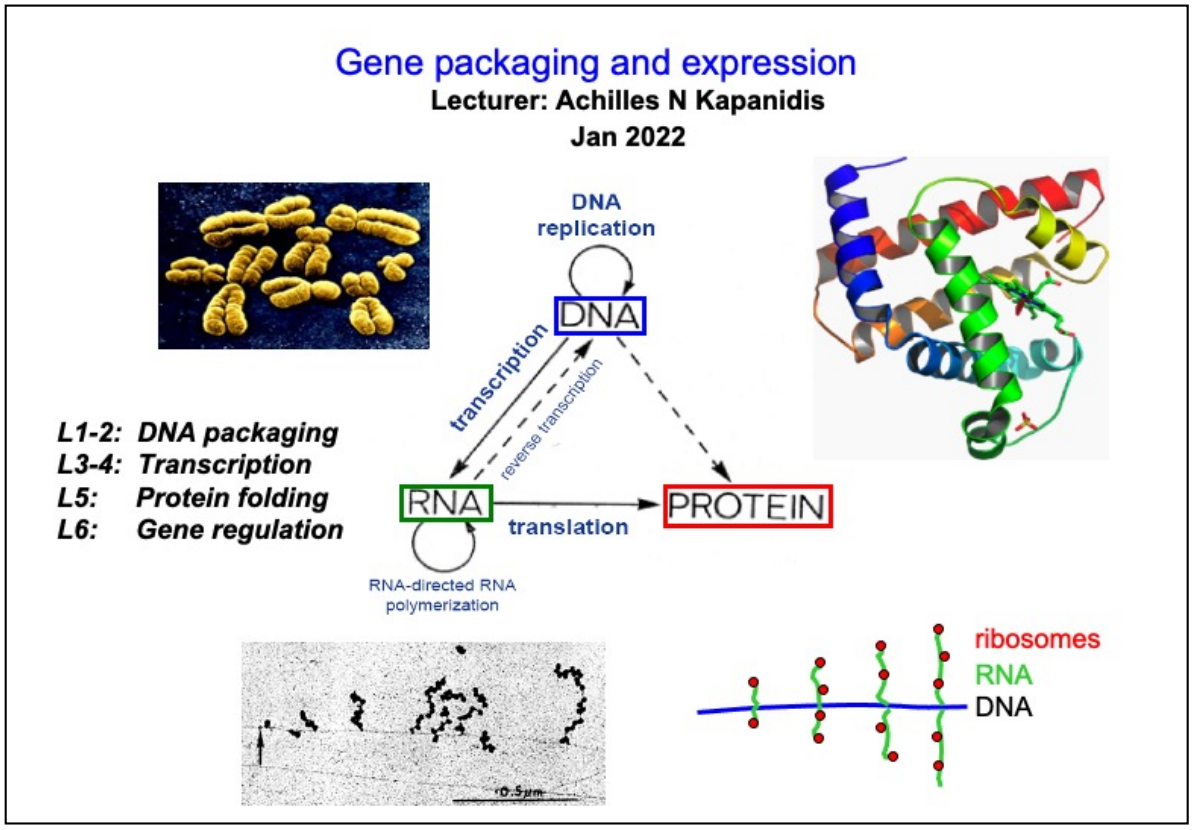


This part of the course will deal with genes and the mechanisms responsible for their packaging and expression.

Such mechanisms are crucial for life, since they allow cells and organisms to store and retrieve information quickly, when/where needed, and with few errors.

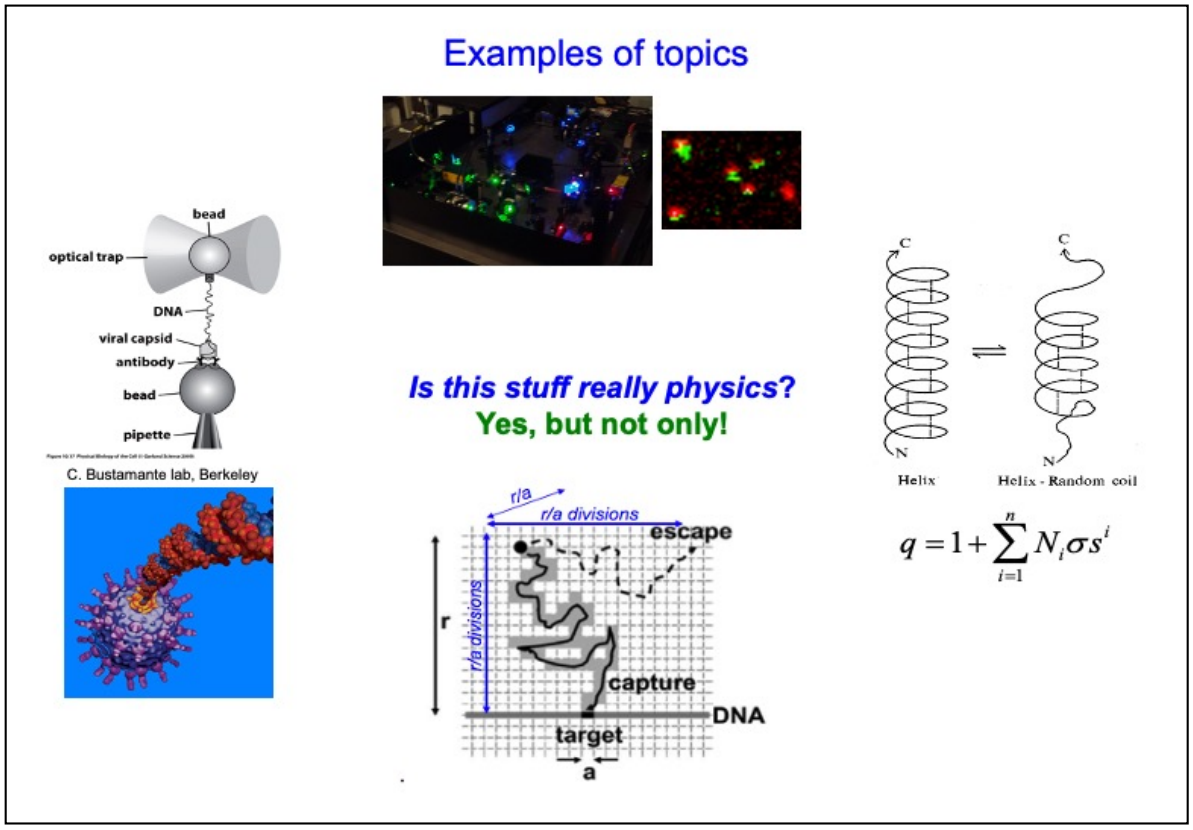
2006 Nobel prizes: transcription for Chemistry (Roger Kornberg for X-ray structures of RNA polymerase), gene regulation (for gene silencing; Andrew Fire and Craig Mello for discovering it) for Medicine.

