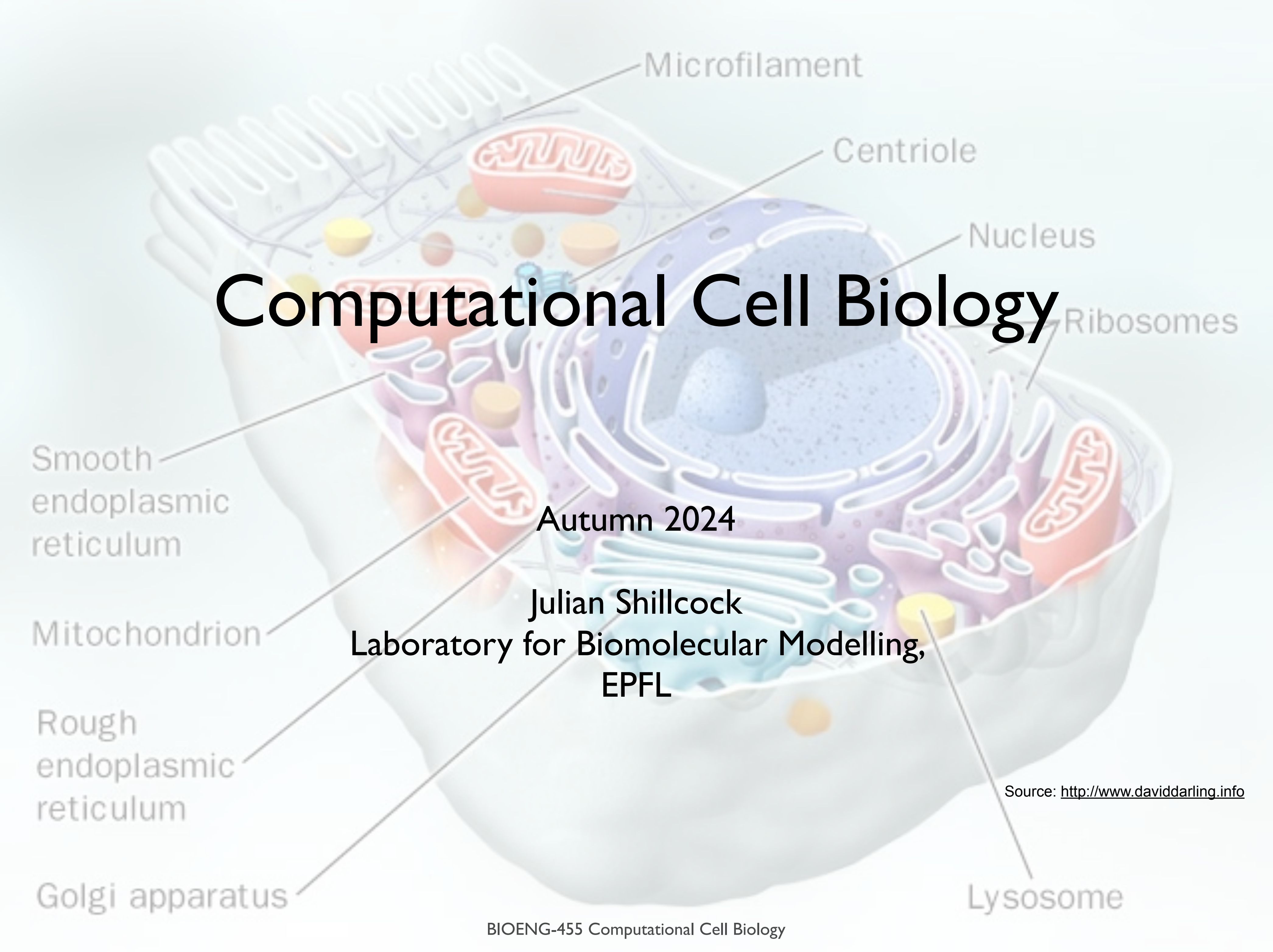
Computational Cell Biology

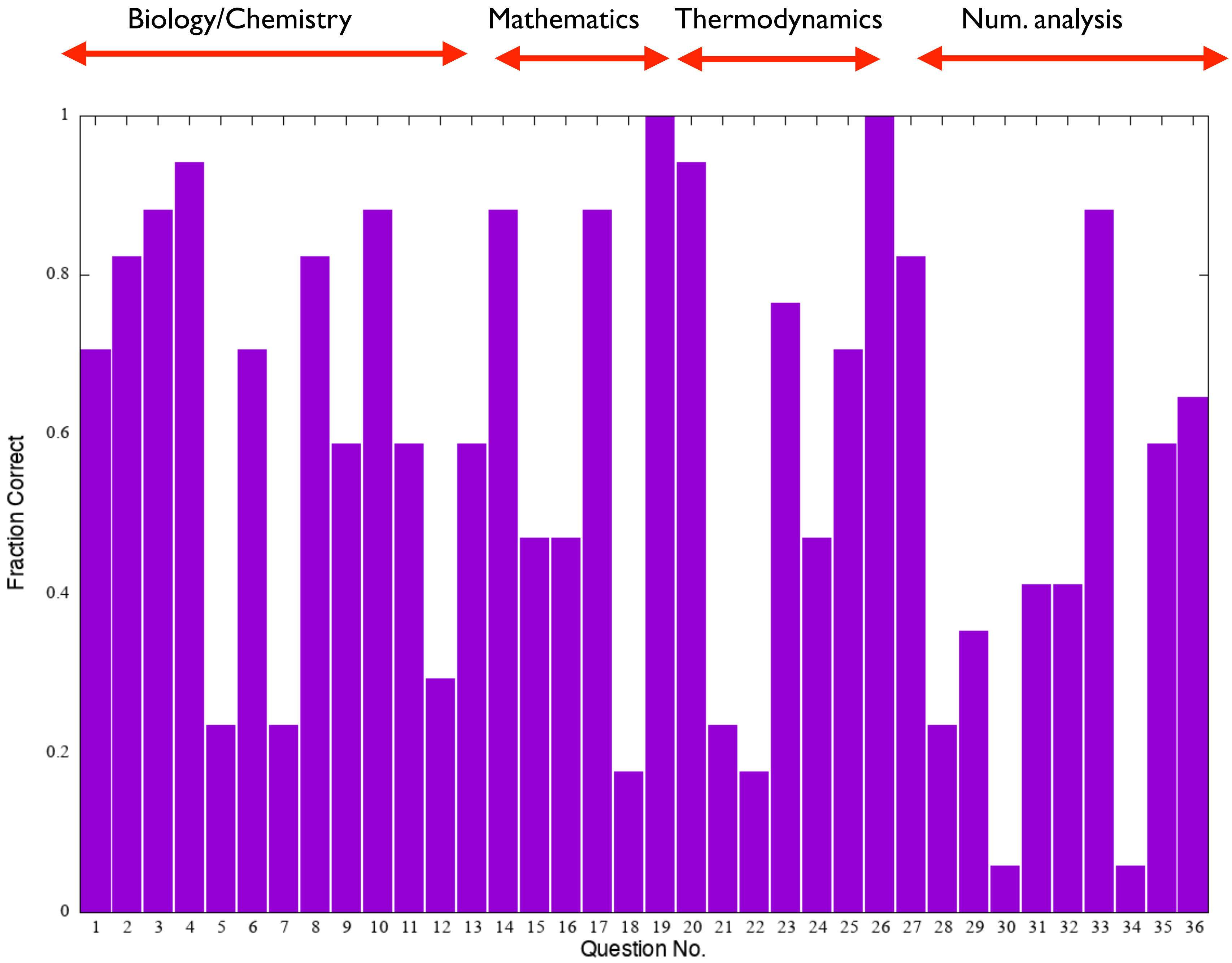
Autumn 2024

Julian Shillcock
Laboratory for Biomolecular Modelling,
EPFL



Source: http://www.daviddarling.info

Background Quiz Histogram



Biology/Chemistry

Mathematics

Thermodynamics

Num. analysis

Core Concepts

Small/folded molecules are dominated by energy

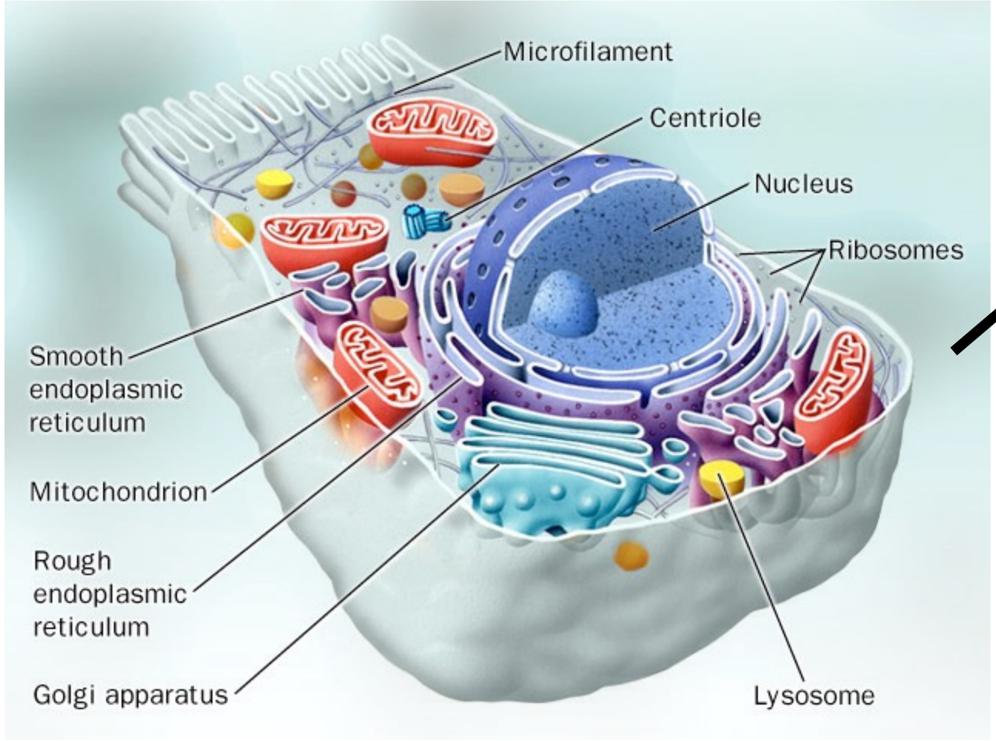
Polymers are dominated by entropy

Biopolymers (Proteins, lipids, DNA, RNA) are more interesting than manufactured polymers

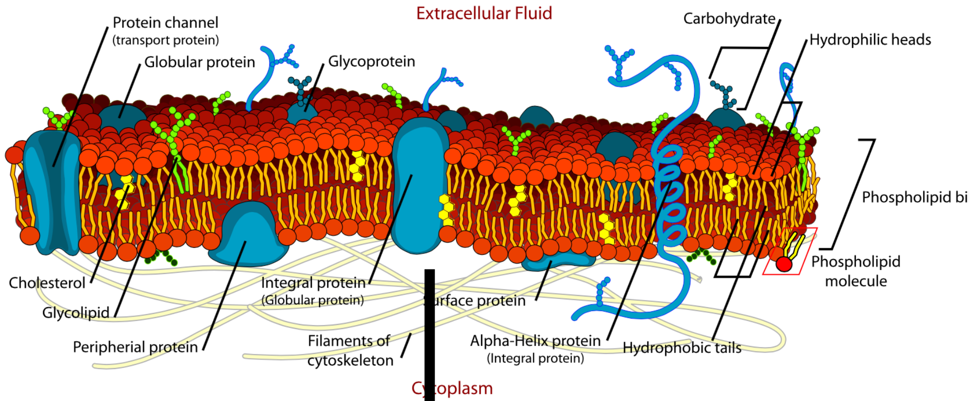
Many small interactions can produce a large force

mentalfloss.com

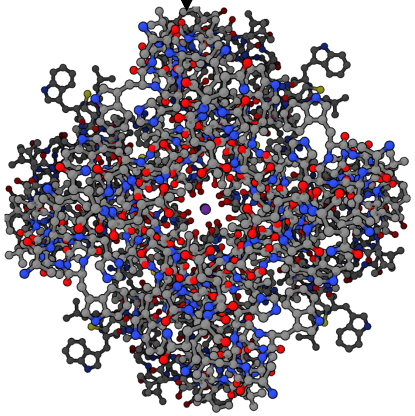
Macromolecules in the cell



www.daviddarling.info



U. Ghent



wikipedia.org

Apart from water, ions, small molecules like ATP, the cell contains many supramolecular structures - membranes, organelles, protein complexes, DNA/chromatin, etc.

