕ= -57˚, ψ=-47˚

- right handed twist of backbone
- **hydrogen bonding** (dashed blue lines) **form parallel to helical axis** (i_{CO}→ $i+4_{NH}$) stabilizing the conformation
- helices are weak dipoles (N-terminal +; C-terminal -)







P and G "break the helix", other amino acids have various tendencies to form helices

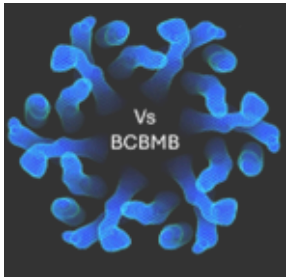
A,R,K,L,M,Q,E are most favorable

Why those? Answer: except A (which has a small sidechain, just a –CH3 group), they all have "long" and relatively "skinny" sidechains that point away from the helix backbone and therefore do not interfere with it, or each other.

coloring refers to hypothetical residues that are charged (blue, red) or nonpolar (yellow); **don't pay attention to the particular choices**…....**but**

- **be aware** of how residues on a helix are correlated! eg 1,4,8,11 or 1,5,8,12
- general: ZxxZxxxZxxZ = sidechains of **every third and fourth residue (Z) point in the same direction** (= compartmentalization!) .. this will become really important for designing membrane proteins!

knowing this, except for dihedral angles, is helpful to understand a great many things about proteins and what they do

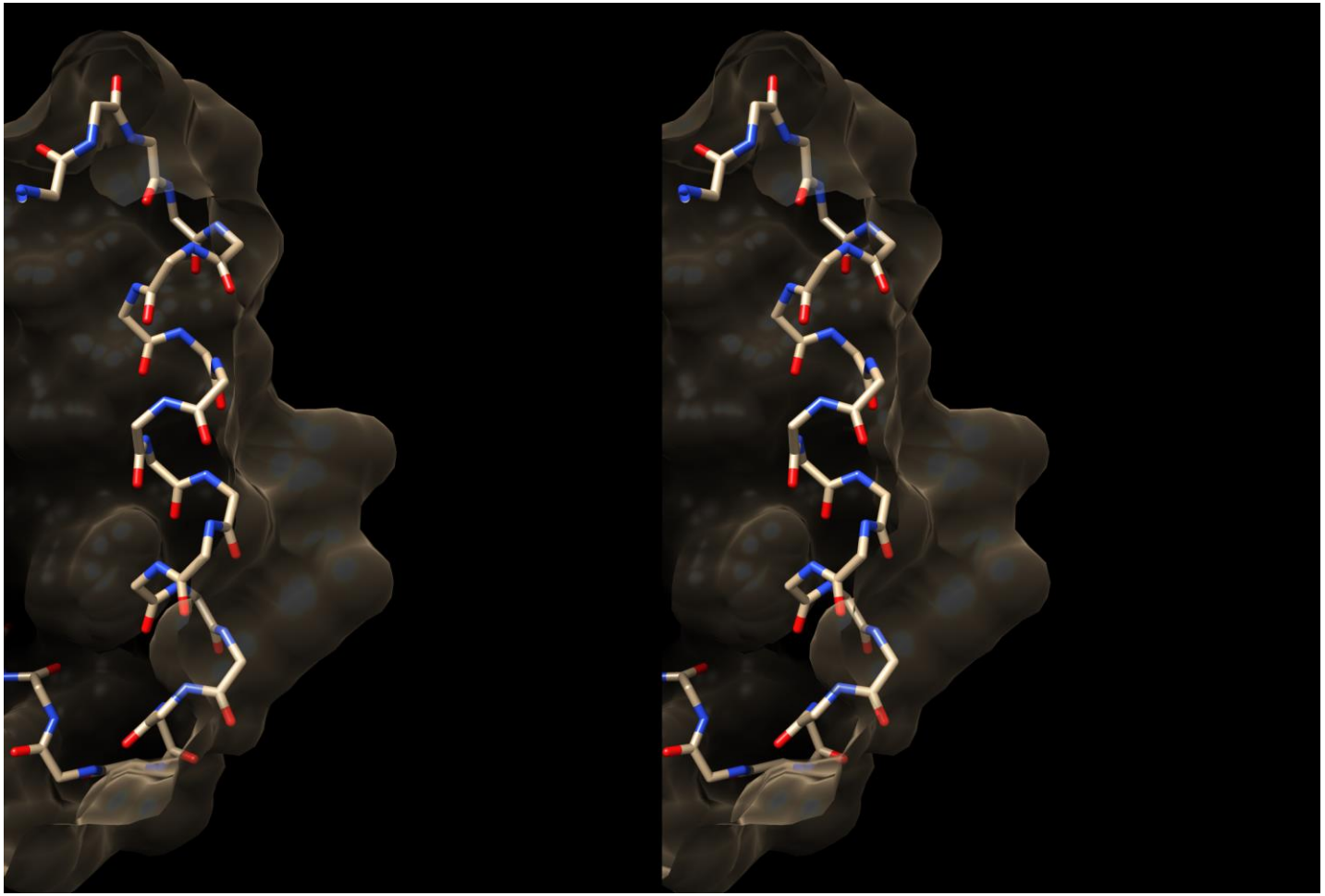


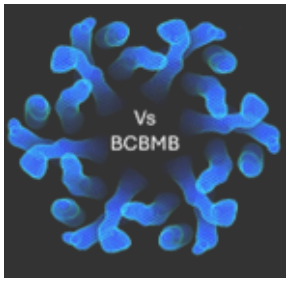
Alpha-Helix – The Views….

instead of cartoons, here are cross-eyed stereo views of a real alpha helical backbone structure (blue=N, red =O)

(taken from the small copper chaperone ATX1 (the full "tour" is linked…consider watching it if you have not worked through the Fundamentals – PROTEINS chapter)

to get the stereo effect – hold images ~30cm away from you, look at them and cross + adjust your eyes until you get the 3D depth (once reached, this is very comfortable and does not hurt your eyes at all)



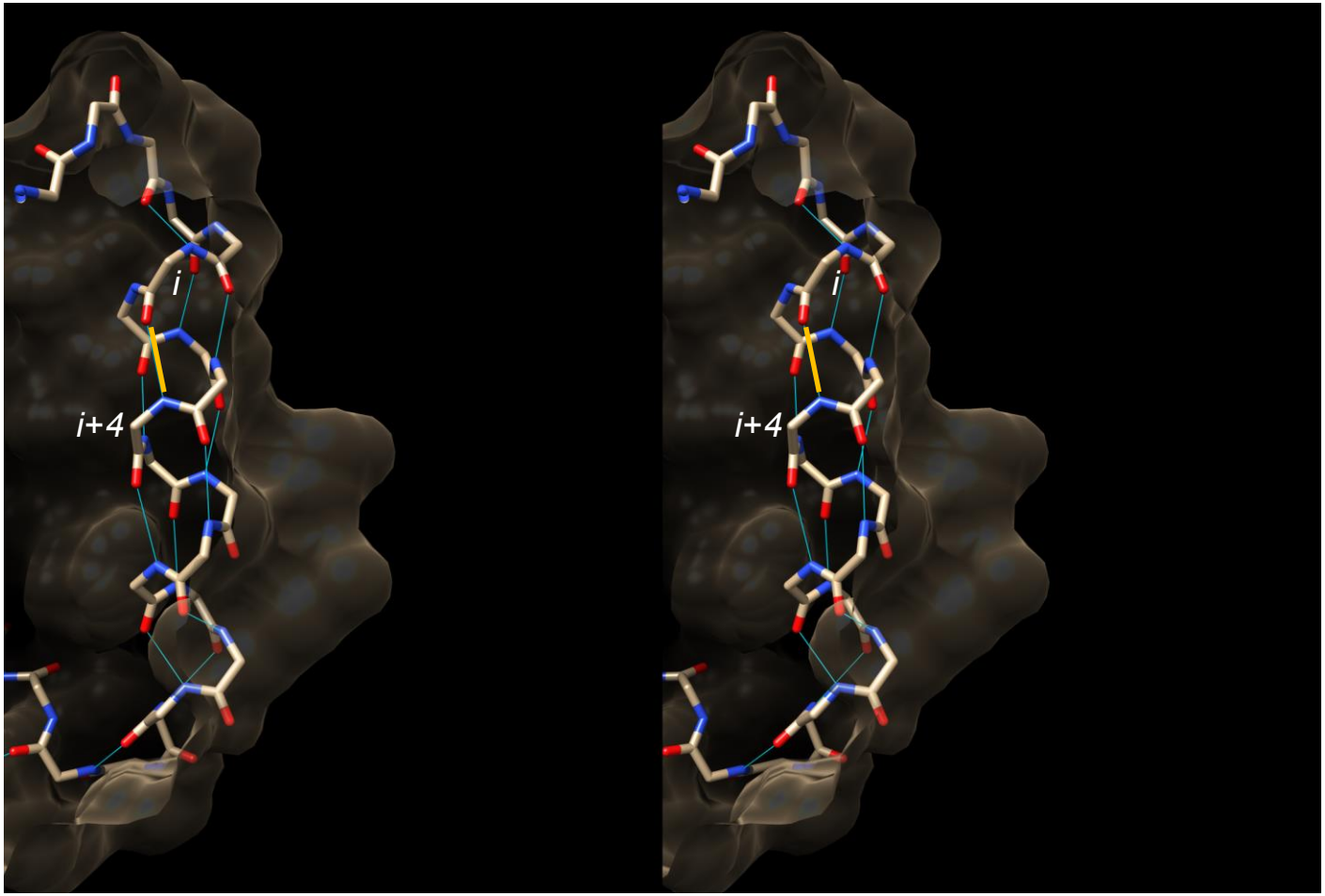


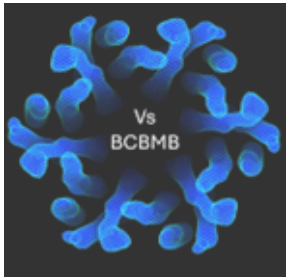
Alpha-Helix – The Views….

the same helix, but this time, the H-bonding pattern between the backbone –CO- and –NH- groups is shown

note how the H-bonds are aligned with the helix axis!

the H-bond highlighted in orange illustrates how a –CO- group of residue "i" - H-bonds to the –NH- group of the residue "i+4"





Beta-Sheets

composed of <u>at least</u> 3 β-strands

all amino acids except for P can be part of a β–strand
nost favorable: V,I,F,Y,W,T … all relatively bulky ... filling the space above and below the plane defined by the backbone atoms

ϕ= -139˚, ψ=+135˚

Antiparallel

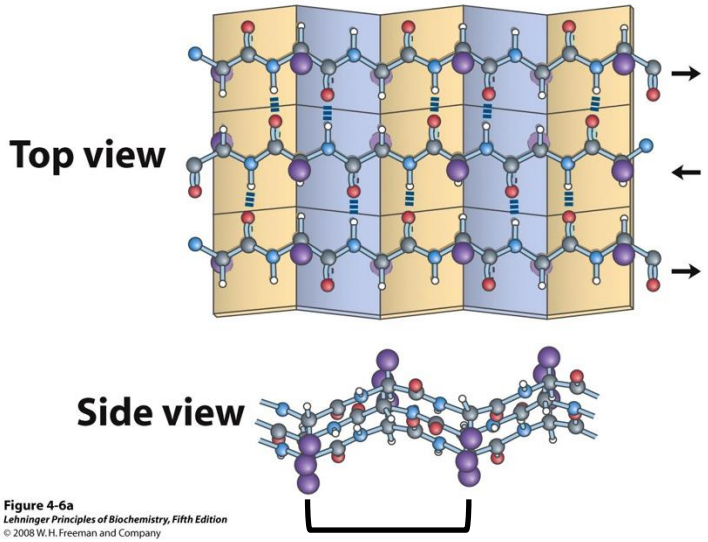


Figure 4-6a
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

2 residues, 7Å

ϕ= -119˚, ψ=+113˚

Parallel

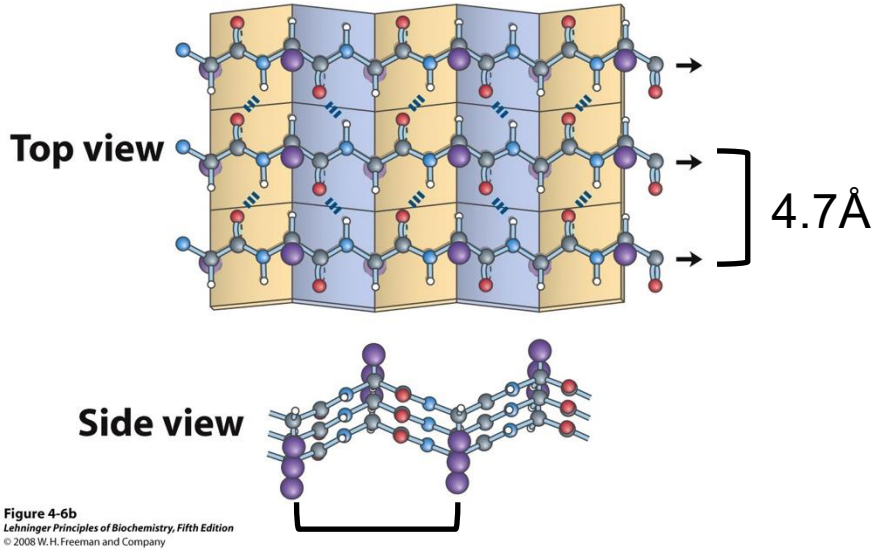
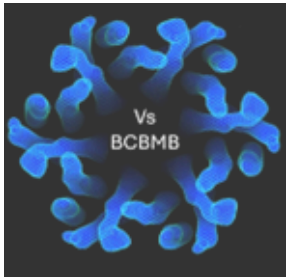


Figure 4-6b
Lehninger Principles of Biochemistry, Fifth Edition
© 2008 W.H. Freeman and Company

<u>2 residues</u>, 6.5Å

note: sidechains (represented by purple spheres) project above and below the sheet
H-bonds are formed BETWEEN strands within the sheet….
H-bonding pattern is different for parallel-antiparallel sheets → while sheets with >3 strands can be "mixed" (=have parallel and antiparallel regions), mixing of the types is less stable than just one consistent type

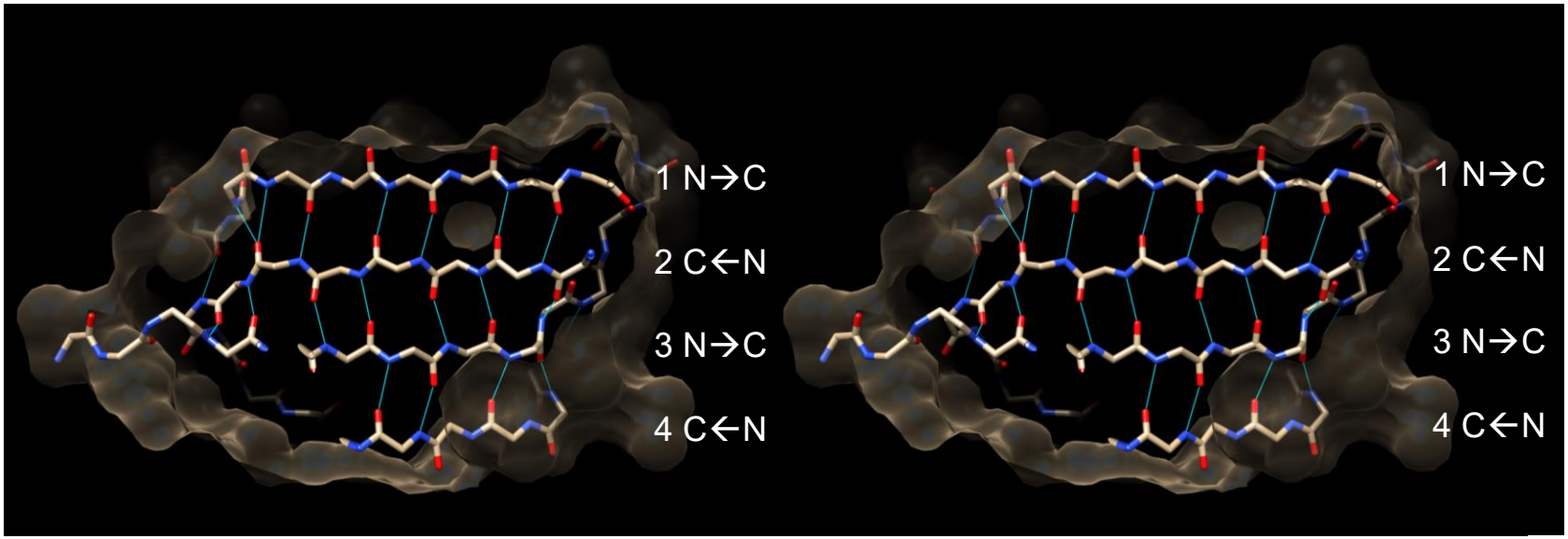
knowing this, except for dihedral angles, is helpful to understand a great many things about proteins and what they do

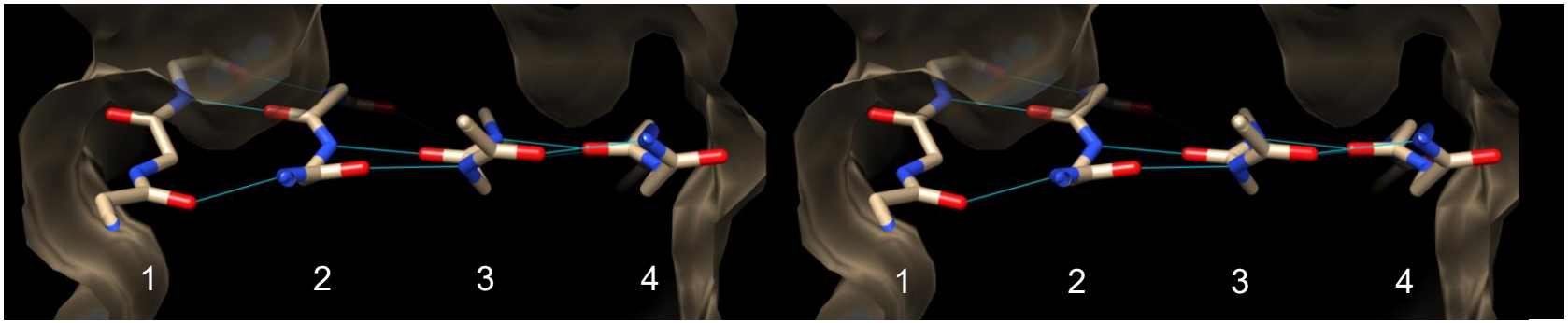


Beta Sheet – The Views….

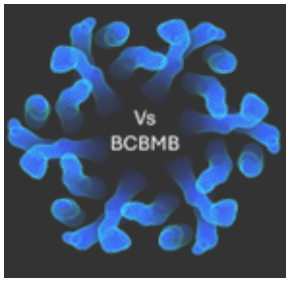
Antiparallel β-sheet in ATX1
Shown with H-bonds

perpendicular to plane of sheet





parallel to plane of sheet



Completing the Basic Inventory of Secondary Structures

β–Turns

(note … **without means to turn ….** polypeptides would just be very long, skinny noodles ….not very useful for anything much ….)

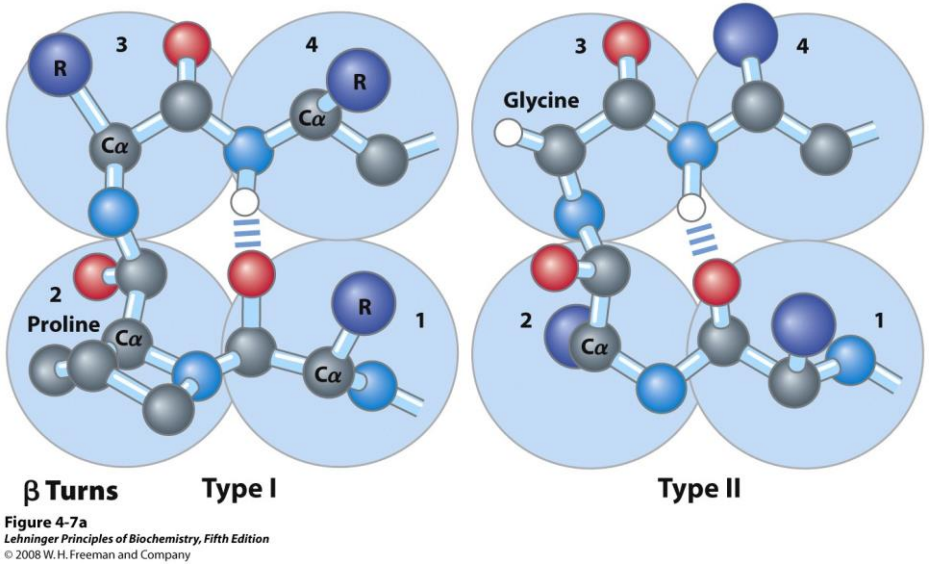


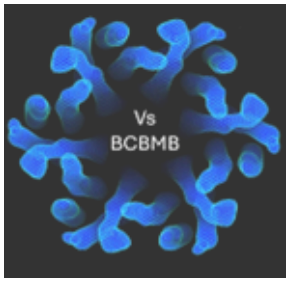
Figure 4-7a
Lehninger Principles of Biochemistry, Fifth Edition

Compare:
in RNA a minimum of 3-4bp are needed to turn a single chain around
= very similar to proteins

often contain P and G

note: flexible, disordered stretches of 4 or more residues can replace β-turns

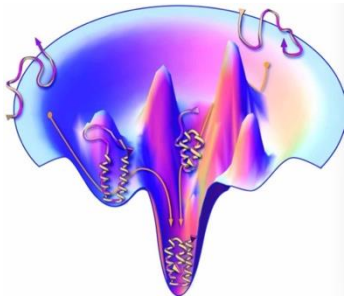
knowing this helps for understanding protein structure and function



Fold Stabilization

with the "hard (dry and boring) stuff" out of the way – let's explore folding and structural properties in a little more depth than in the Fundamentals Chapter.

will a polypeptide always fold into the same native structure?



