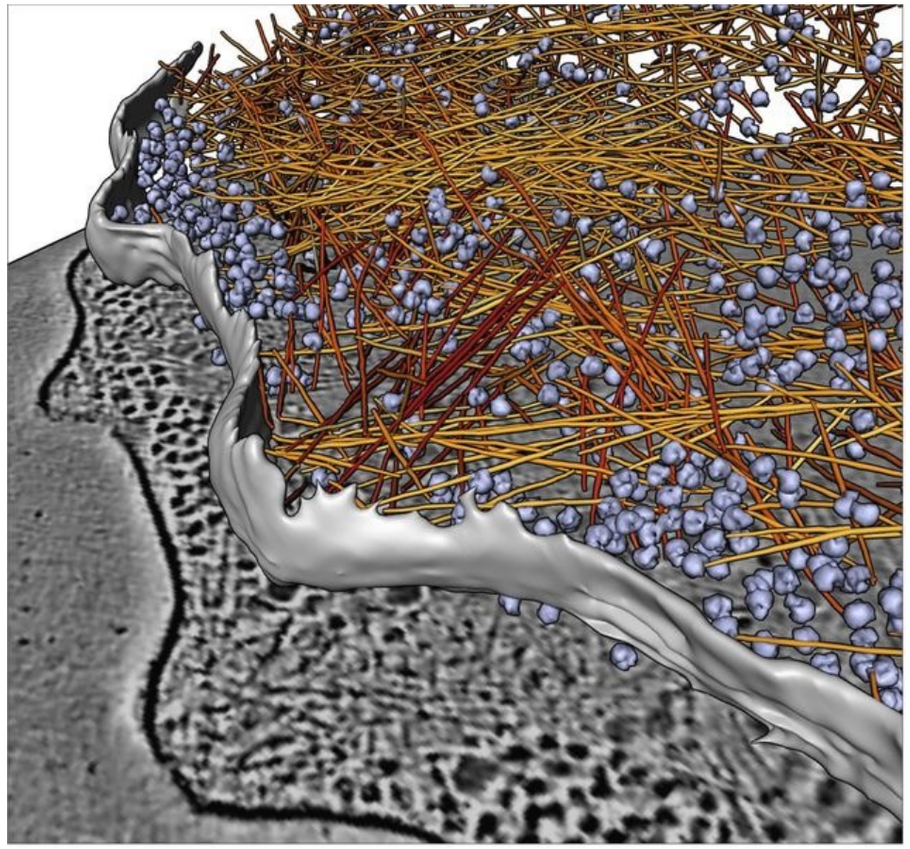
Interactions between proteins, DNA and RNA

—

The energy, length and time coordinate system to find your way in the cell

Karsten Rippe



Coordinates for my lectures:

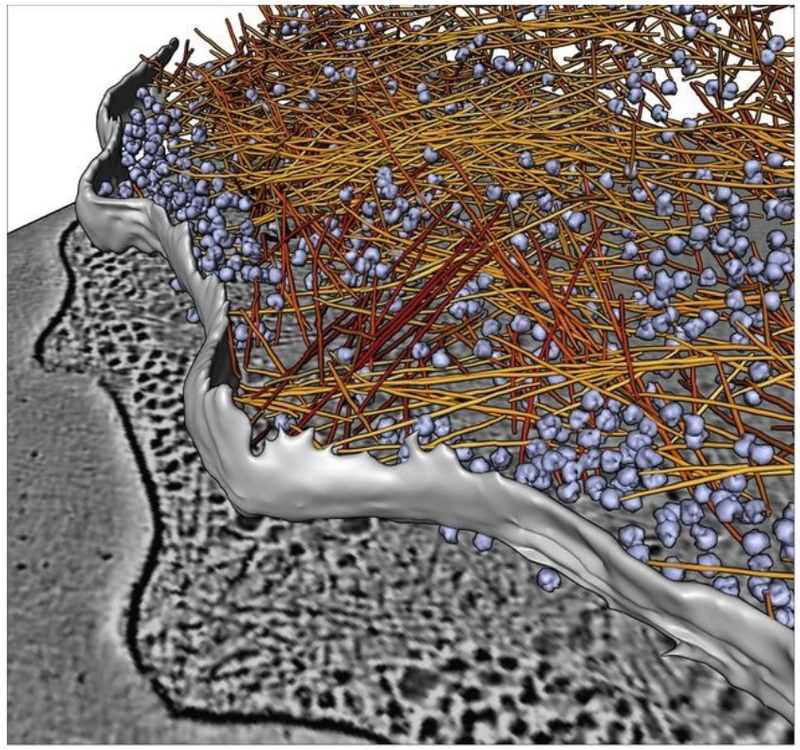
Karsten Rippe
BioQuant
Room 645, 6th floor
Telefon: 54-51376
e-mail: Karsten.Rippe@dkfz.de

Overview on learning Biophysics in Heidelberg
http://malone.bioquant.uni-heidelberg.de/teaching

Material for the lecture: Biophysical concepts and theoretical descriptions
http://malone.bioquant.uni-heidelberg.de/teaching
-> Physico-Chemical Methods in Systems Biology

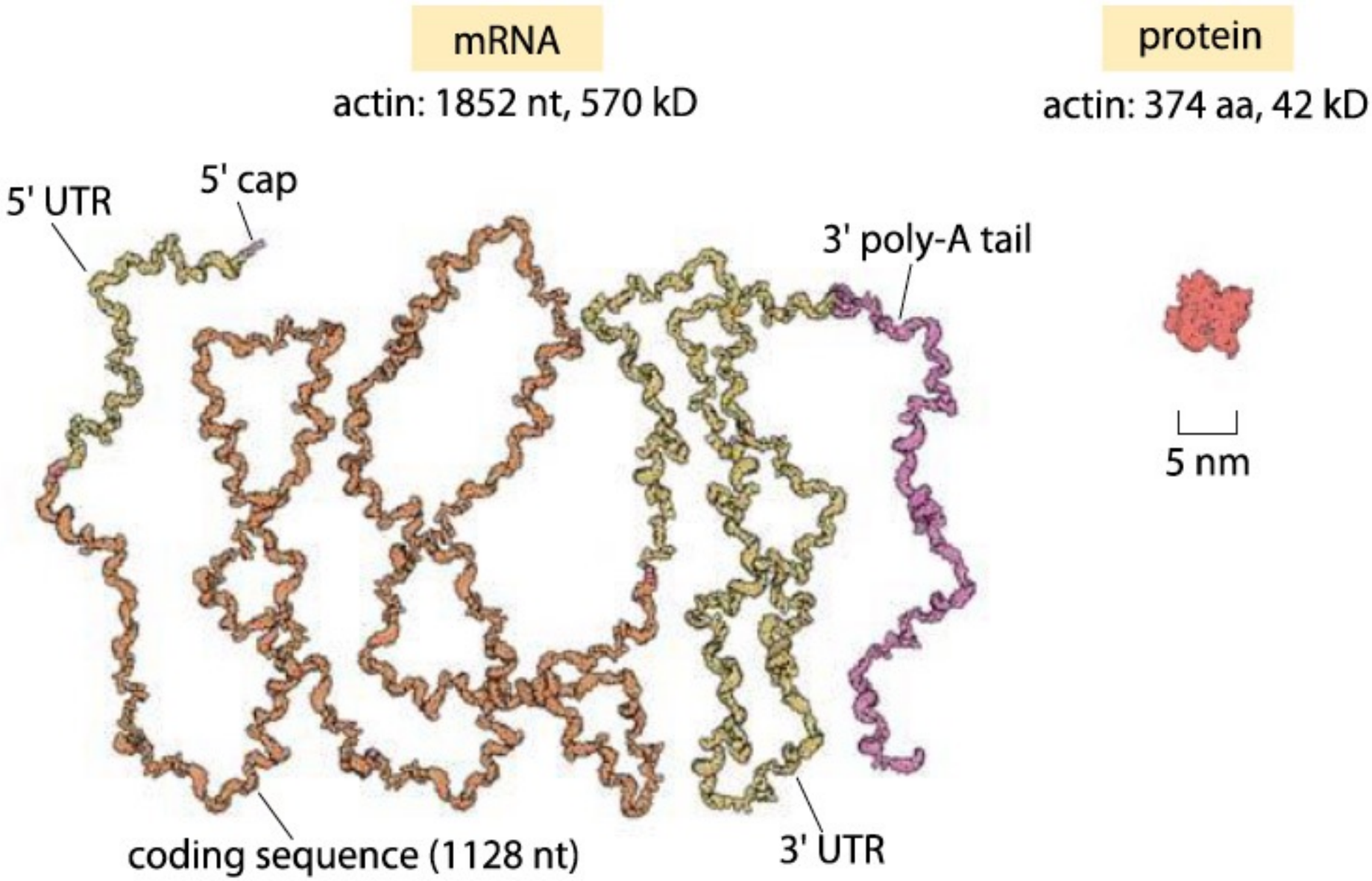
Username: teaching
Password: nonukes

The energy, length and time coordinate system to find your way in the cell

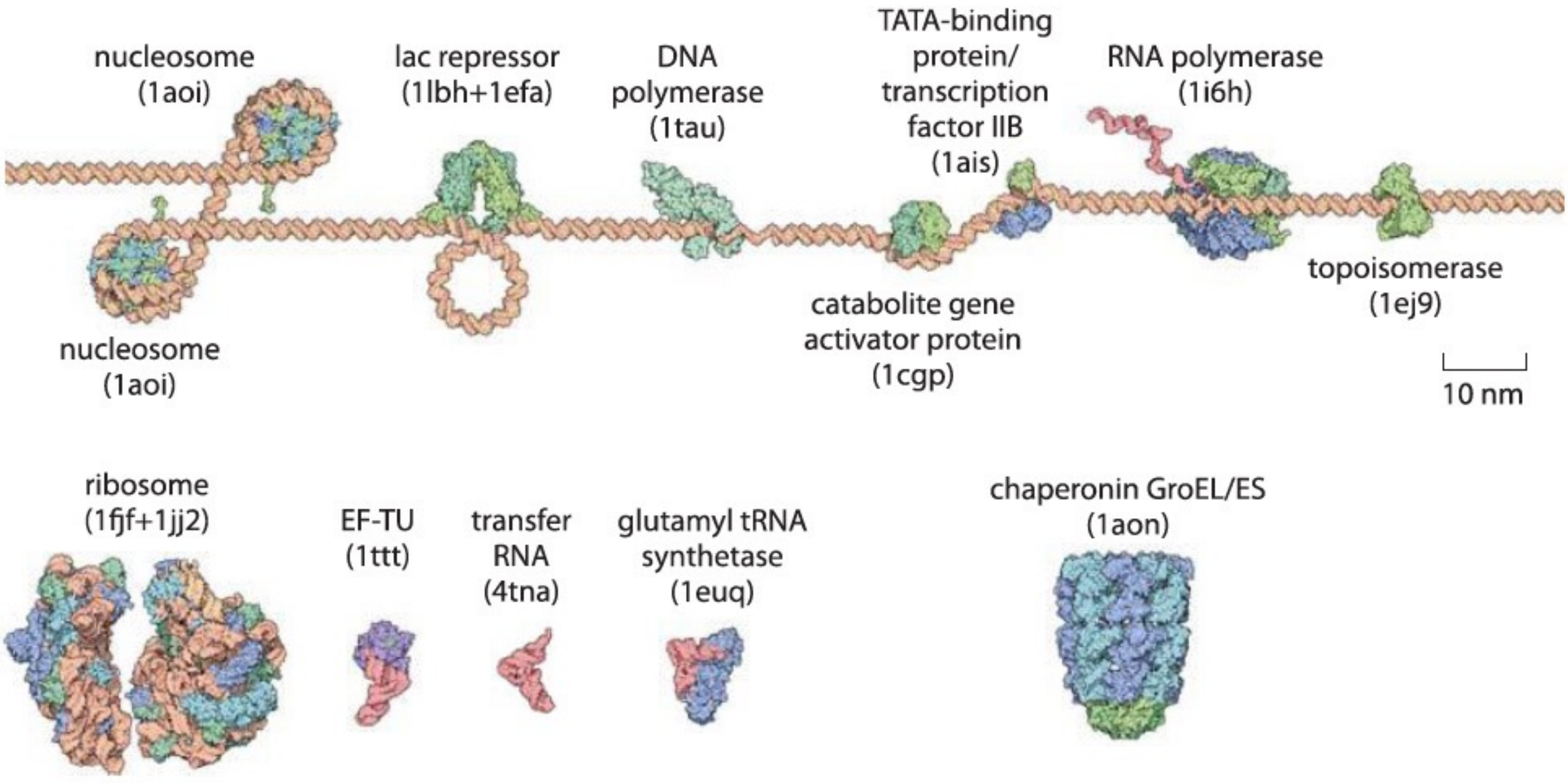


- Length: What is larger: RNA or protein?
- Speed: How fast do transcription factors find their target sequence?
- Energy: How much ATP is hydrolyzed in a cell?