2009 Nobel prizes: translation for Chemistry (Tom Steitz, Ada Yonath, Venki Ramakrishnan for X-ray structures of the ribosome), DNA replication (for telomeres and telomerases, Elizabeth Blackburn, Carol Greider, Jack Szostak) for Medicine.

2012 Nobel prizes (medicine): mature cell reprogramming through DNA and transcription factors (John Gurdon and Shinya Yamanaka)

2020 Nobel prize (chemistry): CRISPR-based genome editing (Emmanuelle Charpentier and Jennifer Doudna)

Mechanistic perspective: we will look at important protein machines and structures responsible for gene-related processes and their mechanisms



Mechanism, molecular machines, and force-based modulation of activity

Single-molecule methods: force, light, and electricity

Simple statistical mechanical models to model transitions in protein folding

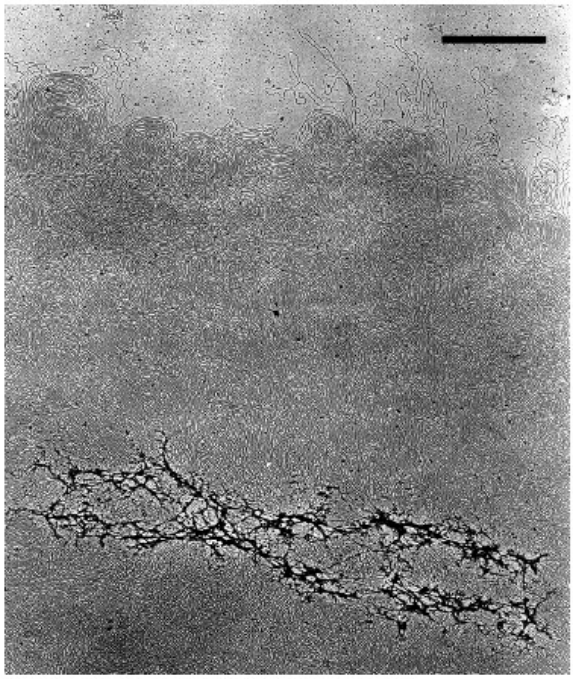
Modeling search processes in cells using simple models of molecular diffusion and random walks

Biological Physics is an interdisciplinary arena; we will need the basic biology and biochemistry to understand the context, the questions, and some of the tools!

DNA compaction and packaging

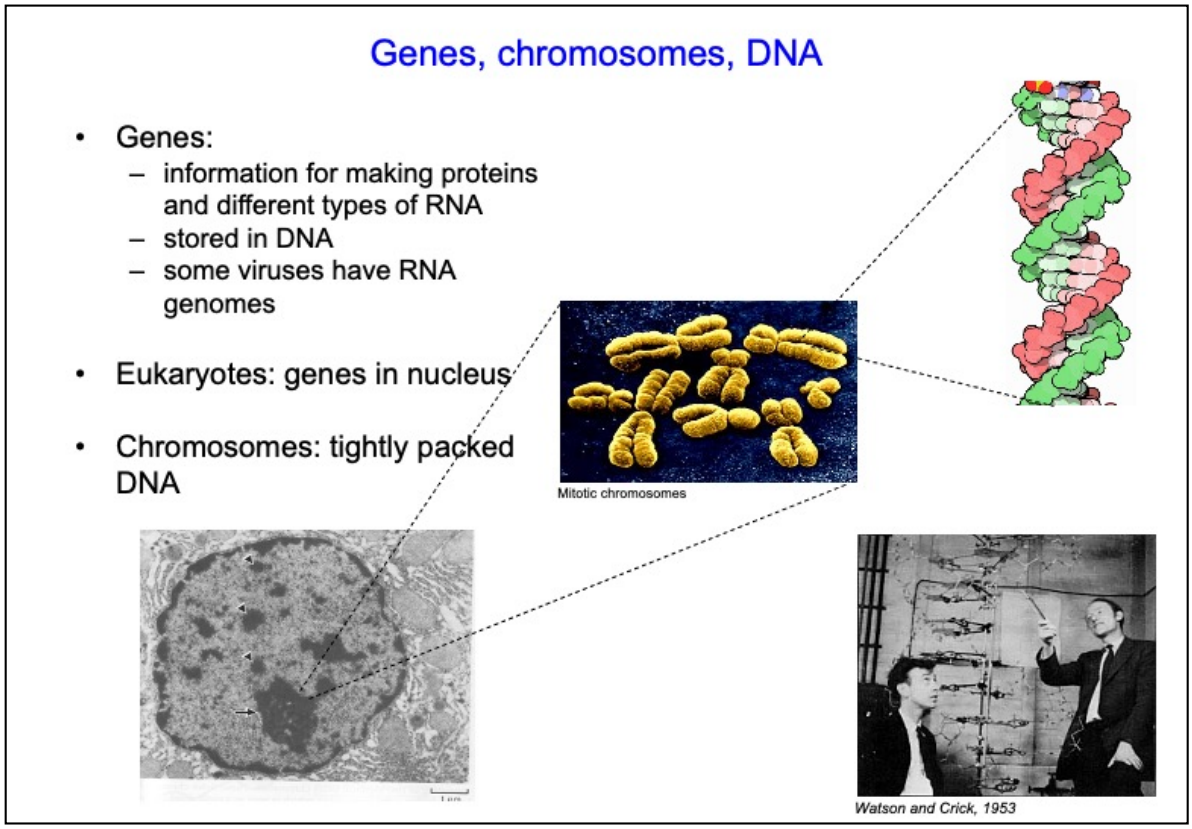
Lecture 1
A N Kapanidis

- Alberts, Chapter 4: DNA and chromosomes (pg 191-216)
- Schiessel H, 2003: The physics of chromatin (*J. Phys.: Condens. Matter* **15,** pg 1-11)
- References in this handout



Paulson and Laemmli, Cell, Vol. 12, 817-828, November 1977.

The big challenge: how to package enormously long DNA strands in biological cells and structures while keeping the DNA accessible for copying and repair.