Answer: under normal circumstances: yes.
to get a better sense of what exactly that means and looks like, here are two computer simulations of how two small proteins fold …. you **really should watch those … as they are super insightful**

collection of movies:
http://www.ks.uiuc.edu/Gallery/Movies/

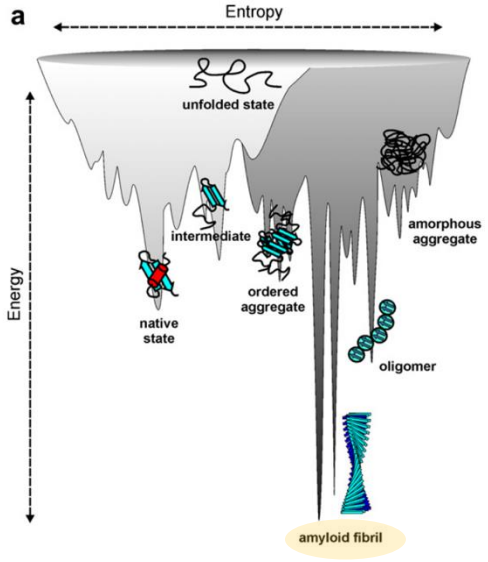
Movie 1 to watch: http://www.youtube.com/watch?v=sD6vyfTtE4U&feature=youtu.be

Movie 2 to watch: https://www.youtube.com/watch?v=XsQgjxMDjNw

although – misfolding does occur:

→ most extreme: **amyloid formation** (eg in Alzheimers)

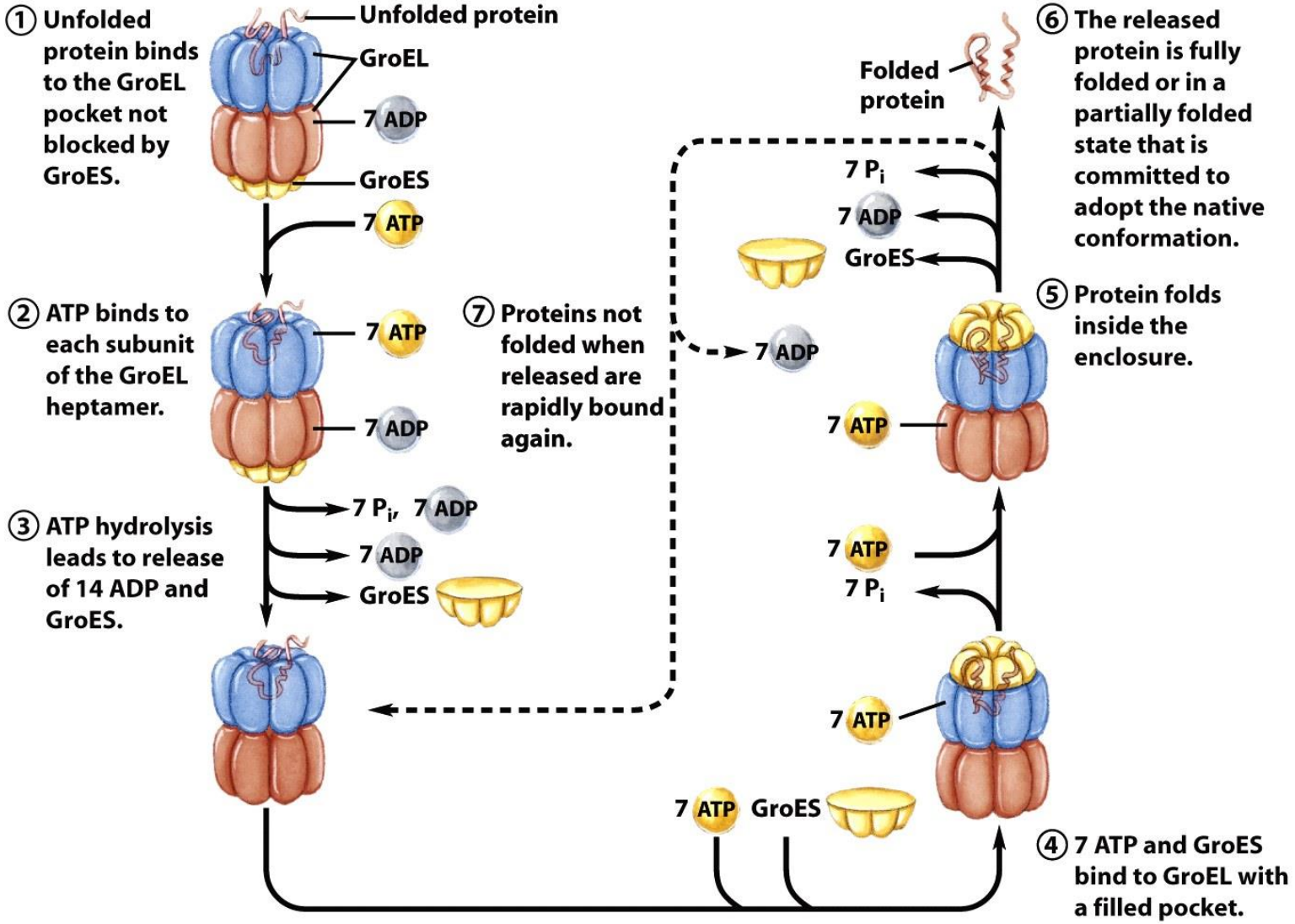
- ➢ far more prevalent: chain gets stuck on local minimum (10-15% in E coli).
- ➢ molecular folding chaperones allow arrested chains to escape from false local minima
- ➢ rescue from local minima requires ATP (next slide)



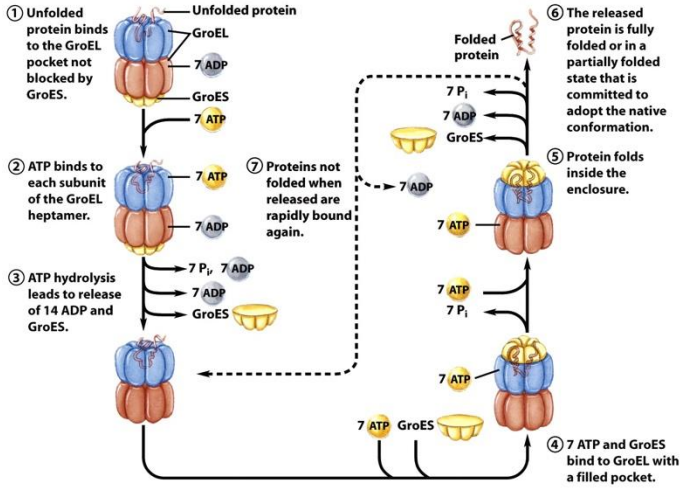
Jahn TR, and Radford SE (2008)

E coli's GroEL/GroES is the Most Famous Folding Chaperone

(Comments: Next Slide)



E coli's GroEL/GroES is the Most Famous Folding Chaperone



note: chaperones burn through a lot of energy! If you watched "movie 1" you will have noticed the moment when the short chain gets stuck. while in this case the chain can resolve the issue itself over time, sometimes chains needs help. folding chaperones provide that help by immobilizing the chain, then pull on it to undo badly formed contacts before disengaging to let the chain "do its thing" ….. that moment of "pulling" is what costs the energy

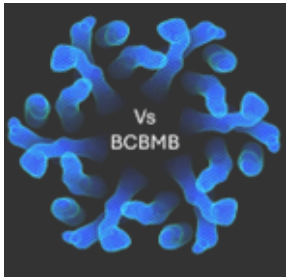
expending this energy is not wasteful but making proteins that end up partially folded is.
spending a handful of ATP molecules to rescue the is WAY more energy efficient than degrading the chain and start from scratch. helping the protein fold also avoids accumulation of "bad chains" that do nothing and/or start aggregating non-specifically

A curious question to ask: how does a chaperone "know" that a protein chain needs help?

Answer: if chain folding – **of a water soluble, "globular" protein -** is arrested prematurely or is at an early stage, then the surface of the protein will present too many hydrophobic patches …. these are the bait that catches the chaperone's interest and causes it to engage. If successful, then these "patches" will reduce in size or disappear, signaling that folding has now succeeded (no further help required).

even if the chain could succeed without the chaperone … engagement with a chaperone at an early stage of folding (where a lot of hydrophobic residues are still exposed at the surface) will just speed up the "search process" for finding the right polypeptide backbone conformation = no harm done.

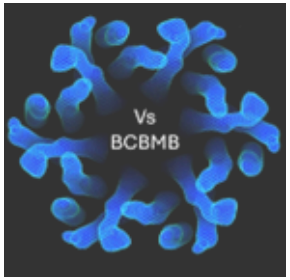
lastly: even if a hydrophobic patch on the surface is the right conformation for this protein (because it may need that patch to function properly) … engaging with a chaperone … getting a little tuck and go back to exactly that conformation after release …..does no "harm". Other than costing a few ATP – this is a "better save than sorry" approach, counting on the fact that the chaperone will "go away" before taking another "bite"



Moving on to a **REALLY IMPORTANT** Question

what is the overall stabilization energy of the native fold?

I know that most of you will "hate" this question but …try … what do you think? …..



Moving on to a **REALLY IMPORTANT** Question
what is the overall stabilization energy of the native fold?

I know that most of you will "hate" this question but …try … what do you think?

you may "hate" this question because you may have no idea what the stabilization energy might be, or you may be completely lost, not knowing what this question is even asking ……and that is OK … most instructors or textbooks will not go there, but we will because once you do understand this, you will **understand** why proteins can do so many things

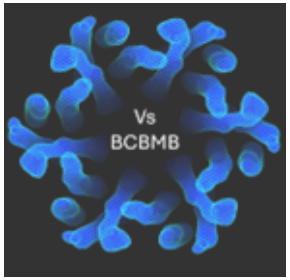
let's start thinking about protein folding again … slides 20-23 ….
take away: folding constrains the backbone atoms….like putting them into a straight jacket …
ask yourself: would you voluntarily put on a straight jacket? Unlikely!

→ folding = restraining the backbone **reduces entropy = costs energy!**
→ That energy comes from forming H-bonds (to stabilize secondary and tertiary structures), ionic interactions and Van der Waals interactions

➔ **what the question about "overall stabilization energy" asks is this**: let's assume that folding the chain costs you "x-amount of $" …. and
as we saw, folding is gradual = you "pay as you go" …

….when you are completely done folding, by how much have you "overpaid" the folding energy cost to make sure this goes all the way **AND** stays there … (only the fully folded state is native = functional)

Answer: Surprisingly small at 10kcal/mol
This is the energy content of only ~2-3 hydrogen bonds (out of hundreds or thousands that are present in the total structure)



Moving on to a **REALLY IMPORTANT** Question
what is the overall stabilization energy of the native fold?

Answer: Surprisingly small at 10kcal/mol

This is the energy content of only ~2-3 hydrogen bonds (out of hundreds or thousands that are present in the total structure)

why making a "fuss" about this?

because the **immediate implication** of this is that the functional protein fold is only marginally stable … **constantly undergoing small changes that undo little bits and pieces here and there** ….

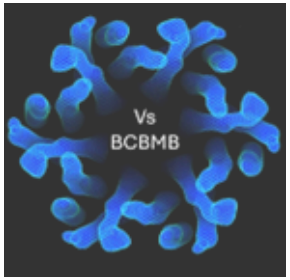
…you could think of it as the whole structure "wiggling/breathing" … constantly.

….in fact, it is so delicate that if you disturb this even just a little, it will "fall apart"

And that is FANTASTIC!!

WHY???!???!

…try to think of an answer …. (**hint:** what do proteins do again … ?)



Moving on to a **REALLY IMPORTANT** Question

What is the overall stabilization energy of the native fold?

Answer: Surprisingly small at 10kcal/mol

this is the energy content of only ~2 hydrogen bonds (out of hundreds or thousands that are present in the total structure)

why making some "fuss" about this?

because the **immediate implication** of this is that the functional protein fold is only marginally stable … **constantly undergoing small changes that undo little bits and pieces here and there** ….

…you could think of it as the whole structure "wiggling/breathing" … constantly.

….in fact, it is so delicate that if you disturb this even just a little, it will "fall apart"

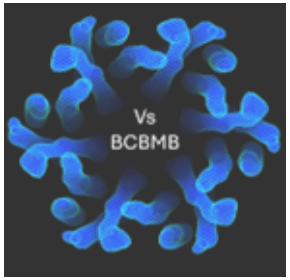
And that is FANTASTIC!!

WHY???!???!

Answer: …because protein function, in most cases, involves changes in 3˚ structure → the structure needs to be "pliable" = <u>MUST NOT</u> be stabilized so much that it turns rigid.

(to be discussed much more in later chapters)

Does this remind you of something….?



Moving on to a **REALLY IMPORTANT** Question

What is the overall stabilization energy of the native fold?

because the **immediate implication** of this is that the functional protein fold is only marginally stable … **constantly undergoing small changes that undo little bits and pieces here and there** ….

…you could think of it as the whole structure "wiggling/breathing" … constantly.

….in fact, it is so delicate that if you disturb this even just a little, it will "fall apart"

And that is FANTASTIC!!

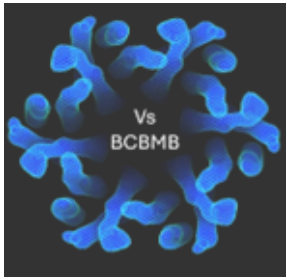
WHY???!???!

Answer: …because protein function, in most cases, involves changes in 3˚ structure → the structure needs to be "pliable" = <u>MUST NOT</u> be stabilized so much that it turns rigid.
(to be discussed much more in later chapters)

Does this remind you of something….?

Hopefully it does …. and hopefully it is MEMBRANES!

In the Advanced Lipids and Membranes Chapter you learned that biological membranes are metastable and constantly "frustrated" (slide 64 &65) and that this property is essential for membranes being able to carry out their different biological functions.



just in case you are (still) trying to get a sense
for what **10kcal/mol** actually means

…here are a few comparisons…

10kcal/mol is: 0.00000000000000000167 cal **per molecule**
(curiously: hydrolysis of 1 ATP molecule yields similar energy @ 0.00000000000000000225 cal)

your daily caloric need is what? ~2000 - 2500 kcal (more if you are an athlete)
= 2,000,000 – 2,500,000 cal ….. **just for you!**

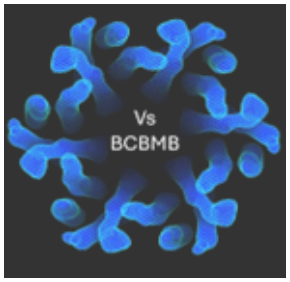
how many calories has a 100g chocolate bar? (hello "midterm stress")

~600,000 cal or (= ~1/3 to 1/4 of your daily recommended calorie intake)

and bringing it back to proteins:

If a 40kDa polypeptide's fold is stabilized by 10kcal/mol, then this corresponds to
~25cal/100g, which would just be enough to warm 25ml of water by 1˚C

compared to
80,000 cal to bring a liter of water to boil from 20˚C (room temperature)



Equally Surprising/Amazing:

the small overall free energy gain associated with the native structure allows folding to occur spontaneously, reproducibly and within a reasonable time frame:

Levinthal's paradox:
assume 100 amino acids, 10 conformations/residue

→ 10^{100} conformations

sample 1 conformation per 10^{-13}s

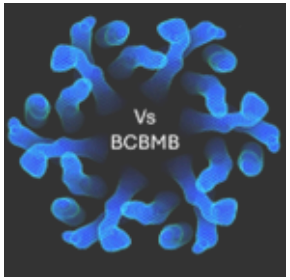
How long would it take to fold if the protein had to explore every possible combination of angles?

→10^{87}s or roughly **10^{80} years** (Earth is 4.5×10^{9}yrs old)

in the cell, such a protein folds within a few seconds!
how is that possible??

Short Answer: the protein chain does not wait to fold until all of it has been synthesized by the ribosome. instead: it begins folding as soon as enough has emerged from the ribosome to allow exploration of different secondary and tertiary structures.

this link is to a short video (less than a minute) that illustrates this really well...you definitely want to watch this …. and if you do … what does that remind you of? (3D-printing ….)

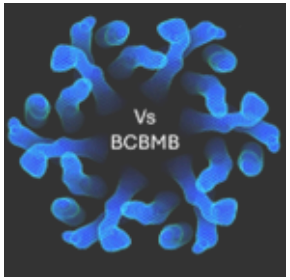


…and on the matter of folds ….

given the complexity of the folding process: **is every fold unique?**

…what do you think?

if you worked through the Fundamentals – PROTEINS chapter …you may remember the answer that, back then, we took for granted ….



…and on the matter of folds ….

given the complexity of the folding process: **is every fold unique?**

…what do you think?
if you worked through the Fundamentals – PROTEINS chapter …you may remember the answer.

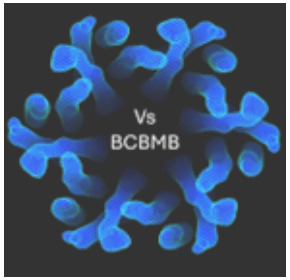
Answer: no – folds are redundant = many different primary structures can adopt the same fold
….why??

Answer:
in the "Fundamentals" chapter you probably just took it for granted when you were told that this has to do with the fact that the backbone prefers certain conformations (helices, strands, turns)

now that we looked at these secondary structures in more detail, you can better appreciate why that answer was true. Looking through slides 36-41 again… you realize that the driving factor for secondary structure formation is the ability to support certain patterns of hydrogen bonding ….(parallel to helix axis …or…parallel to the plane of β–sheet, between its component β–strands)

in other words: whenever you have a short run of amino acids in the primary structure (sequence) that prefer to be in a helical conformation, a helix will form.
Same is true for beta strands if you use amino acids that prefer strand conformation.

not ALL amino acids in that short section of the primary structure have to prefer the same secondary structure element though … what ultimately happens is determined by what the majority of residues in a given sequence window prefer to do … and it also explains why the overall compromise is **only** marginally stable



…and on the matter of folds ….

in the "Fundamentals" chapter you probably just took it for granted when you were told that this has to do with the fact that the backbone prefers certain conformations (helices, strands, turns)

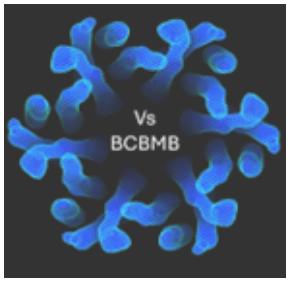
now that we looked at these secondary structures in more detail, you can better appreciate why that answer was true. Looking through slides 34-39 again… you realize that the driving factor for secondary structure formation is the ability to support certain patterns of hydrogen bonding ….(parallel to helix axis …or…parallel to the plane of β–sheet, between its component β–strands)

in other words: whenever you have a (short) run of amino acids in the primary structure (sequence) that prefer to be in a helical conformation, a helix will form.
Same is true for β–strands if you use amino acids that prefer that conformation.

not ALL amino acids in that short section of the primary structure have to prefer the same secondary structure element though… what ultimately happens is determined by what the majority of residues in a given sequence window prefer to do.

that is: **generic backbone conformations create structures that serve as a "blank" canvas. Decoration of that canvas with sidechains determines long-range interactions between the secondary structure elements, which determines the tertiary structure and through that, the function of the protein.**

if you prefer a simple intuitive argument: 20 amino acids → protein made from 200 amino acids could have 20^{200} different primary structures → strictly speaking yields 20^{200} different tertiary structures (= overall 3D-structures), **BUT** since backbone conformations are generic, the backbone of the hypothetical protein will not present in 20^{200} variants …and it doesn't have to because – for instance - sidechains that are on the surface have little constraint … changing those will not make much (if any) difference



Fold Degeneracy – A Case Study (Copper Chaperones ATX vs CCS)

both proteins bind to copper ions but deliver them to different cellular targets. The amino acids directly involved in forming the copper binding site are highlighted orange (more specifically, it is the two "C" = cysteine sidechains that bind the copper ion

In the Fundamentals chapter, we had used ATX1 for our initial exploration of protein structure

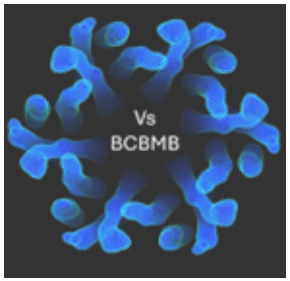
```
ATX     MAEIKHYQ--FNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFILE 58
CCS     MTTNDTYEATYAIPMHCENCVNDIKACLKNV-PGINSLNFDIEQQIMSVESSVAPSTIIN 59
        *:  . *:  : : * *..* . ::  *.:: *.:..:::.:*:*::.* :::. . *::

ATX     KIKKTGKE--VR-SGKQ-------L---------------------------------- 73
CCS     TLRNCGKDAIIRGAGKPNSSAVAILETFQKYTIDQKKDTAVRGLARIVQVGENKTLFDIT 119
        .::: **:  :* :**         *
```

→ **VERY little sequence homology** (only 16 out of 73 residues are identical "*"; but more extensive similarities ":" that refer to sidechain properties (eg both chains have hydrophobic residue, or hydrogen bonding potential)

YET, the overall structure (a pair of α–helices packing against an antiparallel β–sheet) looks (and is) almost identical

intuitively this makes sense ... both proteins bind copper …and if structure determines function, then there should be a structural similarity/identity that encodes "copper binding"

➔ whatever the differences in sequence, they either do not matter …**or** are important for the functional difference (=delivery to different cellular targets)



Fold Degeneracy – A Case Study (Copper Chaperones ATX vs CCS)

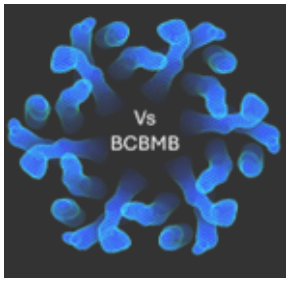
```
ATX  MAEIKHYQ--FNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFILE 58
CCS  MTTNDTYEATYAIRMHCENCVNDIKACLKNV-PGINSLNFDIEQQIMSVESSVAPSTIIN 59
     *:  . *:  : : * *..* . ::  *.:: *.:..:::.:*:*::.* :::. . *::

ATX  KIKKTGKE--VR-SGKQ-------L---------------------------------- 73
CCS  TLRNCGKDAIIRGAGKPNSSAVAILETFQKYTIDQKKDTAVRGLARIVQVGENKTLFDIT 119
     .::: **:  :* :**        *
```

how much variance is permissible?