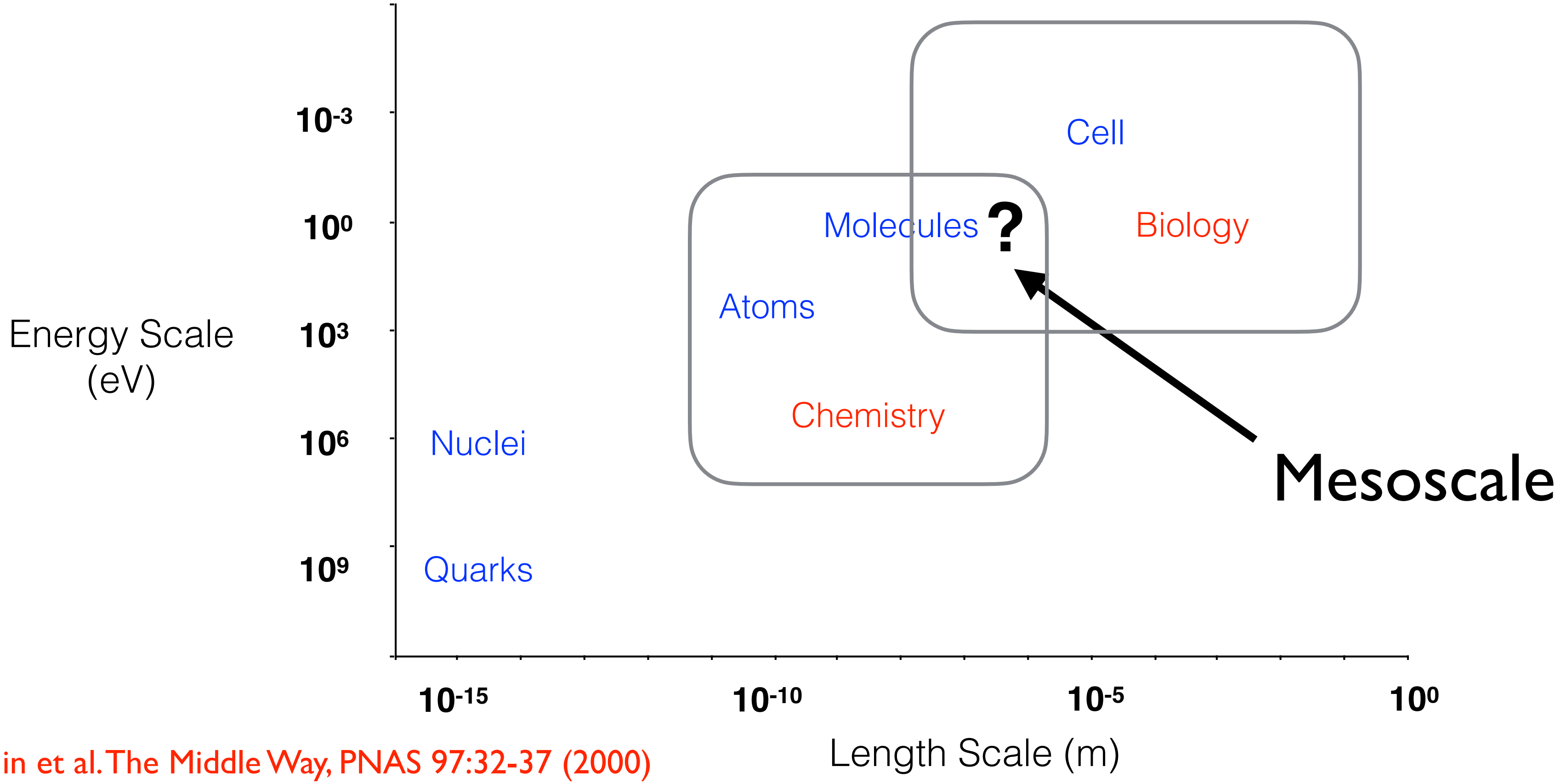
Each structure is composed of smaller structures, which are composed of smaller…. down to molecules and atoms in a hierarchy of structural units on different length scales

Hierarchical structure is why science works

Science works because the world appears to be able to be divided into different levels - defined by relevant length and energy scales - in which entities move and interact according to forces that only refer to, and act on, entities in that layer.

Is there something new in the mesoscale between molecules and cells?

Microscopy and simulations are both exploring this regime (nm < mesoscale < micron)



R. B. Laughlin et al. The Middle Way, PNAS 97:32-37 (2000)

Polymers and Macromolecules

Industrial polymers have been in use since ~ 1945, and consist of simple repeated units:

Polyethylene $(C_2H_4)n$
Polyethyleneglycol (PEG) $(CH_2\text{-}CH_2\text{-}O)_n$

Industrial polymers are important, but dull: plastic bags, door knobs, ….

Biological Polymers

DNA/RNA - linear chain of nucleotides

Lipids - approx. 2 - 4 nm - membranes

Sugars - 2 dim. networks

Proteins - linear chains of amino acids

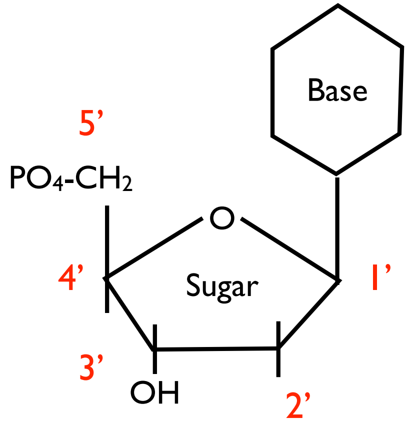
Filament	Monomer	Location	Diameter	Mass
DNA	A, C, G, T Base	Nucleus	2 nm	0.65 kDa
Actin	Actin monomer	cytoplasm	7 nm	43 kDa
Intermediate filaments	α-helical rod	cytoplasm	10 nm	70 - 200 kDa
Microtubules	Tubulin monomer	cytoplasm	25 nm	50 kDa

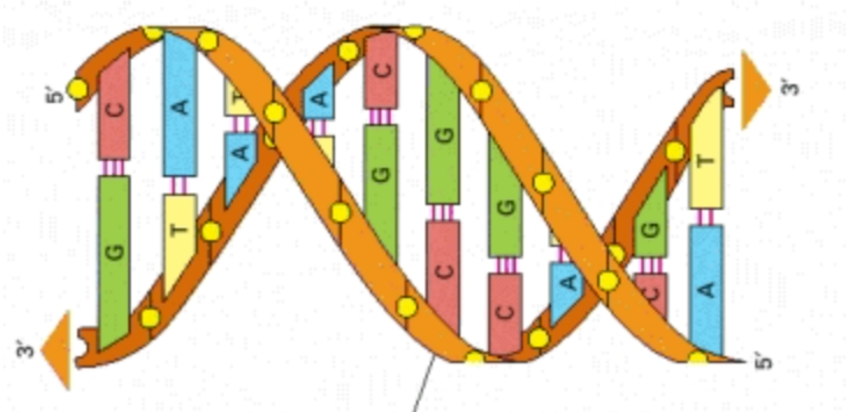
Small, folded proteins can also self-assemble into polymers, e.g., actin filaments, microtubules

Why are biological polymers so much more interesting than industrial polymers?

DNA

Double-stranded helix of nucleotides connected in linear chains





Alberts and Bray et al.

Nucleotide = 5-carbon sugar with an attached base (A, C, G, T) on the 1’ carbon, and a phosphate group on the 5’ carbon.

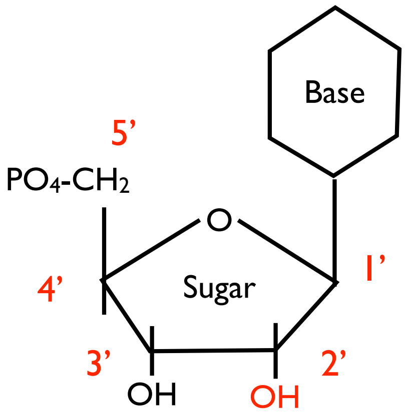
Base = N-containing ring compound, either purine or pyrimidine

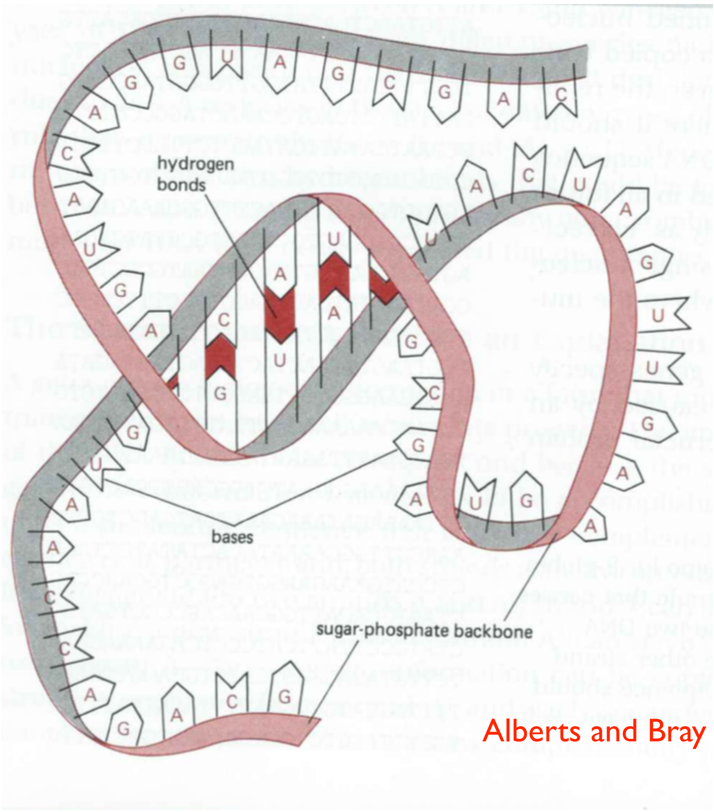
Chain is formed by phosphodiester bonds at 3’ and 5’ carbons, so the polymer has a direction or *polarity*

First reason biopolymers are interesting: the monomers have structure that carries information (industrial polymers are usually one monomer, e.g., CH_2, repeated)

RNA

Single-stranded linear chain of nucleotides - only difference in the nucleotide to DNA is the OH group on 2’ carbon - and T is replaced by U





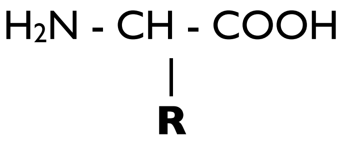
Alberts and Bray et al.

RNA is an intermediate between DNA and proteins (Central dogma: DNA —> RNA —> Protein)

But RNA is also a structural part of cellular machines like Ribosomes, it can catalyse reactions, and has some complementary binding.

Proteins

Linear polymers of amino acids directionally connected by peptide bonds from the amine end (N terminus) to the carboxylic acid end (C terminus):



$$\mathrm{H_2N - \underset{\underset{\displaystyle \mathbf{R}}{|}}{CH} - COOH}$$

R can be: Acidic, Basic, Uncharged polar, Non-polar

Proteins fold into 3D shapes that minimise their energy by arranging charged polar/acidic/basic side-chains to cover the surface and shield non-polar side-chains from water.

Folded state is not covalently bonded, there must be many weak non-covalent interactions to stabilise the folded state agains thermal fluctuations.

Proteins can de-nature if the solvent conditions change (pH, temperature, other proteins, …)

Name	Symbol	Type
Alanine	A	Non-polar
Cysteine	C	Non-polar
Aspartic acid	D	Acidic
Glutamic acid	E	Acidic
Phenylalanine	F	Non-polar
Glycine	G	Non-polar
Histidine	H	Basic
Isoleucine	I	Non-polar
Lysine	K	Basic
Leucine	L	Non-polar
Methionine	M	Non-polar
Asparagine	N	Uncharged polar
Proline	P	Non-polar
Glutamine	Q	Uncharged polar
Arginine	R	Basic
Serine	S	Uncharged polar
Threonine	T	Uncharged polar
Valine	V	Non-polar
Tryptophan	W	Non-polar
Tyrosine	Y	Uncharged polar

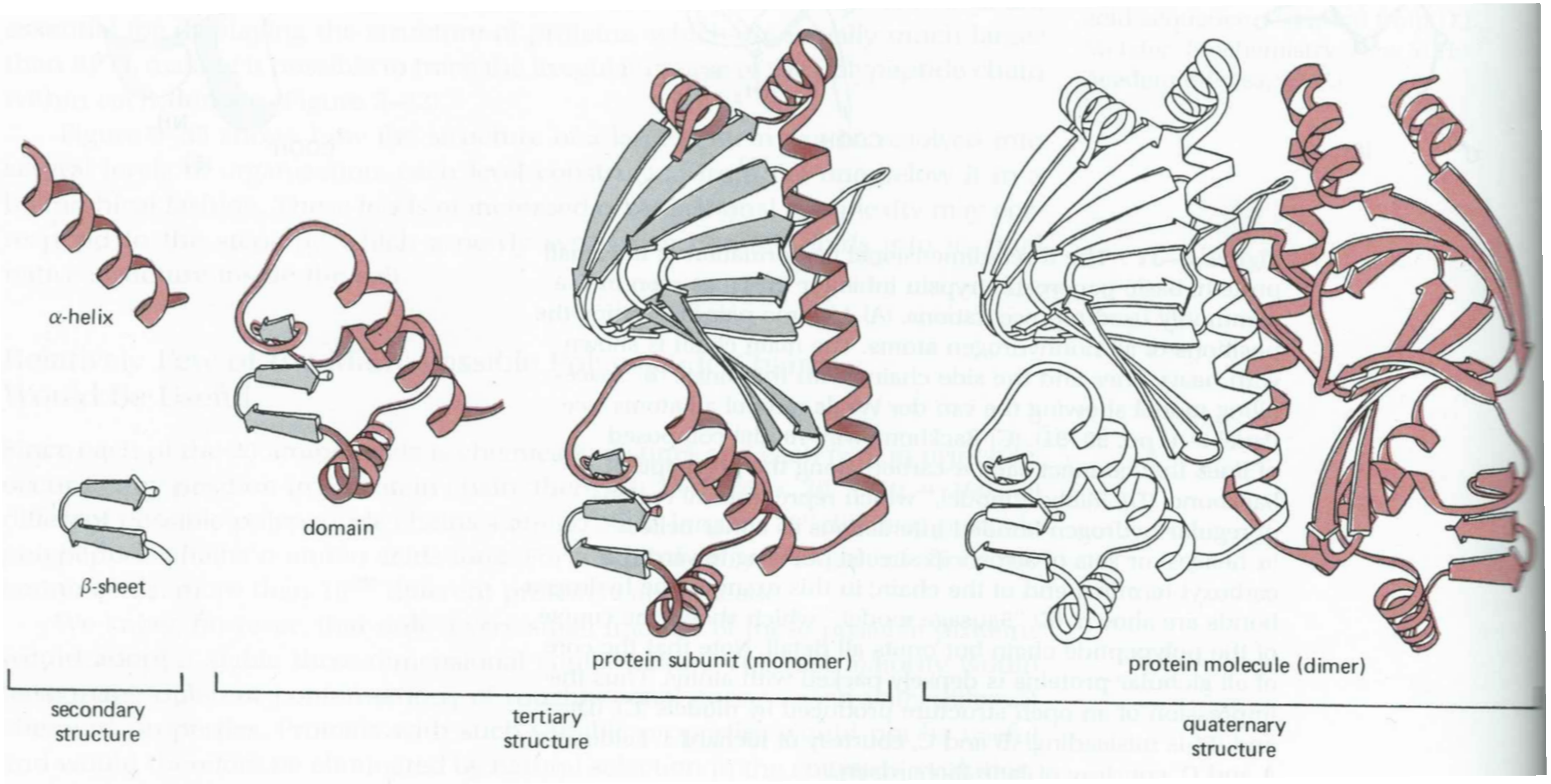
3D structure of proteins

Primary structure = sequence of amino acids

Secondary structure = H-bonding of contiguous aa's into α helices and β sheets

Tertiary structure = domains of globular units

Protein complexes = assembly of several proteins by non-covalent bonds



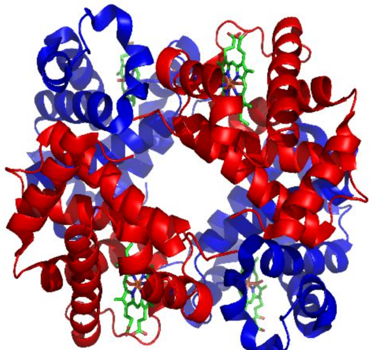
Alberts and Bray et al.