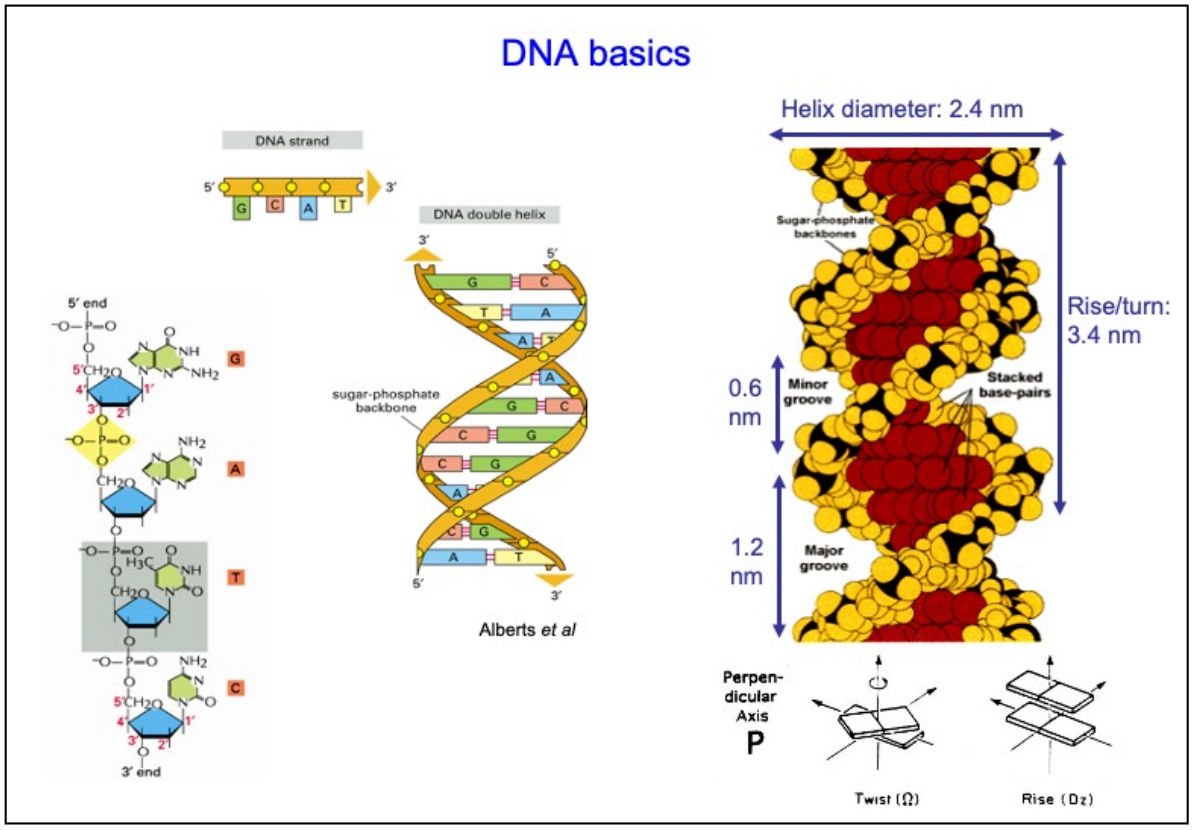


Genes determine the characteristics of species and of individuals and they are passed from generation to generation.

Genes are instructions, written in the language of DNA, that code for the synthesis of proteins and RNA. These instructions control when, where, and how much of the given protein (or RNA) should be prepared.

Genome of an organism: the complement of genetic instruction of an organism.

Mitotic chromosomes are chromosomes that have been fully replicated and are ready for mitosis, i.e., the process of providing each daughter cell with one copy of each parental chromosome.



DNA in cells is made of two strands that form a right-handed spiral: the famous double helix.

Each strand is made of a sequence of 4 deoxyribonucleotides: A, T, C, G

Helix exterior: sugar-phosphate backbone. Due to the negative charge of the phosphate groups, the sugar-phosphate backbone is hydrophilic.

Helix interior: the steps of the spiral staircase; these are called base-pair steps, and they are hydrophobic in nature.

The two polynucleotide chains run in opposite directions

Two grooves of different magnitude; the wide groove known as the "major groove", and a narrower groove known as the "minor groove". Both grooves are used by proteins for recognizing and processing DNA.

To get a right-handed helix, a basepair has to be rotated relative to its closest neighbour; this rotation occurs around the local long axis of the double helix and is known as "twist". Twist per basepair step for B-DNA: ~36°.

Genomes are large and need compaction

- Human genome: 3.2×10^9 bp in 24 different chromosomes
- Extend DNA in a human *diploid* cell:
  2 m contour length
  (6.4×10^9 bp x 0.34×10^{-9} m/bp)
- This should fit in nucleus (~10 µm diameter)
  → the "compaction problem"
- Compaction should allow use and repair of genes





24 different chromosomes: 22 pairs of homologous chromosomes (autosomes) + X + Y for men or X + X for women = **46** chromosomes in total

Immense degree of compaction: equivalent to squeezing a 40 km-long fishing line in a tennis ball!

DNA is also enormously thin compare to its length: for the human genome, the length is ~9 orders of magnitude longer than its width!

Compaction should occur in an organised, hierarchical way, that allows:

-Copying of genes for duplication (DNA→DNA)

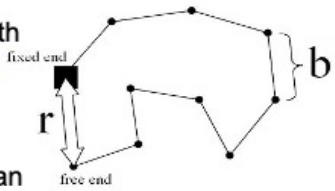
-Transcription of genes for expression (DNA→DNA)

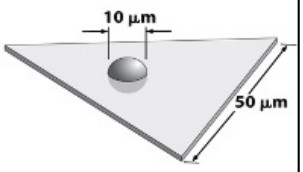
-Repair of genes

Analogy to a library: 1D string of letters stored in an organised way→lines, pages, books, rows, bookcases, sections and floors. 1 book: 10 km text

But DNA behaves like a worm-like chain!

- Worm-like chain: freely jointed chain with segment length b twice the persistence length l_p *(Kuhn length)*





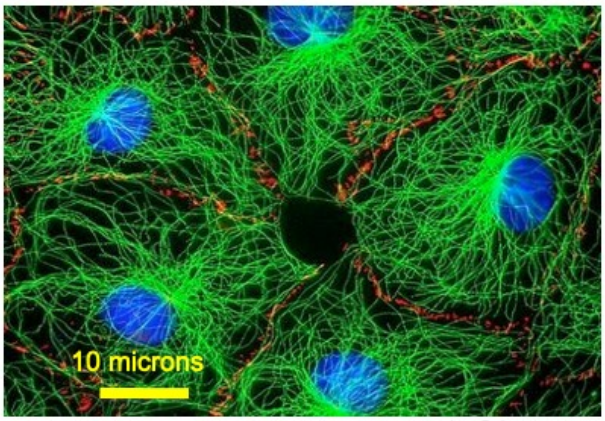
- Average end-to-end distance *r* for human genome:

$RMS_{distance}$= (*Kuhn length* x contour length)$^{1/2}$

~ (2 x 50 nm x 2 m)$^{1/2}$

~ 450 μm

- Even a "WLC genome" won't fit at all in the nucleus!