30% identity in 1˚ structure (sequence): **guaranteed** to have same fold **if identities are scattered throughout sequence**

even at 10% identity strong fold similarity may prevail; in these cases, it is just not obvious based on sequence

take a second to let the implications of this sink in……
…what ARE the implications ….?



Fold Degeneracy – A Case Study (Copper Chaperones ATX vs CCS)

```
ATX   MAEIKHYQ--FNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFILE 58
CCS   MTTNDTYEATYAIPMHCENCVNDIKACLKNV-PGINSLNFDIEQQIMSVESSVAPSTIIN 59
      *:  . *:  : : * *..* . ::  *.:: *.:..:::.:*:*::.* :::. . *::

ATX   KIKKTGKE--VR-SGKQ-------L---------------------------------- 73
CCS   TLRNCGKDAIIRGAGKPNSSAVAILETFQKYTIDQKKDTAVRGLARIVQVGENKTLFDIT 119
      .::: **:  :* :**        *
```

how much variance is permissible?

30% identity in 1˚ structure (sequence): **guaranteed** to have same fold **if identities are scattered throughout sequence**

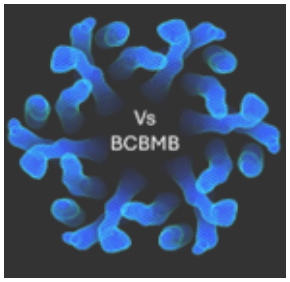
take a second to let the implications of this sink in……
…what ARE the implications ….?

the main take away point is that - despite the complexity of protein folding itself - protein structure (even though it is only marginally stabilized in the end) is **INCREDIBLY robust** ….and resistant to changes in the overall fold

….if you think about this … you realize just how important that was to "evolve" proteins at a time when mutations were going unchecked because DNA repair mechanisms were not fully perfected

Once a "solution to a functional need" was found, it would stay and not be lost because many/most mutations can be tolerated without affecting the structure/function.

Similarly, once you had a solution to a functional need, you could start "playing" with it to see if you can tweak it to do more, or do the same thing…but on a different set of target molecules …(redundancy)



Fold Degeneracy – A Case Study (Copper Chaperones ATX vs CCS)

```
ATX  MAEIKHYQ--FNVVMTCSGCSGAVNKVLTKLEPDVSKIDISLEKQLVDVYTTLPYDFILE 58
CCS  MTTNDTYEATYAIRMHCENCVNDIKACLKNV-PGINSLNFDIEQQIMSVESSVAPSTIIN 59
     *:  . *:  : : * *..* . ::  *.:: *.:..:::.:*:*::.* :::. . *::

ATX  KIKKTGKE--VR-SGKQ-------L---------------------------------- 73
CCS  TLRNCGKDAIIRGAGKPNSSAVAILETFQKYTIDQKKDTAVRGLARIVQVGENKTLFDIT 119
     .::: **:  :* :**        *
```

how much variance is permissible?

30% identity in 1˚ structure (sequence): **guaranteed** to have same fold **if identities are scattered throughout sequence**

so, yes: as expected … folds can be degenerate
= different sequences, same fold

BUT:
even this robustness has a limit in that

functionally and structurally important residues MUST always be conserved!

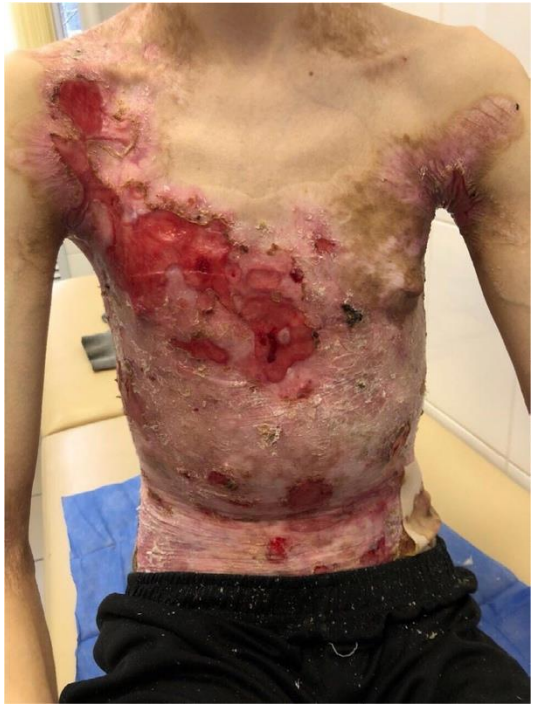
(see orange underlay in sequence above = **MxCxxC** = copper binding
→ mutation of either Cys = non-functional)

Single Point Mutations – Examples For When Things Go Wrong

Progeria

Point mutation that affects splicing in ”Lamin A” a protein that forms a protein support mesh on the surface of the inner nuclear envelope membrane

Epidermolysis Bullosa



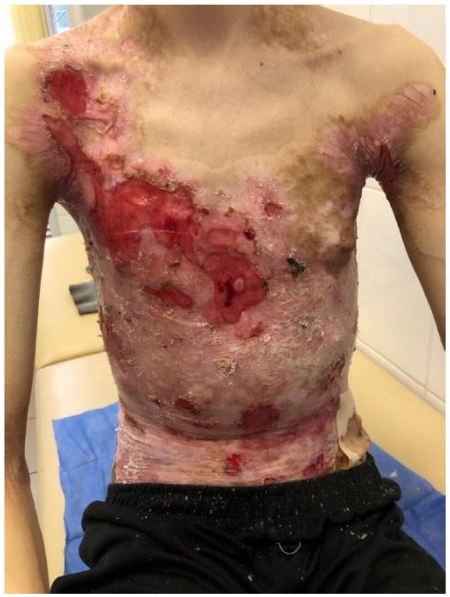
Point mutation in connective tissue protein Collagen VII

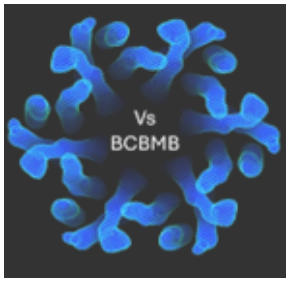
Epidermodysplasia verruciformis (EV) – Treeman disease – caused by a point mutations in a zinc transporter

Single Point Mutations – Examples For When Things Go Wrong

and these are examples for mutations that are still viable ….

… many more mutations that destroy a protein's structure and/or function you will never encounter because those mutations are embryonic lethal = the affected individual will never be born

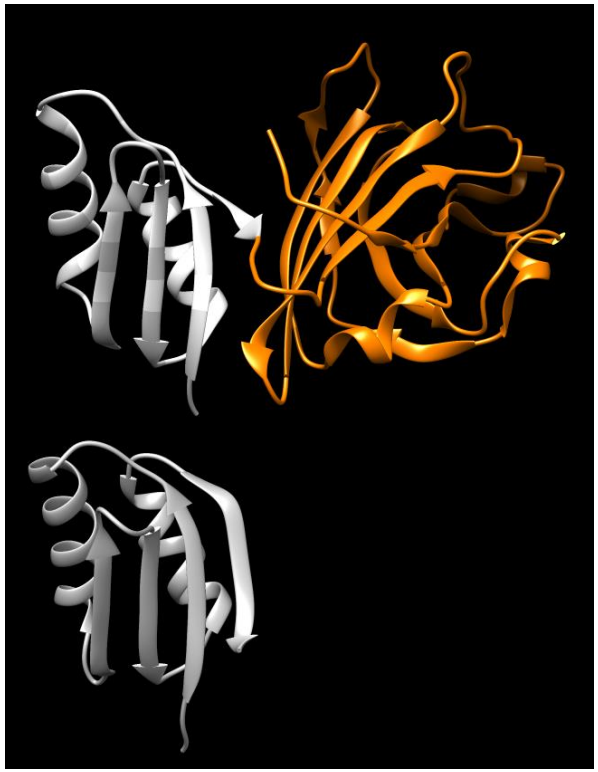




Comparison of Two Copper Binding Chaperones - yATX1 and yCCS - Teaches Us More About Protein Structure

looking at the structures of ATX1 and CCS some more, you notice that CCS has a distinctive second region in addition to the copper binding part (orange color)

moreover, the second region seems to be "independent" in that it does not seem to extensively interact with the copper binding part of the protein.

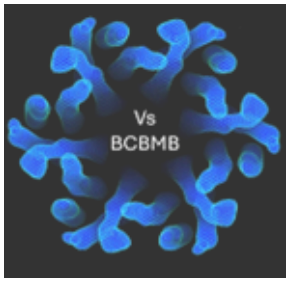


this observation – existence of multiple independent regions in a structure - is not limited to CCS, but is, in fact, very common and is acknowledged by a specific term: "**domain"**

definition: DOMAIN is the smallest part of a protein that can independently adopt a stable fold.

since a fold = shape/structure, **the question arises whether the ability of domains to fold independently is relevant at a functional level?**

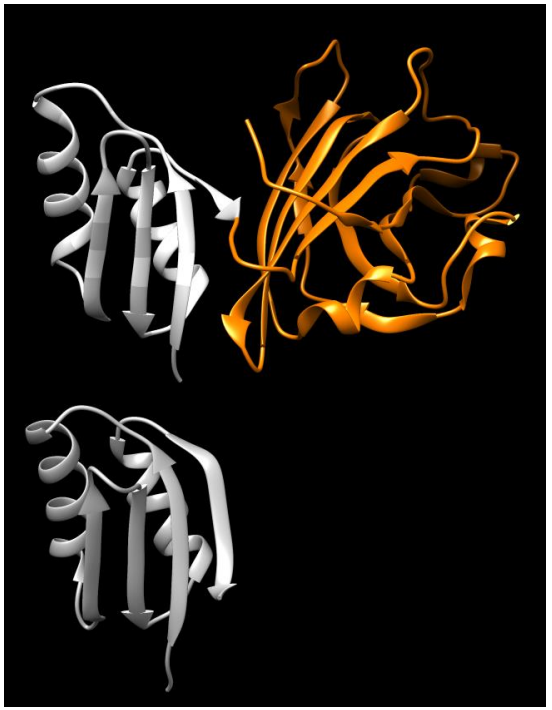
….*what are your thoughts?....*



Comparison of Two Copper Binding Chaperones - yATX1 and yCCS - Teaches Us More About Protein Structure

looking at the structures of ATX1 and CCS some more, you notice that CCS has a distinctive second region in addition to the copper binding part (orange color)

moreover, the second region seems to be "independent" in that it does not seem to extensively interact with the copper binding part of the protein.



this observation – existence of multiple independent regions in a structure - is not limited to CCS, but is, in fact, very common and is acknowledged by a specific term: **"domain"**

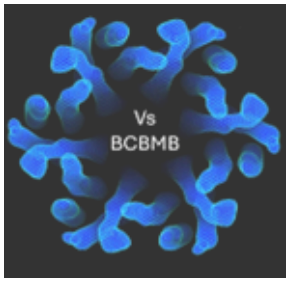
definition: DOMAIN is the smallest part of a protein that can independently adopt a stable fold.

since a fold = shape/structure, **the question arises whether the ability of domains to fold independently is relevant at a functional level?**

Yes: typically, a domain contains a specific function (eg SH3 domains (src-homology 3) recognize P-rich peptides, SH2 domains recognize phopshorylated tyrosine residues, PH-domain (pleckstrin homology) recognizes PIP3 in the membrane).

what is the size of independently folded domains?

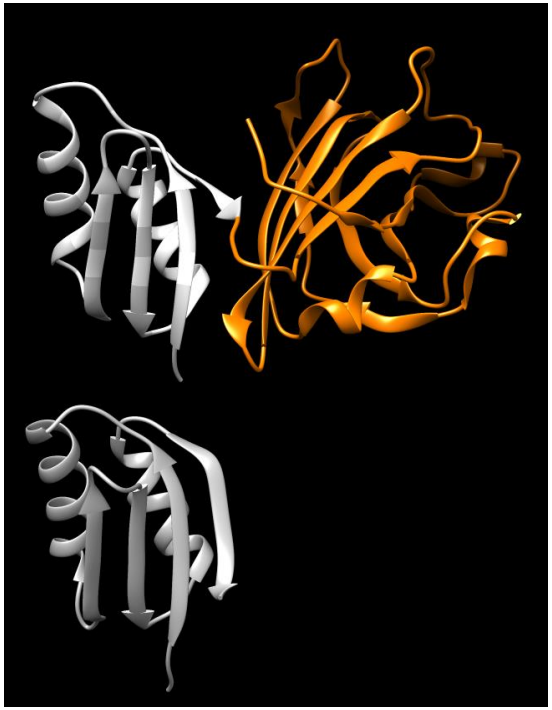
…take a guess!....



Comparison of Two Copper Binding Chaperones - yATX1 and yCCS - Teaches Us More About Protein Structure

Yes: typically, a domain contains a specific function. examples you will come across in your lecture courses are:

- "SH3 domains (src-homology 3)" recognize Pro-rich peptides (important in molecular signaling)
- SH2 domains recognize phopshorylated tyrosine residues (also important in molecular signaling),
- PH-domain (pleckstrin homology) recognizes PIP3 (a triply phosphorylated inositol headgroup) in the membrane.

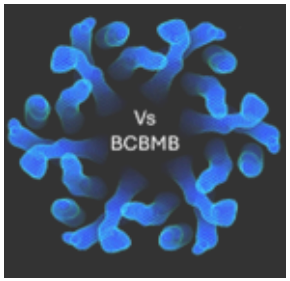


what is the size of independently folded domains?

most domains have a size of ~200 residues (range ~30-600)

put together in our example: CCS is "**multidomain** protein".
more specifically, CCS has two domains – one for copper binding, the other for ….you'll see a little later (slide 67), but there is one more question we want to ask before going there:

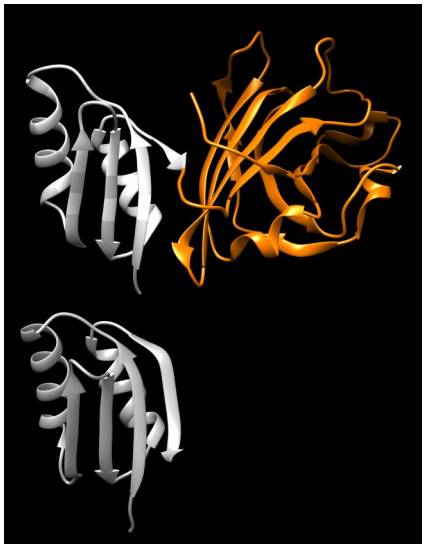
what is the advantage of having/"inventing" domains?
…try to think of reasons …



Comparison of Two Copper Binding Chaperones - yATX1 and yCCS - Teaches Us More About Protein Structure

most domains have a size of ~200 residues (range ~30-600)

what is the advantage of having/"inventing" domains?



Answers:

- makes folding of large, complex proteins possible, efficient, and fast
- allows for **modularity and redundancy** = combinatorial development (modularity) and tuning (redundancy) of protein function

let's "unpack" the second answer a little more since thinking in terms of "conceptual frameworks" may not be too familiar to you…

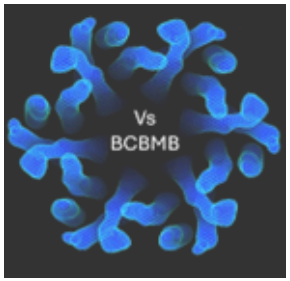
start by thinking what you ideally would want to do once you overcame all odds associated with folding and created a new protein functionality.

You already know that what you created is incredibly robust (slide 57) because, once tamed, the folding process is quite insensitive to small changes in protein sequence here and there.

➔ you can stop worrying about folding/getting a functional protein and start thinking about how to diversify functionality of biological processes/systems at a larger scale by testing what happens to the biological system if you try this "new module" in different situations/contexts.

that is: one thing you definitely want to be able to do is "copy, cut and paste" that exact function into different proteins to see if it creates more efficient, new biological solutions …. shuffling the **same** function **without any changes = modularity**

(Recall "ATP" … exact same molecule with **so many different** functions, Chemical Evolution Chapter, slide 30)



Comparison of Two Copper Binding Chaperones - yATX1 and yCCS - Teaches Us More About Protein Structure

most domains have a size of ~200 residues (range ~30-600)

what is the advantage of having/"inventing" domains?

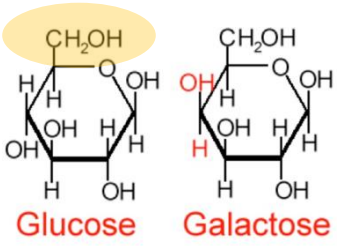
Answers:

- makes folding of large, complex proteins possible and efficient,
- allows for **modularity and redundancy** = combinatorial development (modularity) and tuning (redundancy) of protein function

the second thing you would want to do with your "new functionality" is to "**tweak it**"

meaning: you want to adapt this function to basically do the same thing but use a slightly different target (or do whatever it does at a different speed, or under different cellular conditions)

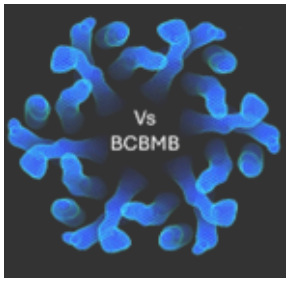
for example: instead of adding a phosphate group to the sugar glucose at its sixth carbon atom (orange underlay), you want to add the phosphate to the sixth carbon atom of another sugar, galactose.



= same reaction, same position, but ever so slightly different targets …..

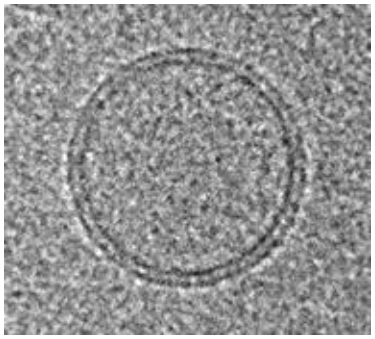
as you will learn in the catalysis chapter … the protein that carries out this phosphate addition for glucose is **very selective … it will not** act on galactose at all at physiological concentrations of that sugar.

➔ what you want to do here then is to slightly change the shape of the protein site that carries out the reaction so that it will accept galactose instead of glucose. This "tweaking" creates a **redundant form of the protein catalyst** (hexokinase → galactokinase) **….** and the robustness of the fold helps you achieve this by changing a few sidechains within the reaction chamber/active site of the protein catalyst

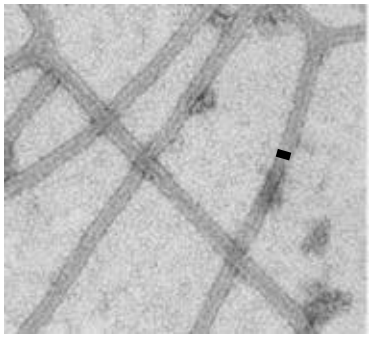


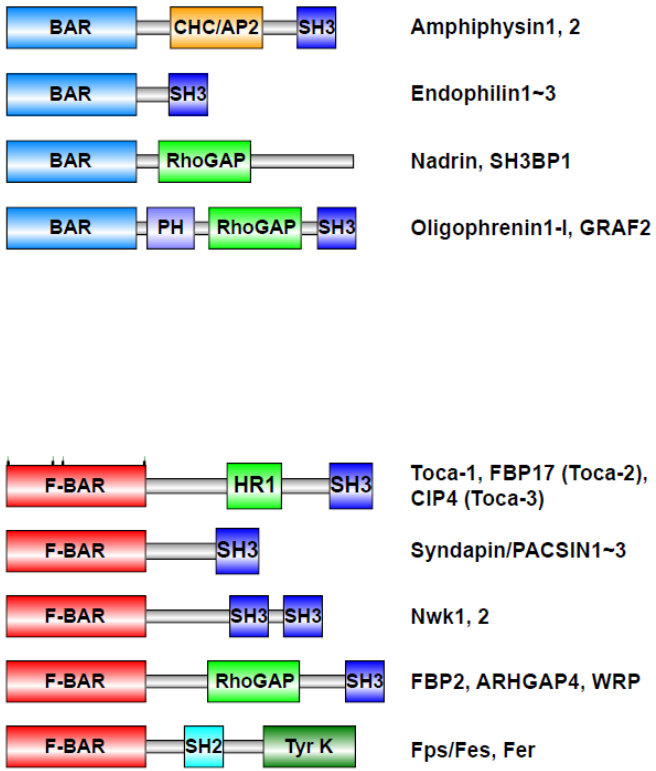
Examples for "Plug and Play" in Protein Structures

(if you worked through the "Membrane Remodeling Case Study" you will remember these)



BAR





CHC/AP2 = clathrin adapter

RhoGAP = domain that inactivates Rho protein kinases involved in signaling

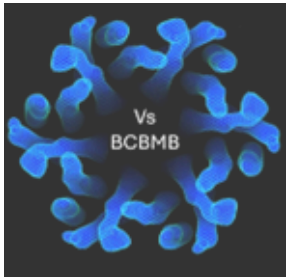
SH2, SH3, PH domains, see slide 62

HR1 = protein:protein or protein:lipid interaction domain

TyrK = tyrosine kinase

BAR and F-BAR domains all function as membrane remodelers = increasing membrane curvature. Which membranes, by how much and in what process is determined by the complement of additional domains that are found in each particular "BAR domain" protein.

for instance: SH3-domins are the most common protein-protein interaction domains present in BAR-/F-BAR-domain proteins and are responsible for interactions with downstream effectors such as dynamin(membrane fission) or N-WASP (cytoskeleton organization/dynamics)

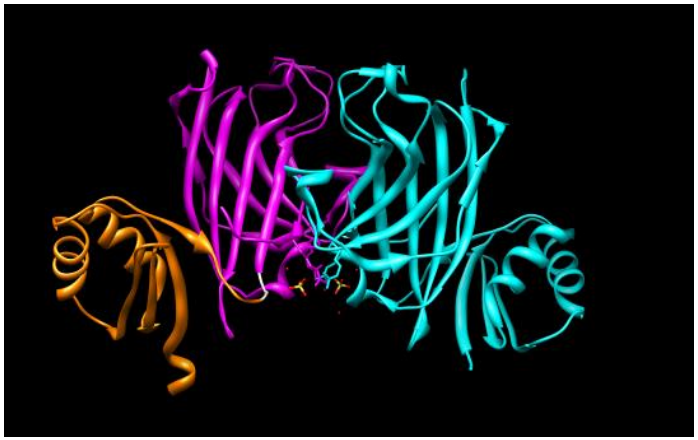


AND MORE…..