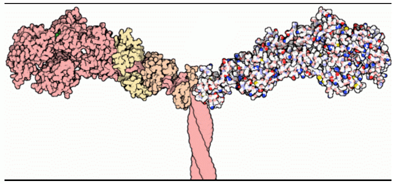
Second reason: biological polymers fold into shapes to perform functions (not just random flapping about)

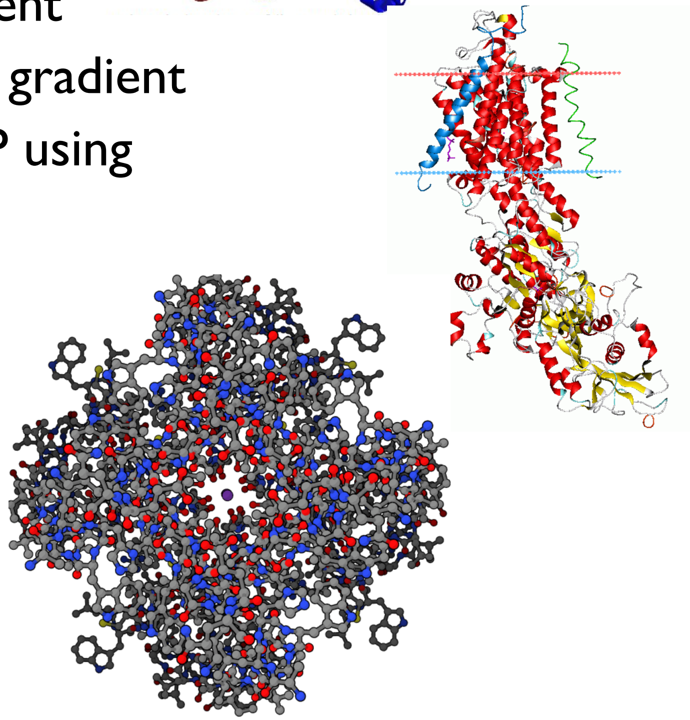
Many proteins are molecular machines

What are some common proteins?

- Haemoglobin - binds oxygen
- Actin - forms filaments, cytoskeleton
- Myosin motor - pulls cargo along a filament
- Na-K ATPase - pump ions against gradient
- Ion channels - allow ions to flow down gradient
- Respiratory chain proteins - makes ATP using energy of e- transport







wikipedia.org

Machines because they are precisely arranged (mutate one aa and it can destroy the protein) and they operate relatively independently of other proteins

Proteins fold into precise, unique, lowest-energy states that are known from crystal structures: but is this true of all proteins?

Think-Pair-Share - 3 minutes

One person argue for, one against, and then discuss which you most agree with.

Proteins are coded for by genes, and genes evolve by random mutations, which may influence the proteins' structure and function; but if the function is critical for life, the protein is generally conserved across species.

Q. If a protein with a given number of residues appears in many species and performs the same function, is the sequences of amino acids (largely) conserved?

Lipids

Lipids are fatty acids that are composed of:

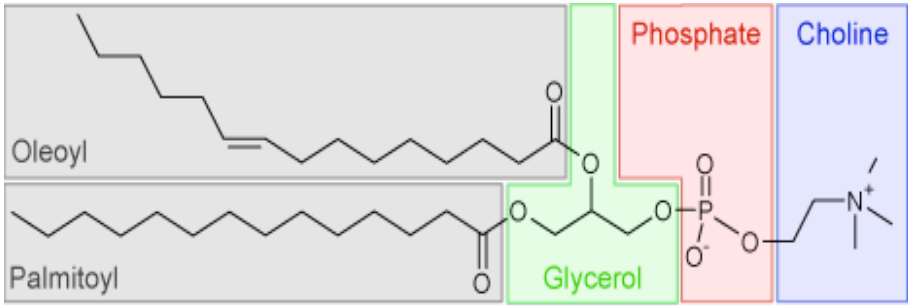
Headgroup - soluble in water, may be charged, uncharged, polar, small or bulky

Oily tails - C-C chains with typical length 8 - 24, saturated or unsaturated, only soluble in organic solvents

Lipids are called:

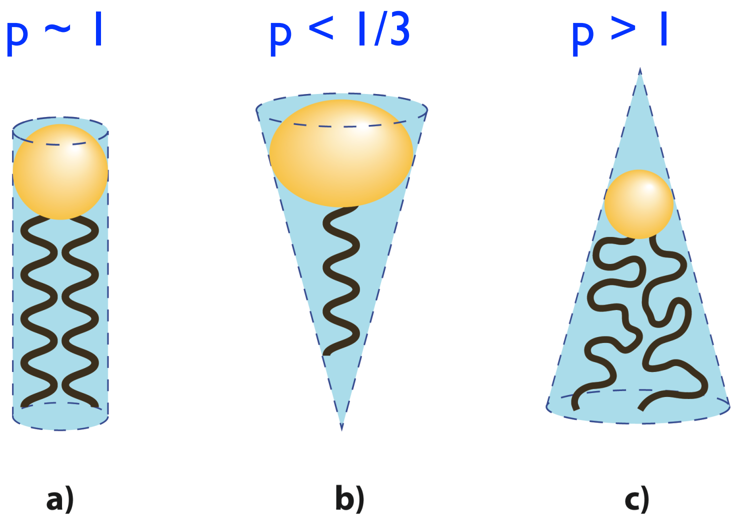
Amphiphiles = love water and oil - frustrated in either bulk solution

Typical phospholipids look like POPC below: palmitoyl-oleoyl phosphatidylcholine



Lipids do not covalently bond to each other, but self-assemble into distinct 3d structures driven by their frustration: **tails want to hide from water, headgroups want to be solvated.**

Lipids have a *shape* controlled by relative size/interactions of their headgroup and tails



J. C. Shillcock, Fig. 3, Ch. 26 in Biomolecular Simulations,
ed. L. Monticelli and E. Salonen, Methods in Mol. Biol. 924, Humana Press 2012

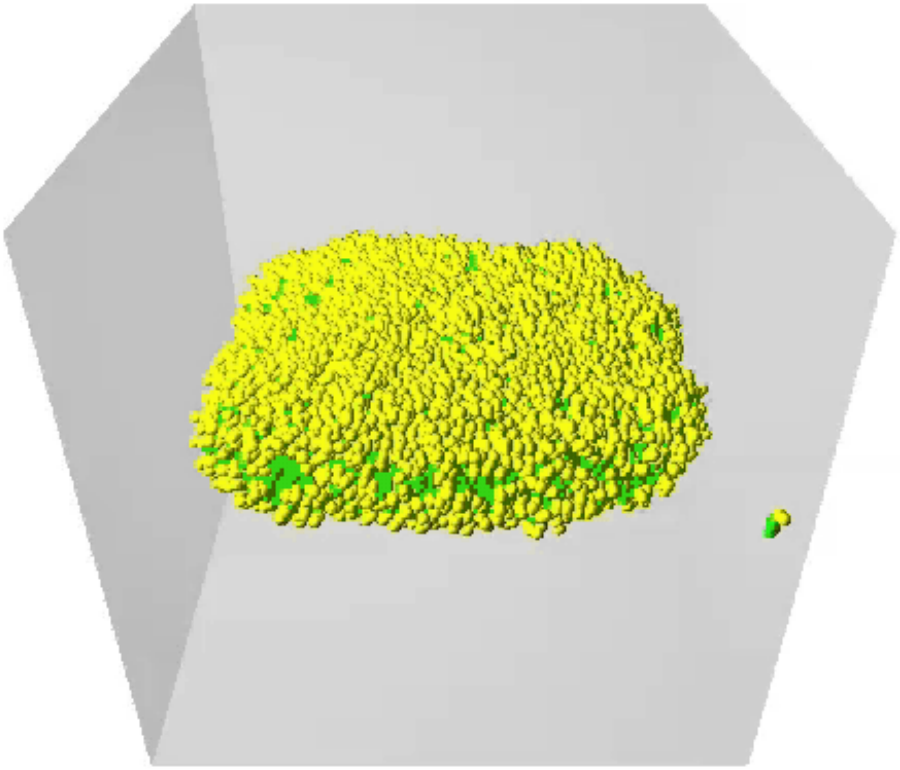
If the headgroup has the same "cross-sectional area" as the tails, the molecule is like a cylinder; otherwise the molecule is like a cone.

Israelachvili devised a *packing parameter* to quantify the "shape" of lipids in the fluid phase:

$$p = v / a_0 l_c$$

v = equilibrium volume of the molecule (depends on environment, temperature, etc!)
a_0 = equilibrium cross-sectional area of the molecule (" ")
l_c = maximum extension of the hydrocarbon chains (see Table 1 in Lecture 1)

Lipids have many phases



DPD simulation of 3428 lipids in a planar disk: the lipids on the edge have tails exposed to water; the patch bends so that it can close off its edge and forms a vesicle (water is invisible).

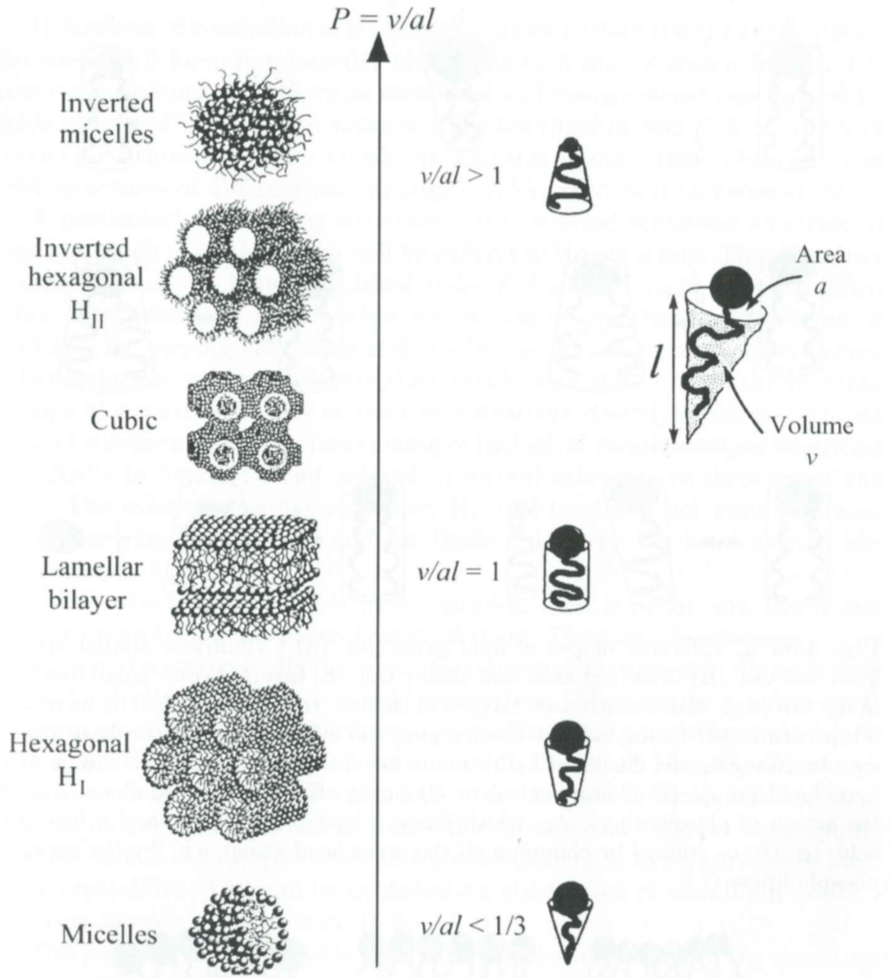


Fig. 4.4. Schematic illustration of lamellar and non-lamellar lipid aggregates formed in water. The different structures have different curvature and are arranged in accordance with the value of packing parameter $P = v/al$

Third/fourth reasons: biopolymers non-covalently form aggregates that can respond to their environment by changing shape or structure.

Life as a Matter of Fat,
O. G. Mouritsen, Springer, Berlin 2005

What's special about lipids?

Lipids are often viewed as just providing a bounding membrane around cells that separates interior from exterior and is a solvent for proteins.