Jan Schmoranzer

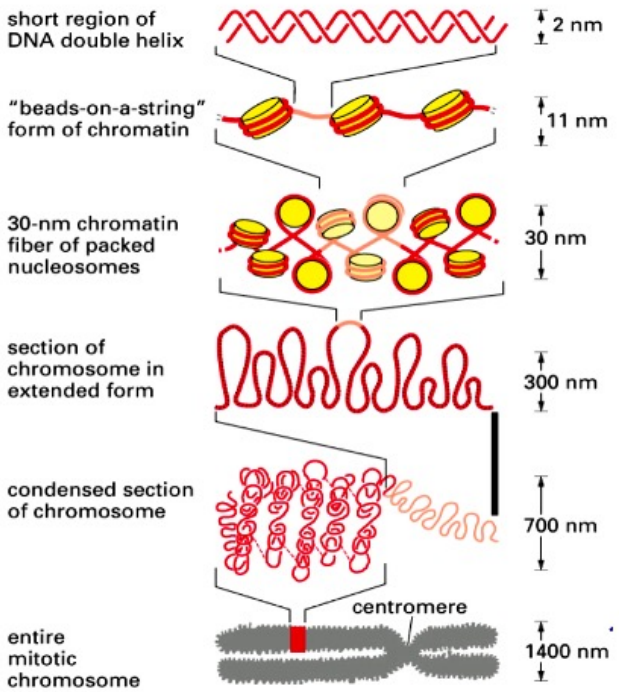
Kratky-Porod model (worm-like chain model): see B1 biophysics lectures.

Persistence length: a measure of stiffness of a linear chain. It is the characteristic distance at which thermal energy starts to affect significantly the direction of the long axis of the polymer.

The **Kuhn length** for double-stranded DNA is 100 nm.

Fibroblasts are cells that produce fibrous tissues and collagen. Can be considered as a "standard cell" in several biophysical discussions.

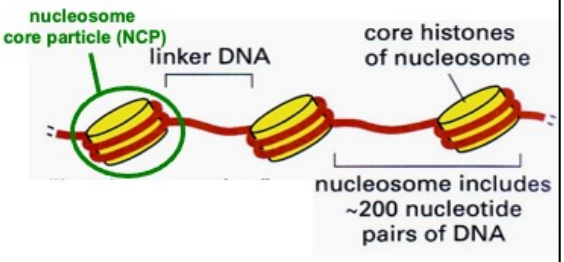
Solution to the compaction problem: hierarchical folding



Example: chromosome 22: stretched is 1.5 cm long; in mitosis, only ~2 µm long: **10^4-fold compaction** How? protein-based DNA folding. Main protein-DNA complex: chromatin

Main unit of compaction: nucleosome

- nucleosome proteins: histones
- chromatin: histone-DNA complex
- histones abundant: $6x10^7$ copies per cell
- first observed using electron microscopy



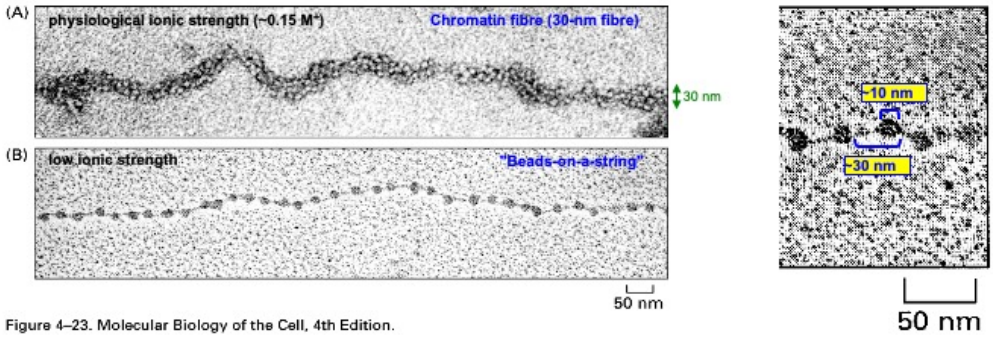


Figure 4–23. Molecular Biology of the Cell, 4th Edition.

Discovered in 1974

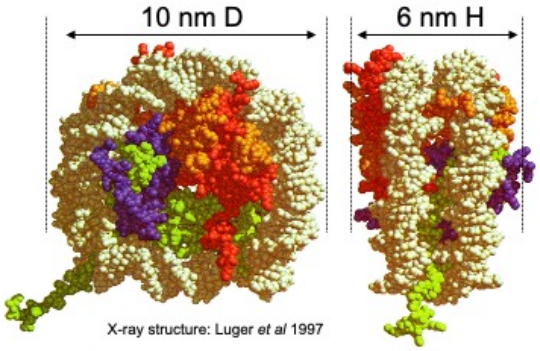
Definitions: nucleosome, histone octamers, linker DNA, nucleosome core particle

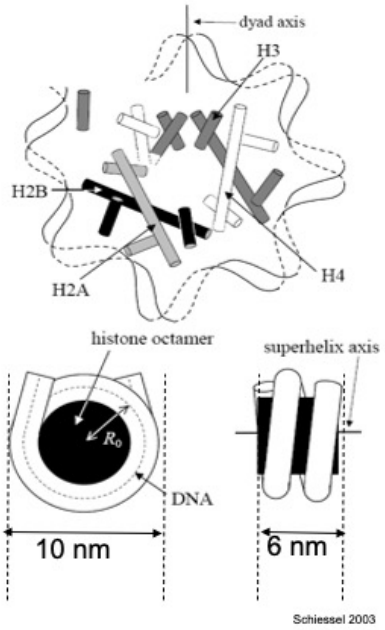
Decrease ionic strength: increase the repulsion between parts of chromatin → chromatin decondenses

1 nucleosome per 200nt; ~3 $x10^7$ nucleosomes per diploid cell

Dissecting the nucleosome

- NCP: left-handed superhelix, 1.7 turns
  - 147 bp DNA: 127 bp bent, 20 bp straight
- 14 points of histone-DNA contact
  - Interactions not sequence-specific
- 147 bp (~50 nm) + linker → ~10 nm + linker → **2- to 3-fold compaction**





NCP: 5 nm radius/6 nm height.

DNA-histone interactions:

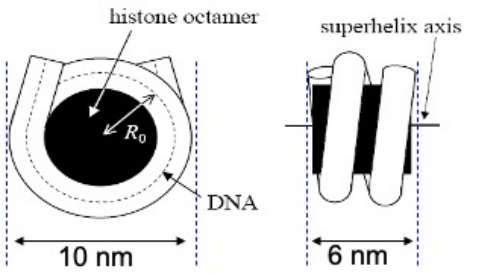
- 142 H-bonds (histone backbone→ DNA backbone)
- Salt bridges (cationic Lys, Arg amino acids→ anionic phosphates in minor groove)
- Hydrophobic interactions

DNA bent by sharp kinks (not smooth bending)

Energy considerations for the nucleosome

Geometric considerations:

- contour length *l*: 127x0.34 =43 nm
- persistence length l_p: 50 nm
- radius of curvature R_o: 4.3 nm



Schiessel 2003

Energy cost for bending 127 bp DNA:

$$E_{elastic} \approx \frac{l_P l}{2R_o^2} k_B T \approx \frac{50 \cdot 43}{2 \cdot 4.3^2} \approx 58 k_B T$$

Distributed among ~14 histone-DNA contacts: ~ +4 k_BT per contact

Since **net** adsorption energy per contact is ~ -2 k_BT *(later slides):*

=> ~ -6 k_BT of pure adsorption energy per contact

Assumption: smooth bending. Not particularly accurate (DNA wrapping in the NCP is kink-based), but good for reliable estimates.