CCS not only is a multidomain protein, it also forms a dimer

QUATERNARY STRUCTURE: association of two or more independently folded polypeptide chains to form a supramolecular assembly. Chains can be identical (homo-oligomers) or different (hetero-oligomers)

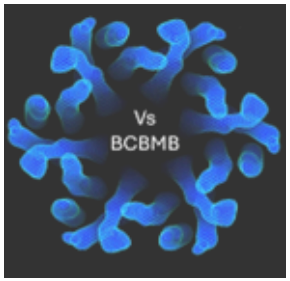
copper binding domain (orange), SOD-like domain (magenta), second monomer (cyan)



in CCS, the "second domain" actually has several functions:

- supports dimerization of CCS
- allows docking of a CCS monomer to its intracellular target (superoxide dismutase, SOD)
- allows CCS to bind to membranes prior to receiving it copper cargo from copper uptake transporters

(something my lab discovered …. if you are interested in this :"scandalous" story, which showed everybody in the field to be wrong about certain things … then follow this link).

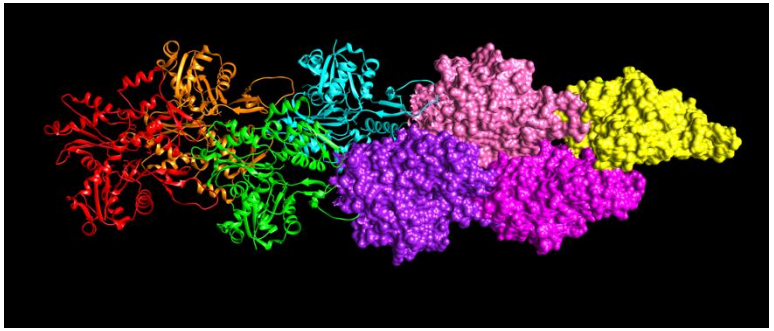


Quaternary Structure – Recall Fundamentals

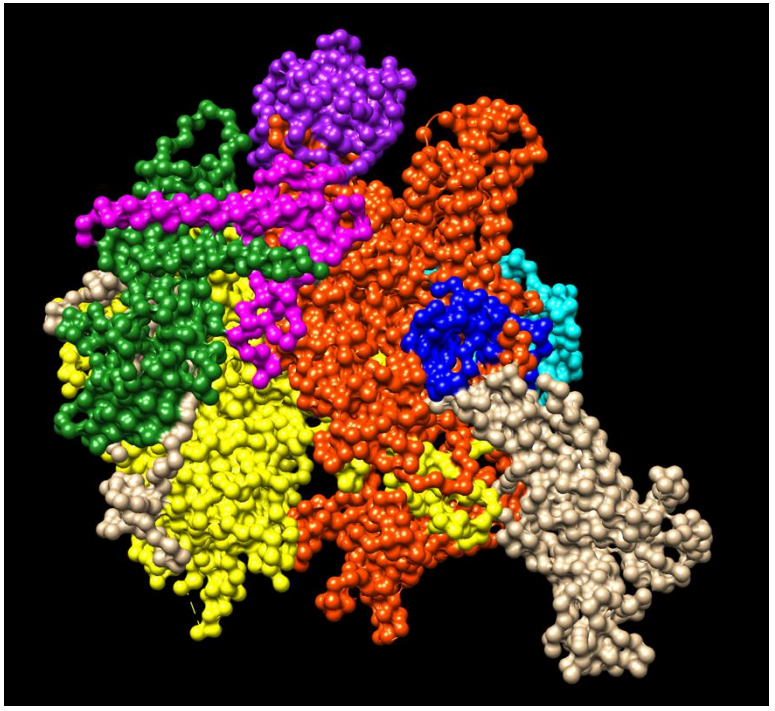
what is biological role of oligomerization?

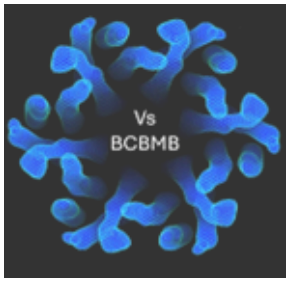
….try and see if you can remember any (or come up with some for the first time)…..

Actin Filament



RNA-Polymerase/TFII complex





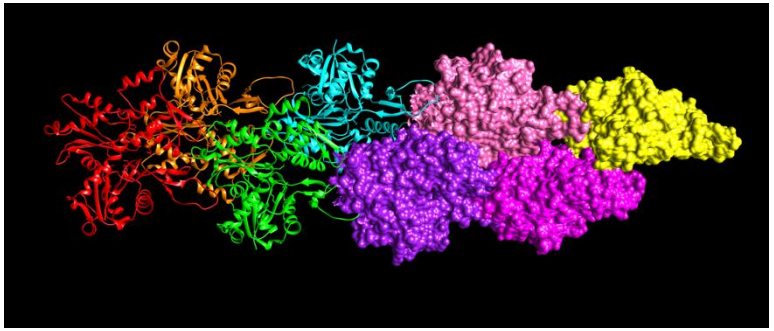
Quaternary Structure – Recall Fundamentals

what is biological role of oligomerization?

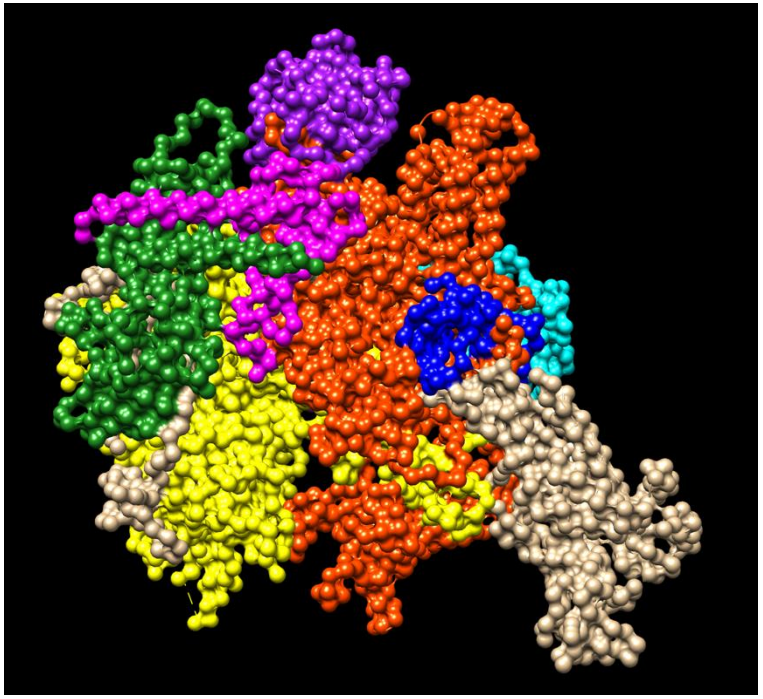
no single answer to that question.
oligomerization supports number of biological purposes:

- formation of structural support elements
(actin fibers, microtubules)
- formation of multifunctional protein complexes
(= alternative to multidomain approach; eg RNA-Polymerase/TFII complex)
- local increase of the number of catalytic centers or substrate binding sites
(e.g. hemoglobin)
- enabling cooperative behavior
(= behavior of each of several identical subunits depends on the state of other subunits)
- formation of membrane permeable pores

Actin Filament



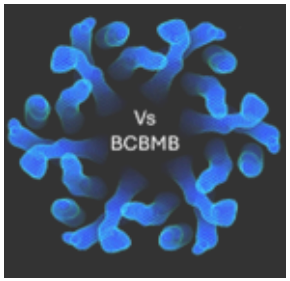
RNA-Polymerase/TFII complex



adding to the structural complexity…..

some proteins contain tightly bound "non-protein" components = **prosthetic groups** (**eg porphyrin** in hemoglobin; **metals** in metalloproteins; **phosphate** groups in phosphoproteins (particularly prevalent in enzymes), and other modifications (too many to list here)

some proteins are conjugated with carbohydrates **(=glycoproteins and proteoglycans),** some are conjugated to lipids **(= lipoproteins).** These modifications are covered in the chapter on carbohydrates

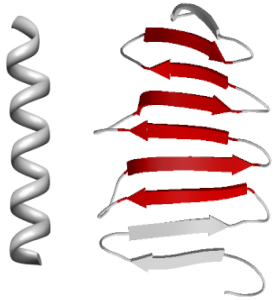


Summary of Protein Structural Organization

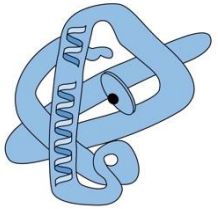
Primary Structure: sequence of amino acids ….for instance:

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

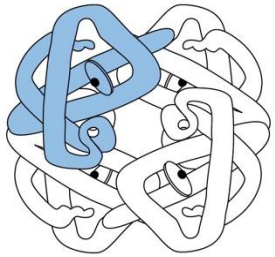
Secondary Structure: local conformation of polypeptide backbone atoms (= amino acid constant core unit)

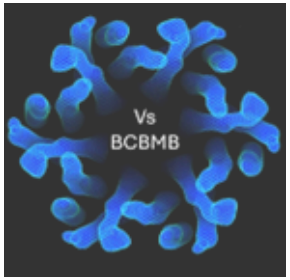


Tertiary Structure: overall spatial arrangement of all of a polypeptide's atoms that is caused and stabilized by long range interactions (mostly) between amino acid sidechains

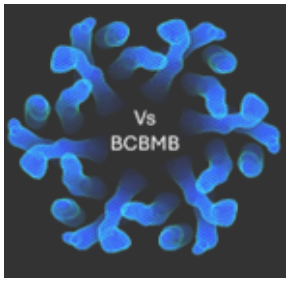


Quaternary Structure: aggregate of ≥2 independently folded (identical) polypeptide chains.





TAKE A BREAK!



Returning to Our Challenge ……
How Can Proteins Help Us With The Impermeability of Biological Membrane to Polar/Charged/Large Solutes?

the simple-minded answer:
make “membrane” proteins that contain pores.

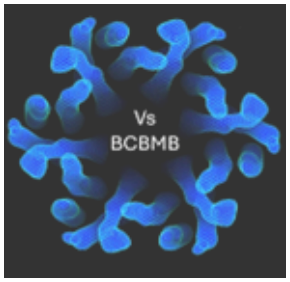
the term "**integral membrane proteins**" was coined, to set such proteins apart from those that are found in aqueous environments and from "peripheral membrane proteins" that only transiently associate with membranes through non-covalent, easily disruptable interactions

now that you have a more detailed understanding of how protein structures come together, you realize that there should not be any fundamental issues with making membrane proteins to meet the cell's functions

because

knowing that generic secondary structure elements rule folding, making membrane proteins simply comes down to finding appropriate "sidechain decorations" for the generic helical and sheet elements.

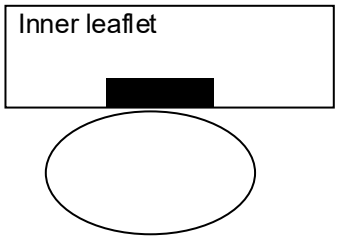
before looking at that, let's get some "terminology and classification" out of the way, and introduce the spread of functions that membrane proteins fulfill



Biology Uses Two Fundamentally Different Types of Integral Membrane Proteins

Integral = **cannot** be removed from a membrane by washing the membrane with salt

Integral Membrane Proteins



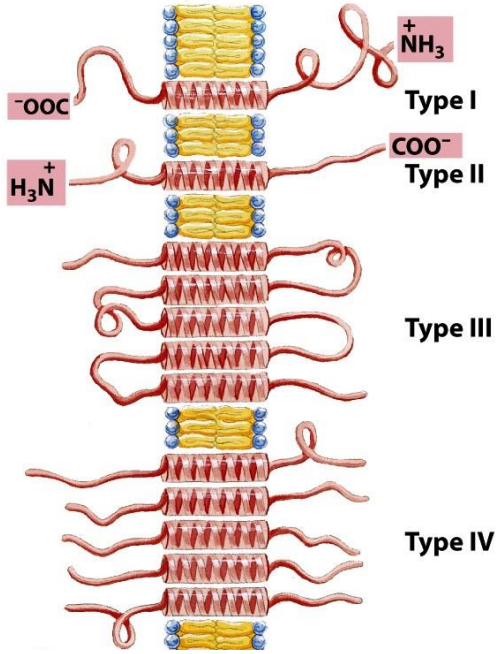
partial penetration (amphipathic alpha-helix – remember the ZxxZxxxZxxZ pattern mentioned on slide 36
→ hydrophobic sidechains in each of the "Z" positions creates a helix that has a hydrophobic strip on its surface that can productively interact with bilayer core)

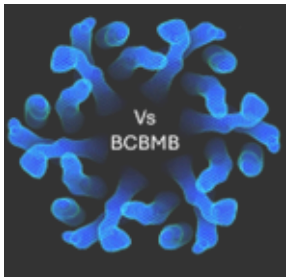
transmembrane

Monotopic
1 membrane spanning segment

Polytopic
>1 membrane spanning segment on same polypeptide

Complex
quaternary structure formed from >1 transmembrane protein





Functional Classification of Integral Membrane Proteins

(Receptors and Transport Proteins are Covered in Later Chapters)

Presenilin (protease)

ENZYMES STRUCTURAL

•Ligand gated
•Voltage gated
•Mechanosensitive

CHANNELS

Integral Membrane Proteins

Facilitated Diffusion

RECEPTORS

•G protein-coupled
•Integrins
•Cytokine
•Receptor tyrosine kinases

TRANSPORTERS

Primary Active/Group Translocation Secondary Active

Transport ATPases using PEP

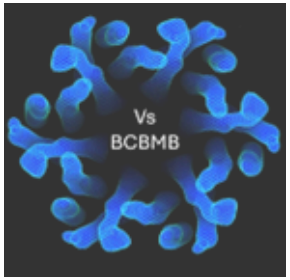
•ABC
•P-type
•V-type
•Heavy Metal

sugar transport (prokaryotes)

Symporter Antiporter

•Na^+/substrate
•H^+/substrate

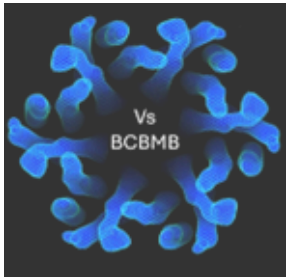
•H^+/Na^+
•H^+/antibiotic



Structure of Integral Membrane Proteins - Energetics

to understand **WHY certain protein structures can be stably integrated into biological membrane**, we – together – want to look at some thermodynamics aspects.

while your instructor likely will be silent on the issue and not ever ask you about this in an exam……what is shown over the next few slides is **REALLY helpful** for understanding anything that is associated with membrane proteins



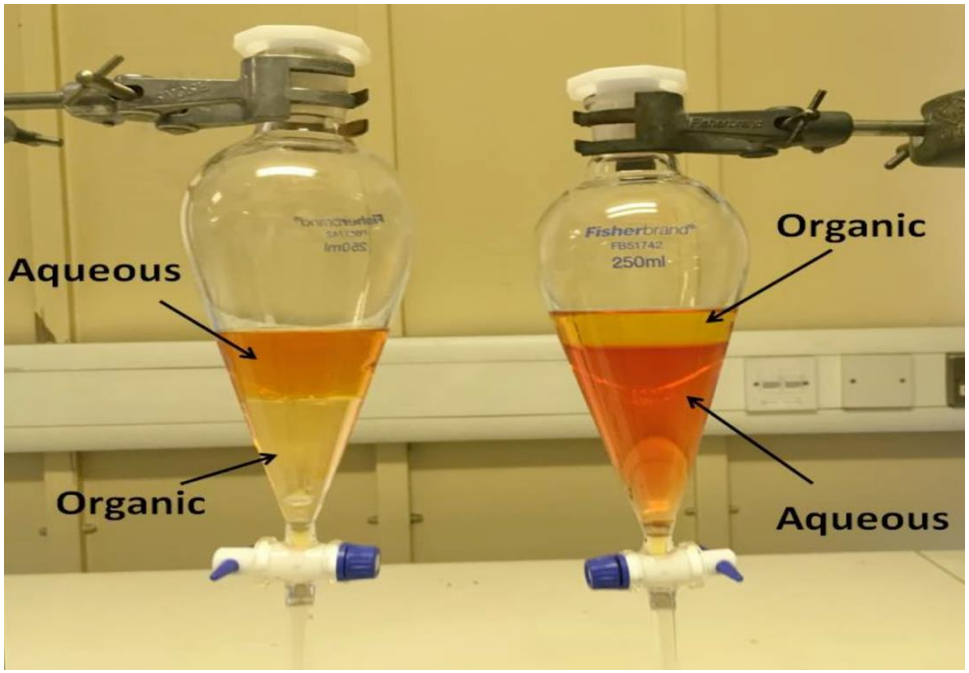
Structure of Integral Membrane Proteins - Energetics

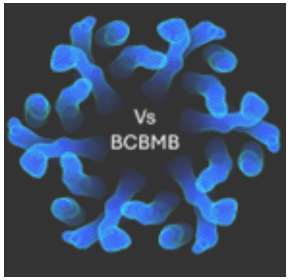
to understand **WHY certain protein structures can be stably integrated into biological membrane**, we – together – want to look at some thermodynamics aspects.

while your instructor likely will be silent on the issue and not ever ask you about this in an exam……what is shown over the next few slides is **REALLY helpful** for understanding anything that is associated with membrane proteins

to start **…**

do you know/remember what is shown in the picture below …. when did you come across it??



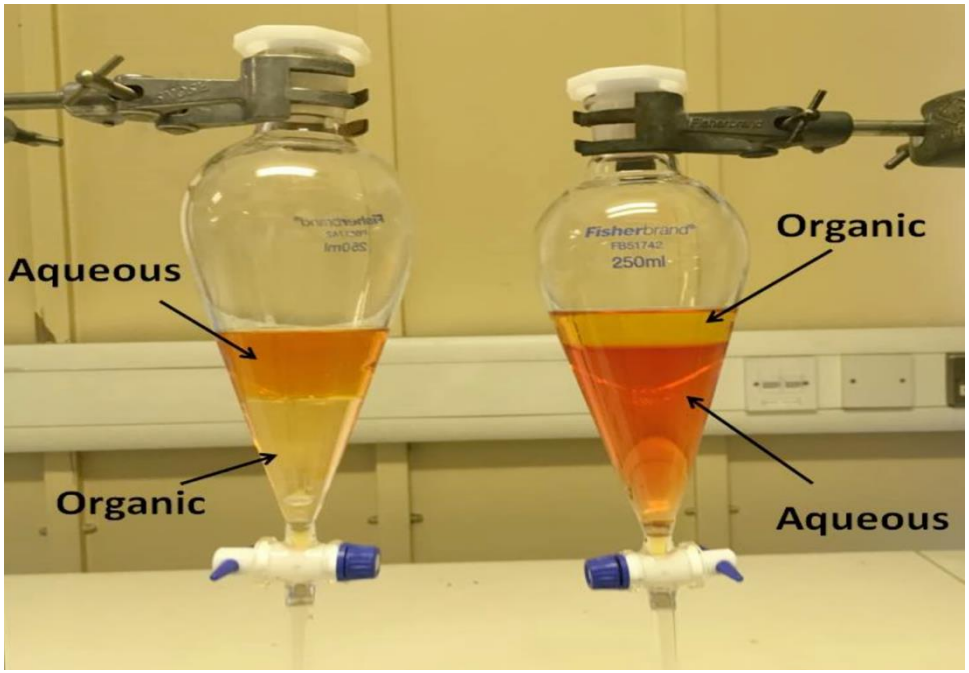


Structure of Integral Membrane Proteins - Energetics

To understand **WHY certain protein structures can be stably integrated into biological membrane**, we – together – want to look at some thermodynamics aspects.

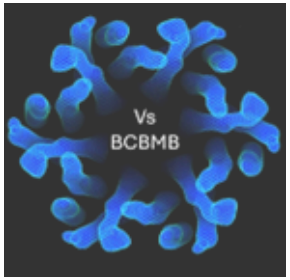
While your instructor likely will be silent on the issue and not ever ask you about this in an exam……what is shown over the next few slides is **REALLY helpful** for understanding anything that is associated with membrane proteins

To start **…**
do you know/remember what is shown in the picture below …. when did you come across it??



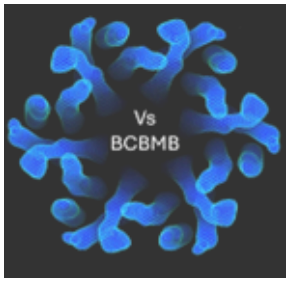
Answer:
most likely during organic chemistry lab … when you were trying to separate two compounds based on their different solubility in two different solvents **….**
….like
water, chloroform or hexane….?

the picture shows two cases to remind you that the "heavier", more dense liquid will be on the bottom…. an aspect that doesn't matter for where we want to go with this (but it is good to remember it anyways).