Actin: mRNA size compared to protein?

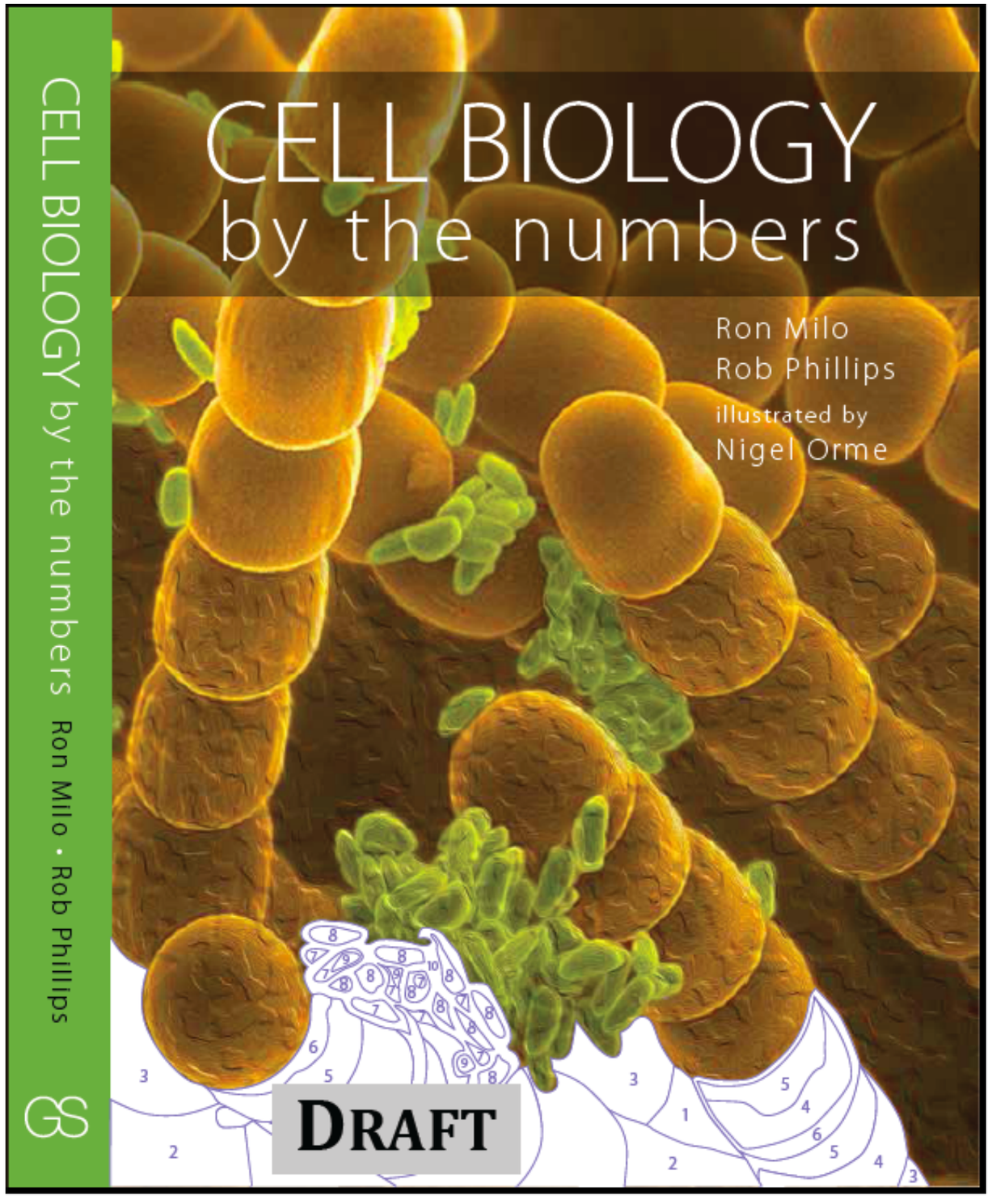


Protein and DNA size scales



Quantitative biology needs numbers…

Cell Biology by the Numbers (upcoming book by Rob Philipps & Ron Milo, Lecture by Ron Milo).
Book draft:
http://book.bionumbers.org



Bionumbers database:
http://bionumbers.hms.harvard.edu

B10NUMB3R5

THE DATABASE OF USEFUL BIOLOGICAL NUMBERS

Home \ Search | Browse | Resources | Cell Biology by the Numbers | About Us | Login \ Submit

Popular BioNumbers | Recent BioNumbers | Key BioNumbers | Amazing BioNumbers

Find Terms

search

e.g., ribosome coli , p53 human , transcription , OD

Help improve by sending Ron Milo your feedback
(data to add, errors found or an unsolicited thumbs up...)

What we would like to do…

- Develop a coordinate system and intuition about the cellular world
- Apply it to “sanity” checks (discussions, design of experiment…)
- Understand principles that rationalize how the cell operates

Some examples for basic numbers in cell biology and how to use them

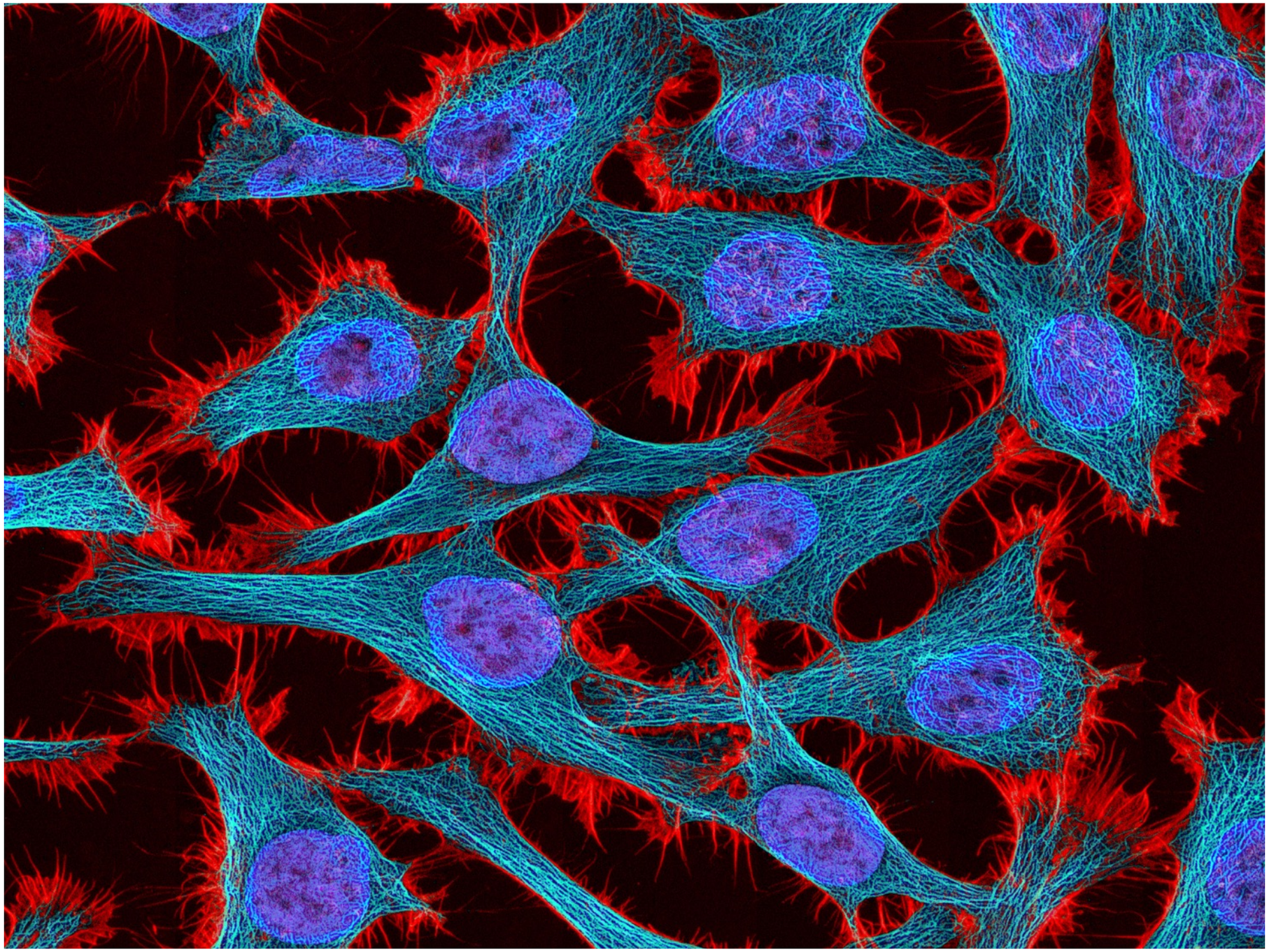
What is the volume of a cell?