The expression for the elastic energy for bending DNA is derived by applying beam theory considerations to double stranded DNA.

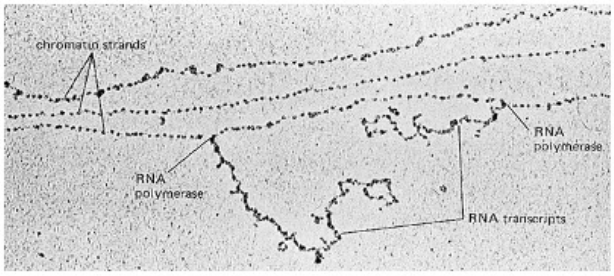
Check equation qualitatively: if persistence length increases, the energetic cost increases; if curvature ($1/R_o$) increases (tighter bend), the energetic cost increases; and the cost scales with the contour length of DNA to be wrapped around the histone core.

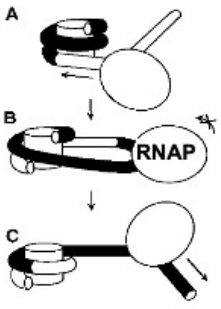
The large, favorable interaction per contact site shows that the non-specific interactions more than compensate for the energetic cost of bending.

"Bendable" sequences will have shorter persistence length, therefore the bending cost will be smaller, and the nucleosome will be more stable.

Nucleosomes are dynamic

- Nucleosome accessible to proteins for:
  - DNA repair
  - DNA replication
  - Gene transcription
  - Transcription regulation
- Nucleosomes moved on DNA by molecular motors





Studitsky et al, (1997) Science 278, 1960

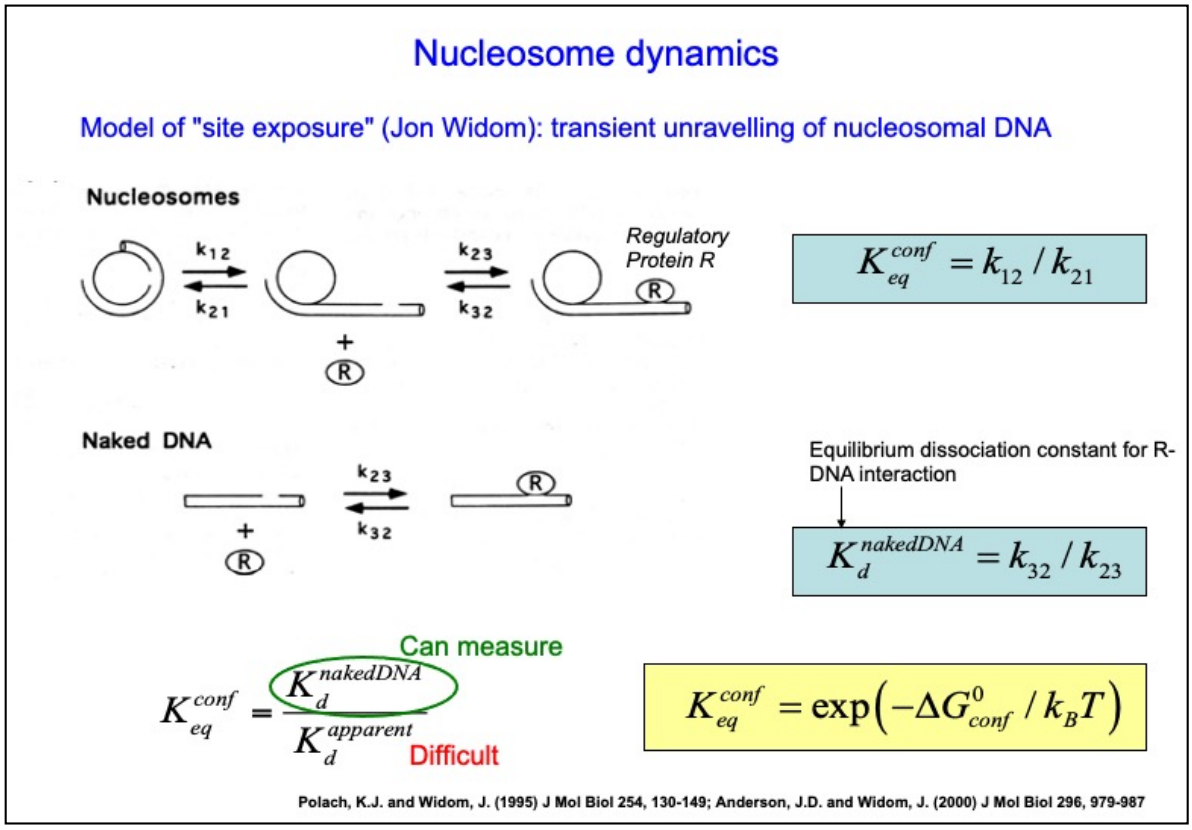
=> Nucleosome **has** to be a dynamic structure

Electron micrograph: two RNA polymerases transcribing DNA on chromatin (beads-on-a-string pattern on DNA); see also RNA forming secondary structures. Therefore, gene transcription takes place without problems on chromatin templates. But how RNA polymerase bypasses the nucleosome?

[Note: electron micrographs can be misleading since sample preparation for EM converts 3D structures into 2D and requires harsh treatments; therefore, conclusions on dynamics of cellular structures based on EM are questionable.]

Model of nucleosome bypass by RNA polymerase (RNAP): RNAP moves towards the nucleosome which partially unwraps, and through a transient loop, it is transferred from a "downstream"-DNA location (i.e., in front of RNAP) to an "upstream"-DNA location (i.e., behind RNAP).

Interesting (and unresolved) question: does RNAP "wait" for nucleosome to be transferred, or actively invades it? (RNAP can exert forces up to 25 pN)