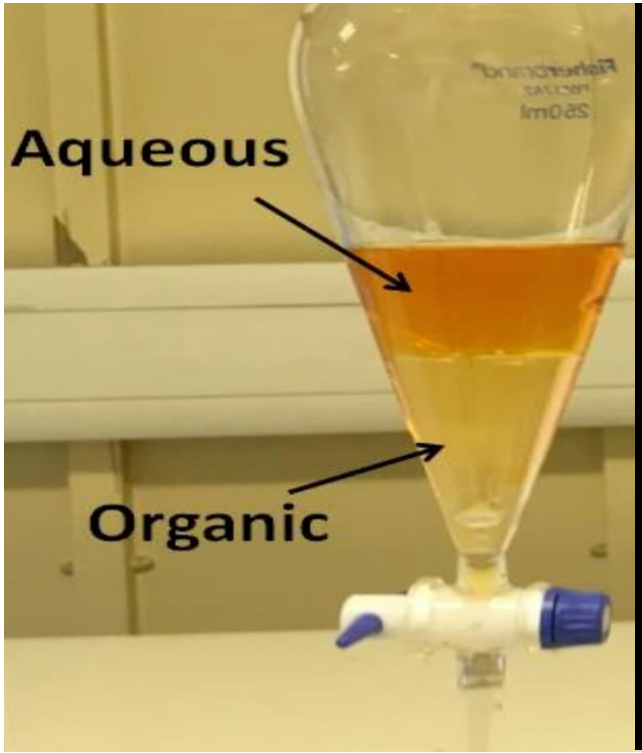
Structure of Integral Membrane Proteins - Energetics

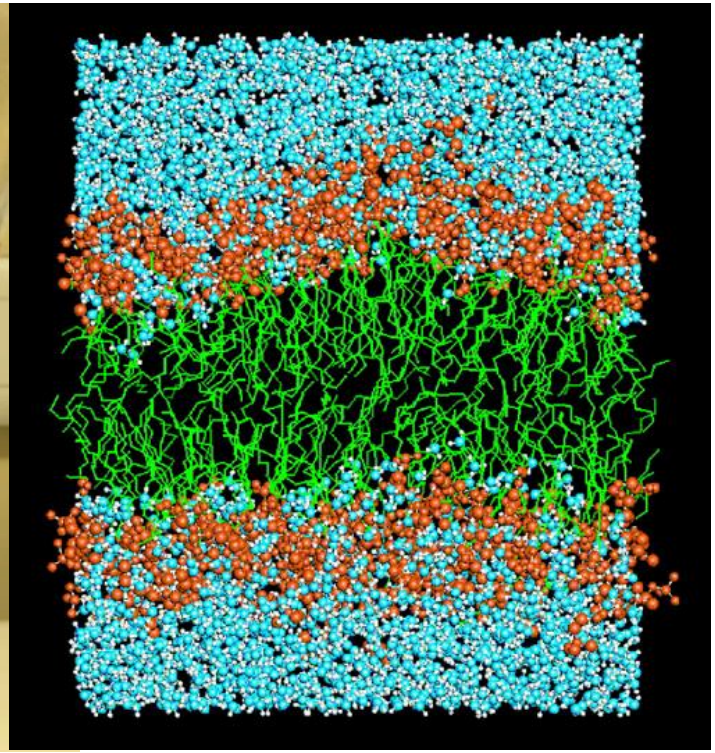
the reason I wanted to remind you of this liquid-liquid extraction setup is to help you understand some experimental data we will shortly look at … but let's take one more intermediary step to prep you some more by taking a look at what is shown in the next slide …..



Structure of Integral Membrane Proteins - Energetics

the reason I wanted to remind you of this liquid-liquid extraction setup is to help you understand some experimental data we will shortly look at … but let's take one more intermediary step to prep you some more by taking a look at this ….

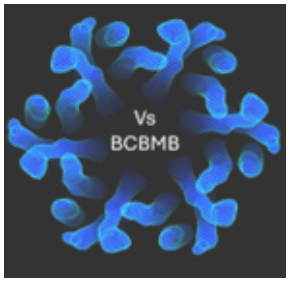




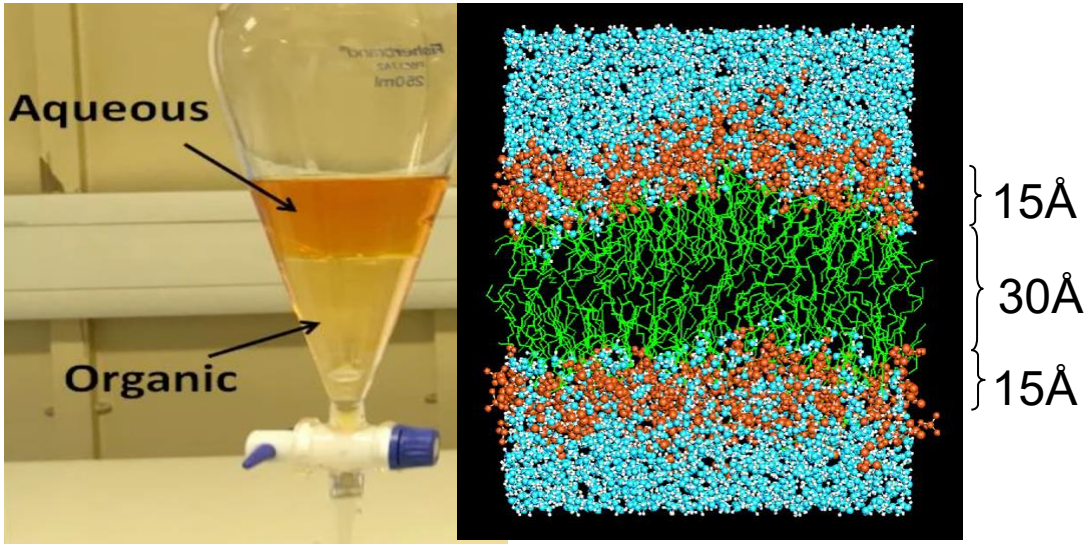
do you start seeing where this is going ….? If so, **awesome!!**

if not, here is the hint: the hydrophobic core of the membrane and the adjacent interface/aqueous regions (like the cytosol) act like a "two solvent system" …. and that means protein structures are subjected to **very different** regimens depending on where they reside.

exposing sidechains to the membrane core … which ones would you choose?
which ones for the aqueous region?

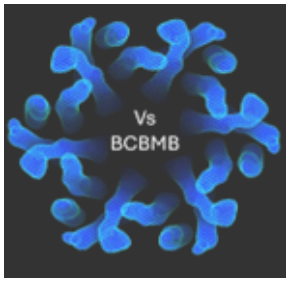


Structure of Integral Membrane Proteins - Energetics



do you start seeing where this is going ….? If so, **awesome!!**
If not here is the hint: the hydrophobic core of the membrane and the adjacent interface/aqueous regions (like the cytosol) act like a "two solvent system" …. and that means protein structures are subjected to **very different** regimens depending on where they reside.
Exposing sidechains to the membrane core … which ones would you choose? Which ones for the aqueous region?

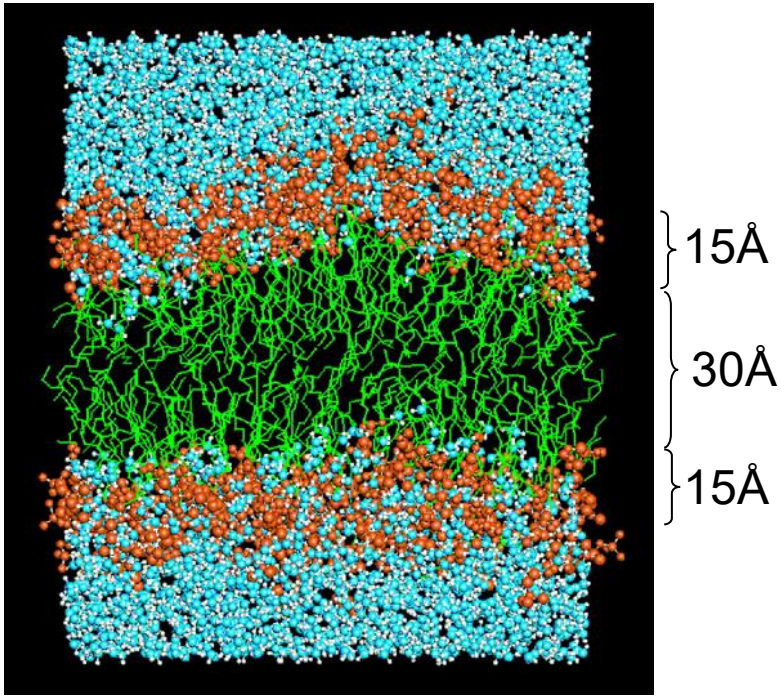
Easy: like on like
use hydrophobic for exposure to membrane core (hydrophobic on hydrophobic)
use polar/charged for exposure to aqueous environments (hydrophilic on hydrophilic)
BUT
what about the backbone??? It makes for a significant part of the polypeptide = how does the backbone (-CO- and –NH- groups) take to the two different environments?
…and there you have it … we are ready to look at some experimental data …..



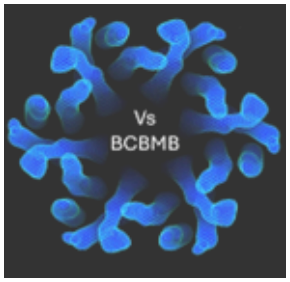
Structure of Integral Membrane Proteins - Energetics

thought Experiment: if you were to put a **single peptide** bond into a separation funnel with water as one solvent and an organic solvent as the other phase …

… what do you think would happen? **Where would the peptide bond go? To the water …or the organic phase?**

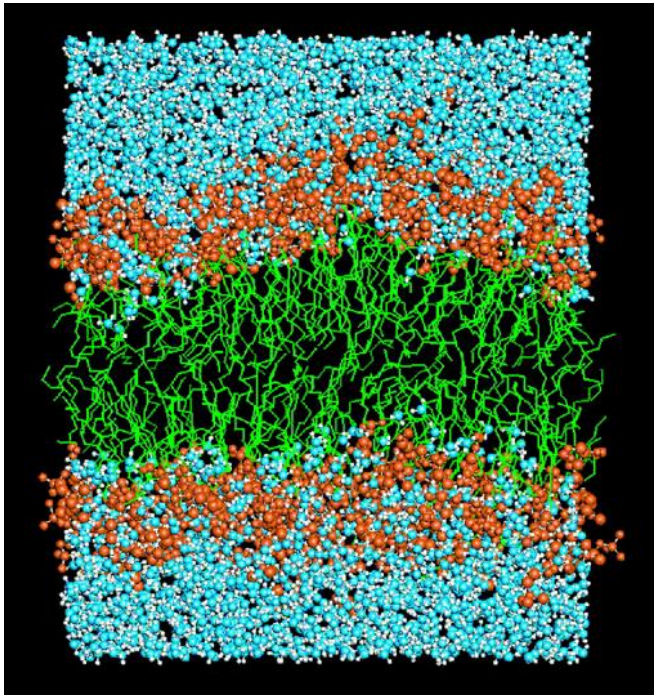


From: Popot & Engelman (2000), Annu Rev Biochem 69:881-922

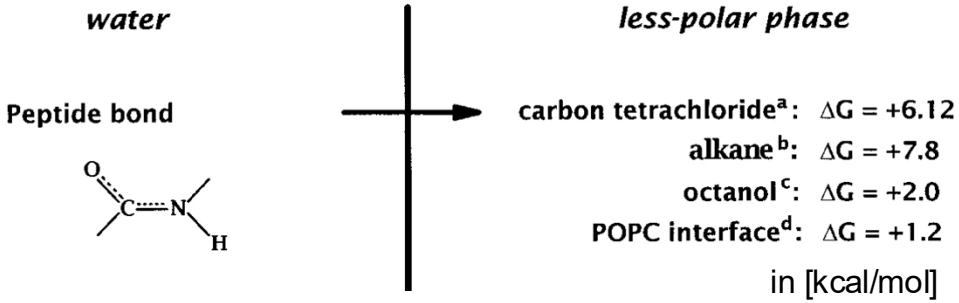


Structure of Integral Membrane Proteins - Energetics

Thought Experiment: if you were to put a **single peptide** bond into a separation funnel with water as one solvent and an organic solvent as the other phase …
… what do you think would happen? **Where would the peptide bond go? To the water …or the organic phase?**



From: Popot & Engelman (2000), Annu Rev Biochem 69:881-922

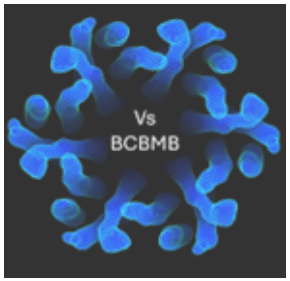


for each of these matchups

water: carbon tetrachloride
water:alkane
water:POPC interface
(POPC: palmitoyl-oleoyl-phosphatidylcholine = a phospholipid)

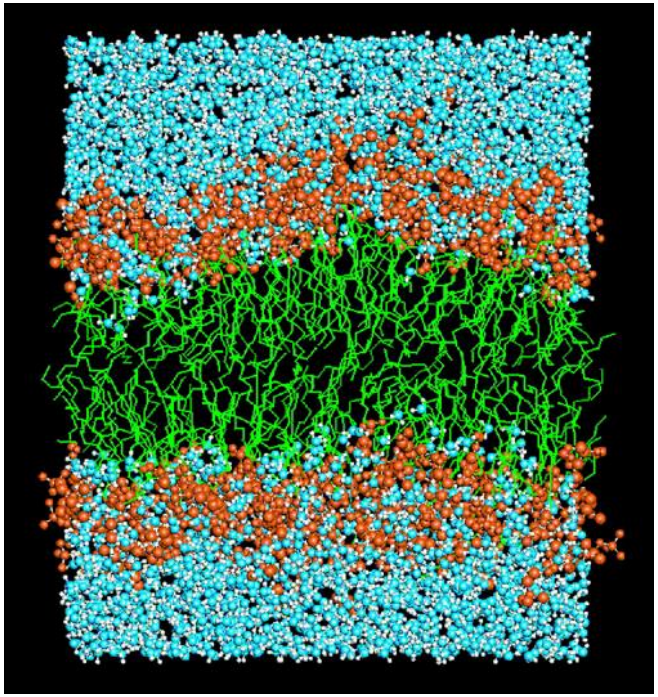
look at the change in Gibbs Free Energy ….

what do these numbers tell you?



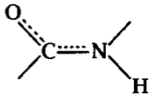
Structure of Integral Membrane Proteins - Energetics

Thought Experiment: if you were to put a **single peptide** bond into a separation funnel with water as one solvent and an organic solvent as the other phase …
… what do you think would happen? **Where would the peptide bond go? To the water …or the organic phase?**



water → *less-polar phase*

Peptide bond



carbon tetrachloride[a]:	ΔG = +6.12
alkane[b]:	ΔG = +7.8
octanol[c]:	ΔG = +2.0
POPC interface[d]:	ΔG = +1.2

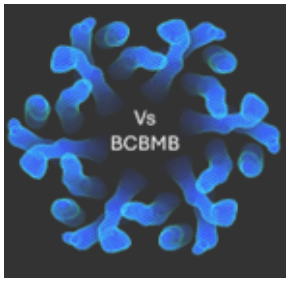
in [kcal/mol]

for each of these matchups

water: carbon tetrachloride
water:alkane
water:POPC interface
(POPC: palmitoyl-oleoyl-phosphatidylcholine = a phospholipid)
look at the change in Gibbs Free Energy ….
what do these numbers tell you?

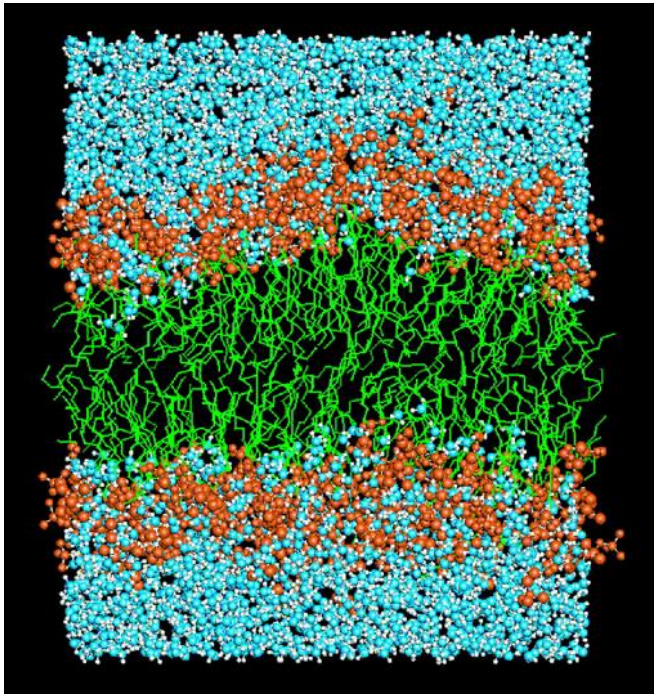
From: Popot & Engelman (2000), Annu Rev Biochem 69:881-922

Answer: **the ΔG values are ALL positive = the peptide bond prefers water over any of the organic phases.** That said: the penalty is the lowest for exposing the peptide bond to the headgroup region of the phospholipid (for simplicity thinking of it as a liquid … it sort of is, but water is less concentrated there than in the bulk aqueous solution)

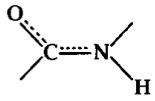
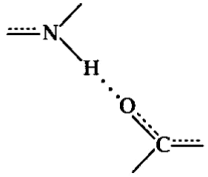
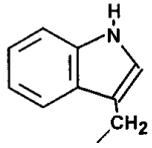


Structure of Integral Membrane Proteins - Energetics

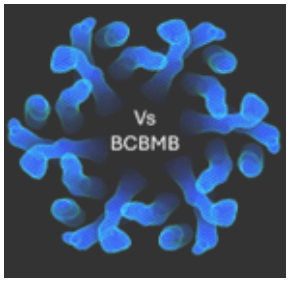
now that you are an expert in reading these data ….here are a few more hypothetical matchups …. What do these tell you?



From: Popot & Engelman (2000), Annu Rev Biochem 69:881-922

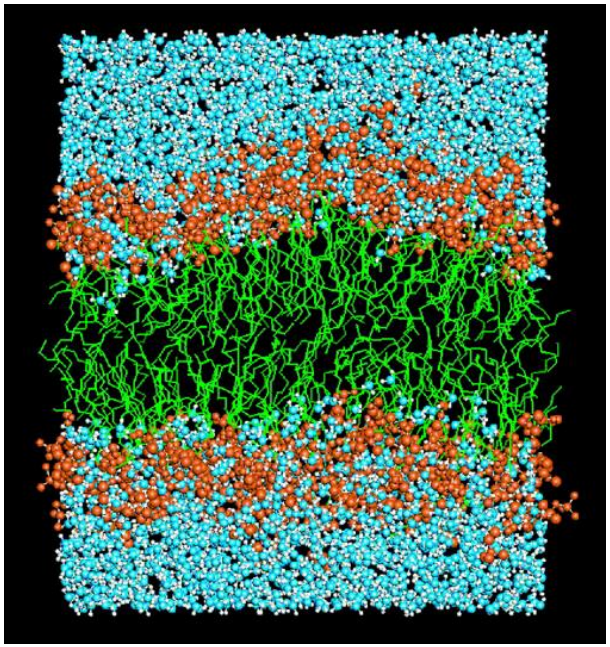
	water	*less-polar phase*
Peptide bond		carbon tetrachloride[a]: ΔG = +6.12 alkane[b]: ΔG = +7.8 octanol[c]: ΔG = +2.0 POPC interface[d]: ΔG = +1.2
H-bonded peptide bond		carbon tetrachloride[a]: ΔG = +0.62 alkane[b]: ΔG = +3.8 POPC interface[e]: ΔG ≈ +0.5
Trp sidechain		octanol[c]: ΔG = –3.24 POPC interface[d]: ΔG = –3.05 in [kcal/mol]

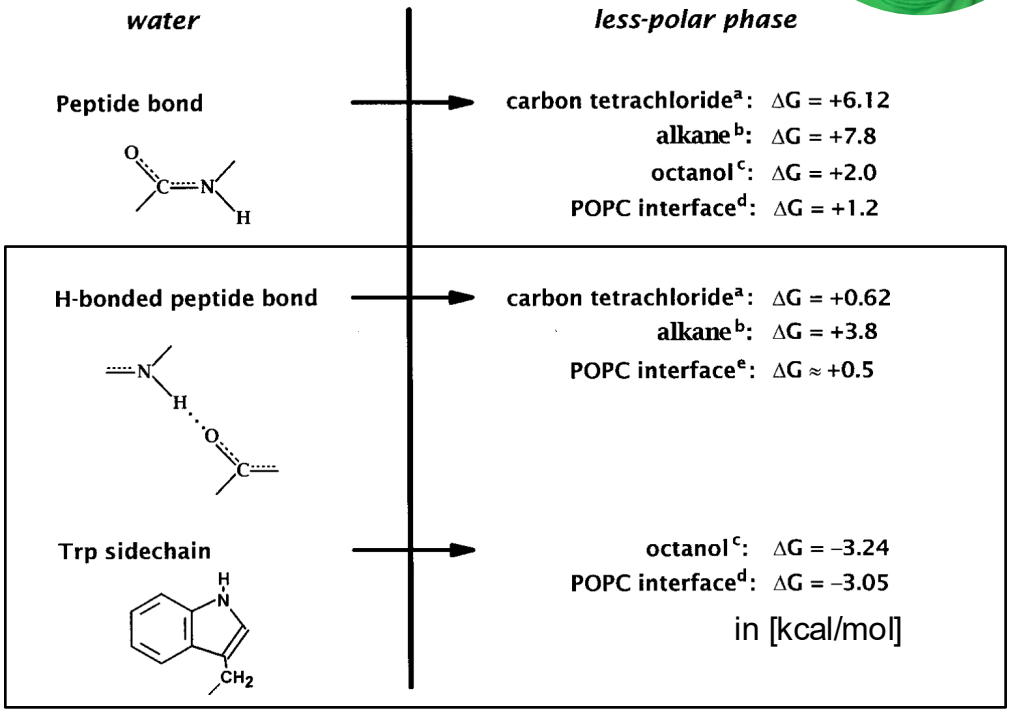
Answer: … be bold … try!