But they have important active roles too:

Signalling - endocannabinoids and steroid hormones

Disease - important lipid-binding proteins: PLAs (inflammation), ApoE (Alzheimer's), alpha synuclein, PUFAs (Parkinson's)

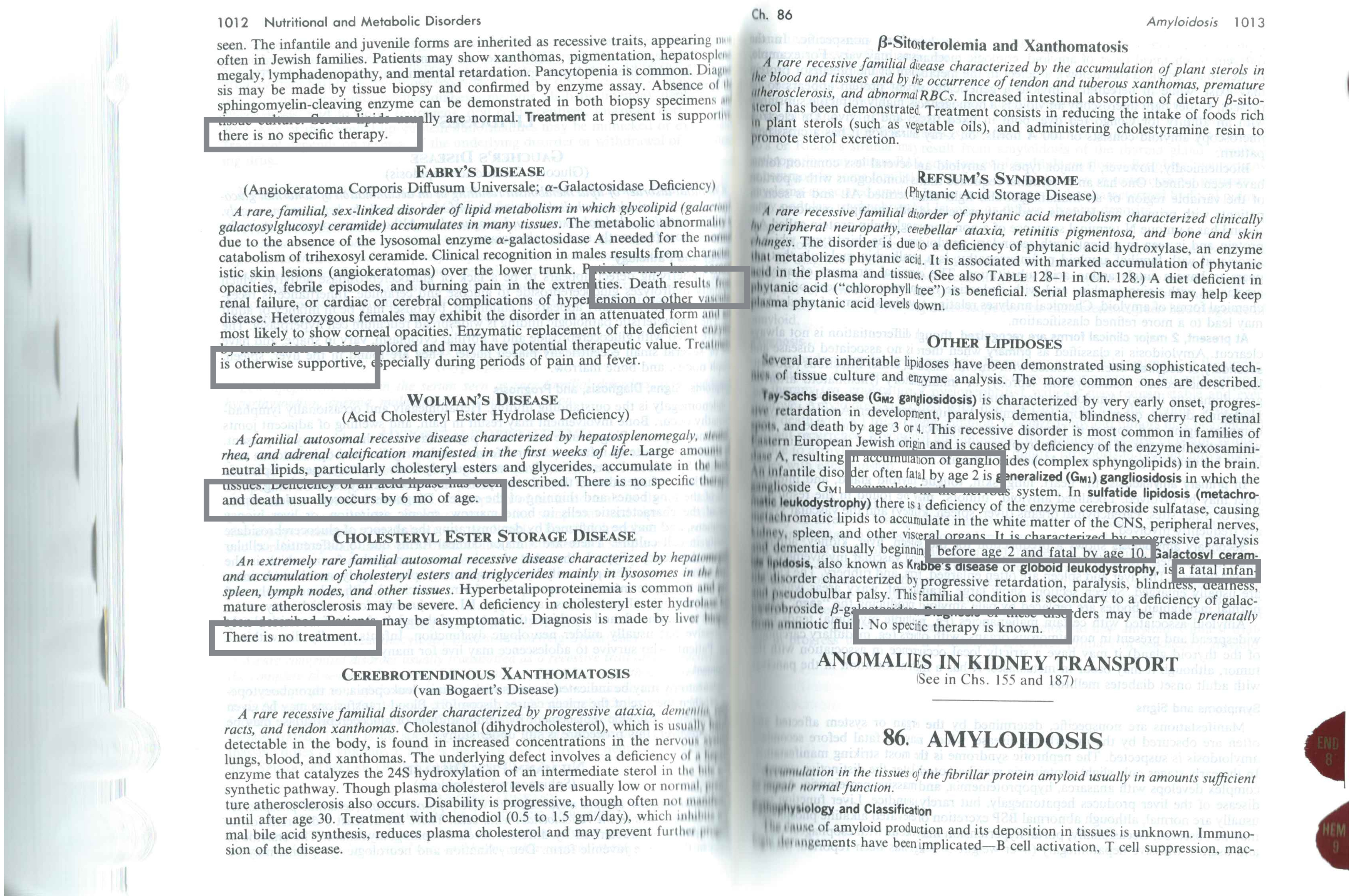
Synaptic dynamics - modify ion channel dynamics, receptor currents

B. Davletov and C. Montecucco, Lipid function at synapses. Curr. Op. Neurobiol. 20:543 (2010)

Why are there so many different types of lipid in the body?

- If all nature needed was a barrier, a few lipid types would be sufficient - but there are 10,000s of different types of lipid in cell membranes (cp. 4 bases in DNA and 20 amino acids)
- Membrane is a fluid so all components could diffuse freely - but different proteins need distinct environments due to thickness, tension, and their function depends on local lipid composition: membrane is not an ideal mixture

Lipid-related diseases are fatal



seen. The infantile and juvenile forms are inherited as recessive traits, appearing most often in Jewish families. Patients may show xanthomas, pigmentation, hepatosplenomegaly, lymphadenopathy, and mental retardation. Pancytopenia is common. Diagnosis may be made by tissue biopsy and confirmed by enzyme assay. Absence of the sphingomyelin-cleaving enzyme can be demonstrated in both biopsy specimens and tissue culture. Serum lipids usually are normal. **Treatment** at present is supportive; there is no specific therapy.

FABRY'S DISEASE

(Angiokeratoma Corporis Diffusum Universale; α-Galactosidase Deficiency)

A rare, familial, sex-linked disorder of lipid metabolism in which glycolipid (galacto galactosylglucosyl ceramide) accumulates in many tissues. The metabolic abnormality is due to the absence of the lysosomal enzyme α-galactosidase A needed for the normal catabolism of trihexosyl ceramide. Clinical recognition in males results from characteristic skin lesions (angiokeratomas) over the lower trunk. Patients may have co opacities, febrile episodes, and burning pain in the extremities. Death results fro renal failure, or cardiac or cerebral complications of hypertension or other vascu disease. Heterozygous females may exhibit the disorder in an attenuated form and most likely to show corneal opacities. Enzymatic replacement of the deficient enzy by transfusion is being explored and may have potential therapeutic value. Treatment is otherwise supportive, especially during periods of pain and fever.

WOLMAN'S DISEASE

(Acid Cholesteryl Ester Hydrolase Deficiency)

A familial autosomal recessive disease characterized by hepatosplenomegaly, stea rhea, and adrenal calcification manifested in the first weeks of life. Large amounts neutral lipids, particularly cholesteryl esters and glycerides, accumulate in the tissues. Deficiency of an acid lipase has been described. There is no specific therapy and death usually occurs by 6 mo of age.

CHOLESTERYL ESTER STORAGE DISEASE

An extremely rare familial autosomal recessive disease characterized by hepato and accumulation of cholesteryl esters and triglycerides mainly in lysosomes in the spleen, lymph nodes, and other tissues. Hyperbetalipoproteinemia is common and mature atherosclerosis may be severe. A deficiency in cholesteryl ester hydrolase been described. Patients may be asymptomatic. Diagnosis is made by liver bi There is no treatment.

CEREBROTENDINOUS XANTHOMATOSIS

(van Bogaert's Disease)

A rare recessive familial disorder characterized by progressive ataxia, dementia racts, and tendon xanthomas. Cholestanol (dihydrocholesterol), which is usually detectable in the body, is found in increased concentrations in the nervous lungs, blood, and xanthomas. The underlying defect involves a deficiency of a enzyme that catalyzes the 24S hydroxylation of an intermediate sterol in the bile synthetic pathway. Though plasma cholesterol levels are usually low or normal, ture atherosclerosis also occurs. Disability is progressive, though often not manif until after age 30. Treatment with chenodiol (0.5 to 1.5 gm/day), which inhibits mal bile acid synthesis, reduces plasma cholesterol and may prevent further pro sion of the disease.

β-Sitosterolemia and Xanthomatosis

A rare recessive familial disease characterized by the accumulation of plant sterols in the blood and tissues and by the occurrence of tendon and tuberous xanthomas, premature atherosclerosis, and abnormal RBCs. Increased intestinal absorption of dietary β-sitosterol has been demonstrated. Treatment consists in reducing the intake of foods rich in plant sterols (such as vegetable oils), and administering cholestyramine resin to promote sterol excretion.

REFSUM'S SYNDROME

(Phytanic Acid Storage Disease)

A rare recessive familial disorder of phytanic acid metabolism characterized clinically by peripheral neuropathy, cerebellar ataxia, retinitis pigmentosa, and bone and skin changes. The disorder is due to a deficiency of phytanic acid hydroxylase, an enzyme that metabolizes phytanic acid. It is associated with marked accumulation of phytanic acid in the plasma and tissues. (See also TABLE 128–1 in Ch. 128.) A diet deficient in phytanic acid ("chlorophyll free") is beneficial. Serial plasmapheresis may help keep plasma phytanic acid levels down.

OTHER LIPIDOSES

Several rare inheritable lipidoses have been demonstrated using sophisticated techniques of tissue culture and enzyme analysis. The more common ones are described.

Tay-Sachs disease (G_{M2} gangliosidosis) is characterized by very early onset, progressive retardation in development, paralysis, dementia, blindness, cherry red retinal spots, and death by age 3 or 4. This recessive disorder is most common in families of Eastern European Jewish origin and is caused by deficiency of the enzyme hexosaminidase A, resulting in accumulation of gangliosides (complex sphyngolipids) in the brain. An infantile disorder often fatal by age 2 is **generalized (G_{M1}) gangliosidosis** in which the ganglioside G_{M1} accumulates in the nervous system. In **sulfatide lipidosis (metachromatic leukodystrophy)** there is a deficiency of the enzyme cerebroside sulfatase, causing metachromatic lipids to accumulate in the white matter of the CNS, peripheral nerves, kidney, spleen, and other visceral organs. It is characterized by progressive paralysis and dementia usually beginning before age 2 and fatal by age 10. **Galactosyl ceramide lipidosis,** also known as **Krabbe's disease** or **globoid leukodystrophy**, is a fatal infantile disorder characterized by progressive retardation, paralysis, blindness, deafness, and pseudobulbar palsy. This familial condition is secondary to a deficiency of galactocerebroside β-galactosidase. **Diagnosis** of these disorders may be made *prenatally* from amniotic fluid. No specific therapy is known.

ANOMALIES IN KIDNEY TRANSPORT

(See in Chs. 155 and 187)

---

86. AMYLOIDOSIS

Accumulation in the tissues of the fibrillar protein amyloid usually in amounts sufficient to impair normal function.

Pathophysiology and Classification

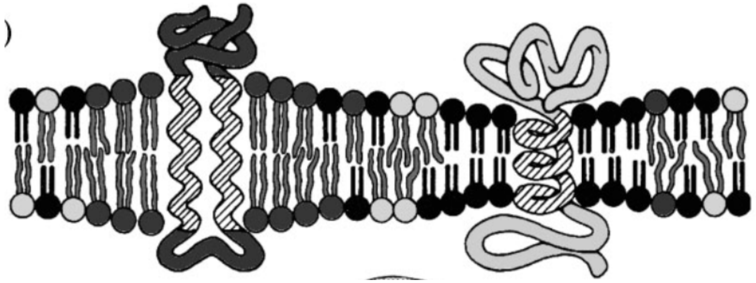
The cause of amyloid production and its deposition in tissues is unknown. Immunologic derangements have been implicated—B cell activation, T cell suppression, mac-

Merck Manual, 15th ed. https://en.wikipedia.org/wiki/Merck_Manual_of_Diagnosis_and_Therapy

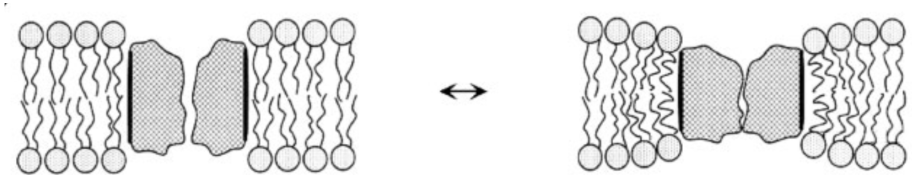
Lipids influence membrane material properties

Thickness, bending modulus, stretch modulus, fluidity

Protein hydrophobic region must match the membrane **thickness** **(or pay high energy cost)**, hence they prefer to be surrounded by lipids with matching thickness



Channel proteins can have their open/closed equilibrium changed by **local membrane constituents**



Membrane **bending stiffness** (or degree of saturation of lipids) controls their shape; flexible lipids can form curved membranes, vesicles, tubes, rigid lipids form flat bilayers;

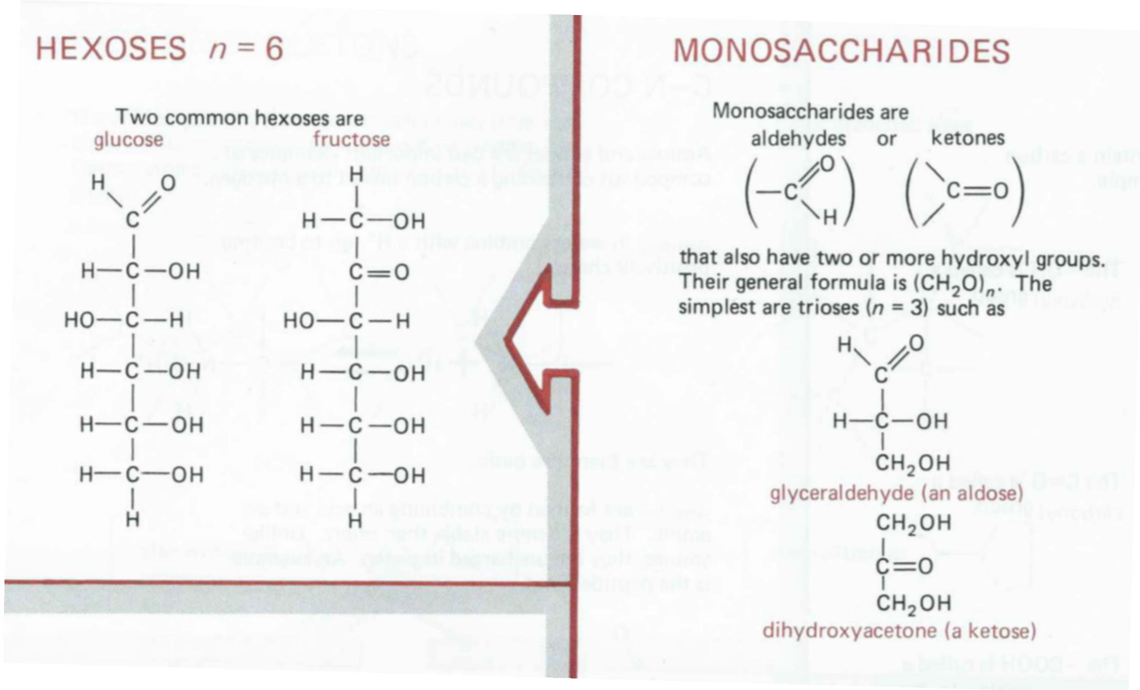
Stretch modulus is high which maintains surface area constant

Fluidity of membrane components allows cell to create/remove lipids that associate by diffusion; transient domains can form to aid signalling, tension can be relieved by lipid flow