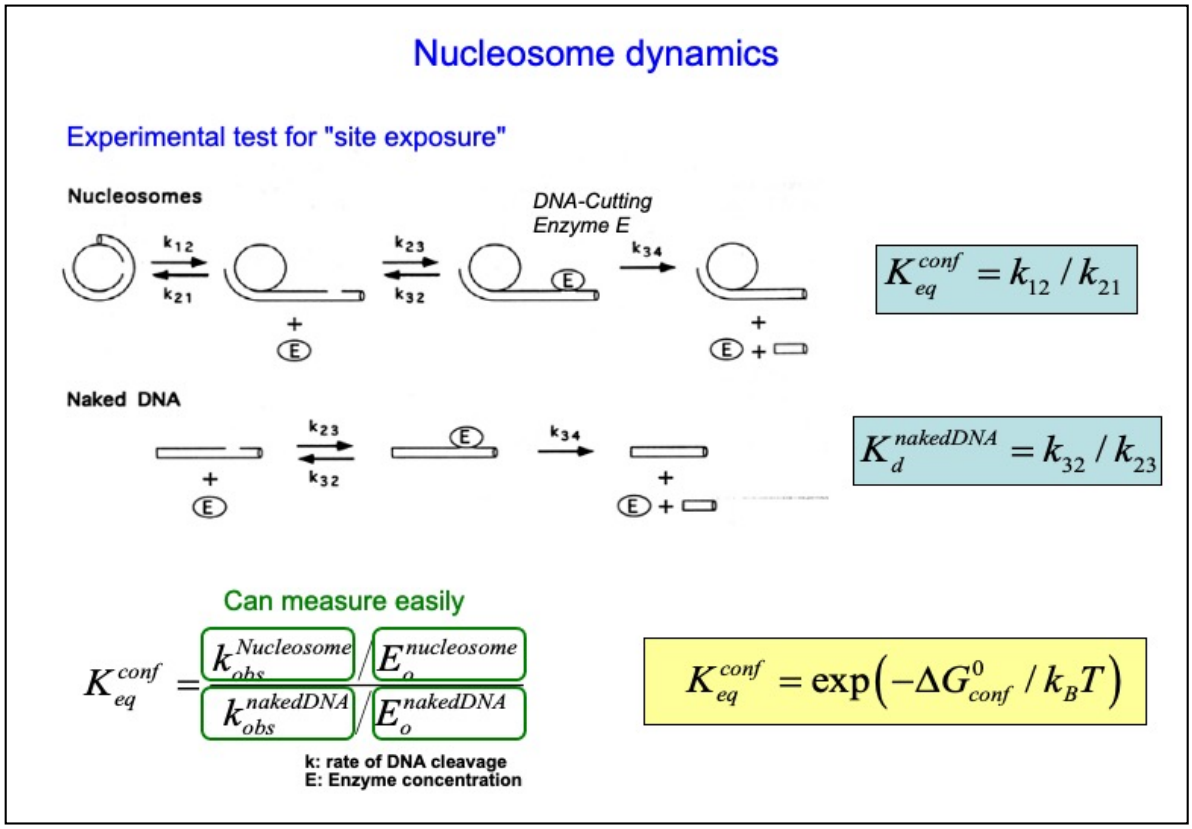
Site-exposure provides a window of opportunity for a regulatory protein R (e.g. a protein that controls transcription) to bind to DNA; then, this binding can help another protein to bind, leading to cooperativity.

The nucleosome will, of course, decrease the binding of R as compared to naked DNA; this will INCREASE greatly the apparent equil dissociation constant K_{d_appar}, since K_{conf_eq} is much smaller than 1. However, it is difficult to measure K_{d_appar} .

We need to find K_{conf_eq} in order to:

a. show that unwrapping occurs

b. to calculate the energetic cost of unwrapping

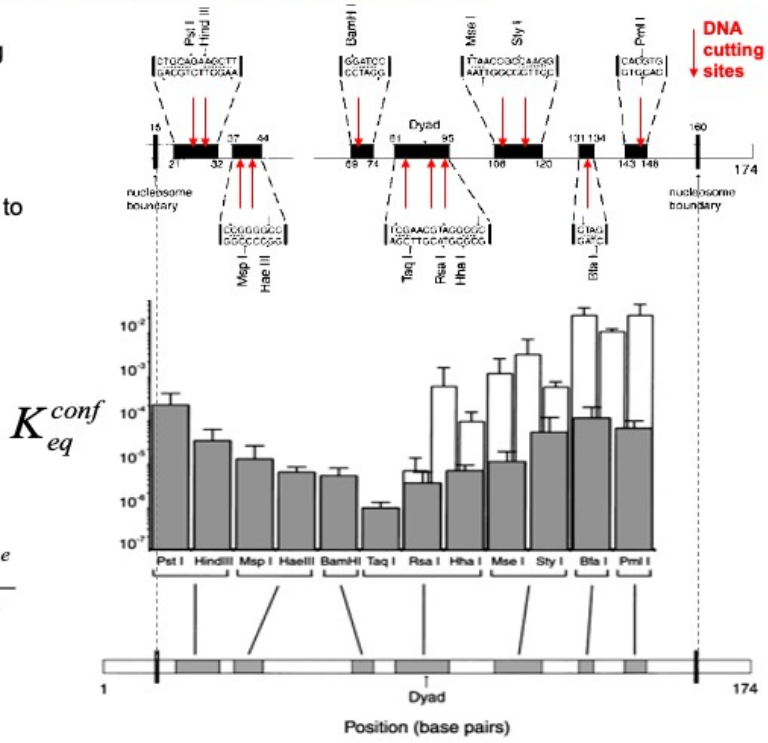


An experimentally tractable system was set based on "DNA-restriction" enzymes (enzymes that cleave DNA at specific sites). It was shown that the unwrapping equil constant K_{conf_eq} can be measured by measuring rates of DNA digestion at specific enzyme concentration.

Nucleosome dynamics: experiments

- Nucleosomes with DNA having many sites for DNA-cutting enzymes
- cuts everywhere within nucleosomal DNA, even close to dyad axis!
- rate constants, enzyme concentrations → K_{eq}^{conf}

$$K_{eq}^{conf} = \frac{k_{obs}^{Nucleosome} / E_o^{nucleosome}}{k_{obs}^{nakedDNA} / E_o^{nakedDNA}}$$



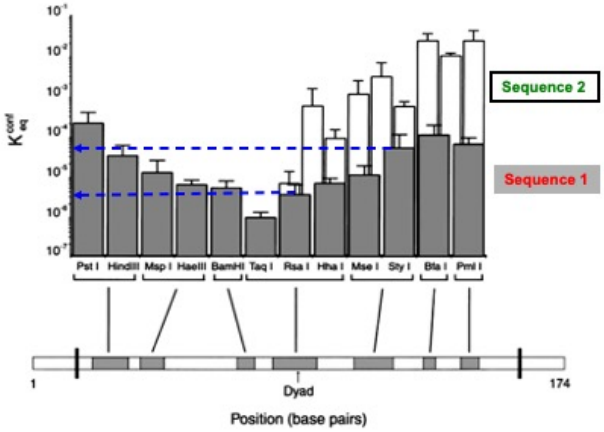
The DNA-cutting reaction (known as "DNA restriction") was monitored by monitoring the electrophoretic mobility of DNAs in polyacrylamide gels: intact DNA migrates in polyacrylamide gels more slowly than the cut DNA (since due to large size, it is slowed down by the small pores of the gel).

Note that the enzymes were used one at a time.

"Dyad": the two-fold symmetry axis of the nucleosome, therefore the site further away from the nucleosome boundaries.