Structure of Integral Membrane Proteins - Energetics

now that you are an Expert in reading these data ….here are a few more hypothetical matchups …. What do these tell you?



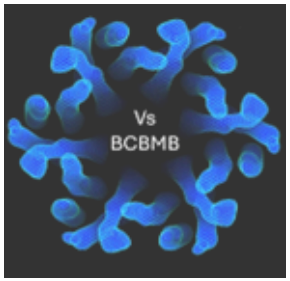


From: Popot & Engelman (2000), Annu Rev Biochem 69:881-922

Answer:

the ΔG values are still positive for "two hypothetical half peptide bonds" that are hydrogen bonding = this hypothetical situation still prefers water … **but** note how the penalties have **decreased by a full order of magnitude** (+6.1 → +0.62) for transfer into the least polar solvent carbon tetrachloride [closest polarity match to fatty acid hydrophobic tail in this set of experiments]

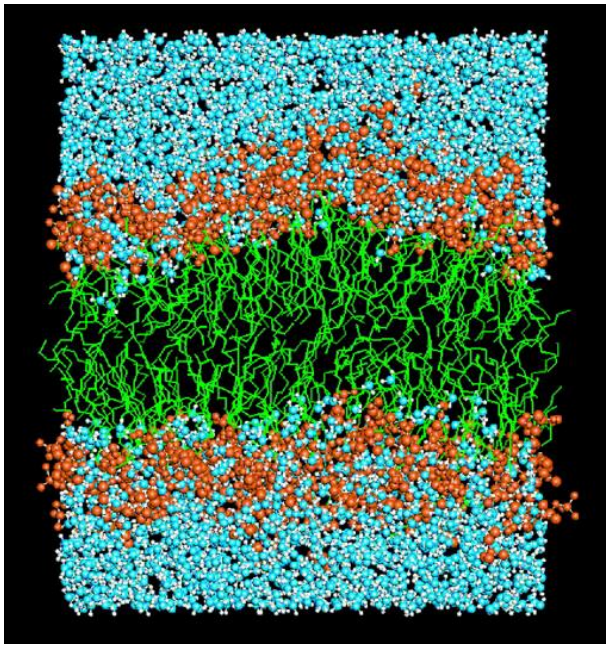
the ΔG values for the sidechain of the amino acid tryptophan are significantly negative in both octanol and the POPC interface = the sidechain of Trp prefers the organic phase over water!!!!



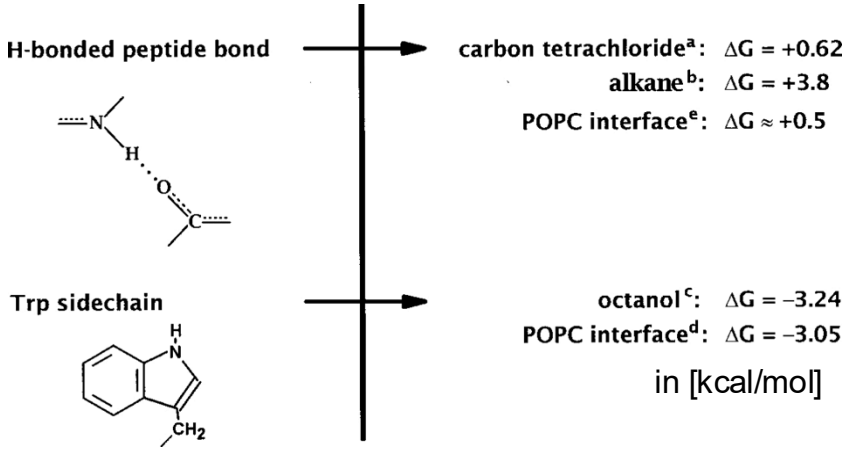
Structure of Integral Membrane Proteins - Energetics

the sidechain of Trp prefers the organic phase over water!!!!

this is a **GAMECHANGER!** Why?



From: Popot & Engelman (2000), Annu Rev Biochem 69:881-922

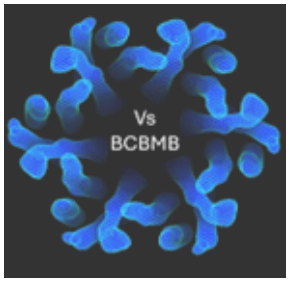


it is a game changer because ΔG values in coupled systems are additive .... meaning ….if you try to place a polypeptide into the bilayer …. the free energy changes for the backbone atoms and the sidechains need to be added together to gauge whether this can stably exist…..

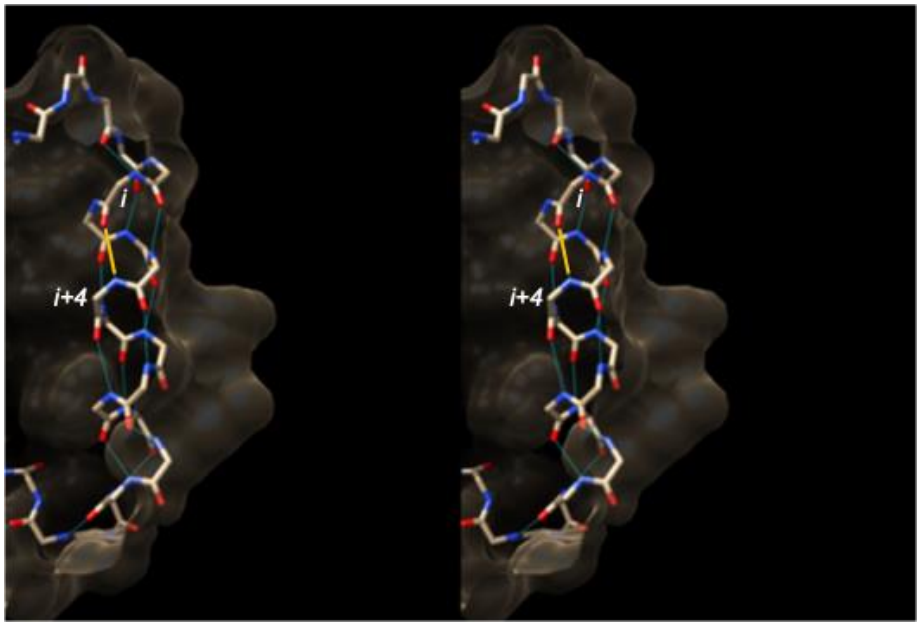
in a polypeptide – each peptide bond comes with a sidechain

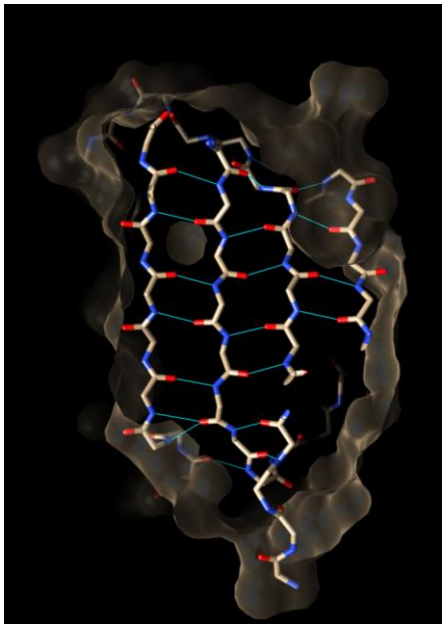
→ **if you want to put this in a bilayer you need to pay the cost for accommodating the polarity of the –CO- and –NH- groups, but you also earn "credit" for any hydrophobic sidechain you bring along.**

→ if this sidechain were **Trp … you earn A LOT, especially if you can somehow make the peptide bonds being engaged in H-bonds (**low Free Energy cost of less than 1 kcal/mol) … and **how could you do that?**



SECONDARY STRUCTURE!

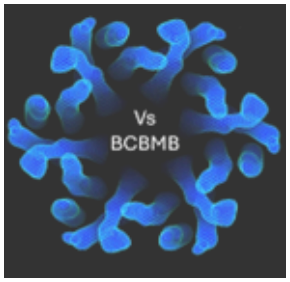




that simple! The secondary structure of the polypeptide is built by making the –CO- and –NH- groups of the polypeptide engage in H-bonding patterns

- parallel to the helix axis for α–helices
- between strands in a β–sheet

or if you want to use other words: **the hydrophobic membrane core ENFORCES secondary structure of the polypeptide** (like a straight jacket) because the energetic cost of unpaired hydrogen bonds would be prohibitive. **Note**: how this is the same as what we found for folding of water soluble proteins, where the hydrophobicity of the "molten globule" induced formation of secondary structures ….(see slide 21)



The Previous Few Slides Where The Hardest Part Of It All

if you followed and understood all that…then the rest will be super easy …

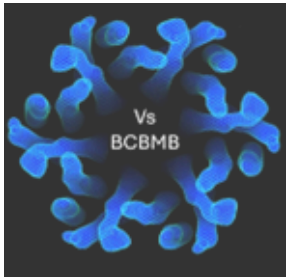
because

we now just need to think about how to utilize the findings to put together functional structures

even if you know nothing about integral membrane proteins….

….you can intuitively deduce what the two fundamentally different groups of structures will look like ….

one group will use …. (complete sentence)
the other group will use …. (complete the sentence)



Making of a Membrane Protein Pt-1

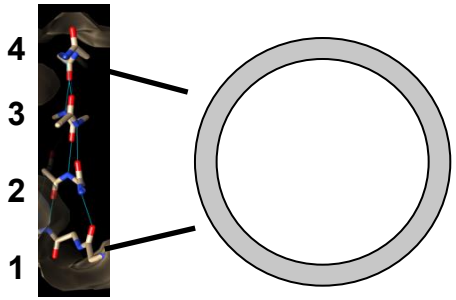
β–barrel

- need ~8-12 amino acid residues per strand to span hydrophobic core of membrane
  (distance between residues on b-strand ~3.5Å → 30Å core is ~9)

- at least every other sidechain **must** be hydrophobic
  the small recap picture (slide 40) of beta sheet illustrates how the wall of the barrel is made by the backbone atoms = the sidechains will alternate pointing inward-outward as you along each strand

- need to create a closed barrel (8 strands minimum)

as otherwise you would have an open sheet with two edges that display unpaired hydrogen bonding partners …not possible)



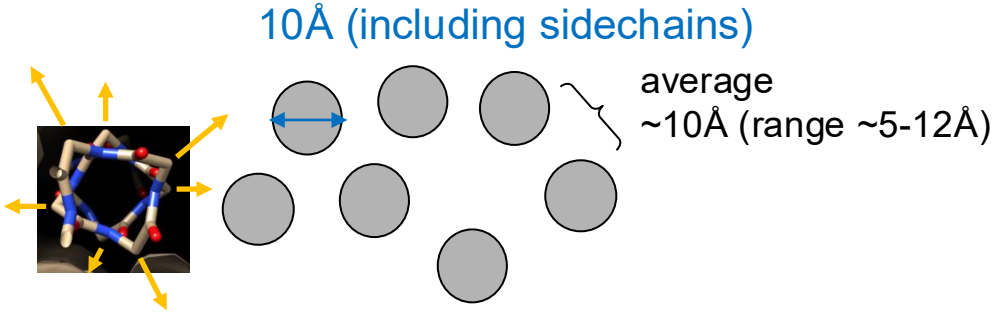
α–helical

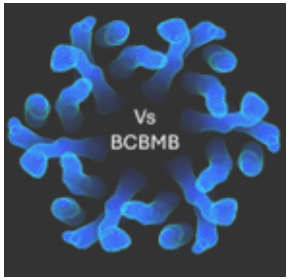
- need ~20 residues to span hydrophobic core of membrane.
  In the "recap picture" of the end-on view of a helix, sidechains project in the direction of arrows; arrow length indicates distance away from POV (long = closer)

- every 3-4th sidechain (remember: ZxxZxxxZ) exposed to lipid → generally must be hydrophobic (few exceptions), but additional hydrophobic residues are required to allow helix-helix packing

- helices are self-contained

→ need no minimum number to achieve stability within membrane as long as no polar or charged sidechains are exposed within the hydrophobic membrane core.





Making of a Membrane Protein Pt-2

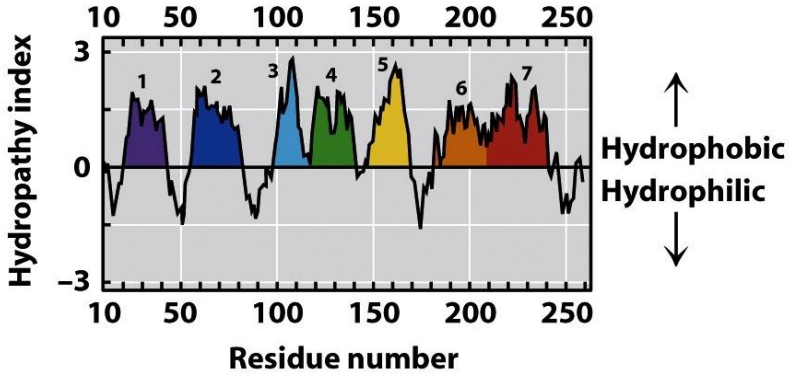
consequences of structural design

β–barrel

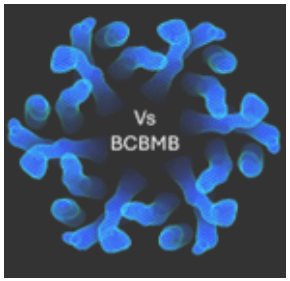
- severely limited in structural diversity
- function usually contained within monomer
- only found in outer membrane of gram negative bacteria and mitochondria
  = these membrane proteins were first, before α–helical membrane proteins began to dominate the scene
- hard to predict from amino acid sequence because a short stretch of sequence that alternates hydrophobic-hydrophilic sidechains could be found almost anywhere in any protein structure

α–helical

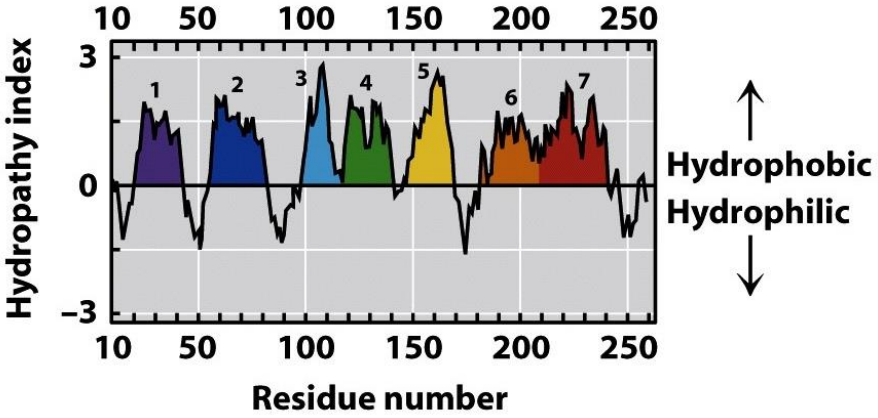
- no limitations to what can be build
- function not constrained to monomer but can also be created by oligomerization (eg. potassium channels, gap junctions, IP3-receptors).
- by far the most abundant type of membrane protein
- transmembrane helices can be predicted from sequence (works often, but not always!)



Hydropathy Plot



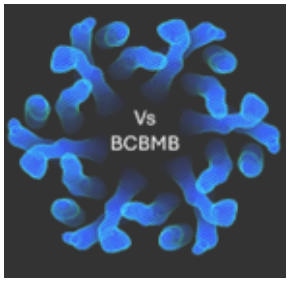
Predicting α–Helical Transmembrane Segments



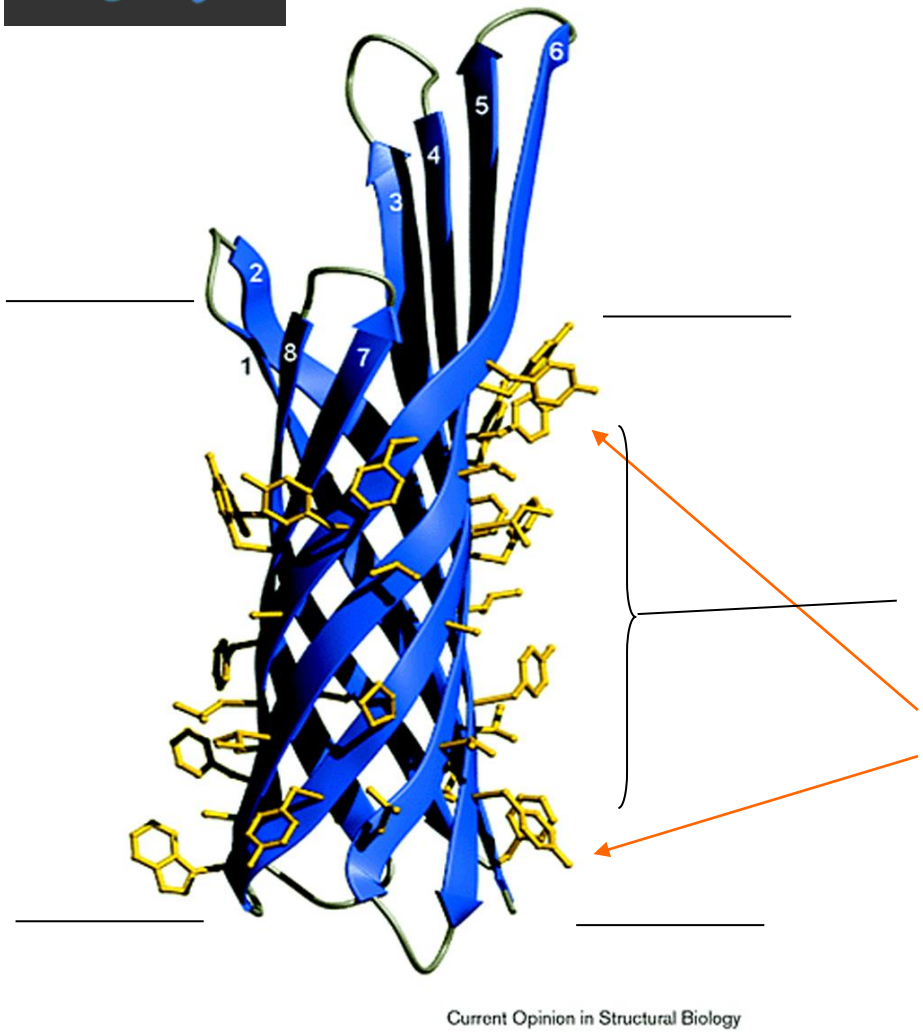
- **Hydropathy Plot**
- in this method you choose a boxsize (let's say 8 amino acids) and place it over the first 8 residues in the sequence
- you calculate how hydrophobic this group is (using tabulated values that tell you how hydrophobic each of the sidechains is) and plot the hydrophobicity index as shown above
- then you advance the box by 1 amino acid along the sequence and repeat the calculation+plotting
- ….and repeat this previous step until the box contains the last 8 amino acids of the sequence
- The aggregated plot looks like what is shown in the picture.

one curious and fortuitous detail: do you remember which amino acids prefer to be in helical secondary structure?

A, L, M = all hydrophobic/non-polar sidechains, and RKQE (and those have handy uses too like determining the positioning of the helix within the membrane, or voltage sensing in voltage gated channel proteins…).