E. coli	1 μm^3
budding yeast	30 μm^3
HeLa cell line	3 000 μm^3

Cell volumes

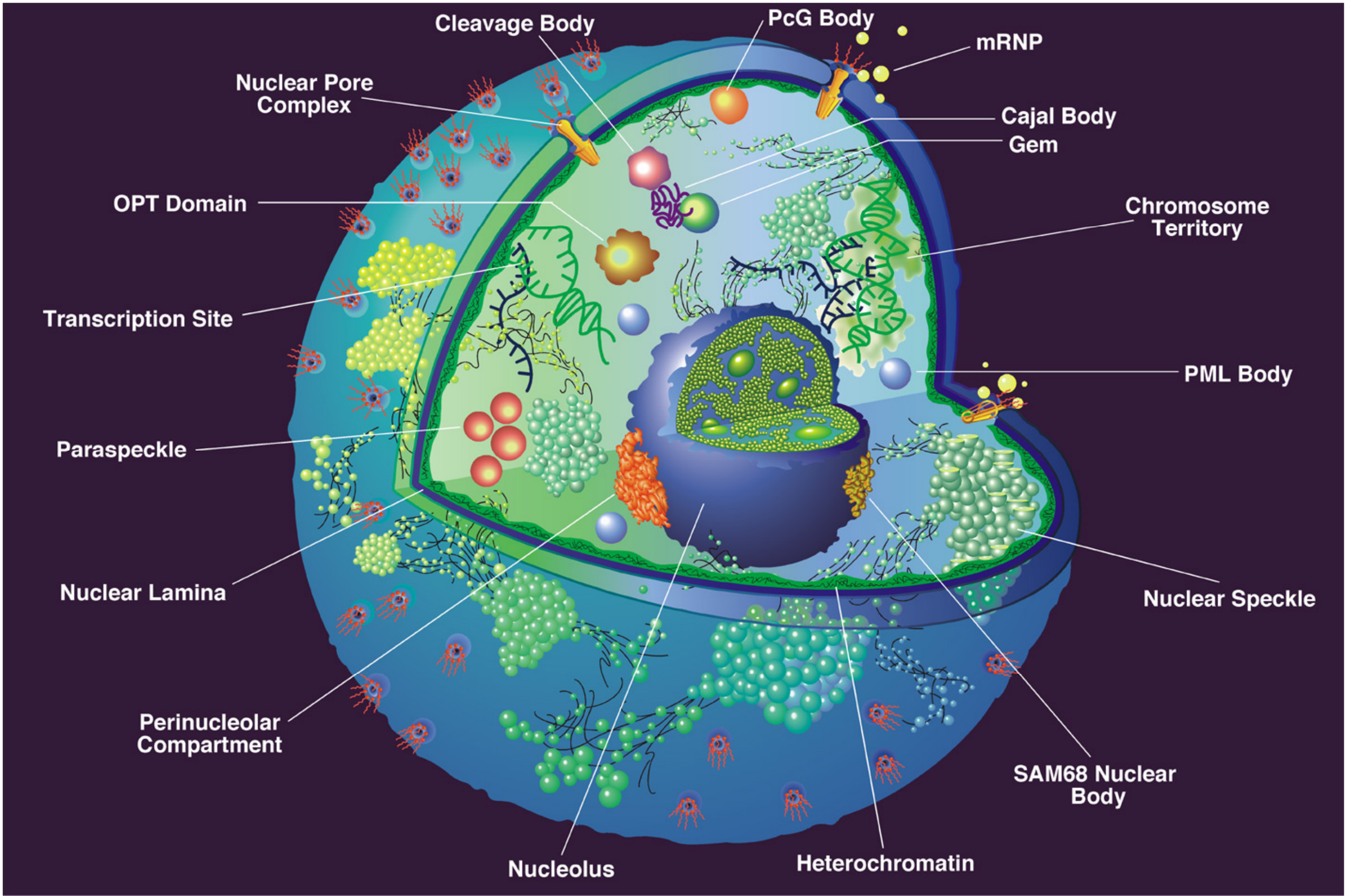
cell type	average volume (μm^3)	BNID
sperm cell	30	109891, 109892
red blood cell	100	107600
lymphocyte	130	111439
neutrophil	300	108241
beta cell	1,000	109227
enterocyte	1,400	111216
fibroblast	2,000	108244
HeLa, cervix	3,000	103725, 105879
hair cell (ear)	4,000	108242
osteoblast	4,000	108088
alveolar macrophage	5,000	103566
cardiomyocyte	15,000	108243
megakaryocyte	30,000	110129
fat cell	600,000	107668
oocyte	4,000,000	101664

HeLa cell (wikipedia): DNA, blue; microtubuli, cyan; actin, red

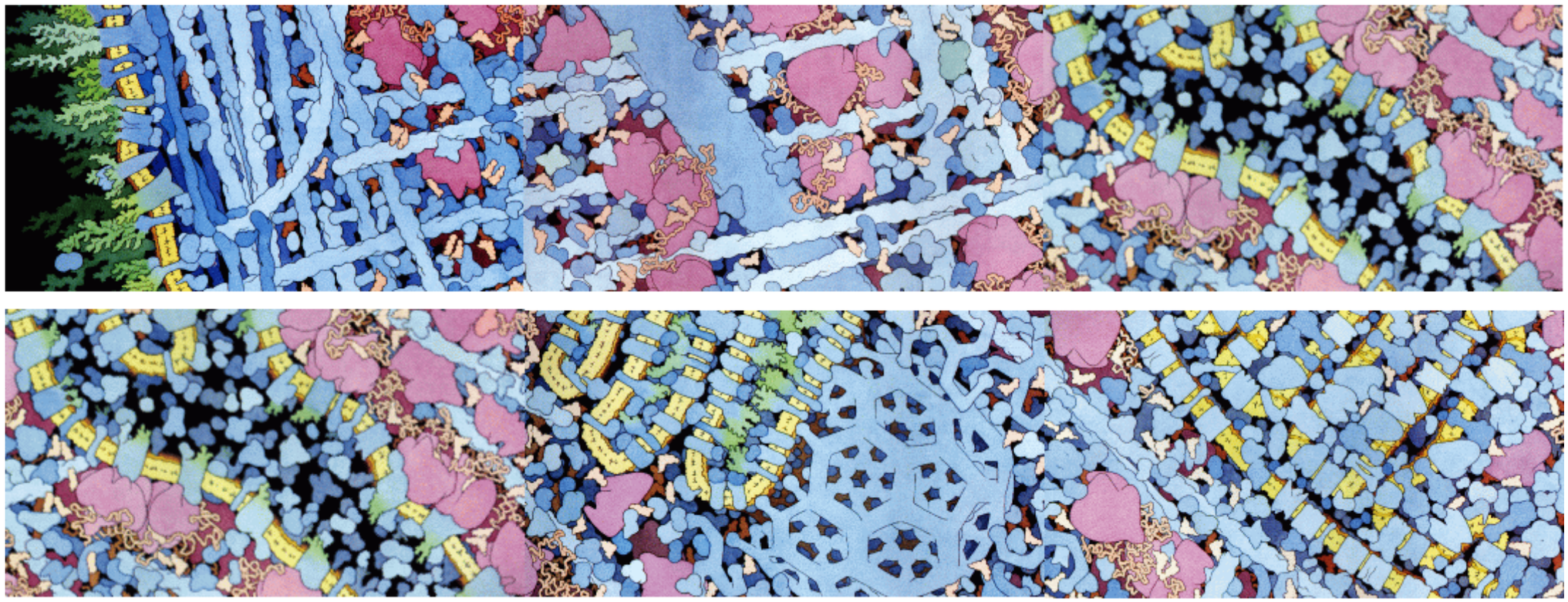


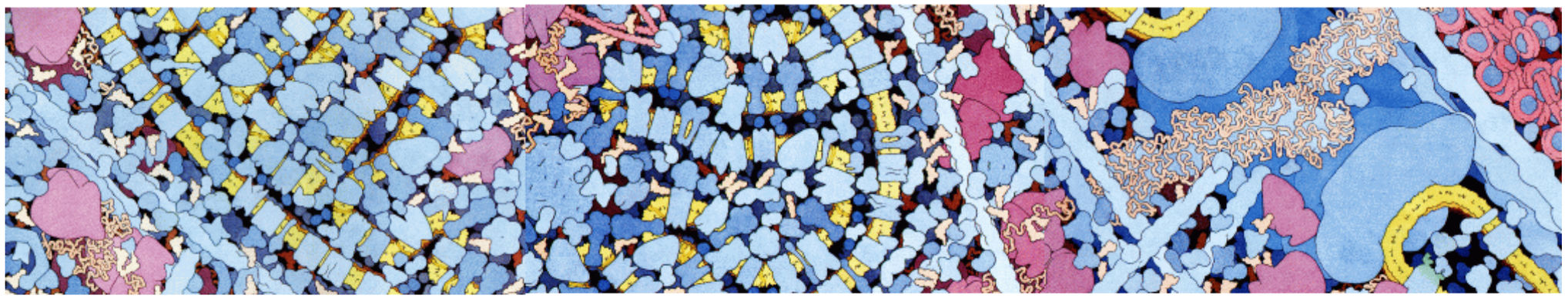
Nucleus diameter: ~10 μm, 0.4 pl volume

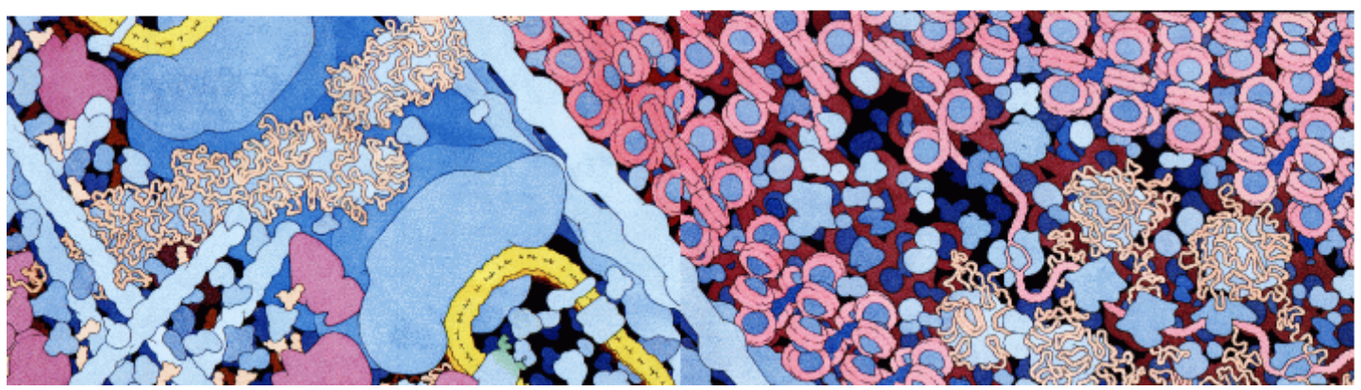
The mammalian cell nucleus

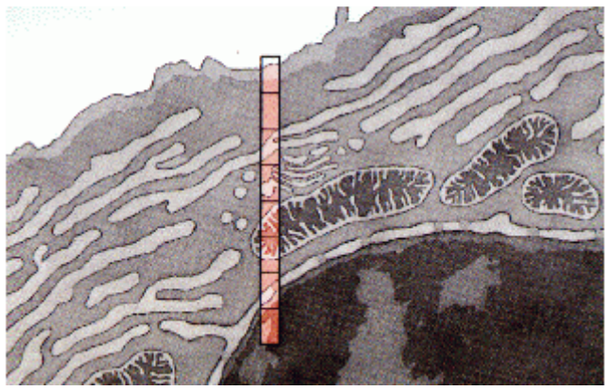


The cell is a very crowded place (David Goodsell)









from left to right:

cell surface

cytoplasm

synthesis of proteins from the endoplasmic reticulum

Golgi apparatus,

coated vesicle

mitochondrion

nucleus

proteins: blue

DNA and RNA: red and orange

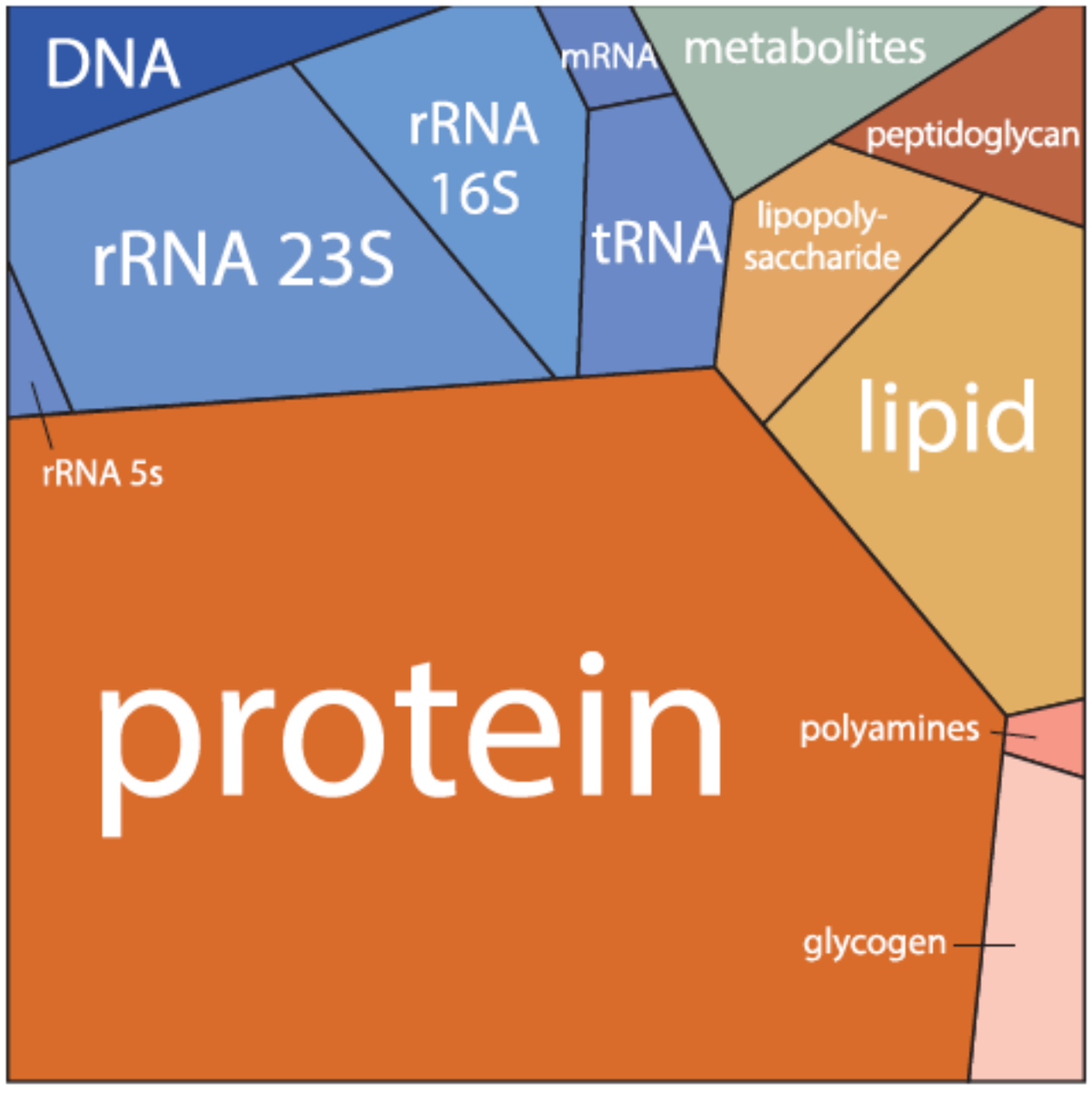
lipids: yellow

carbohydrates:

green

Ribosomes: magenta

Mass fractions in dry mass of E. coli growing at doubling time of 40 min



Concentration of proteins and DNA/RNA in the nucleus

DNA	**RNA**	**Protein**
~ **15mg/ml** (6pg DNA per cell, [19] nucleus ~1/10 of cell volume $4x10^{-9}$ cm^3 typical) [20] **~18.5mg/ml** (56mM nucleosome concentration, [21] 200 bp/nucleosome, 2bases/bp,1Mbase/330g. [22] **~19 mg/ml** [23] ~**20-31 mg/ml** (8.1-12.5pg/cell, [24] nucleus ~1/10 of cell volume $4x10^{-9}$ cm^3 typical) [20]	**~11 mg/ml** (5-25pg RNA per cell, [25] 18% in nucleus, [26] nucleus ~1/10 of cell volume $4x10^{-9}$ cm^3 typical). [20] **~12-15mg/ml** (27.1-33.1pg/cell, [24] 18% in nucleus, [26] nucleus ~1/10 of cell volume $4x10^{-9}$ cm^3 typical). [20]	**~106-215 mg/ml** in various regions of the nucleus. [27] **~108mg/ml** (6pg DNA per cell, [20] protein mass 72X DNA mass and cell volume $4x10^{-9}$ cm3 typical). [20] **~200-300mg/ml** in E.coli. [28]

How many mRNA molecules are in an *E. coli* cell?

a) 10^3 - 10^4

b) 10^5 - 10^6

c) 10^7 - 10^8

d) 10^9 - 10^{10}

and in a human cell?

a) same

b) 10x

c) 100x

d) 1000x

Strategies to answer the question

- Calculate the number of proteins per cell
- Estimate the protein synthesis rate *R* from the division time
- Estimate the protein production rate *r* per mRNA
- calculate the number of mRNA = *R*/*r*

How many proteins are in a cell?

protein mass per volume (≈ 0.2 g/ml)

number of proteins per cell volume:

$$\frac{N}{V} = \frac{C_p}{l_{aa} \times m_{aa}}$$

m_{aa}: mass aa (≈ 100 Da)

l_{aa}: aa per protein ($\approx 400 \frac{aa}{protein}$)

Avogadro's number

$$\frac{N}{V} = \frac{0.2\,[g/ml] \times 6 \times 10^{23}\left[\frac{Da}{g}\right] \times 10^{-12}\left[\frac{ml}{\mu m^3}\right]}{400\left[\frac{aa}{protein}\right] \times 100\left[\frac{Da}{aa}\right]} \approx 3 \times 10^6 \frac{proteins}{\mu m^3}$$

organism	characteristic volume	number of proteins
E. coli	1 μm^3	$\approx 3 \times 10^6$
budding yeast	$\approx 30\ \mu m^3$	$\approx 100 \times 10^6$
HeLa cell line	$\approx 3{,}000\ \mu m^3$	$\approx 10 \times 10^9$

Estimating the protein synthesis rate *R*

$$\frac{N_{protein}}{V} \approx 3 \times 10^6 \text{ proteins}/\mu m^3$$

rate of protein production per cell

$$R = \frac{N_{protein}}{\tau}$$

doubling time

bacteria

$V \approx 1\,\mu m^3$,
$\tau \approx 1$ h

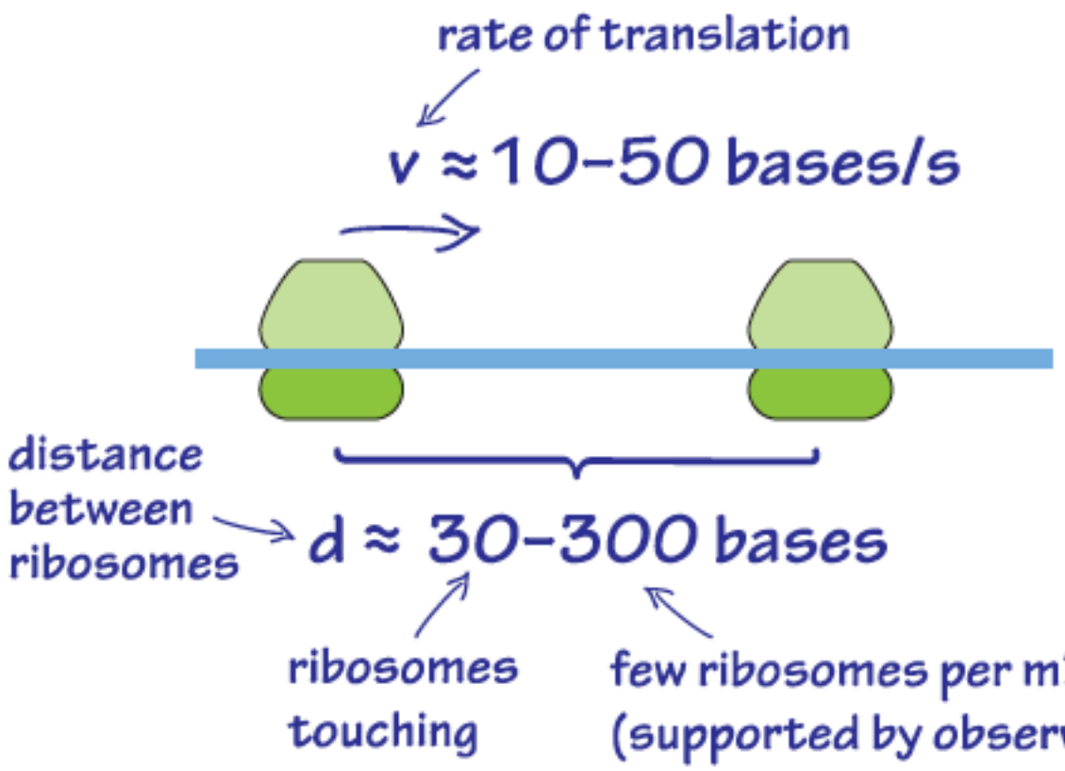
$$\approx \frac{3 \times 10^6 \text{ proteins}}{3{,}000 \text{ s}} \approx 10^3 \text{ proteins/s}$$

mammalian cell

$V \approx 3{,}000\,\mu m^3$,
$\tau \approx 24$ h

$$\approx \frac{10^{10} \text{ proteins}}{10^5 \text{ s}} \approx 10^5 \text{ proteins/s}$$

Estimate the protein production rate *r* per mRNA



rate of protein production per mRNA

$$r = \frac{v}{d} \approx 0.1\text{–}1 \text{ protein/mRNA/s}$$

Calculate the number of mRNAs

$$N_{mRNA} = \frac{R}{r}$$

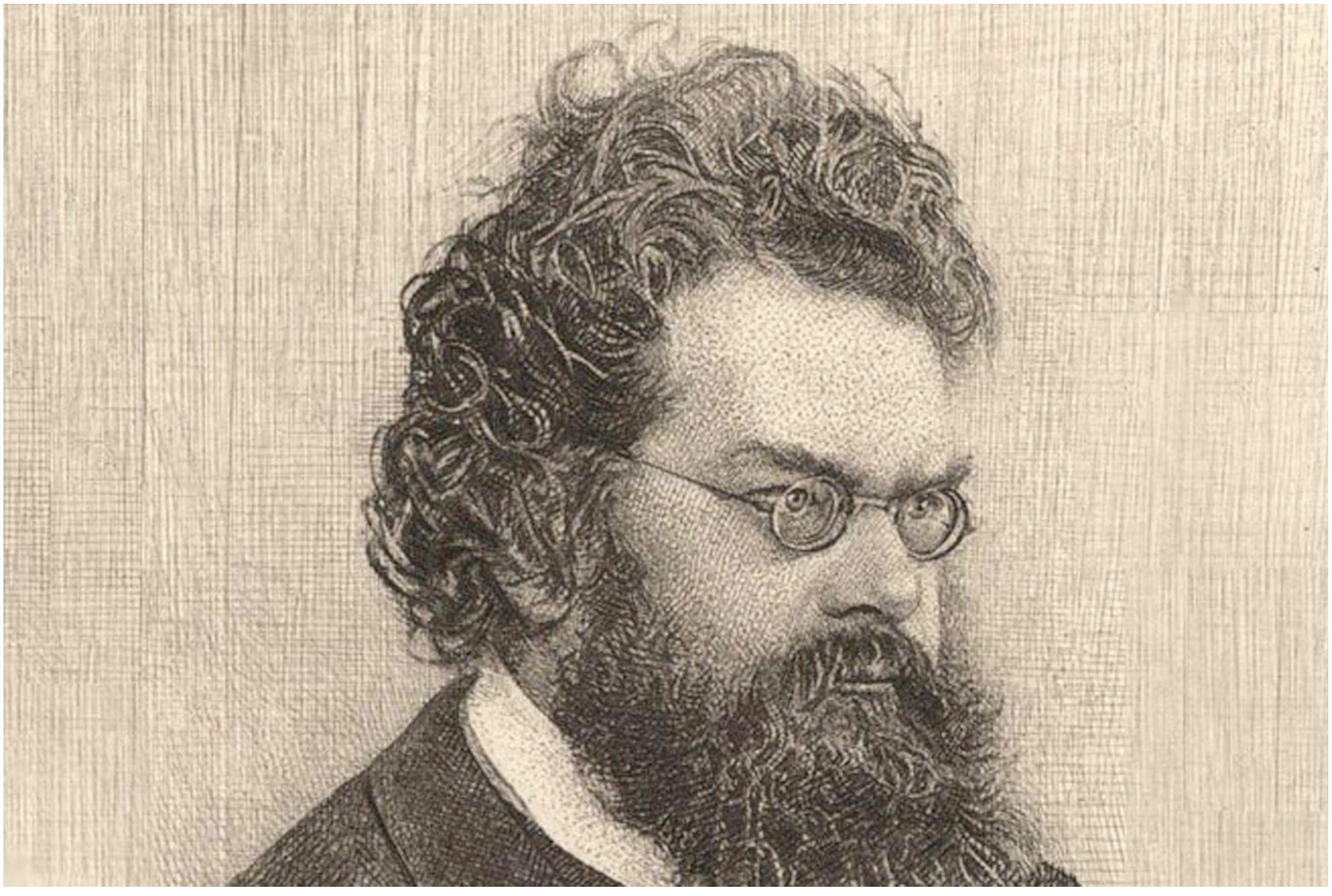
bacteria $\approx \dfrac{10^3 \text{ proteins/s}}{0.1\text{–}1 \text{ protein/mRNA/s}} \approx 10^3\text{-}10^4 \text{ mRNA/bacterial cell}$

mammalian cell $\approx \dfrac{10^5 \text{ proteins/s}}{0.1\text{–}1 \text{ protein/mRNA/s}} \approx 10^5\text{-}10^6 \text{ mRNA/mammalian cell}$