Sugars

Monosaccharides have formula $C_n-H_{2n}-O_n$

Glucose is a hexose - $C_6-H_{12}-O_6$
Ribose is a pentose - $C_5-H_{10}-O_5$



HEXOSES $n = 6$

Two common hexoses are glucose fructose

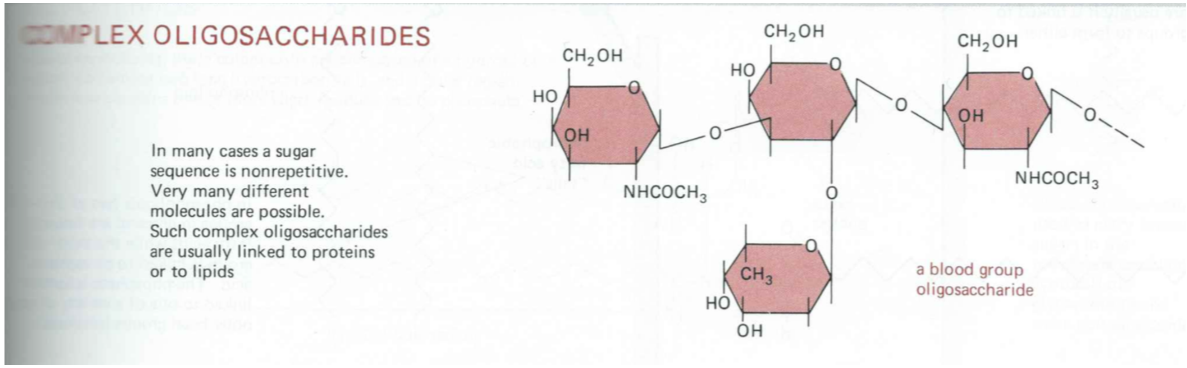
MONOSACCHARIDES

Monosaccharides are aldehydes or ketones that also have two or more hydroxyl groups. Their general formula is $(CH_2O)_n$. The simplest are trioses ($n = 3$) such as

glyceraldehyde (an aldose)

dihydroxyacetone (a ketose)

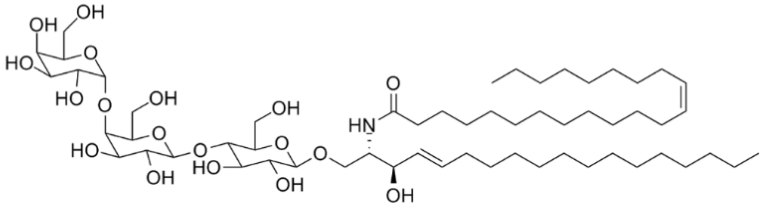
Complex oligosaccharides often bind to proteins or lipids (glycoprotein, glycolipid) to encode information in the immune system, signalling, pathogen entry to cells.



COMPLEX OLIGOSACCHARIDES

In many cases a sugar sequence is nonrepetitive. Very many different molecules are possible. Such complex oligosaccharides are usually linked to proteins or to lipids

a blood group oligosaccharide

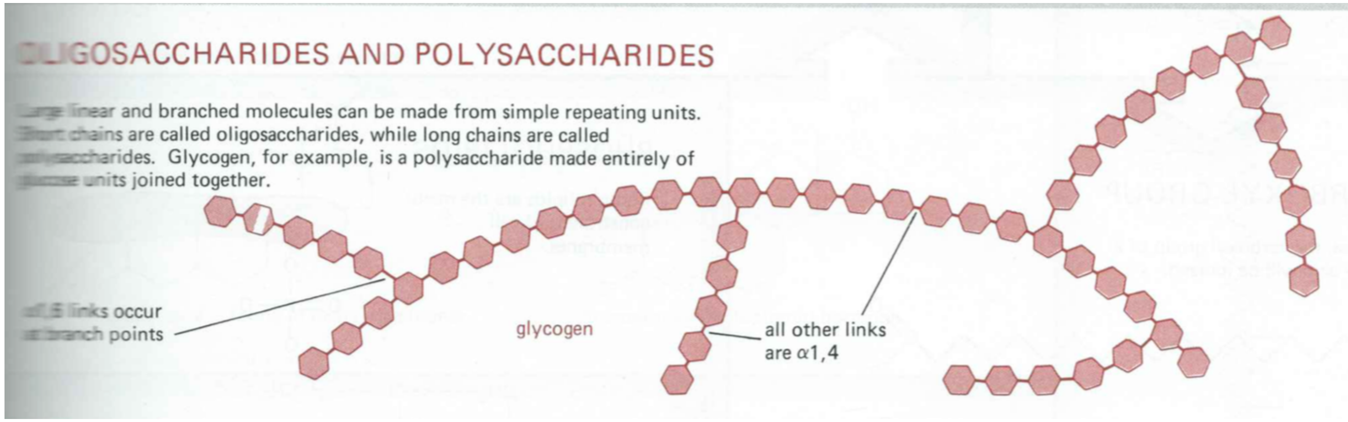


Shiga toxin binds to Gb3 to enter cell

Alberts and Bray

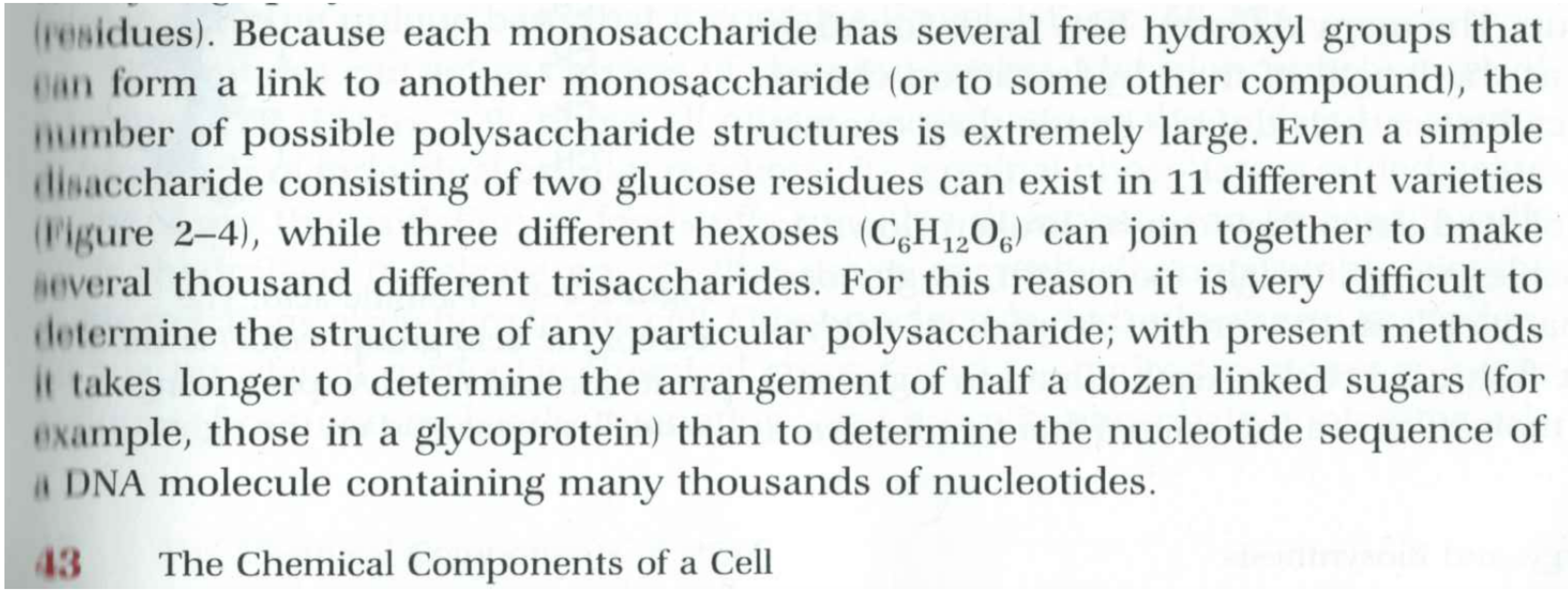
Sugars can be structurally complex

Sugars can be linear or ring-like, and can link to form long linear chains (e.g., glycogen), or branched networks (via OH groups).



OLIGOSACCHARIDES AND POLYSACCHARIDES

Large linear and branched molecules can be made from simple repeating units. Short chains are called oligosaccharides, while long chains are called polysaccharides. Glycogen, for example, is a polysaccharide made entirely of glucose units joined together.



(residues). Because each monosaccharide has several free hydroxyl groups that can form a link to another monosaccharide (or to some other compound), the number of possible polysaccharide structures is extremely large. Even a simple disaccharide consisting of two glucose residues can exist in 11 different varieties (Figure 2–4), while three different hexoses ($C_6H_{12}O_6$) can join together to make several thousand different trisaccharides. For this reason it is very difficult to determine the structure of any particular polysaccharide; with present methods it takes longer to determine the arrangement of half a dozen linked sugars (for example, those in a glycoprotein) than to determine the nucleotide sequence of a DNA molecule containing many thousands of nucleotides.

43 The Chemical Components of a Cell

Alberts and Bray

Macromolecules form complex aggregates

DNA/RNA - helical chains, DNA (2 nm) has hierarchical folding of histones (11 nm), packed nucleosomes (30 nm), chromosome (300 nm); RNA makes supramolecular machines with proteins

Proteins - rigid rods, floppy polymers, globular structures, non-covalently bonded aggregates, RNA/protein complexes

Lipids - micelles, membranes, vesicles, lipid-protein aggregates (e.g., LDL, HDL)

Sugars - flexible chains, 2 dim. networks, bonded to proteins/lipids to form glycoproteins/glycolipids

Biopolymers are interesting because:

- monomers carry information
- spontaneously fold into a 3d shape
- non-covalently self-assemble into 1d, 2d, and 3d aggregates
- respond to their environment by changing shape

Are all proteins folded?

Do all proteins fold into lowest-energy states? ~20 years ago answer would have been yes.

Common picture of proteins was precisely-folded molecular *machines* that made lock-and-key binding with ligands or other proteins. Specific sequences of amino acids determined who bonded to whom, and that was it.

Structure-Function relationship dominates this picture.

Very mechanical - very Newtonian - not thermodynamically accurate

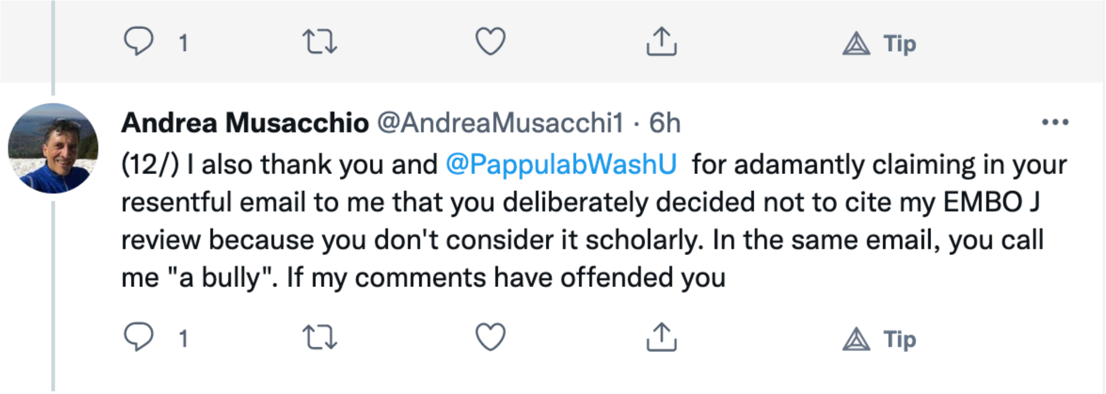
There are other types of protein - *intrinsically-disordered proteins (IDP)*- that have no stable structure because of their amino acid sequence, but do exhibit biological activity, and play a fundamental role in organising cellular dynamics: we'll look at them next week and lectures 12/13.

Examples

- Casein in milk, disordered structure makes digestion easier
- Calcineurin activates immune T cells on Ca binding, needs accessible binding sites, and surrounds its binding partner
- PSD-95, scaffold protein in neuronal spines

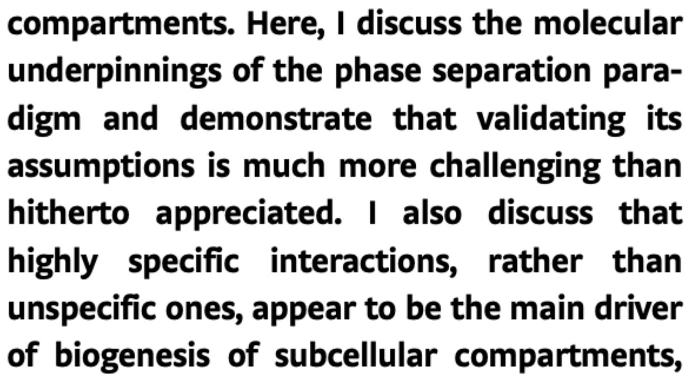
Some references on intrinsically-disordered proteins (including a controversy …)

- R. van der Lee et al., Chemical Reviews 114:6589-6631 (2014) (see here for non-conserved sequences with the same elastic properties)
- C. J. Oldfield and A. K. Dunker, Ann. Rev. Biochem. 83:553-584 (2014)
- H-X Zhou et al. , Trends Biochem. Sci. 43:499-516 (2018)
- V. Marx, Nature Methods 17:567-570 (2020)
- J. Z. Zhang et al. Trends Pharma. Sci. 42:845-856 (2021)
- A. Musacchio, EMBO Journal 41:e109952 (2022) - very interesting review that summarises a current debate about the relevance of in vitro studies to in vivo.