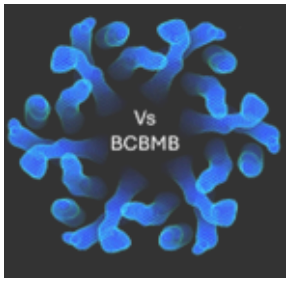
Design of β-Barrel Proteins Pt1



the 9 Rules

- **number of β-strands is even** and the N- and C-termini are at the periplasmic end
- all **β-strands are antiparallel** and **connected** locally to their **next neighbors along the chain**
- the strand **connections** at the **external** end of the barrel are usually **long**
- the strand **connections** at the **periplasmic** end are **short**
- the β-strand **tilt is always around 45˚** and corresponds to the natural β-sheet twist
- the surface contacting the nonpolar membrane interior consists of aliphatic sidechains, forming a **nonpolar ribbon with a width of ~22Å**
- the aliphatic ribbons are lined by **two girdles of aromatic side chains**
- **sequence variability** of all parts of the barrel **during evolution is high** when compared to soluble proteins
- the **external loops show exceptionally high sequence variability** and are usually mobile

From Schulz GE (2000), Curr Opin Struct Biol 10:443-7

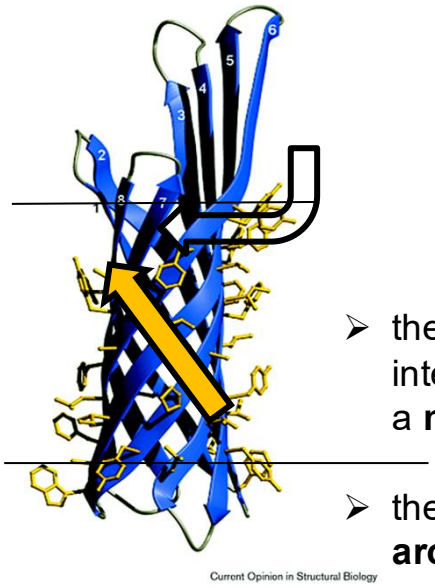


β-Barrel Proteins

9 Rules … how arbitrary can it get??
do any of them make sense?
yes, they do (well … most of them ….); here is a brief explanation for the most important ones

- **number of β-strands is even….**
- All **β-strands are antiparallel**

these two go together … the strands are antiparallel because that makes for better geometry of the H-bonds, which are aligned parallel to the barrel surface (roughly in the direction of the orange arrow)
= do **not point into the bilayer core**
➔ if the number were odd, you would get a seam where two strands run parallel to each other … if you look back at slide 39, you will notice that the H-bond pattern is different for the two types of β-sheet – such a seam would be very unstable and unable to hold up inside the membrane



- the surface contacting the nonpolar membrane interior consists of aliphatic sidechains, forming a **nonpolar ribbon with a width of ~22Å**

makes sense "hydrophobic facing hydrophobic

- the aliphatic ribbons are lined by **two girdles of aromatic side chains**

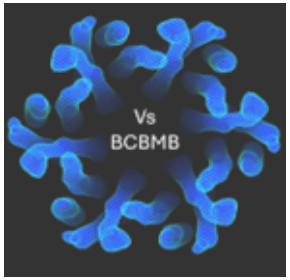
the π-electron clouds of the aromatic rings can interact favorably with water and parts of the headgroups (= these sidechains "fix" the position of the barrel within the membrane)

- **sequence variability** of all parts of the barrel **during evolution is high** when compared to soluble proteins

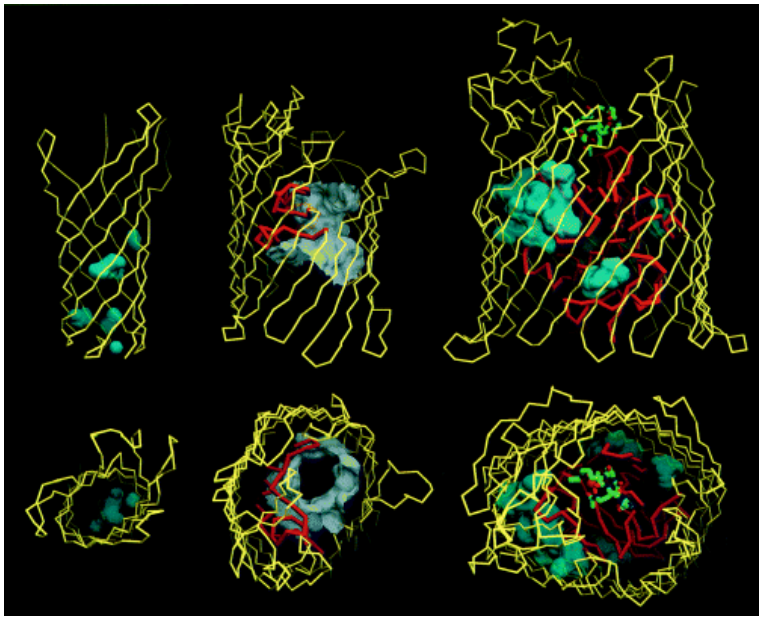
possible because 50% of the sidechains face the membrane core = as long as they are hydrophobic all will be well

- The **external loops show exceptionally high sequence variability** and are usually mobile

these loops determine specificity for biological function = **must** be highly variable



Examples β-Barrel Proteins



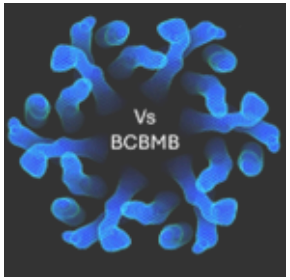
OmpA: 8 strands, physical linkage between outer membrane and peptidoglycan layer; monomeric

OmpF: 16 strands, diffusion pore for ions and other small molecules; homotrimer

FhuA: 22 strands, uptake of Fe-siderophore complex, signal transduction; monomer

Fig. 1. Bacterial β–barrel membrane proteins OmpA, OmpF and FhuA (from left to right), as seen from the plane of the membrane (top) and from the top of the membrane (bottom). Internal cavities (cyan) and channel surfaces (white) are indicated. The polypeptide backbone is shown in yellow, and **protein segments that constrict the barrel interior** (loop L3 of OmpF, N-terminal plug domain of FhuA) are shown in red. The ferrichrome molecule that binds on top of the plug domain is coloured in green. All figures were prepared using the program DINO
Source: Koebnik et al Mol Microbiol (2000) 37:239-53

protein segments that constrict the barrel interior !! This is the "problem" with barrels … as more strands are inserted the "hole" just gets bigger = at some point you lose the ability to control what can go through **unless** you design a "plug" that will plug the open space **and** contains the passageway for what will be transported ... this complexity is one reason helical membrane proteins took over



Design Rules of α-Helical Membrane Proteins

compared with -barrel protein no firm "rules" have been derived that describe the design of helical membrane proteins. Reason: α-helix is a more versatile building block compared to β-sheet and the number of functions carried out by helical proteins is much larger than that of β-barrel protein (= no clear correlation)

However, a few design principles are evident:

- the **hydrophobic effect, favorable partition of hydrophobic residues into membrane core and cooperative formation of hydrogen bonds upon helix formation allow incorporation of "helix breakers" (P,G,T,S,Y,I,V)** = the excessive free energy penalty for leaving the backbone unstructured inside the bilayer core will force these residues into a helical structure whether they "like it or not" (think: "my way or the highway") → they will comply …grudgingly….

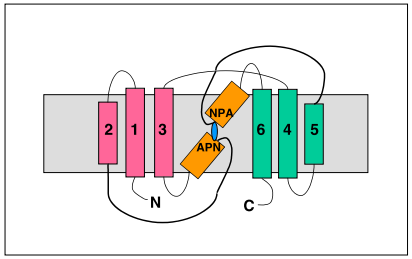
➔ the upside of this: larger diversity of structures possible + can place functional groups within membrane.

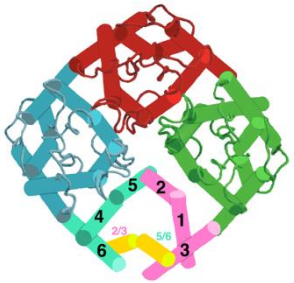
- **GxxxG-motif (GG4 motif)** is over-represented in transmembrane spanning helices → allows tight helix association

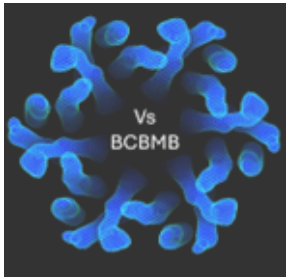
(note ZxxxZ = both glycine residues are aligned above each other along the helical axis ….the lack of a sidechain generates a very flat and smooth surface that allows neighboring helices to come very close to each other!)

- average tilt of transmembrane helix with respect to membrane plane is 21˚. This keeps the average number of amino acids needed to form that helix at a reasonable level (more tilt = more amino acids needed to span membrane) and is significant because evolving these segments is harder (more rules) than evolving strands for a beta barrel….and sticking to those more complex rules consistently gets harder as the length increases …

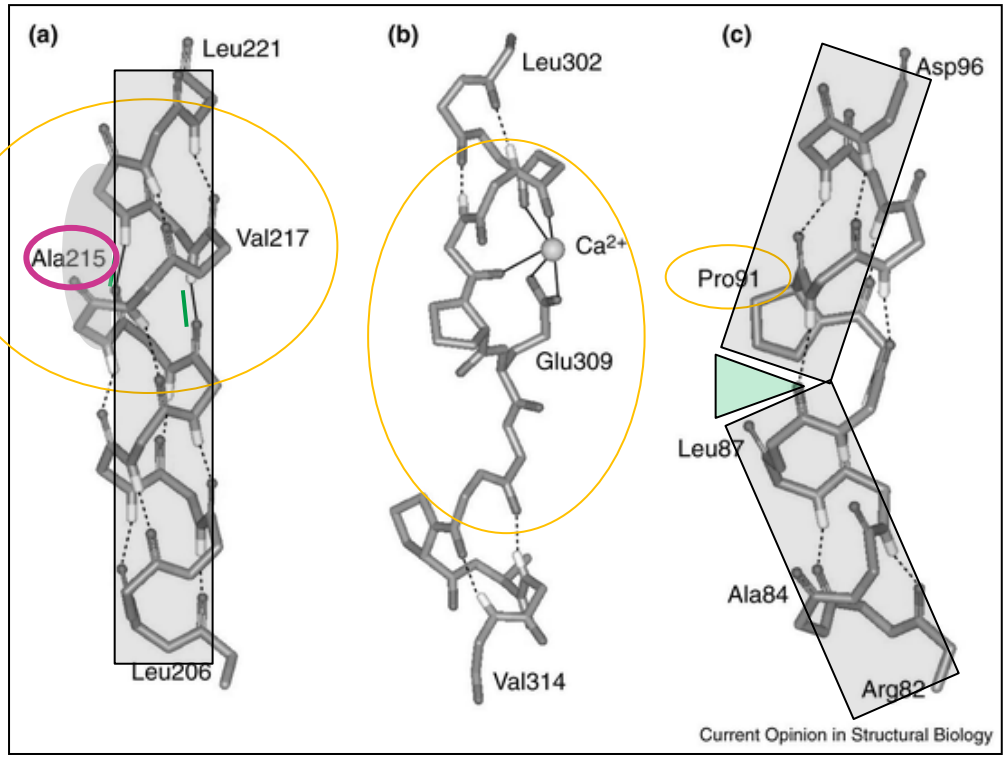
- in many cases: helices consecutive in amino acid sequence are next neighbors in bundle. However, this may not be true once connecting loop is >30 residues. One example is water channel, aquaporin … which also is weird in that it has "two half helices" …. Which we will look at in more detail in the Membrane Transport Chapter







By Enforcing the Helical Structure, the Hydrophobic Effect Allows Disruptions of α-Helical Structure in Transmembrane Helices



π-bulge

= "insertion of a residue" into a helical turn

unwinding

proline-kink

(a) **π-Bulge** in helix G of bacteriorhodopsin ([13•]; PDB entry code 1C3W). The α-helical (i, i + 5) hydrogen bonds are shown by solid lines. Note that the carbonyl group of Ala215 does not participate in backbone hydrogen bonding.

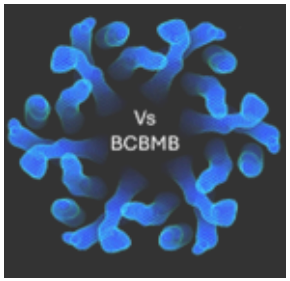
(b) **Unwinding** of helix M4 in the calcium ATPase of the sarcoplasmic reticulum ([14•]; PDB entry code 1EUL). **Note** that the unwinding exposes several backbone carbonyl groups. Some of these carbonyls are involved in the coordination of calcium (denoted by solid lines). **IMPORTANT NOTE:** this segment is surrounded by "regular" helices to shield the Ca-ion binding site from the hydrophobic membrane core!
(more about it in the Membrane Transport Chapter)

(c) **Proline kink** in helix C of bacteriorhodopsin ([13•]; PDB entry code 1C3W). Note the increased spacing of the backbone below the proline and the carbonyl of Leu87 devoid of hydrogen bond. The canonical α-helical (i, i + 4) hydrogen bonds are denoted by dashed lines.

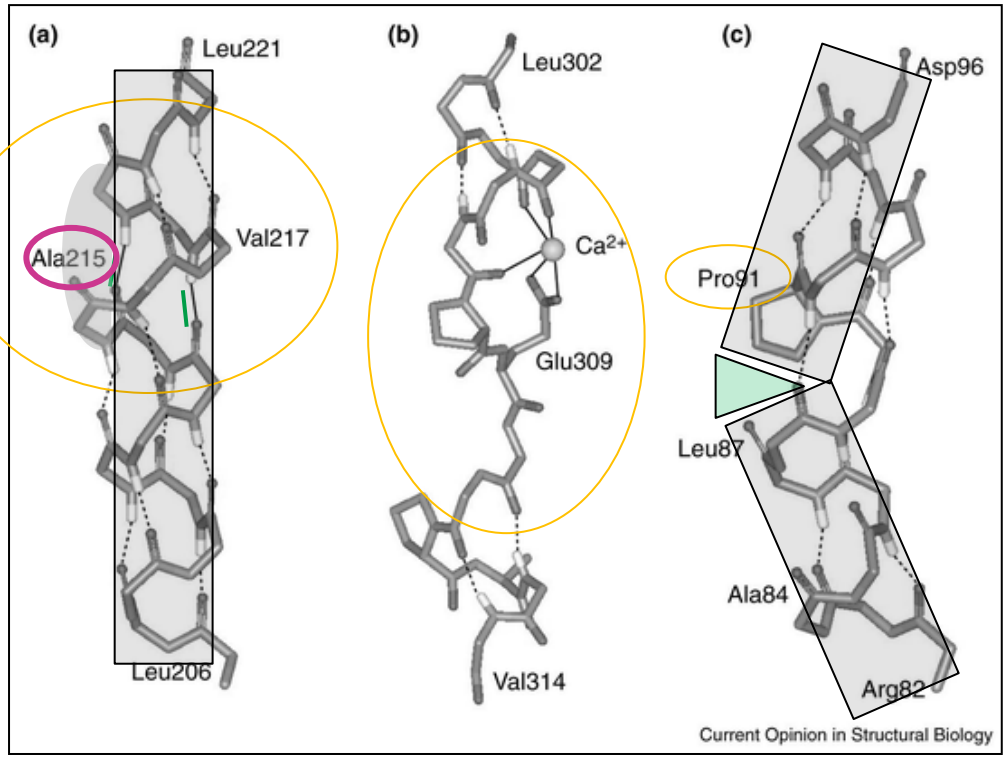
From: Ubarretxna-Berandia & Engelman (2001) Curr Opin Struct Biol 11:370-6

---

why do these irregularities exist? ….*try to think of reasons ….*



By Enforcing the Helical Structure, the Hydrophobic Effect Allows Disruptions of α-Helical Structure in Transmembrane Helices



π-bulge

= "insertion of a residue" into a helical turn

unwinding

proline-kink

(a) **π-Bulge** in helix G of bacteriorhodopsin ([13•]; PDB entry code 1C3W). The α-helical (i, i + 5) hydrogen bonds are shown by solid lines. Note that the carbonyl group of Ala215 does not participate in backbone hydrogen bonding.

(b) **Unwinding** of helix M4 in the calcium ATPase of the sarcoplasmic reticulum ([14•]; PDB entry code 1EUL). **Note** that the unwinding exposes several backbone carbonyl groups. Some of these carbonyls are involved in the coordination of calcium (denoted by solid lines). **IMPORTANT NOTE:** this segment is surrounded by "regular" helices to shield the Ca-ion binding site from the hydrophobic membrane core!
(more about it in the Membrane Transport Chapter)

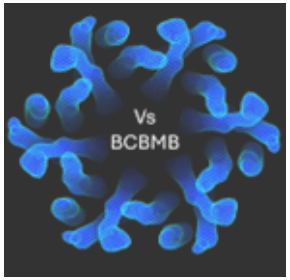
(c) **Proline kink** in helix C of bacteriorhodopsin ([13•]; PDB entry code 1C3W). Note the increased spacing of the backbone below the proline and the carbonyl of Leu87 devoid of hydrogen bond. The canonical α-helical (i, i + 4) hydrogen bonds are denoted by dashed lines.

From: Ubarretxna-Berandia & Engelman (2001) Curr Opin Struct Biol 11:370-6

why do these irregularities exist?

- create non-hydrogen bonded carbonyls/amines which can interact with ligands or neighboring helices.
- proline kinks act as "hinges" during conformational changes (which are necessary for function!)

More about this in Membrane Transport Chapter



The Following Few Slides Show You Examples For the Structural Diversity of Helical Membrane Proteins

just **look at them** and let the impressions sink in ….

….as you do, you will hopefully **realize why these types of proteins came later than the beta barrel membrane proteins.**

briefly:
the grouping of helices in the proteins are called "helical bundle".
to design a helical bundle you must

- satisfy the requirement for a hydrophobic surface all around to allow embedding into the membrane core, and maintain that requirement over a 20+ residue long stretch of amino acids for each helix
- you also must design contact regions between the helices in the bundle to allow them to pack/form the bundle in the first place (=tertiary structure!)
- in many cases, you also must design interfaces to allow quaternary structures to form (2+ membrane proteins associating with each other inside the membrane)
- beyond that, you must design the bundle for whatever its function is = the sidechains that allow passage of polar molecules need to be chosen and placed properly

AND

- you must design all of this in ways that allow for conformational changes if they are required for the protein to function (you will learn in the Membrane Transport Chapter … not all membrane transport processes require extensive movement of the helices)

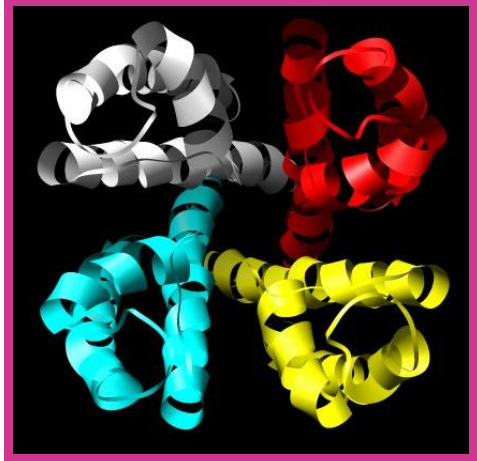
to make a functional beta barrel

- you only need 8 short stretches (or more as long as it is an even number) of at least 8 amino acids (no more than 12) that like to make beta sheets (**high likelihood of happening).** The hydrogen bonding pattern of the backbone will assemble **and** close the barrel at the secondary structure level (= **not tertiary structure interactions like in helical packing) = much easier**

➔ all you need to make sure of is that every other sidechain is hydrophobic to meet the bilayer core + that the inward pointing sidechains form a proper passageway (and – where necessary – plug)

Helical Membrane Proteins: CHANNELS

KcsA Potassium channel(1BL8)

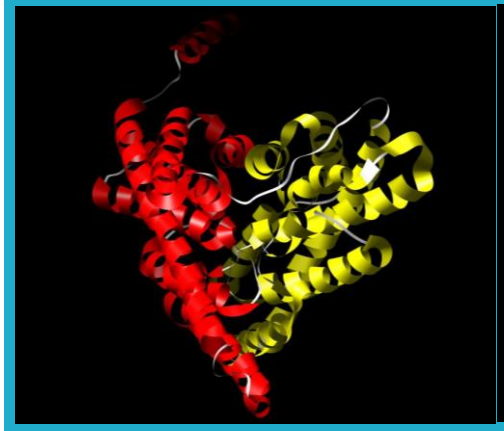


MscL Mechanosensitive Channel (1MSL)

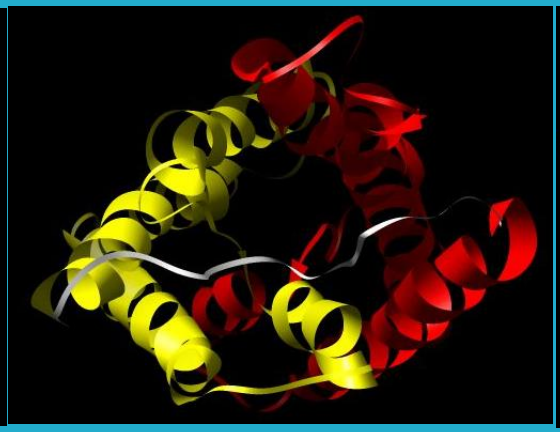
Acetylcholine Receptor (2BG9)

Oligomers
(=quaternary structures)
pore formed between subunits and positioned along central symmetry axis

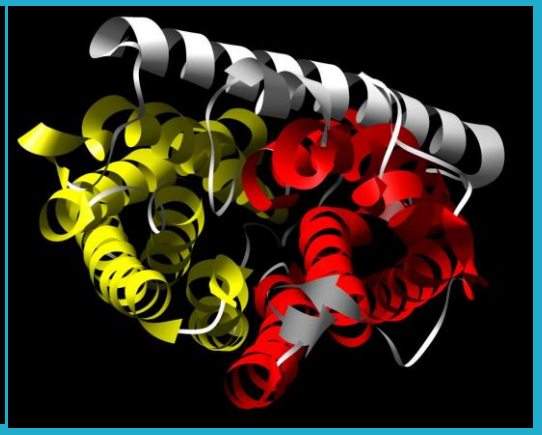
Cl^- channel (1KPL)



Human Aquaporin 1 (1FQY)



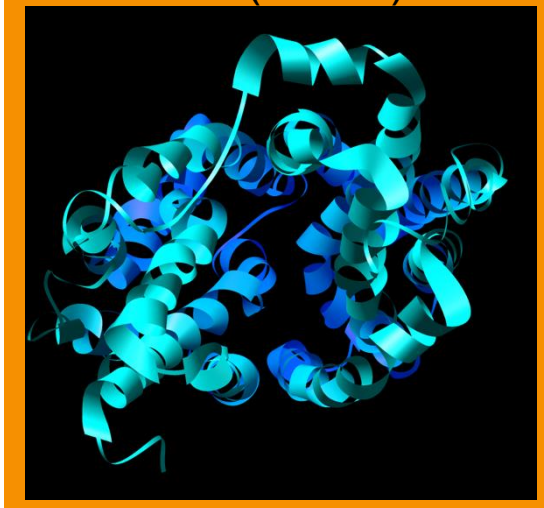
AmtB ammonia channel (1U77)



monomer with internal duplication in sequence leads to a structure with pseudo-twofold symmetry in the membrane plane (= left/right side of molecule look like mirror images)

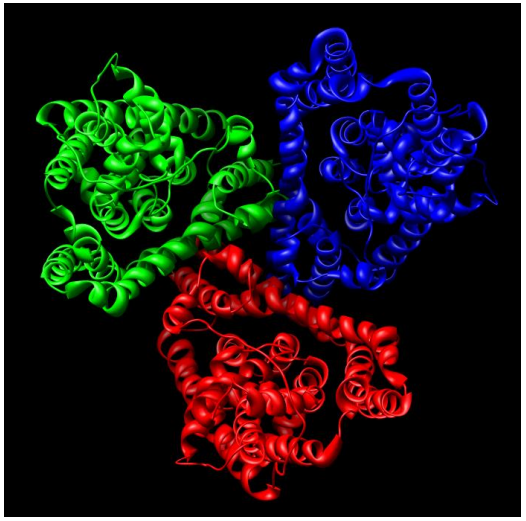
Helical Membrane Proteins: Transporters

LacY (2CFQ)



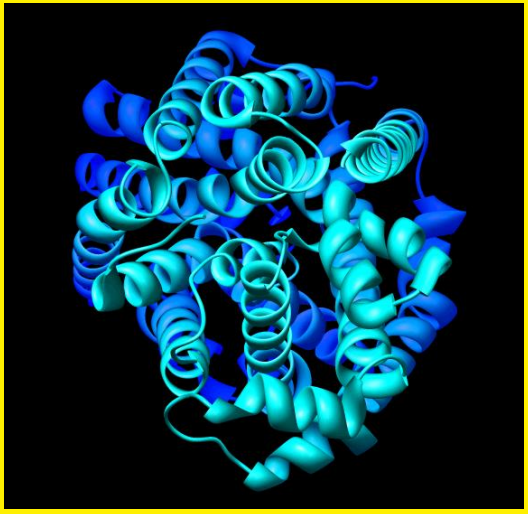
pseudo-twofold symmetry
substrate transport occurs along central pore

Glutamate Transporter (1XFH)

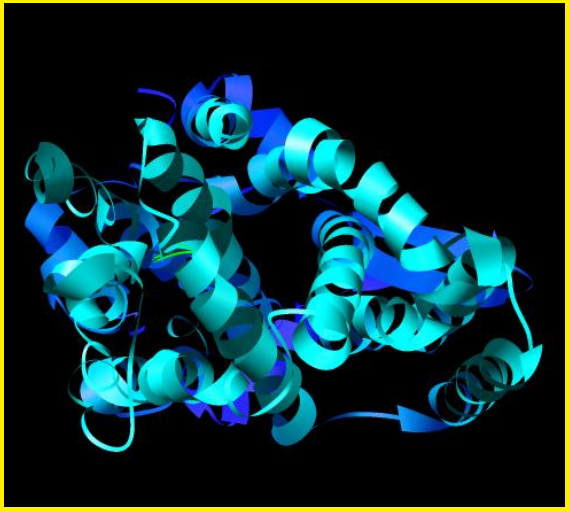


trimer, but substrate transport occurs within each monomer.