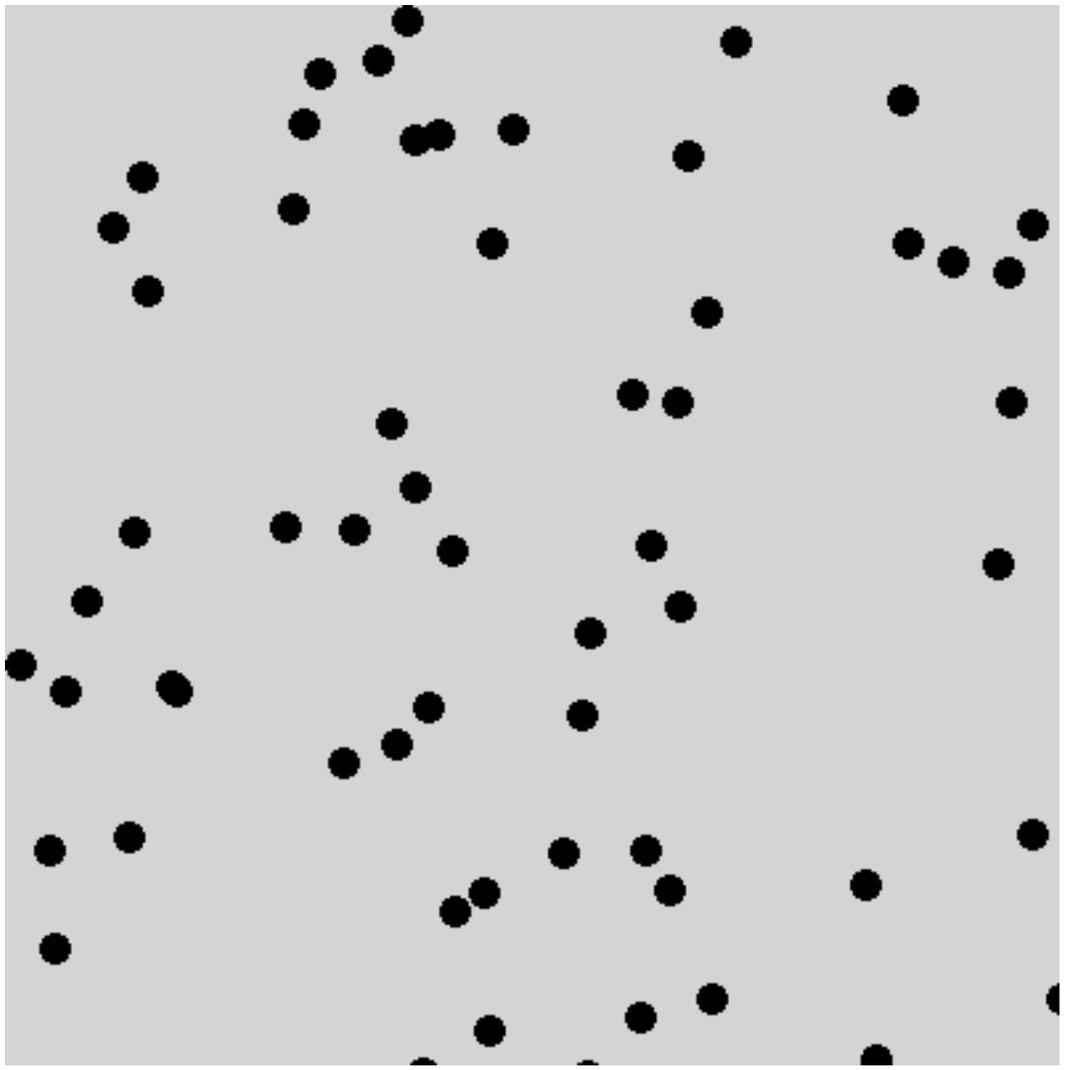
Cell parameters

property	*E. coli*	budding yeast	mammalian (HeLa line)
cell volume	0.3–3 μm^3	30–100 μm^3	1,000–10,000 μm^3
proteins per μm^3 cell volume		$2–4 \times 10^6$	
mRNA per cell	10^3-10^4	10^4–10^5	10^5–10^6
proteins per cell	~10^6	~10^8	~10^{10}
mean diameter of protein		4–5 nm	
genome size	4.6 Mbp	12 Mbp	3.2 Gbp
number protein coding genes	4300	6600	21,000
regulator binding site length		10–20 bp	
promoter length	~100 bp	~1000 bp	~1000 bp
gene length	~1000 bp	~1000 bp	~10^4–10^6 bp (with introns)
concentration of one protein per cell	~1 nM	~10 pM	~0.1–1 pM
diffusion time of protein across cell ($D \approx 10\ \mu m^2/s$)	~0.01 s	~0.2 s	~1–10 s
diffusion time of small molecule across cell ($D \approx 100\ \mu m^2/s$)	~0.001 s	~0.03 s	~0.1–1 s
time to transcribe a gene	<1 min (80 nts/s)	~1 min	~30 min (incl. mRNA processing)
time to translate a protein	<1 min (20 aa/s)	~1 min	~30 min (incl. mRNA export)
typical mRNA lifetime	2–5 min	~10 min to over 1 h	5-100 min to over 10 h
typical protein lifetime	1 h	0.3–3 h	10–100 h
minimal doubling time	20 min	1 h	20 h
ribosomes/cell	~10^4	~10^5	~10^6
transitions between protein states (active/inactive)		1–100 μs	
timescale for equilibrium binding of small molecule to protein (diffusion limited)		1–1000 ms (1 μM–1 nM affinity)	
timescale of transcription factor binding to DNA site		~1 s	
mutation rate		10^{-8}–10^{-10}/bp/replication	

Before Ludwig Boltzmann got depressed and killed himself, he did some really great things, like inventing the Boltzmann constant



Molecules in an ideal gas



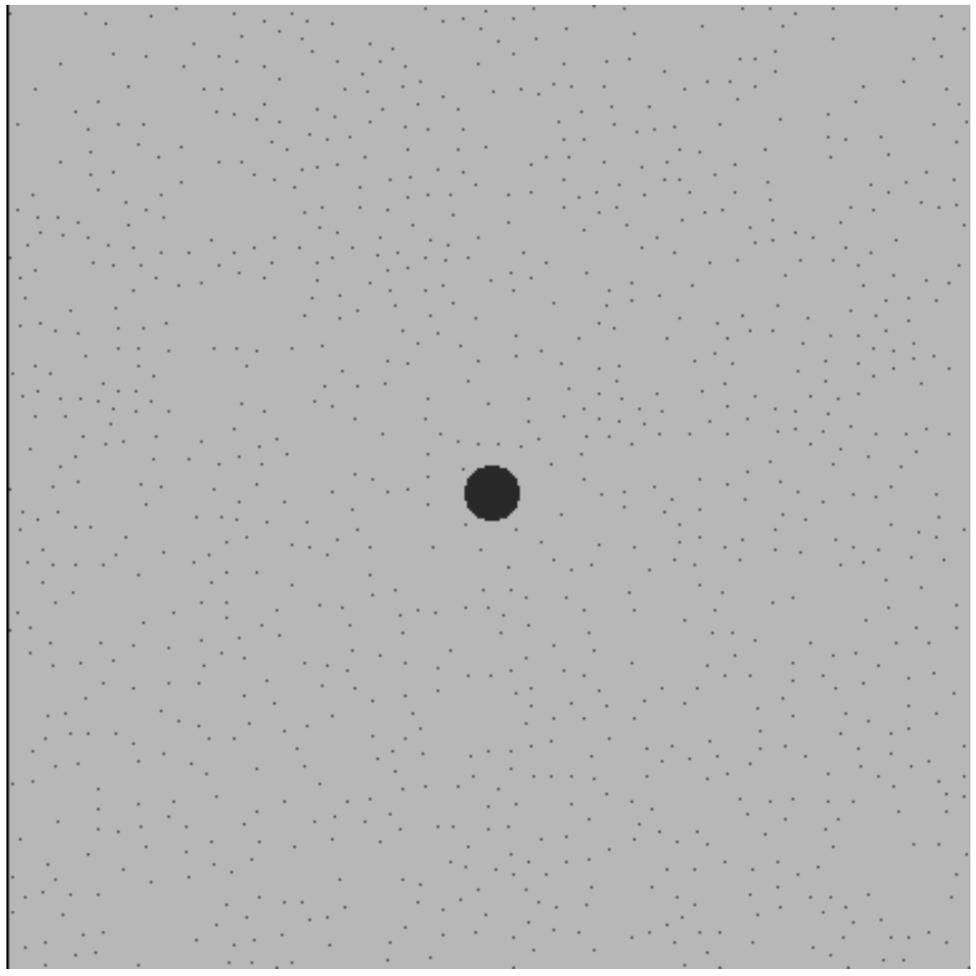
the average kinetic energy of a particle with mass *m* along the x-axis is related to $k_B T/2$

$$\left\langle \frac{m\, v_x^2}{2} \right\rangle = \frac{kT}{2} \qquad \left\langle v_x^2 \right\rangle = \frac{kT}{m}$$

$$\sqrt{\left\langle v_x^2 \right\rangle} = \sqrt{\frac{kT}{m}}$$

A protein like lysozyme (M = 14 kDa) at room temperature would have a speed of $\langle v_x^2 \rangle^{1/2}$ = 13 m/sec **in vacuum**.

Movements of a macromolecule in solution



- The macromolecule collides with water molecules and moves in a “random walk”.
- The length of the free path is much smaller than the diameter of the particle.
- The average distance from the starting point is proportional to the square root of time.