Calculation of net free energy per contact site

- Sites close to dyad axis cleaved less efficiently (less accessible)
- Different sequences of nucleosomal DNA: different K_{eq}^{conf}
  - DNA sequence sets accessibility through "bendability"
- Cost of 2-3 k_BT per contact site exposed

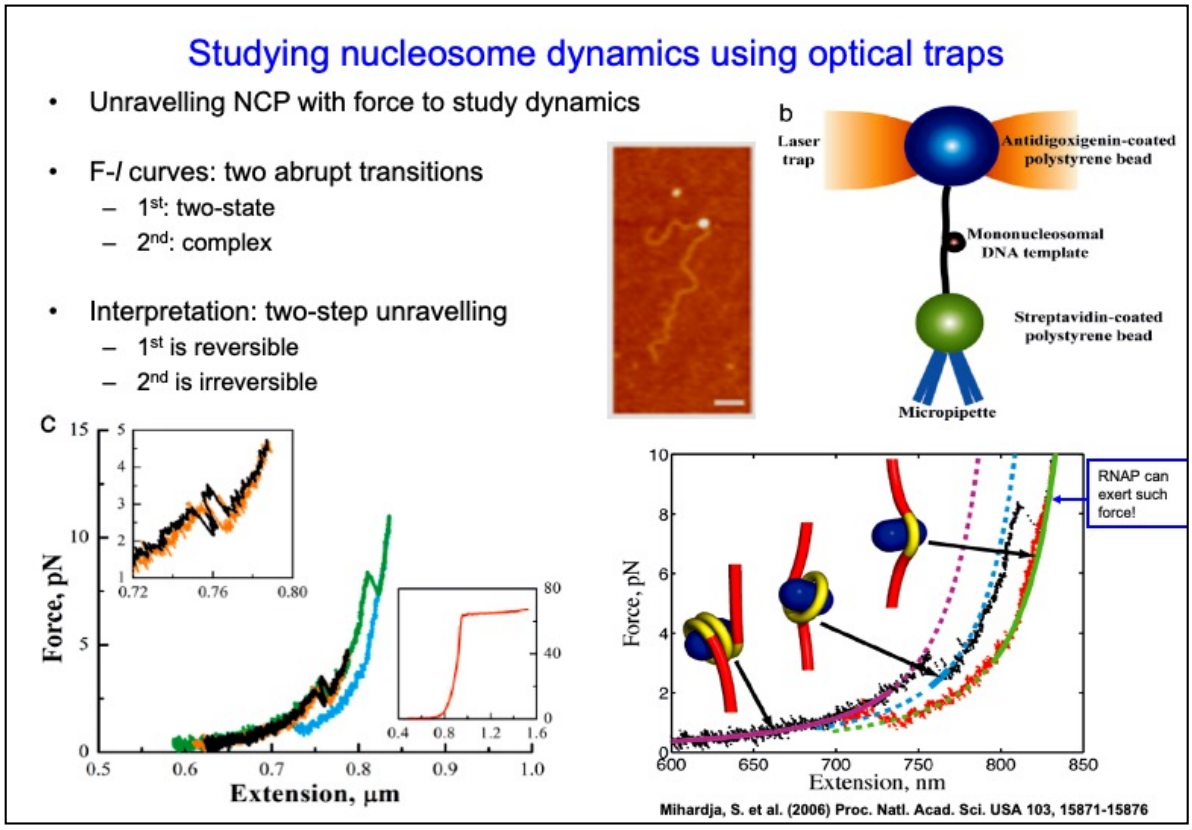


$$\Delta G_{conf}^{0} = -k_B T \ln K_{eq}^{conf}$$

For *StyI* site: $K_{eq}^{conf} \sim 5 \cdot 10^{-5} \Rightarrow \Delta G_{conf}^{0} \sim 10 k_B T$ **~2.5 kT per contact** (~40 bp from end)

For *RsaI* site: $K_{eq}^{conf} \sim 3 \cdot 10^{-6} \Rightarrow \Delta G_{conf}^{0} \sim 13 k_B T$ **~2 kT per contact** (~70 bp from end)

Note: sequences also determine nucleosome positioning, linked to the positioning-dependent DeltaG values for histone-DNA interactions (Anderson & Widom J Mol Biol 2000, 296, 979-987)



Force-extension analysis of structural transitions has been covered during the discussion of RNA folding studies using optical tweezers (see R Berry lectures).

Long (3600bp) DNA with specific sequence that "positions" nucleosome reproducibly and stably

Low-force rip: ~21nm, reversible. The histone core does not dissociate completely from nucleosomal DNA.

High-force rip: ~22nm, irreversible.

Each rip: ~contour length of 1/2 of total DNA wrapped in the nucleosome (total DNA wrapped: 127 x 0.34 nm ~43 nm). In other words, each transition removes ~1 DNA wrap from the nucleosome.

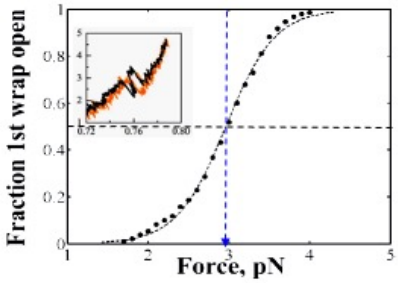
Free energy for first unwrapping

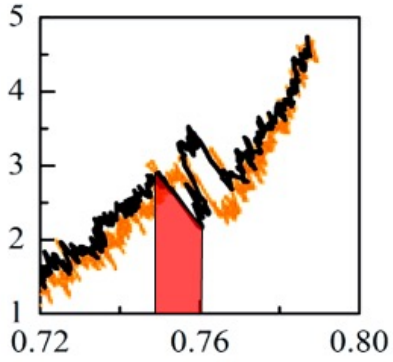
- Method 1: since 1st transition is a stochastic two-state process, probability of 1st wrap to open at a given force;

$$p(open) = \frac{1}{1 + e^{(\Delta G_{unwrap} - F\Delta x)/k_B T}}$$

- Best fit → ~25 kJ/mol

- Method 2: from reversible work $F\Delta x$ (measure area under force-extension curve) → ~22 kJ/mol





Mihardja, S. et al. (2006) Proc. Natl. Acad. Sci. USA 103, 15871-15876

Both methods were used first with RNA hairpins (pioneering work of the Bustamante lab, discussed by Dr Berry).

The midpoint of the sigmoidal curve provides the pulling force that counteracts the large positive term $DeltaG_{unwrap}$ (as unwrapping is unfavoured energetically).