Na/Cl-dependent neurotransmitter transporter homolog (2A65)

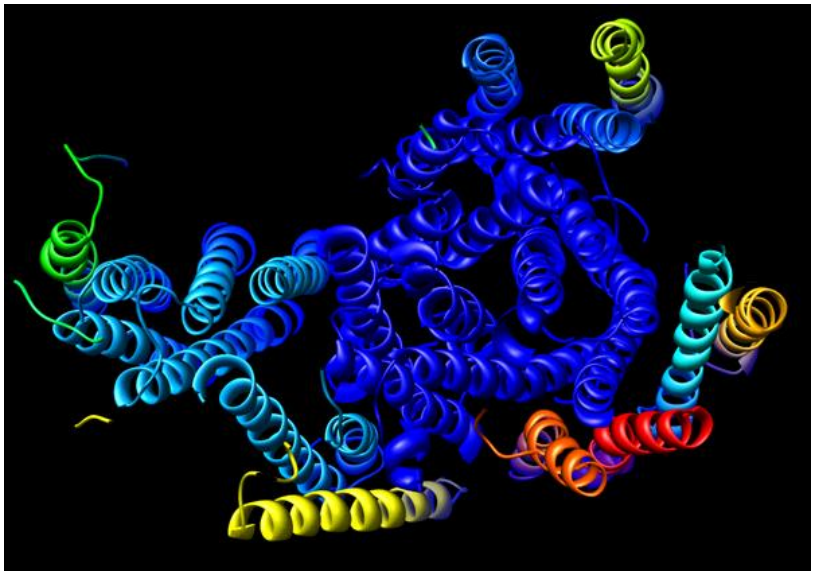


Na^+/H^+-antiporter (1ZCD)



monomer, asymmetric

Helical Membrane Proteins: Enzymes



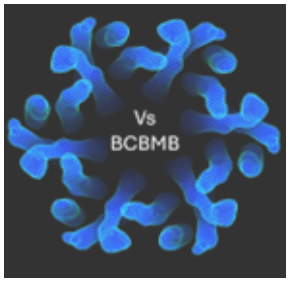
cytochrome c Oxidase (2OCC)
multiple subunits are indicated by separate colors

conclusions:

- transmembrane helices afford great flexibility in structural design

- there is no correlation between appearance and function
  (= not all channels are oligomers (eg KcsA vs AmtB)
  not all transporters are monomers (eg AmtB vs Glu-Transporter)
  not all membrane proteins with pseudo-twofold symmetry are channels (eg AQP vs LacY)

- association of multiple different membrane helical membrane proteins can give rise to highly complex membrane embedded macromolecular machines (eg Cytc-Oxidase)



Switching Gears: How Are Membrane Proteins Synthesized and Inserted into Membranes?

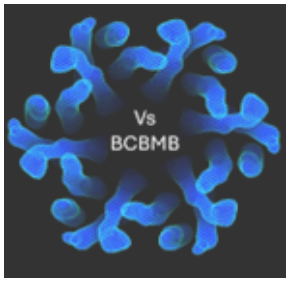
our discussion so far explains how polypeptides can be made to “happily” reside within a bilayer environment – but it does not explain how the membrane proteins are inserted into the membranes.

for our purposes, we will limit the discussion to helical membrane proteins because they are far more abundant than β–barrels. We also will keep this discussion at a basic level to keep this chapter at a reasonable length (if you are interested in learning more about this fascinating topic you can start here)

assume you have a polytopic membrane protein with 4TM segments

to synthesize this you need…..?

what is the one thing you do **NOT** want to do? Why?



Switching Gears: How Are Membrane Proteins Synthesized and Inserted into Membranes?

our discussion so far explains how polypeptides can be made to "happily" reside within a bilayer environment – but it does not explain how the membrane proteins are inserted into the membranes.

for our purposes, we will limit the discussion to helical membrane proteins because they are far more abundant than β–barrels. We also will keep this discussion at a basic level to keep chapter at a reasonable length (if you are interested in learning more about this fascinating topic you can start here)

Assume you have a polytopic membrane protein with 4TM segments

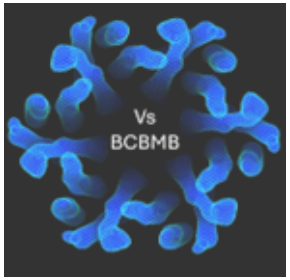
to synthesize this you need…..?

mRNA and ribosomes

what is the one thing you do **NOT** want to do? Why?

do **NOT** want to translate this in cytosol like a water-soluble protein because the TM segments would try to hide their hydrophobicity by aggregating! → no insertion into membrane

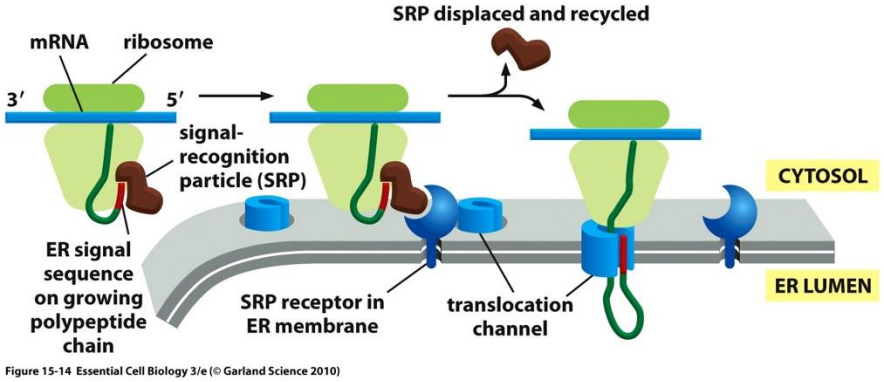
what is the implication of the last answer?

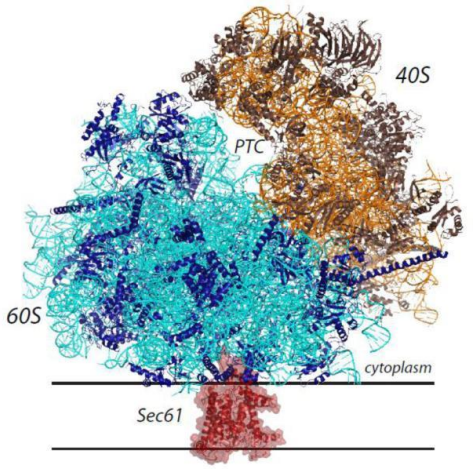


Membrane Protein Synthesis and Insertion

→ need a signal to tell the ribosome that, unlike a "regular protein", this protein should not be synthesized in the cytosol ➔ **Signal Sequence**

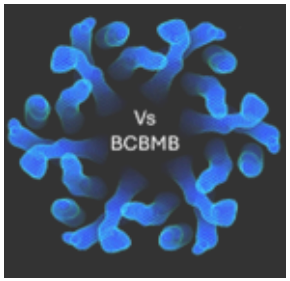
- signal Sequences are found in both secreted and membrane proteins.
- emergence of the signal sequence causes the ribosome to stall,
- stalled mRNA:ribosomes complex is recruited to the ER surface and transferred to the "translocon" (called Sec61 in eukaryotes)





Cymer F, Von Heinje G and White SH (2015)

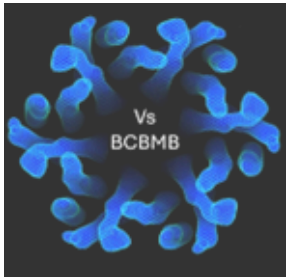
an experimentally determined structure of a Sec61:ribosome complex (shown to the right) gives you an idea of the size differences between key components in this process



Membrane Protein Synthesis and Insertion

once engaged to the translocon, protein synthesis resumes – and with that, what problems present themselves?

…try to think about the logistics of the process from the perspective of slide 72….



Membrane Protein Synthesis and Insertion

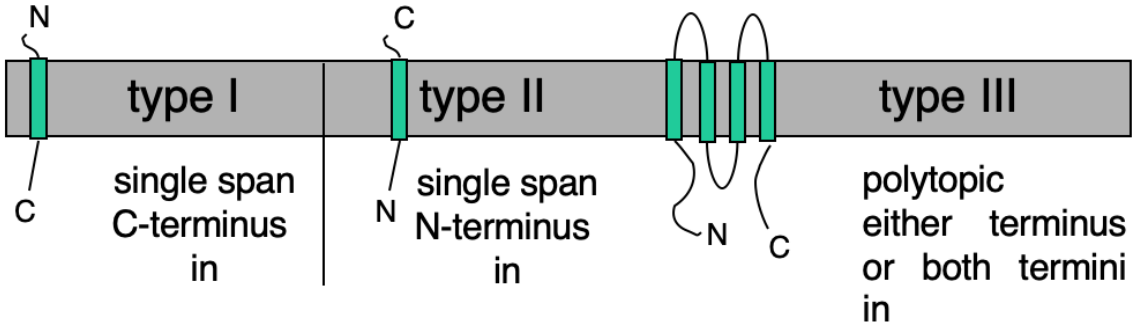
once engaged to the translocon, protein synthesis resumes – and with that, what problems present themselves?

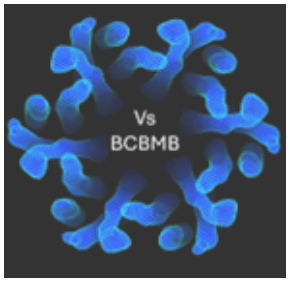
Answers:

(1) need to establish the membrane protein's topology = place N-terminus on the correct side of the membrane

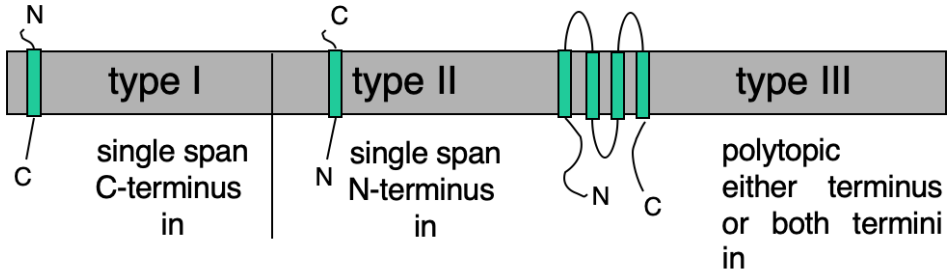
(2) need to somehow supervise co-translational membrane insertion of membrane spanning segments.

why are these the key challenges?





Membrane Protein Synthesis and Insertion



Answer: (1) need to establish the membrane protein's topology = place N-terminus on the correct side of the membrane; **(2)** need to somehow supervise co-translational membrane insertion of membrane spanning segments.

Why are these the key challenges?

responding to Answer (1) above

(1) incorrect topology will yield a non-functional protein

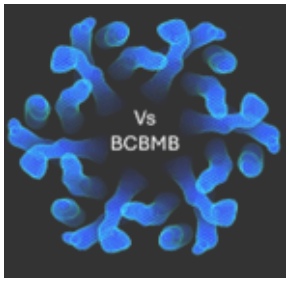
(eg in receptors, a wrong topology will misplace hormone/ligand binding site as well as the surfaces for intracellular interaction partners; in transporters access sites for molecules to be transported or domains that use ATP to generate the energy needed for the transport process would be on the wrong side of membrane; in channels: functionally important elements like voltage sensors would be in the wrong position (see Membrane Transport Chapter));

responding to Answer (2) above

(2) membrane spanning regions of TM-segments are typically 20-24 amino acids long → synthesis proceeds @ up to 8 aa/second (in eukaryotes) ➔ **translocon has ~2-3 seconds to evaluate** any given stretch of amino acids **and** – if called for - **initiate co-translational membrane insertion** by some means.

LET THAT THINK IN by snapping your fingers every 2 seconds …. any mistake, ever, and the protein is "dead"

how are these two major challenges met?



Membrane Protein Synthesis and Insertion

Answers

topology of single span membrane proteins is resolved through structural signals that reside within the signal sequence itself (eg clustering of positive charges at the beginning of the signal sequence will result in N-terminal disposition to the cytoplasmic side of the membrane).

for polytopic membrane proteins it's …..

a STOP and GO

(colors chosen to match the cartoon on the next slide)

how exactly translocons recognize TM-segments is not completely understood.

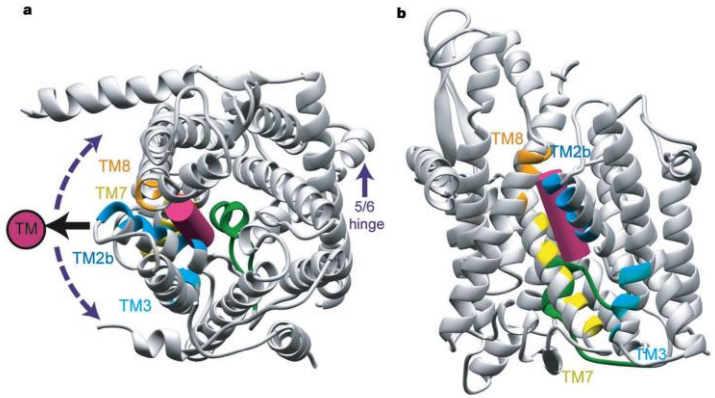
the going hypothesis is that translocons “compute” the overall hydrophobicity of helical segments that enter the translocon channel in a pre-folded state (the ribosomal exit tunnel is wide enough for TM-helices to fold prior to exiting from the ribosome).

- helices hydrophobic enough to be a TM-segment, serve as a series of **“internal signal sequence"** and **”stop transfer”** sequences.
- presence of internal signal and stop transfer sequences trigger a lateral gating mechanism that releases the preformed helix into the membrane.
- in addition, stop transfer sequences also trigger a gating mechanism that closes the channel preventing translocation of additional polypeptide into the ER-lumen.

The cartoon on the next slide helps digesting this ….

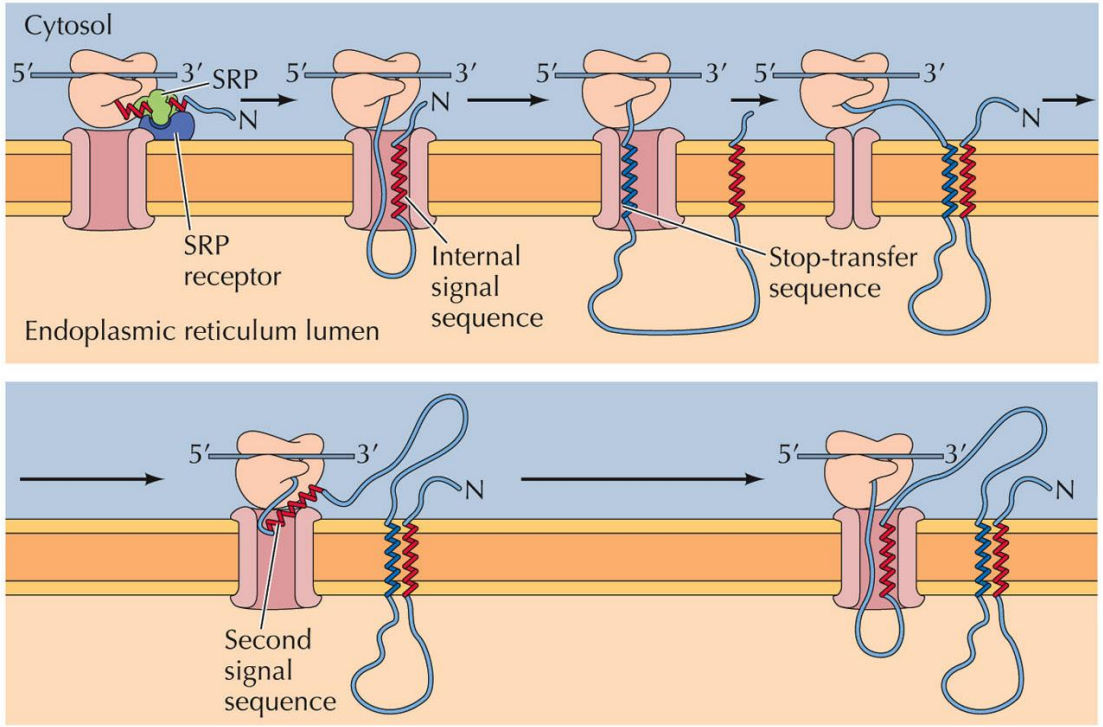
Membrane Protein Insertion Involves A Lateral Gate in the Translocon

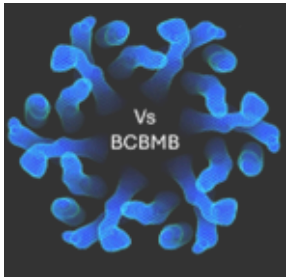
the gate opens laterally to let a helix escape into the bilayer. helices thus released stay close to the translocon awaiting release of their "packing partners" to sequentially assemble the bundle



Van den Berg et al (2004)
Nature. 427:36-44

Polytopic Membrane Protein Insertion = STOP and GO





Membrane Proteins SUMMARY

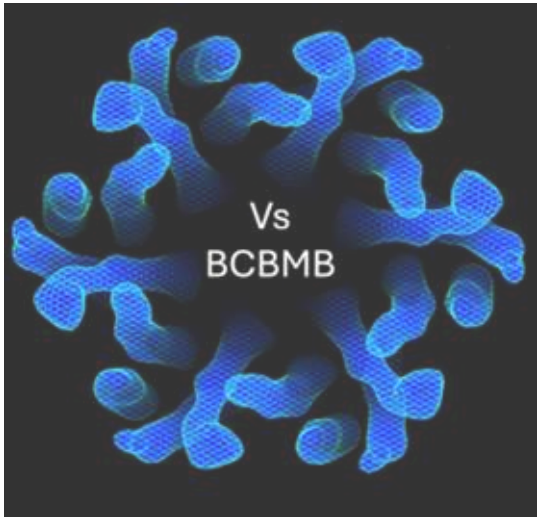
- membrane proteins fall into two groups: peripheral membrane proteins (loosely associated with membrane) and integral membrane proteins
- integral membrane proteins are either inserted into just one bilayer leaflet (typically the intracellular one) or completely traverse the bilayer. The latter are called "transmembrane proteins"
- transmembrane proteins can have a spread of topologies (= relation between N-/C-terminus as well as inside/outside of cell)

at a structural level,

- transmembrane proteins belong to one of two classes: α–helical (bundles) or β–barrels
- in both classes, the hydrophobic effect enforces secondary structure inside the hydrophobic membrane core. In helical membrane proteins **(only),** this allows inclusion of normally helix breaking amino acids into the structure (always structurally/functionally relevant)
- because of the H-bonding pattern, α–helices are self sufficient and can occur as single transmembrane spans. In contrast, single β–strands cannot stably exist in the bilayer core, a minimum of 8 β–strands is required to form a closed β–barrel
- the structural design of β–barrels is simpler than that of helix bundles because the formation of the β–barrel occurs entirely at the level of secondary structure, while helix bundles require intricate helix:helix packing interfaces at the tertiary structure level.

at a functional level,

integral membrane proteins serve many functions as structural components, enzymes, channels, transporters and signaling receptors



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