$k_B T$ is the energy available for spontaneous reactions

$$P_i \propto g_i \cdot \exp\left(\frac{-E_i}{k_B T}\right)$$

The Boltzmann equation yields the probability P_i to find a molecule with energy E_i

- g_i: number of different states with energy E_i
- k_B: Boltzmann constant
- T: Temperature

probability to find a particle with an energy

- of $k_B T$ or larger: 0.37 => processes that requires an energy of $k_B T$ occur spontaneously
- of 10 $k_B T$ *or larger*: 0.00005 => these processes will not occur spontaneously

at 25 ºC $k_B T = 4.1 \cdot 10^{-21}$ J or **$k_B T$ = 4.1 pN·nm**

$k_B T$ refers to a single molecule
for of 1 mol of particles one has to use $k_B T$ x $6.022 \cdot 10^{23} = RT$

at 25 ºC with = 8.3 J · mol^{-1} · K^{-1} => **RT = 2.5 kJ/mol or 0.6 kcal/mol**

Hydrolysis of ATP: ~20 $k_B T$/ATP, 10 - 15 kcal/mol or 40 - 60 kJ/mol

The free energy ΔG

- **At constant pressure *P* and constant temperature *T* the system is described by the Gibbs free energy:**

$$G \equiv H - TS \qquad \Delta G = \Delta H - T\Delta S$$

- ***H* is the enthalpy or heat content of the system, *S* is the entropy of the system**
- **a reaction occurs spontaneously only if $\Delta G < 0$**
- **at equilibrium $\Delta G = 0$**
- **for $\Delta G > 0$ the input of energy is required to drive the reaction**

ΔG of an reaction in equilibrium

$$a\mathrm{A} + b\mathrm{B} + \ldots \rightleftarrows g\mathrm{G} + h\mathrm{H} \ldots$$

$$0 = \Delta G^0 + RT \ln\left(\frac{[\mathrm{G}]^{\mathrm{g}}[\mathrm{H}]^{\mathrm{h}} \ldots}{[\mathrm{A}]^{\mathrm{a}}[\mathrm{B}]^{\mathrm{b}} \ldots}\right)_{\mathrm{Eq}}$$

$$\Delta G^0 = -RT \ln\left(\frac{[\mathrm{G}]^{\mathrm{g}}[\mathrm{H}]^{\mathrm{h}} \ldots}{[\mathrm{A}]^{\mathrm{a}}[\mathrm{B}]^{\mathrm{b}} \ldots}\right)_{\mathrm{Eq}} = -RT \ln K$$

$$K = \left(\frac{[\mathrm{G}]^{\mathrm{g}}[\mathrm{H}]^{\mathrm{h}} \ldots}{[\mathrm{A}]^{\mathrm{a}}[\mathrm{B}]^{\mathrm{b}} \ldots}\right)_{\mathrm{Eq}} = \exp\left(\frac{-\Delta G^0}{RT}\right)$$

Free energy of ATP hydrolysis under physiological conditions

$$ATP + H_2O \rightleftharpoons ADP + P_i$$

equilibrium concentrations, $[\,]_{eq}$*, define* $\Delta G'^{o}$ *= the standard free energy*

$$K'_{eq} = \frac{[ADP]_{eq}/[1M] \times [P_i]_{eq}/[1M]}{[ATP]_{eq}/[1M] \times [H_2O]_{eq}/[55M]} \; ; \; \Delta G'^{o} = -RT \ln (K'_{eq}) \approx -35 \text{ to } -40 \frac{kJ}{mol}$$

correcting for physiological concentrations, $[\,]_{phys}$

$$Q' = \frac{[ADP]_{phys}/[1M] \times [P_i]_{phys}/[1M]}{[ATP]_{phys}/[1M]} \; ; \; \Delta G' = \Delta G'^{o} + RT \ln Q' \approx -50 \text{ to } -70 \frac{kJ}{mol}$$

or 20-30 $k_B T$

Weak forces maintain macromolecular structure

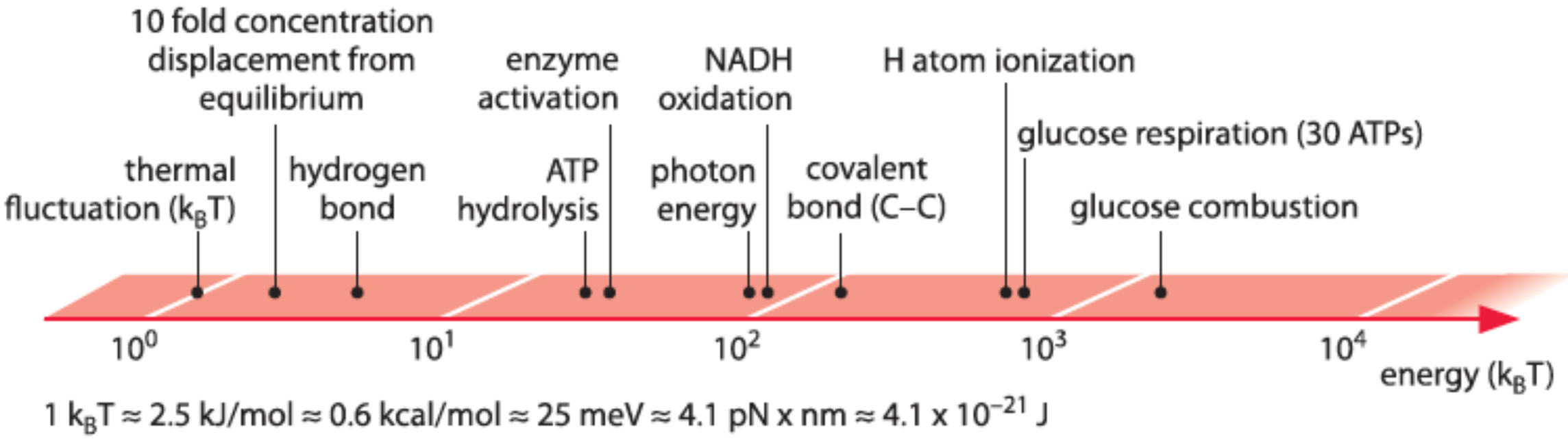
At 25 ^{o}C, 0.6 kcal/mol or 2.5 kJ/mol thermal energy: enough to break non-covalent bonds

=> Noncovalent interactions constantly forming and breaking

- Van der Waals attractive forces: $k_B T$ (0.4 - 4 kJ/mol), $1/r^6$
- Hydrogen bonds: 2-12 $k_B T$, 6 - 30 kJ/mol, $1/r^3$
- Ionic interactions: 5 kcal/mol physiological, $1/r$
- Covalent bonds: 25-300 $k_B T$, 15 - 750 kJ/mol, $1/r^{\text{very high}}$
- Hydrolysis of ATP: 20-30 $k_B T$, 50 - 70 kJ/mol

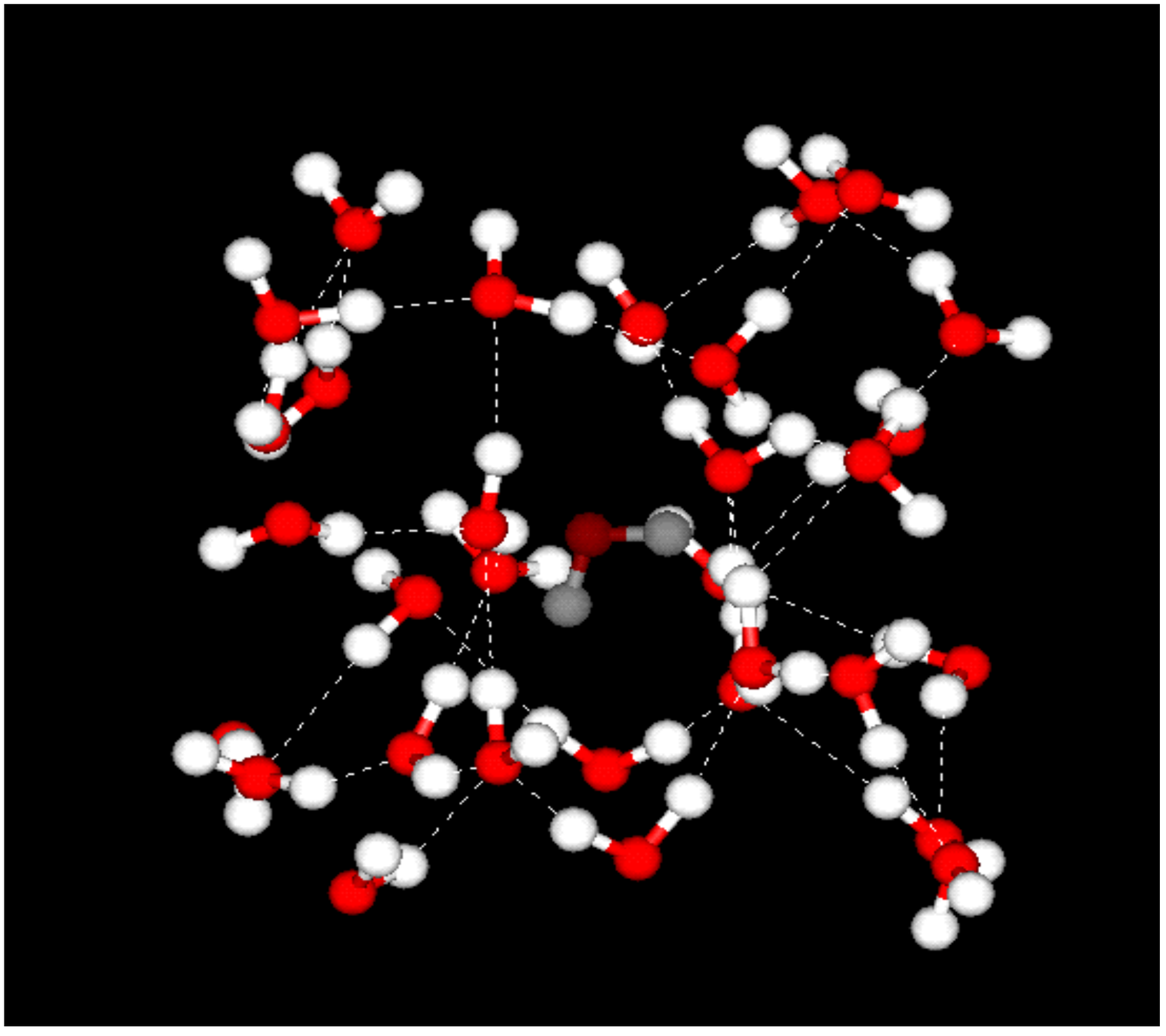
but noncovalent interactions can occur between macromolecule and water/ions just as well...

Energy scale

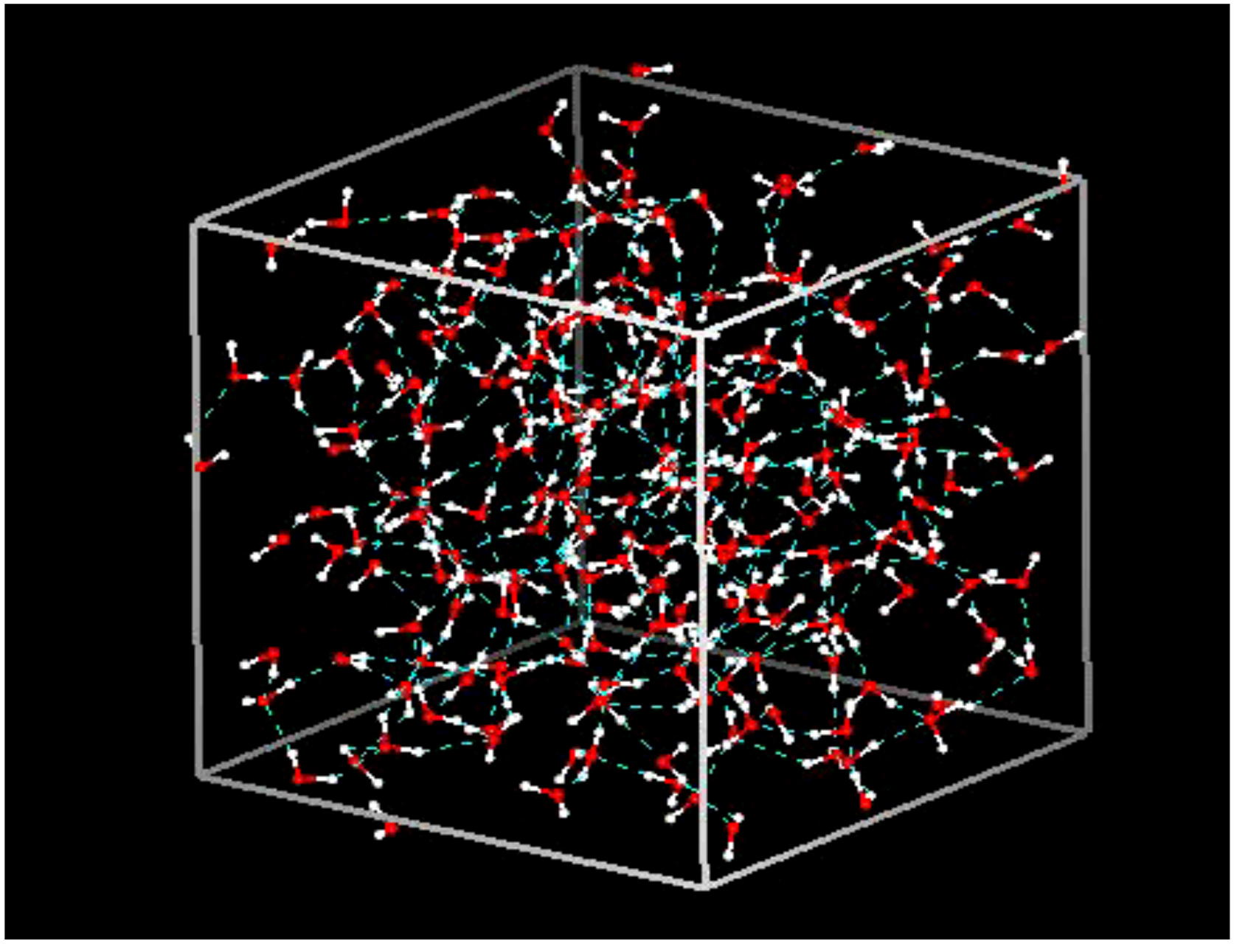


1 $k_BT \approx 2.5$ kJ/mol ≈ 0.6 kcal/mol ≈ 25 meV ≈ 4.1 pN x nm $\approx 4.1 \times 10^{-21}$ J