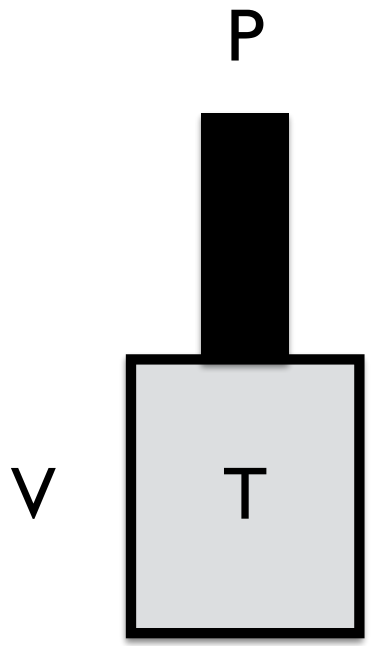
Andrea Musacchio @AndreaMusacchi1 · 6h

(12/) I also thank you and @PappulabWashU for adamantly claiming in your resentful email to me that you deliberately decided not to cite my EMBO J review because you don't consider it scholarly. In the same email, you call me "a bully". If my comments have offended you

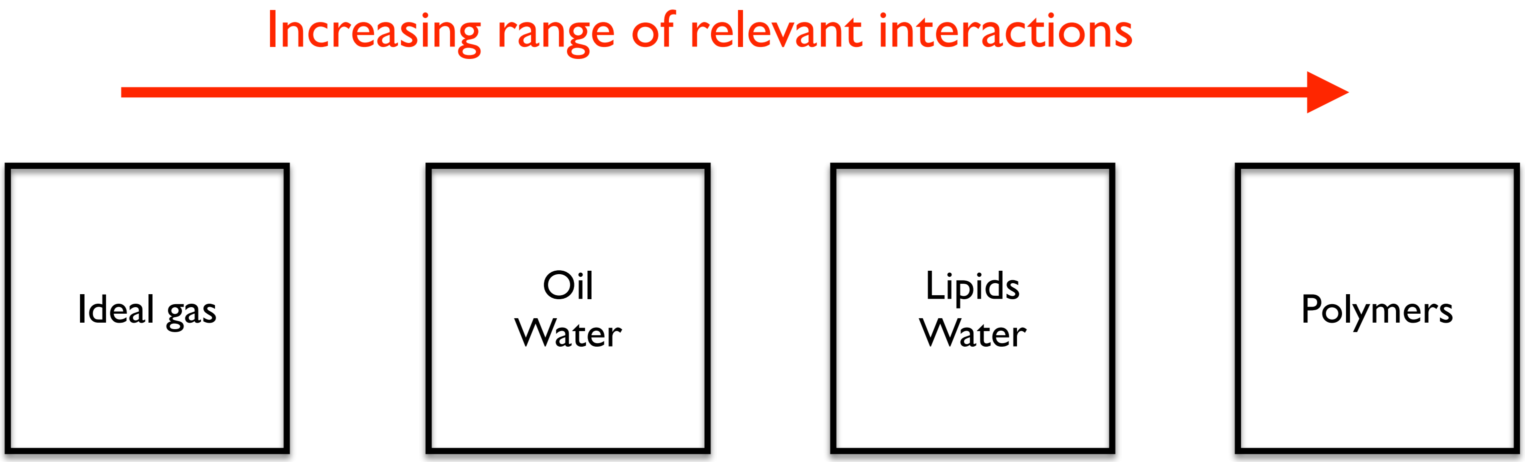

compartments. Here, I discuss the molecular underpinnings of the phase separation paradigm and demonstrate that validating its assumptions is much more challenging than hitherto appreciated. I also discuss that highly specific interactions, rather than unspecific ones, appear to be the main driver of biogenesis of subcellular compartments,

Constructing a model of a polymeric fluid

What is it made of?
How do its molecules interact?
How many degrees of freedom does it have?
How many d.o.f are uncorrelated?
What is the **density of the d.o.f in space?**
How many measurable physical properties does it have?



$pV = Nk_BT$



Increasing range of relevant interactions

Ideal gas

Oil
Water

Lipids
Water

Polymers

Blackboard derivations

During the course, I derive some results on the blackboard.

These are typically fundamental results that are useful in building models in cell biology

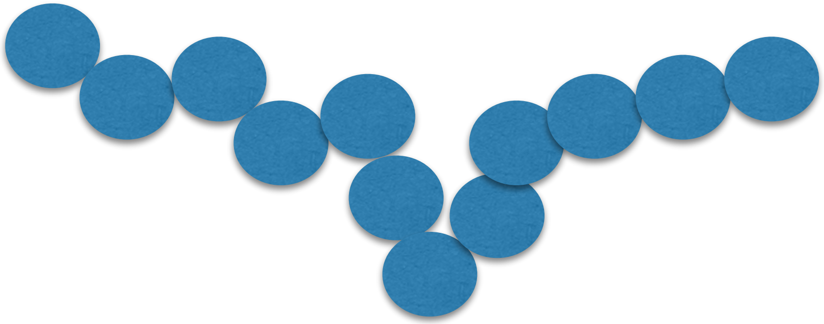
The results may appear in the tests

Q. Why don't I just put the equations on the slides?

Simple model of flexible polymers

Recall from last week that the total energy of a cellular system is shared among all its atoms and molecules (Equipartition theorem); here we seen a consequence of that:

- Continual thermal motion of all monomers in a polymer ($k_BT > 0$)
- Impossibility of completely isolating one part from another (unlike a car)



Consider a polymer as a chain of (identical) monomers bonded together in a linear chain

Monomers move around freely (subject to remaining connected) with no energy cost, so there are an enormous number of configurations available to the molecule

Huge conformation space means entropy dominates polymers' behaviour, and the identity of the monomers is relatively unimportant beyond a few nm (i.e., monomer diameters); explains why stretches of aa sequence in IDPs are not conserved: they do not control their function

Freely-jointed chain polymer

Polymer is a chain of N+1 identical monomers connected in a linear chain by bonds of fixed length **a** that are free to point in any direction independently of all others (aka *ideal* chain, *phantom* chain):

The conformation of the chain is given by N+1 position vectors: $\{\mathbf{R}\} = (\mathbf{R}_0, \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \ldots \mathbf{R}_N)$ or, alternatively, by the N bond vectors $\mathbf{r}_i = \mathbf{R}_i - \mathbf{R}_{i-1}$:

$$\{\mathbf{r}_i\} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \ldots \mathbf{r}_N)$$

The size of the polymer can be represented by the end-to-end length:

$$\mathbf{R}_{ee} = \mathbf{R}_N - \mathbf{R}_0 = \sum \mathbf{r}_i$$ where the sum is over i = 1 … N

By assumption, $< \mathbf{r}_i > = 0$, but $< \mathbf{R}_{ee}{}^2 >$ is not zero. (Blackboard calculation)

$$< R_{ee}^2> = N.a^2 + < \sum r_i \, . \, r_j >$$ where the sum is over i≠ j = 1 … N.

which for the freely-jointed chain reduces to:

$$< R_{ee}^2> = N.a^2$$

as $< r_i \, . \, r_j > = 0$ because there are **no correlations** between monomers (inc. no steric repulsion)

Quantifying the size of a polymer

$$< R_{ee}^2> = N.a^2 + < \sum r_i . r_j >$$

The behaviour of long real polymers often does NOT depend on the chemical details of the monomers, but only on their length (and self-avoidance); the second term represents correlations between monomer positions due to interactions, and is often zero or small.

We can extend the model by assuming a form for $< \sum r_i . r_j >$, e.g., bond angles are fixed, bond angles have a distribution, etc. NB Sum is over all pairs not just adjacent monomers.

The *Characteristic Ratio* of a polymer is defined as:

$$C_N = < R_{ee}^2> / N.a^2$$

Centre of mass of a polymer (with N monomers):

$$\mathbf{R}_{\mathbf{cm}} = 1/N \sum \mathbf{R}_i$$

Radius of gyration of polymer:

$$<\mathbf{R}_{\mathbf{g}}^{\mathbf{2}}> = 1/N \sum (\mathbf{R}_i - \mathbf{R}_{cm})^2 = 1/2N^2 \sum_{i \neq j} \mathbf{R}_{ij}^2 \qquad \mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$$

The ideal chain has: $< R_{ee}^2> = N.a^2$, so $C_N = 1$, and $<R_g^2> \sim 1/6. N.a^2$ as N goes to infinity.

Self-avoiding polymers

We know for a freely-jointed chain polymer that:

$$< R_{ee}^2 > = N.a^2$$

But real polymers cannot pass through each other, so this is only an approximation.

If we impose the condition that monomers cannot intersect each other in space, we can no longer calculate the static properties of the polymer by summing over all conformations as the self-avoidance introduces non-local correlations into monomer positions.

But we can generate *self-avoiding walks* (SAW) on a computer and count all configurations of N monomers. We find that self-avoiding polymers satisfy the relation: (Note: dimensions and scaling)

$$< R_{ee}^2 > = N^{2\nu}.a^2$$

The exponent ν is called the *Flory Exponent*, and is 1/2 for freely-jointed chain polymers, and has the following value in d dimensions for SAWs:

$$\nu = 3 / (d + 2)$$

Soft Condensed Matter Physics in Molecular and Cell Biology, WCK Poon and D Andelman, Taylor and Francis, USA, 2006.

Think-Pair-Share - 5 minutes

One person argue for, one against, and then discuss which you most agree with.

Q1 Modelling a real polymer like DNA as a "phantom chain" is useless because real polymers can never intersect themselves, which is the main property of the phantom chain, so you never get good agreement between model and experiment.

Q2 If you have only the two lengths associated with a polymer — its contour length **L**, and mean square end-to-end length $\mathbf{<R_{ee}^2>}$. How many different ways can you define a new length that is smaller than L?

What are the formulas for these new lengths?

Relating a real polymer to a phantom chain

How do we relate the scaling law $< R_{ee}^2> = N.a^2$ to a real polymer?

Is **N** the same as the number of monomers?

Is the monomer size equal to **a**? (No!)

Define a new length (the *Kuhn length*) $\mathbf{l_k}$ as the ratio of the real polymer's $<R_{ee}^2>$ to its contour length L, and N_k as the ratio of the contour length to this new length:

$$l_k = <R_{ee}^2> / L$$

$$N_k = L / l_k$$

Therefore

$$< R_{ee}^2> = N_k . l_k^2$$

So, a real polymer is equivalent to a RW with step size l_k and number of steps equal to the number of *Kuhn lengths* in its contour length.

Because real monomers are highly correlated in space (backbones have some stiffness) and succeeding steps in a RW must be uncorrelated, the Kuhn length is roughly the distance required for a segment of the polymer to be uncorrelated in direction with the previous segment.

Visualising the Kuhn length

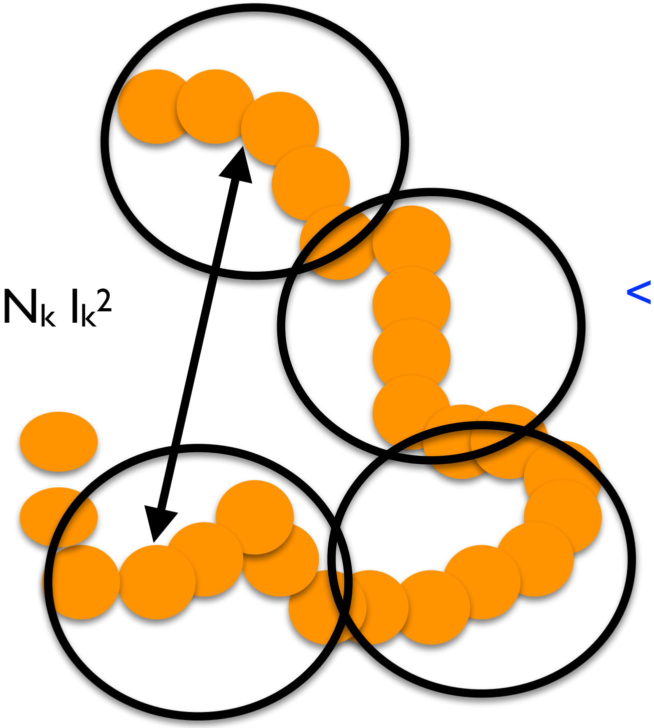
What is the relation between a real polymer and the equivalent phantom one?

A real polymer has *correlations* between adjacent monomers- some *stiffness* along the chain

A phantom polymer has *no correlations* (aka freely-jointed chain)

But if the real polymer is sufficiently long, we can represent it — to some degree of accuracy — as a phantom chain with a smaller number of monomers each of which is larger: this is the Kuhn length.

$< R_{ee}^2> = N_k \, l_k^2$

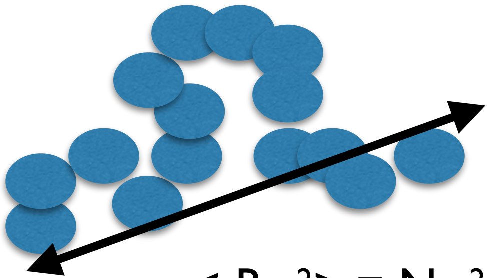


$l_k = <R_{ee}^2> / L$

$N_k = L / l_k$

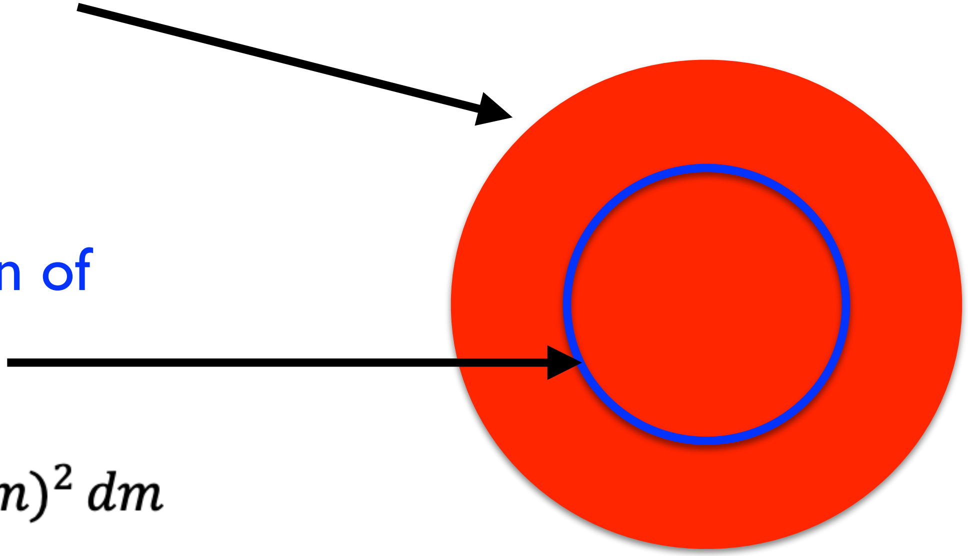
$< R_{ee}^2> = N_k \, . \, l_k^2$

Random walk



$< R_{ee}^2> = N.a^2$

How big is a hard sphere?