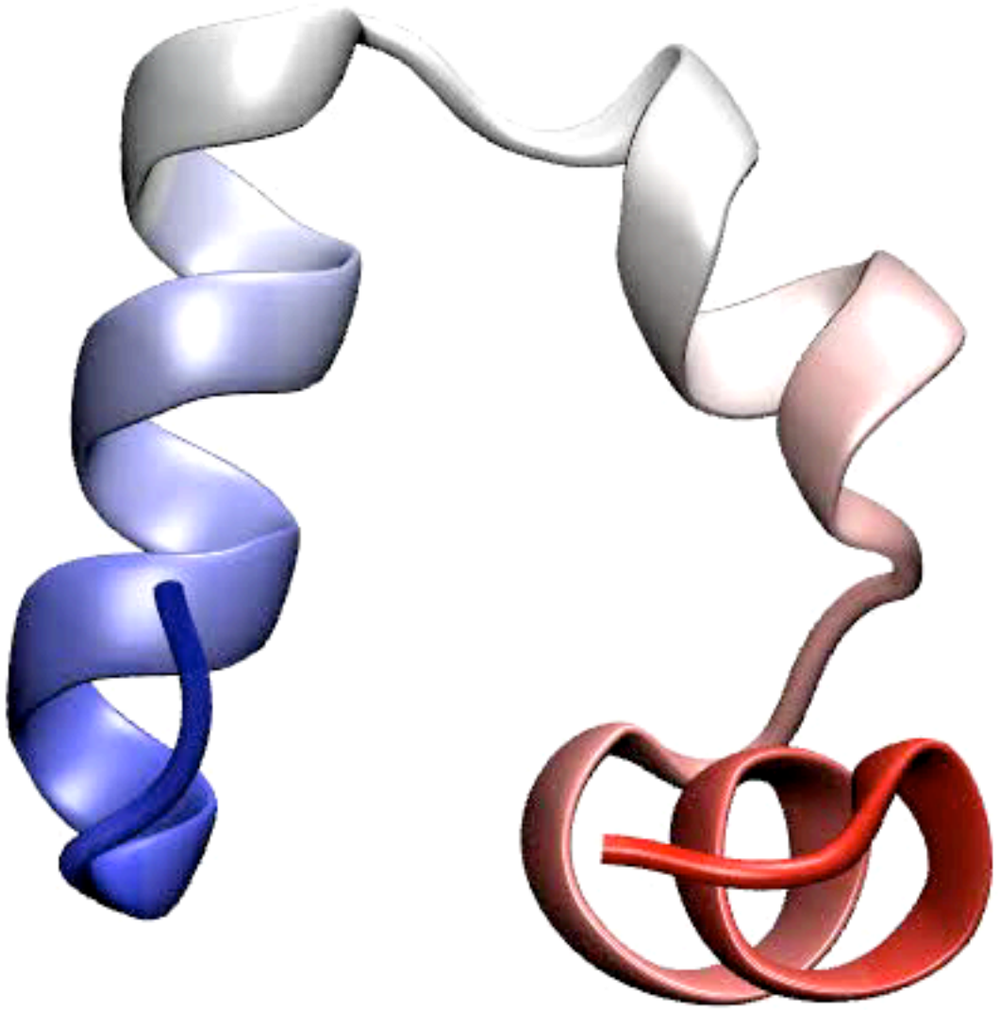
Hydrogen bonding of liquid water



Hydrogen bonding of liquid water



Protein folding as seen in molecular dynamics simulations of the villin protein headpiece (~10 μs time scale)



Peter L. Freddolino and Klaus Schulten. *Biophysical Journal*, 97:2338-2347, 2009

Calculating entropy: how probable or disordered is the final state?

Entropy provides that measure (Boltzmann)…

$$S \equiv k_B \ln W$$

- S: Molecular Entropy
- k_B: Boltzmann Constant
- W: Number of microscopic ways in which a particular outcome (macroscopic state) can be attained

For Avogadro number's of molecules…

$$S = \underbrace{(N_{Avogadro} k_B)}_{R \text{ (gas constant)}} \ln W$$

Therefore: the most probable outcome maximizes entropy of <u>isolated systems</u>

Criterion for Spontaneity:

$\Delta S > 0$ (spontaneous)

$\Delta S < 0$ (non-spontaneous)

Unfavorable conformation entropy for protein folding

for the folded state: ~ 1 conformation

$$S_{\text{folded}} = R\ln(1) = 0$$

for the unfolded state: x is the number of flexible points per residue and z is the number or possible orientations of equal energy at each point.

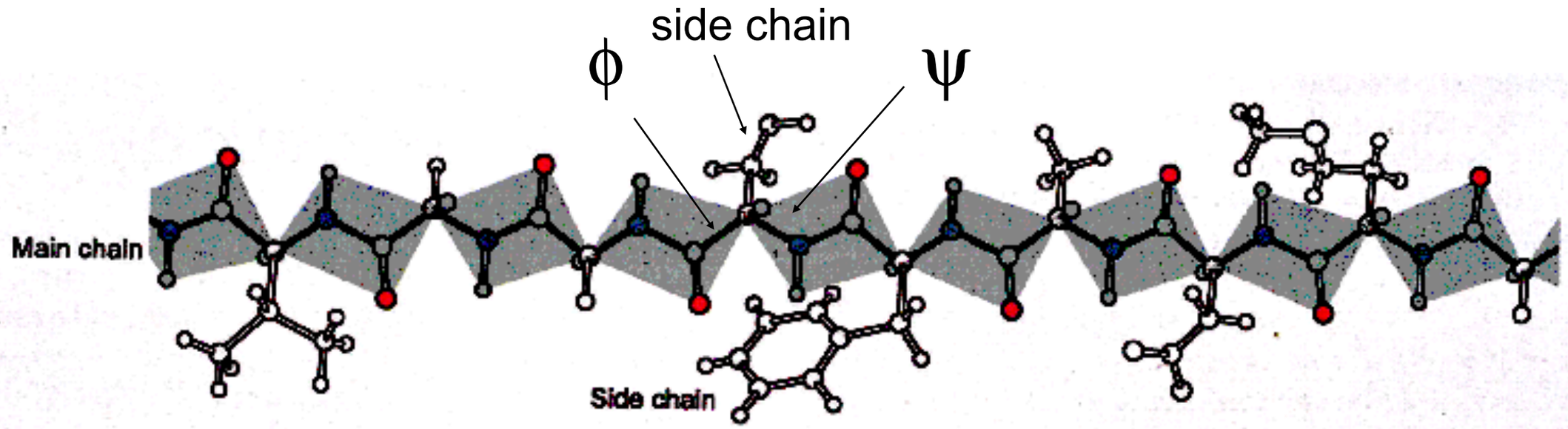
$$S_{\text{unfolded}} = R\ln\left(z^{x}\right)$$

Estimating the unfavorable conformational entropy for protein folding

$$S_{\text{conf}} = R\ln\left(z^{x}\right) \qquad \Delta S_{\text{fold}} = R\ln\left(\frac{W_{\text{unfolded}}}{W_{\text{folded}}}\right)$$

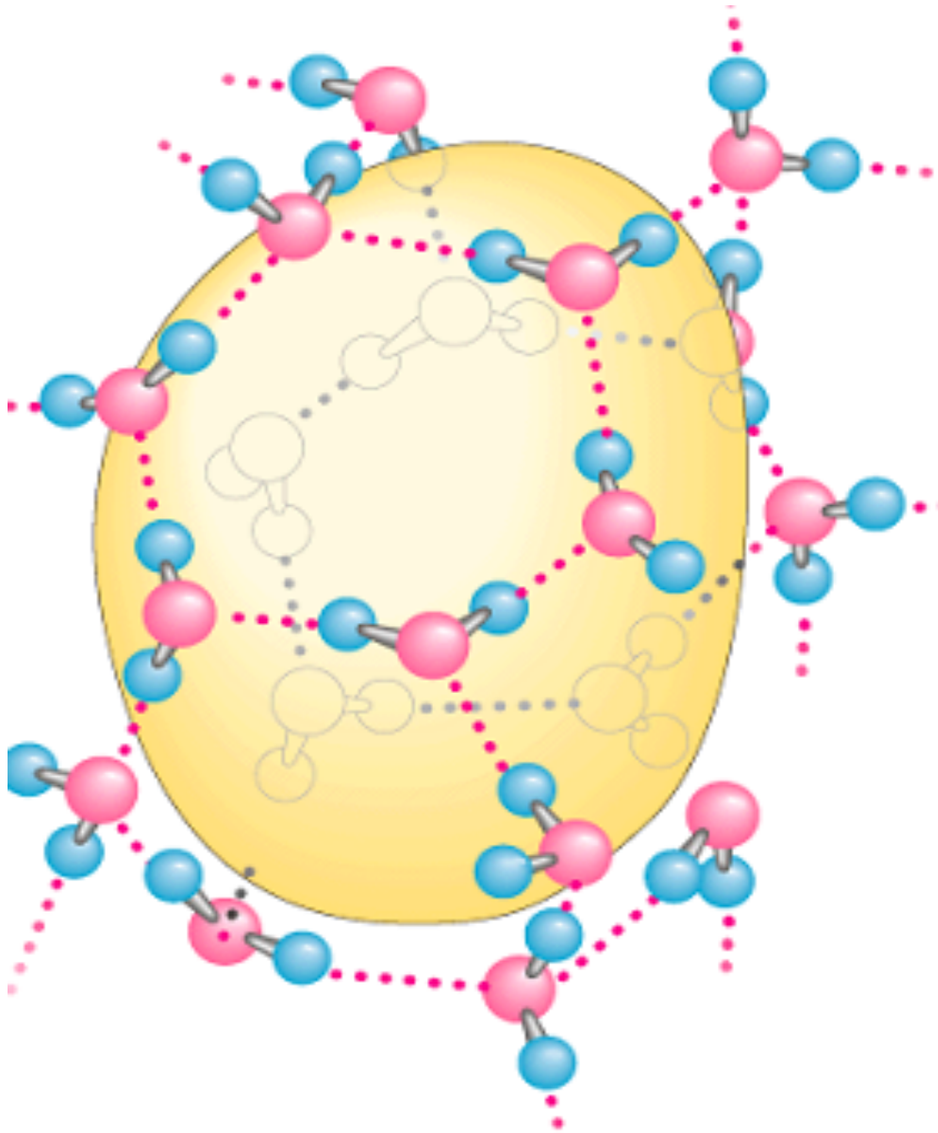
Tanford: For three flexible positions (ϕ , ψ, side chain) with two possible orientations each we have 2^3 conformations per residue:

-TΔS = 1.2 kcal/mol or 5.0 kJ/mol (estimate Tanford 1962)
1.7 kcal/mol (Spolar and Record, Science)