R_h = actual sphere radius

R_g ~ distribution of mass in space

$$M R_g^2 = \int r(m)^2 \, dm$$

e.g.,

Spherical shell: $R_g^2 = R^2$

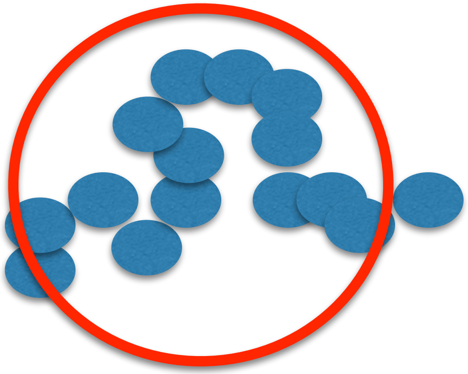
Solid sphere: $R_g^2 = 3/5 \, R^2$ which is smaller than for the shell because the interior mass pulls Rg to smaller values

How big is a polymer?

For a polymer we have the following "sizes":

$< R_{ee}^2> = N.a^2 + < \sum r_i . r_j >$ end to end length

$<R_g^2> = 1/N \sum (R_i - R_{cm})^2 = 1/2N^2 \sum R_{ij}^2$ radius of gyration



hydrodynamic radius

all are useful in some circumstances, but can be misleading

Example: PEG

We have

$$< R_{ee}^2 > = N_k \cdot l_k^2$$

and for polyethylene glycol (PEG) with a M.Wt of 20,000 Da, we have:

M.Wt ~ 20,000 Da
Monomer size (-CH_2CH_2O-) ~ 0.44 nm and mass 44 Da
Kuhn length = 1.8 nm

So, PEG has 20,000/44 = 455 real monomers, and contour length L = 455*0.44 = 200 nm.

There are $N_k = L / l_k = 200/1.8 = 111$ Kuhn lengths in the polymer.

So, we predict the polymer's mean size to be:

$$\sqrt{<R_{ee}^2>} = \sqrt{111} * 1.8 \text{ nm} \sim 19 \text{ nm}$$

And this is the size we would expect to measure in a light scattering experiment, for example.

Example: Bacterial DNA

We have

$$< R_{ee}^2 > = N_k \, . \, l_k^2$$

For the E. Coli chromosome:

Size = 4.64 Mbp
Monomer size (1 base pair) ~ 0.34 nm
Kuhn length = 100 nm

So, the DNA backbone length is $L = 4.64.10^6 * 0.34$ nm = $1.6.10^6$ nm = 1.6 mm!

There are $N_k = L / l_k = 1.6.10^6/100 = 16{,}000$ Kuhn lengths in the chromosome. So, we predict

$\sqrt{<R_{ee}^2>} = \sqrt{16{,}000} * 100$ nm ~ 12,600 nm = 12.6 microns > E. Coli itself (~ 2 microns)

The bacterium must use active packing to store its DNA as a simple collapse of the RW is too large (interesting project possibility…)

Summary

- Small molecules in a cell have only short-range forces (electrostatic forces are screened); but polymers and proteins can have long-range interactions because of their length and thermal fluctuations

- Polymers make thousands of interactions that change all the time as the polymer's shape fluctuates (⇒ each interaction is less important than the total number)
- Phantom chain model represents long, flexible polymers well

- Lipids are special because they are frustrated, and this leads to new phases and behaviour. They are intermediate between small molecules and large polymers: they are small, but have a lot of entropy in their tails.

- Everything in a cell is moving, so the life of a cell (at least its macromolecules) depends on *entropy* as much as *energetic* interactions

- Cells use many small interactions to drive many processes not few large ones

Break - 15 mins.

Exercise

1) First test in 2 weeks (2nd October), I'll put a practise test on moodle next week. Topics will be taken from first 3 lectures.

2) Journal club in 3 weeks (9th October), pick a paper from the moodle set. Work in a group, each person speaks, 10 - 15 minutes with slides.

3) I have to miss first hour on 9th October, so Benedikt will give the lecture that day, and I should be back in time for the exercise period and Journal Club Or shall we move it to Lecture 6 (16th October)?

4) DPD simulation of equation of state of water.

While that's happening …

5) Introduce homework exercises.

6) Download visualisation software - Paraview

7) Download the polymer in water input file and visualize it

8) Plot the equation of state of DPD water

Equation of state of a fluid

The ideal gas EOS is: $pV = N\, k_B T$

For a real gas, it can be written as a Virial expansion ($\rho = N/V$ = density)

$$p = \rho\, k_B T(1 + B_2(T)\,\rho + B_3(T)\,\rho^2 + \ldots)$$

where B_2, B_3 are independent of density. What is the EOS for "DPD water"?

$$p = p(a_{ij}, \rho, T)$$

p = Pressure
a_{ij} = Conservative interaction parameter
ρ = Bead density
T = Temperature

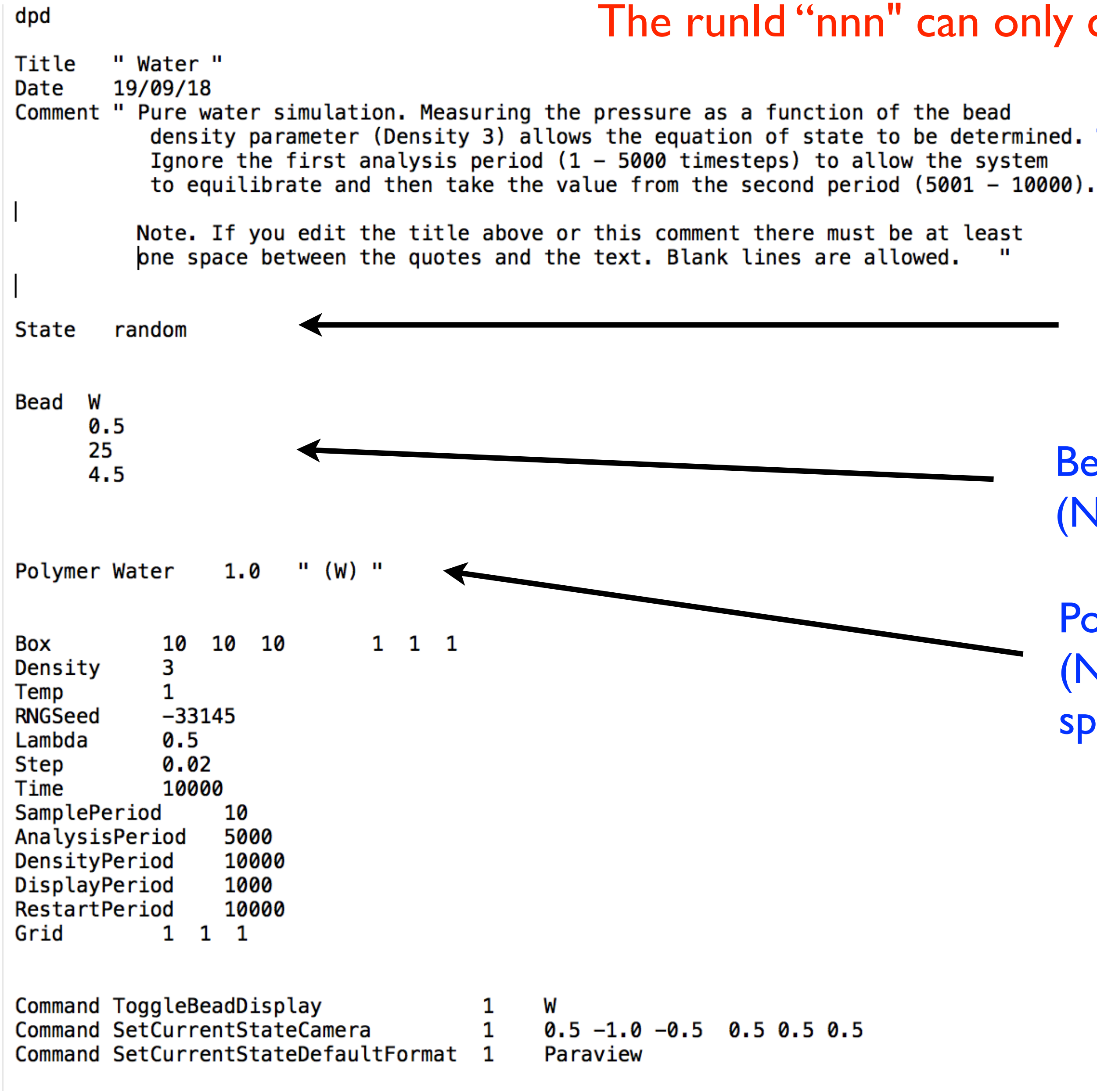
Equation of State Exercise

Goal: to set up and run a DPD simulation of pure water, check that it's equilibrated, and measure the equation of state.

- Description of dmpci.nnn input file and output files
- Simulate pure water and measure T, P against time from dmpchs
- Measure P as a function of bead density and plot the equation of state of dpd "water"
- When is a simulation equilibrated?

dmpci.nnn input file

The runId “nnn" can only contain letters, numbers, - and _



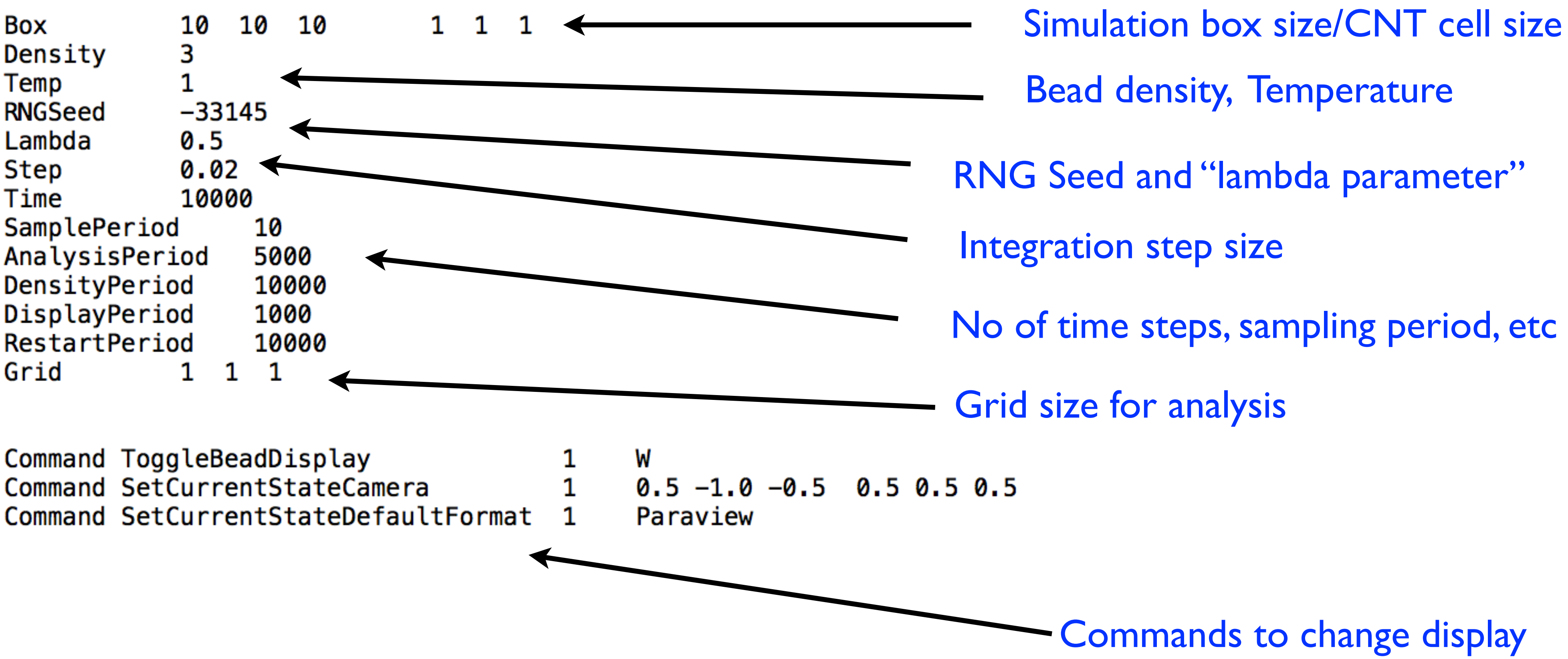
```
dpd

Title   " Water "
Date    19/09/18
Comment " Pure water simulation. Measuring the pressure as a function of the bead
           density parameter (Density 3) allows the equation of state to be determined.
           Ignore the first analysis period (1 - 5000 timesteps) to allow the system
           to equilibrate and then take the value from the second period (5001 - 10000).
|
          Note. If you edit the title above or this comment there must be at least
          one space between the quotes and the text. Blank lines are allowed.    "
|

State   random


Bead  W
      0.5
      25
      4.5


Polymer Water    1.0    " (W) "


Box         10  10  10       1  1  1
Density     3
Temp        1
RNGSeed     -33145
Lambda      0.5
Step        0.02
Time        10000
SamplePeriod     10
AnalysisPeriod   5000
DensityPeriod    10000
DisplayPeriod    1000
RestartPeriod    10000
Grid        1  1  1


Command ToggleBeadDisplay            1    W
Command SetCurrentStateCamera        1    0.5 -1.0 -0.5  0.5 0.5 0.5
Command SetCurrentStateDefaultFormat 1    Paraview
```

Title, Date and description of run - there MUST be space between text and “ “

Initial state type

Bead type definitions (Name, radius, cons. int., diss. int.)

Polymer (or molecule) type definitions (Name, number fraction, shape) - note spaces between shape and “ ”



```
Box           10  10  10        1  1  1
Density       3
Temp          1
RNGSeed       -33145
Lambda        0.5
Step          0.02
Time          10000
SamplePeriod      10
AnalysisPeriod    5000
DensityPeriod     10000
DisplayPeriod     1000
RestartPeriod     10000
Grid          1  1  1

Command ToggleBeadDisplay            1    W
Command SetCurrentStateCamera        1    0.5 -1.0 -0.5  0.5 0.5 0.5
Command SetCurrentStateDefaultFormat 1    Paraview
```

Simulation box size/CNT cell size

Bead density, Temperature

RNG Seed and "lambda parameter"

Integration step size

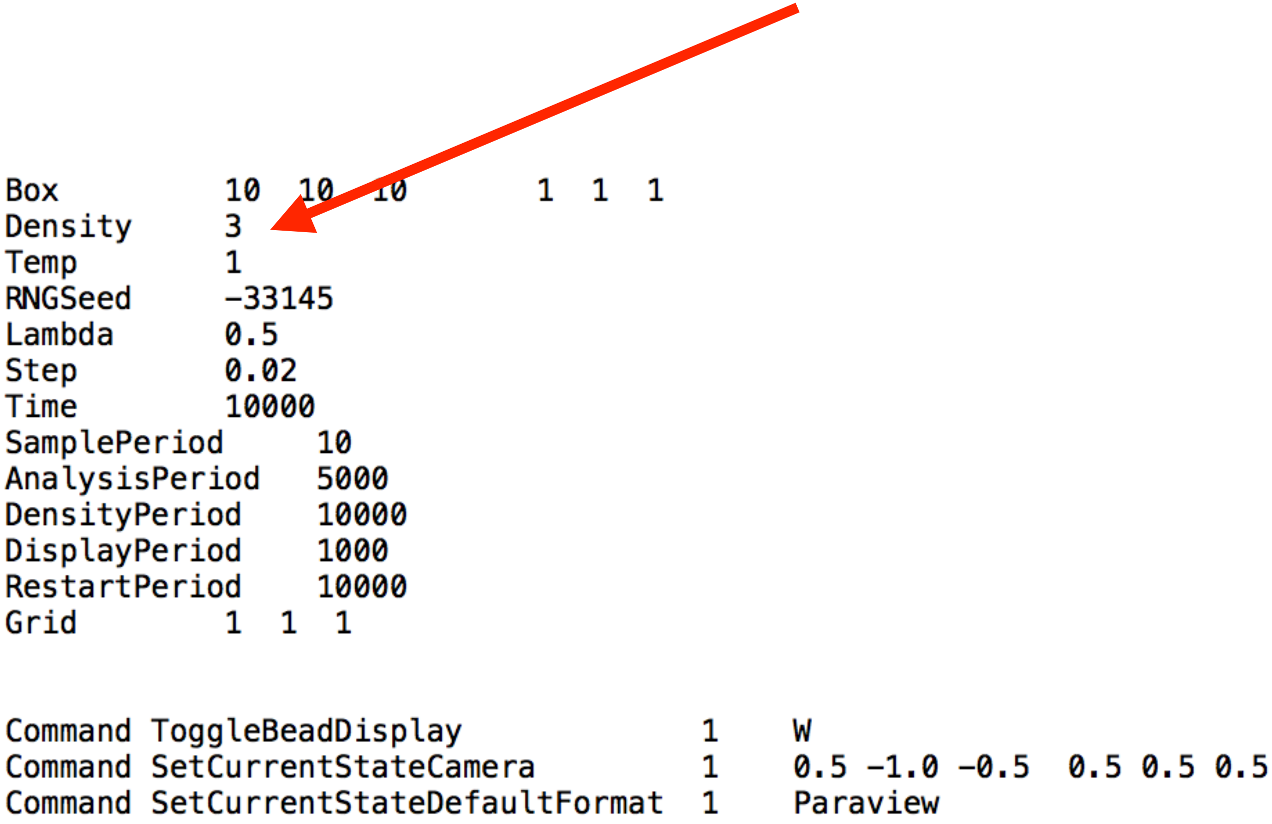
No of time steps, sampling period, etc

Grid size for analysis

Commands to change display

Commands must be time-ordered

Each pair/group set the density to one value:



```
Box           10  10  10        1   1   1
Density       3
Temp          1
RNGSeed       -33145
Lambda        0.5
Step          0.02
Time          10000
SamplePeriod      10
AnalysisPeriod    5000
DensityPeriod     10000
DisplayPeriod     1000
RestartPeriod     10000
Grid          1   1   1


Command ToggleBeadDisplay             1    W
Command SetCurrentStateCamera         1    0.5 -1.0 -0.5   0.5 0.5 0.5
Command SetCurrentStateDefaultFormat  1    Paraview
```

Group	Density
1.	3
2.	4
3.	5
4.	6
5.	8
6.	10
7.	12
8.	14
9.	16
10.	20

Output files

The code produces a set of output files: they all start with “dmpc” and have a suffix identifying the data they contain and the same extension as the input file.

dmpcas.999	Time-averaged analysis data
dmpchs.999	Time series data of T, P diffusion, end-to-end lengths
dmpcis.999	Copy of input data for verification
dmpcls.999	Logfile of commands, error messages, etc
dmpcrs.999.con.1000.dat	Restart state file
dmpccs.999.con.100.pov	Povray snapshot files used for movies, images
dmpccs.999.con.200.pov	(can also output vtk files for Paraview)
…	

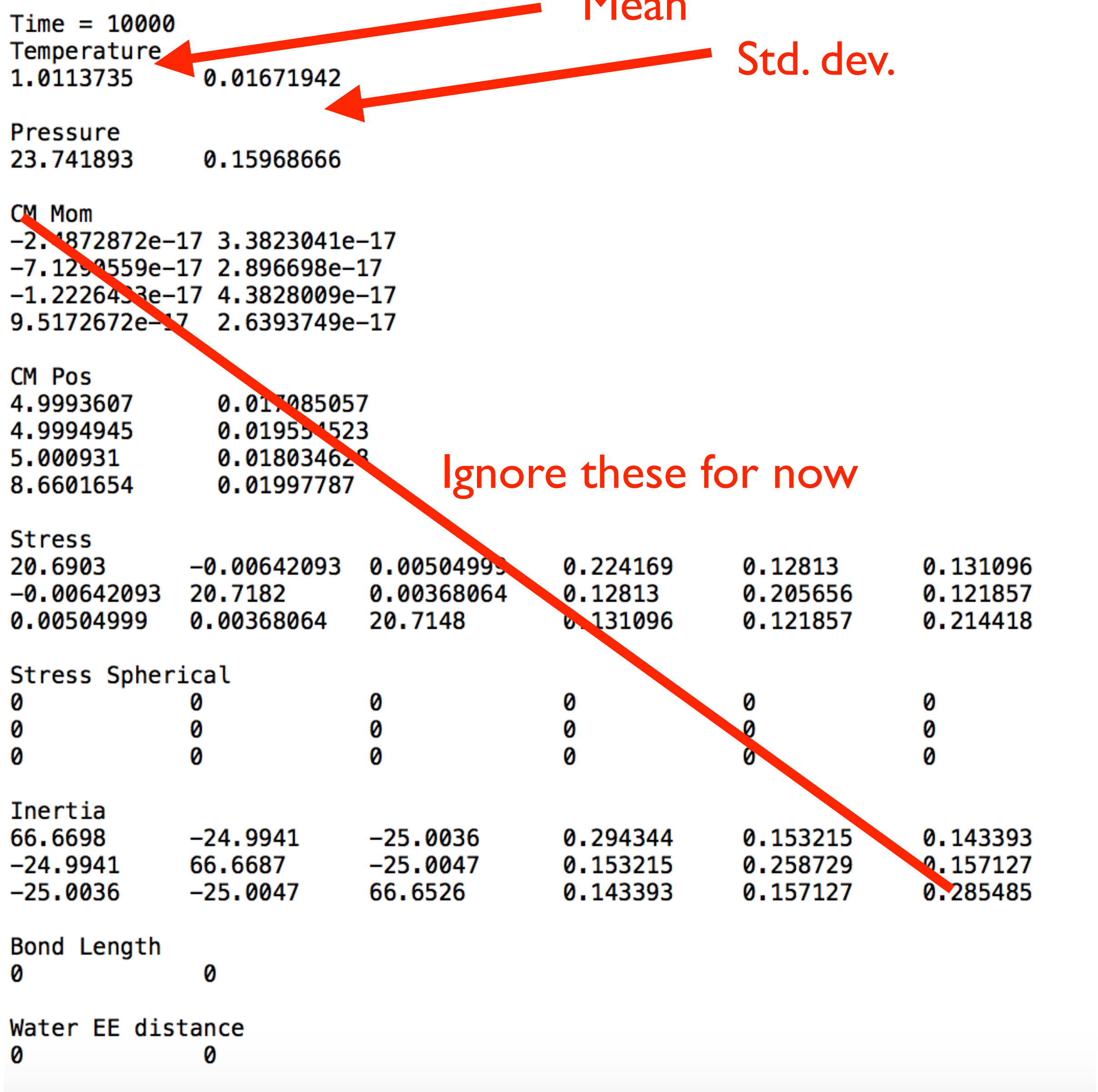
Files produced repeatedly (display and restart states) have time encoded in their names.

T, P are in dmpcas.nnn file

Use second analysis period

Mean

Std. dev.



```
Time = 10000
Temperature
1.0113735     0.01671942

Pressure
23.741893     0.15968666

CM Mom
-2.4872872e-17 3.3823041e-17
-7.1290559e-17 2.896698e-17
-1.2226433e-17 4.3828009e-17
9.5172672e-17  2.6393749e-17

CM Pos
4.9993607      0.017085057
4.9994945      0.019554523
5.000931       0.018034628
8.6601654      0.01997787

Stress
20.6903       -0.00642093  0.00504999    0.224169     0.12813      0.131096
-0.00642093   20.7182      0.00368064    0.12813      0.205656     0.121857
0.00504999    0.00368064   20.7148       0.131096     0.121857     0.214418

Stress Spherical
0             0            0             0            0            0
0             0            0             0            0            0
0             0            0             0            0            0

Inertia
66.6698       -24.9941     -25.0036      0.294344     0.153215     0.143393
-24.9941      66.6687      -25.0047      0.153215     0.258729     0.157127
-25.0036      -25.0047     66.6526       0.143393     0.157127     0.285485

Bond Length
0             0

Water EE distance
0             0
```

Ignore these for now

Mean / Std. dev

When is a simulation equilibrated?

History State File - dmpchs.nnn

The History file contains information on the time evolution of observables; we use this to determine if the simulation is in equilibrium, is unstable, or if numerical errors are large.

To Do:

1. Run a simulation of pure water in a 10^3 box for 10,000 steps, sample every 10 steps (or use one from last week in ~/BIOENG455 Exercises/Exercise 1)
2. Plot time series of temperature and pressure from the dmpchs file
3. Why are there large fluctuations initially?
4. Increase the integration step size to dt = 0.1 (or 0.2) and repeat: what happens?
5. Increase the temperature (set dt = 0.01 here) and repeat: what happens?

Visualising simulations with Paraview

Simulations are boring without pictures and movies:

Go to: https://www.paraview.org

and download the Windows/Mac/Linux executable as needed

https://www.paraview.org/download/

Visualising a simulation

The **dmpccs.nnn.con.ttt.vtk** files contain snapshots of the simulation state (x, y, z coordinates and an integer identifying the type of all beads). These can be exported in Paraview (*.vtk) or povray (*.pov) format.

NB. Add the command "ToggleBeadDisplay 1 W" to the dmpci file to make the water invisible otherwise you won't see anything.

Two options for making images and movies:

- Paraview allows import of a sequence of *.vtk files to make movies.
- Or convert the *.pov files into *.gro and *.xtc files, and use VMD to view them (ask me for the script).

The dmpccs files can also be used for off-line analysis that requires particle coordinates as they are written in plan ascii text.

Visualising a polymer in water

Download the input file dmpci.p001 from moodle

```
State   random


    Bead  W
    0.5
    25
    4.5

    Bead  B
    0.5
    25    25
    4.5   4.5

    Bond  B B 128.0 0.5


Polymer Water    0.9995  " (W) "
Polymer PEG      0.0005  " (B B B B B B B B B B B B B B B B) "
```

New bead type to make monomers

You need 2 more interaction parameters now; one for WB and BB

and a bond to tie them together

Connect the monomers into the polymer

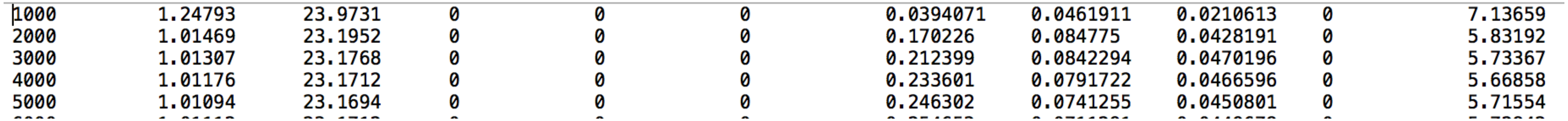
Exercise - Time Series Data

Log State File - dmpcls.nnn

- Sequence of time-ordered information, warning, error messages.
- Shows results of commands executed during a run

History State File - dmpchs.nnn

- Time series data of observables saved at a frequency of: *SamplePeriod*
- Time, Temperature, Pressure, (ignore 4-6), Bead diffusion, Polymer end-end length
- 1 diffusion column for each bead type defined in input file
- 1 end-end length column for each polymer defined in input file (head and tail beads are defined as first and last in the polymer's shape string)



```
1000        1.24793     23.9731     0           0           0           0.0394071   0.0461911   0.0210613   0           7.13659
2000        1.01469     23.1952     0           0           0           0.170226    0.084775    0.0428191   0           5.83192
3000        1.01307     23.1768     0           0           0           0.212399    0.0842294   0.0470196   0           5.73367
4000        1.01176     23.1712     0           0           0           0.233601    0.0791722   0.0466596   0           5.66858
5000        1.01094     23.1694     0           0           0           0.246302    0.0741255   0.0450801   0           5.71554
```