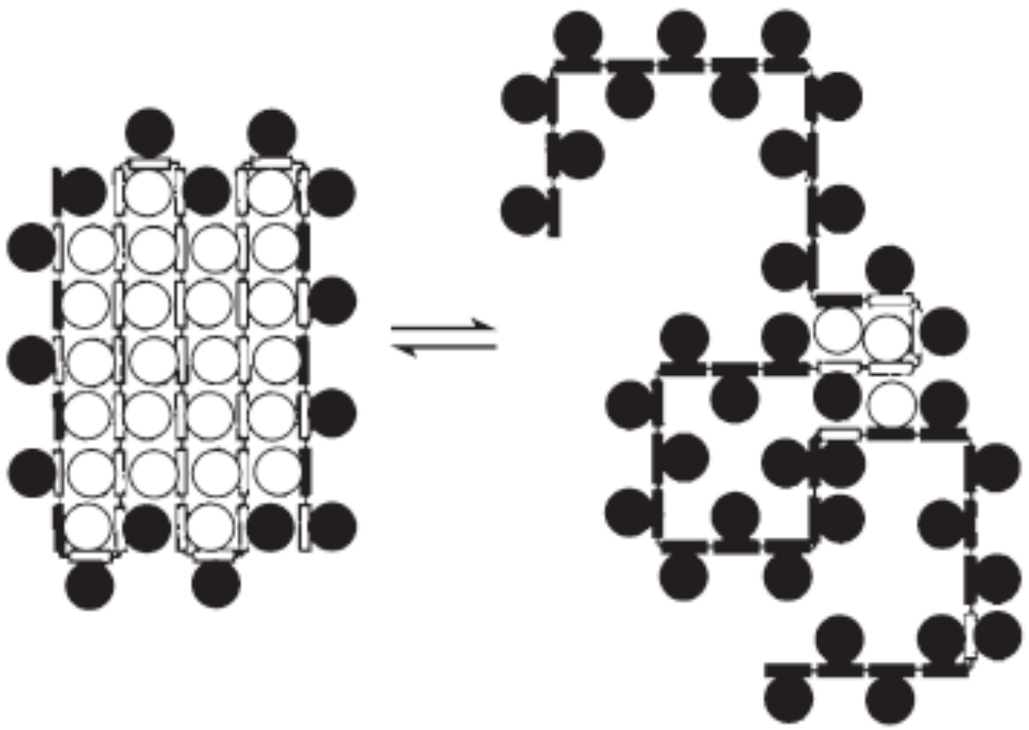


The hydrophobic effect drives macromolecular interactions

- Minimization of non-polar/water surface area leads to stability
- Complex mixture of physical properties
- Entropic contribution most significant
- Water must form a “cage” structure around non-polar surfaces



Scheme of protein folding with solvent accessible amino acids in black



folded state

random coil

Table 1. Folded globular proteins.

	buried[a]
non-polar side chains (Ala, Val, Ile, Leu, Met, Phe, Trp, Cys)	83%
peptide groups (−CO–NH−)	82%
ca. 1.1 intramolecular hydrogen bonds formed per residue[b]	

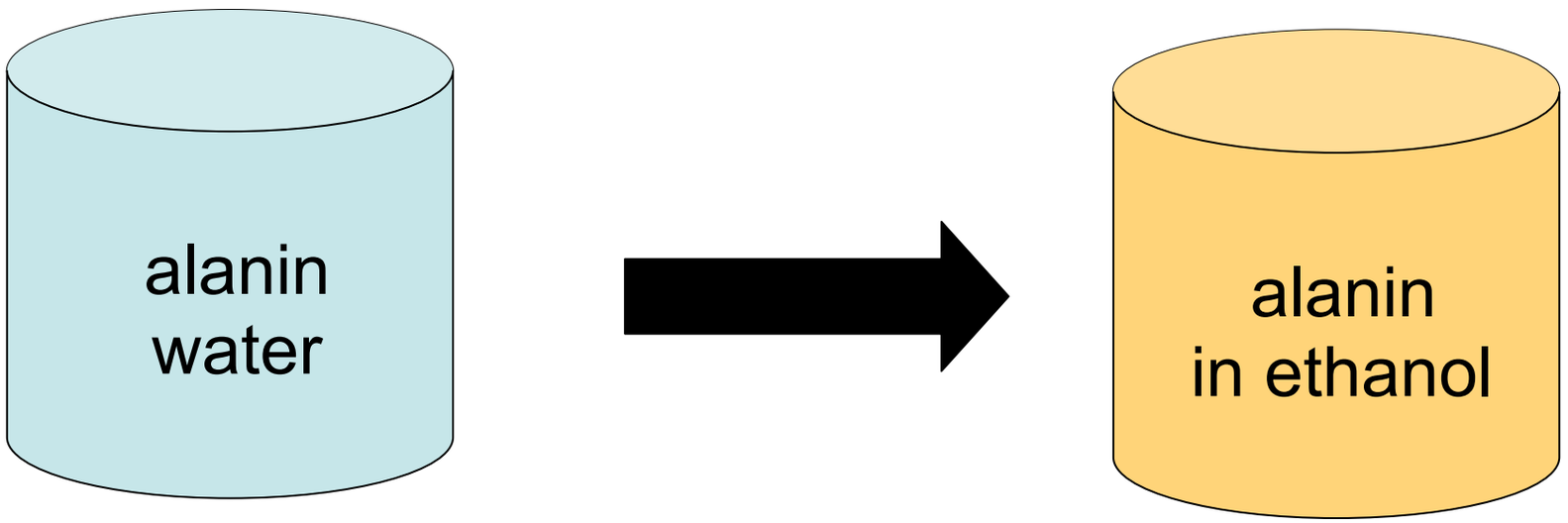
[a] Lesser & Rose (1990).
[b] Stickle *et al.* (1992).

Measure solubility of amino acid in ethanol (= inside the folded protein) and in water (= unfolded state or at the protein surface)

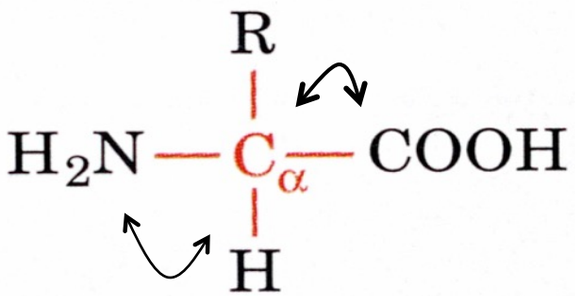
$$\text{partition coefficient } K_{\mathrm{D}} = \frac{\text{solubility } (Alanin_{\mathrm{EtOH}})}{\text{solubility } (Alanin_{\mathrm{H_2O}})}$$

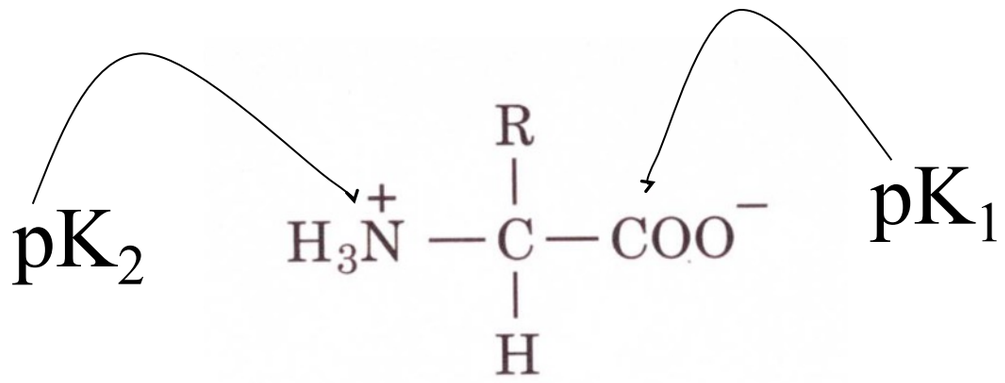
Calculate the free energy from transferring an amino acid from water to ethanol

$$\Delta G_{\mathrm{tr}} = -RT\ln\left(\frac{N_{\mathrm{EtOH}}}{N_{\mathrm{H_2O}}}\right) = -RT\ln\left(K_{\mathrm{D}}\right)$$



Free amino acids carry a positive and a negative charge that is not present in the peptide chain





α amino acids because of the α carboxylic and α amino groups

pK_1 and pK_2 respectively pK_R is for R group pK's

$pK_1 \approx 2.2$ while $pK_2 \approx 9.4$

In the physiological pH range, both carboxylic and amino groups are completely ionized

[CONTRIBUTION FROM THE DEPARTMENT OF BIOCHEMISTRY, DUKE UNIVERSITY, MEDICAL CENTER, DURHAM, NORTH CAROLINA]

Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins

BY CHARLES TANFORD

RECEIVED APRIL 9, 1962

Tanford, C. *J Am Chem Soc* **84,** 4240–4247 (1962).

Tanford 1962

TABLE I[a]

FREE ENERGY CHANGE IN CALORIES PER MOLE FOR TRANSFER FROM ETHANOL TO WATER AT 25°

	ΔF_t, whole molecule	Δf_t, side chain contribution	
Non-polar side chains			
Glycine	−4630	0	
Alanine	−3900	+ 730	
Valine	−2940	+1690	
Leucine	−2210	+2420	
Isoleucine	−1690[b]	+2970[b]	
Phenylalanine	−1980	+2650	
Proline	−2060[c]	+2600[c]	
Other side chains			
Methionine	−3330	+1300	
Tyrosine	− 930[d]	+2870[d]	
Threonine	−4190	+ 440	
Serine	−4590	+ 40	
Asparagine	−4640	− 10	
Glutamine	−4730	− 100	
Aspartic acid[e]	−4090	+ 540	uncharged
Glutamic acid[e]	−4080	+ 550	uncharged

Burying a charged amino acid in the interior (Born expression)

$$W_B = \frac{q^2}{4\pi\varepsilon_0 r}\left(\frac{1}{\varepsilon_1} - \frac{1}{\varepsilon_2}\right)$$

W_B is the free energy of transfer in moving a charged body from a region with a relative dielectric constant ε_2 to a medium with a with a relative dielectric constant ε_1. The parameter r is the radius of the charge.

q (charge of an electron) = 1.60 x 10^{-19} C
dielectric constant in vacuum ε_0 = 8.85 x 10^{-12} C^2 J^{-1} m^{-1}
r is ionic radius, with is typically 1-2 Å

Sharp, K.A. and Honig, B. (1990) Electrostatic interactions in macromolecules: theory and applications. Annu Rev Biophys Biophys Chem, 19, 301-332.

Tanford 1962

TABLE III

CONTRIBUTION OF THE MOST IMPORTANT HYDROPHOBIC INTERACTIONS TO THE FREE ENERGY OF UNFOLDING AT 25°

Side chain	Δf_u per side chain, cal./mole	Number present in myo-globin[a]	β-lacto-globulin[b]	ribo-nuclease[c]
Tryptophan	3000	2	2	0
Isoleucine	2970	9	10	3
Tyrosine	2870	3	4	6
Phenylalanine	2650	6	4	3
Proline	2600	4	8	4
Leucine	2420	18	22	2
Valine	1690	8	10	9
Lysine	1500	19	15	10
Methionine	1300	2	4	4
Alanine	730	17	14	11
Arginine	730	4	3	4
Threonine	440	5	8	10
Total number of residues		153	162	124
$-T\Delta S_{conf}$, kcal./mole		−184	−194	−149
$\Sigma\Delta f_u$, kcal./mole		+173	+192	+100

conformation entropy
hydrophobic effect

Energy scale

- Using numbers for cellular processes
- Energy scales
  - $k_B T$ (single molecule) or RT (per mole) reference scale
  - Equilibrium free energy ΔG or equilibrium binding constant
  - Entropy from changes of number of states
- Macromolecular interactions in aqueous solutions
  - non-covalent interactions
  - hydrophobic effect