Free energy for first unwrapping

- Method 3:
  - force-clamp system at different forces around critical force
  - get kinetics of wrapping-unwrapping transitions (k_{unwrap}, k_{wrap})
  - get K_{eq}:

$$K_{eq} = k_{unwrap} / k_{wrap}$$

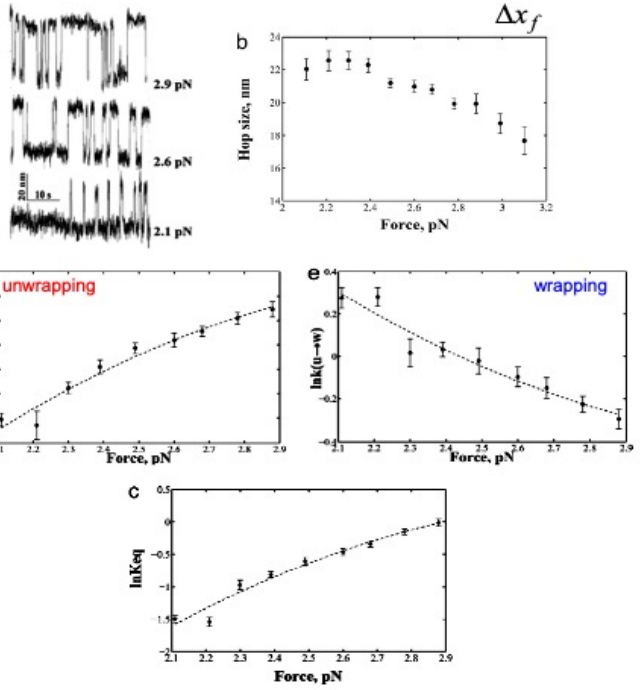
  - Extrapolate to F=0 to get K_{eq} in the absence of force:

$$K_{eq} = e^{(F\Delta x_f - \Delta G_{unwrap} - \Delta G_{stretch}^{w \to u})/k_B T}$$

(Small correction term)

$$\Rightarrow K_{eq}^{F=0} \approx 6.6 \cdot 10^{-6}$$

$$\Rightarrow \Delta G^o \approx 11.9 k_B T \approx 30 kJ/mol$$



Mihardja, S. et al. (2006) Proc. Natl. Acad. Sci. USA 103, 15871-15876

Two-state hopping behaviour (as with the experiments with RNA hairpins by the same lab (discussed by Dr Berry)

Unwrapping rate k(w→u): $1/t_w$, where t is the dwell time in the wrapped state (smallest extension)

Wrapping rate k(u→w): $1/t_u$, where t is the dwell time in the unwrapped state (largest extension)

$DeltaG_{stretchw \to u}$: a small correction term due to tethering and stretching of the molecule

Ensemble vs. single-molecule measurements

- Single-molecule (optical traps):

$$\Delta G^o \approx 22-30 KJ/mol \Rightarrow 9-12 k_B T$$

- Bulk (enzyme accessibility):

$$\Delta G^o \approx 10-13 k_B T$$ (depends on position)

Good agreement!

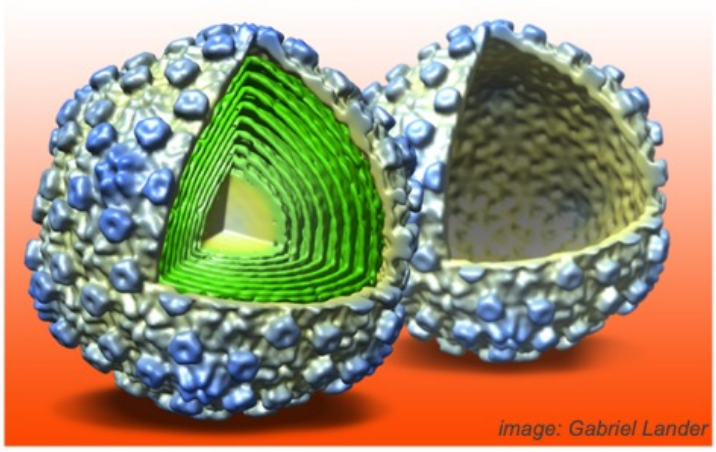
Mihardja, S. et al. (2006) Proc. Natl. Acad. Sci. USA 103, 15871-15876

DNA packaging by bacteriophages

Lecture 2

Achilles N Kapanidis

January 2022



50 nm (!!!)

How and why do they do it???

DNA Packaging in phage lambda. Shown are the structures of the filled (left) and empty(right) capsids of bacteriophage lambda as determined by cryoEM.

Reading for this lecture: Physical Biology of the Cell, Rob Phillips et al, pages 372-381

Phage: the best DNA packers

- DNA Packing ratio *v* for phage lambda:
  - 48,500 bp DNA
  - Spherical head of ~60 nm diameter

L_{genome}	5×10^4 bp	5×10^6 bp	10^9 bp	10^9 bp
ν	0.6	0.1	0.02	0.002

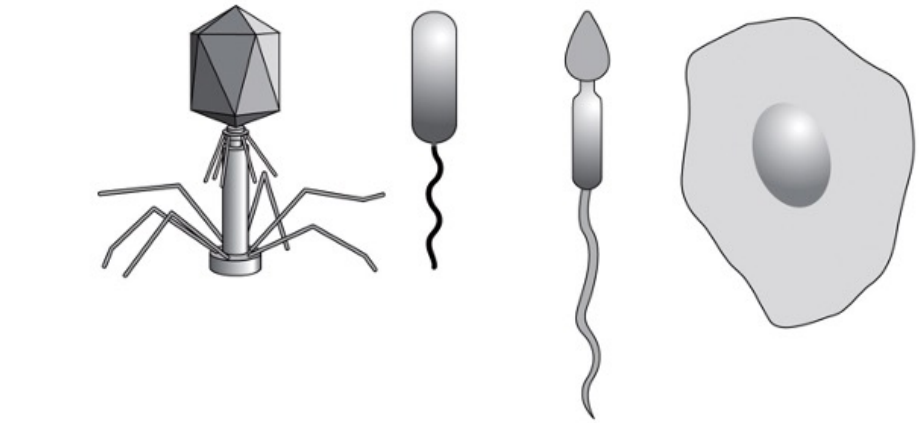
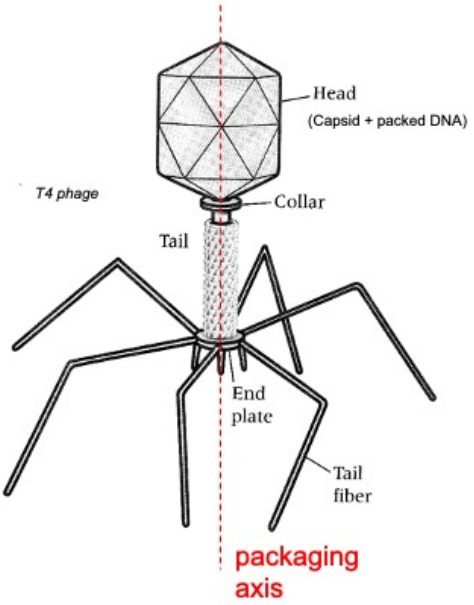


Figure 10.13 Physical Biology of the Cell (© Garland Science 2009)

Packing ratio: fraction on container volume occupied by DNA

Phage vital statistics



Bacteriophages or phages: viruses that infect bacteria

Short genomes (10-50 kb)

5-20 proteins

capsid: protein shell of the DNA compartment of phage

most capsids are helical or icosahedral

Phages are the first model organisms of molecular biology; some of the founders of molecular biology (early 40's; mainly physicists and biologists) were actually known as the "phage group".

A **capsid** is the protein shell of a virus head. Icosahedral capsids have 20 equilateral triangular faces and approximate a sphere. Helical capsids are cylindrical.

Phages are being developed for bionanotechnology uses and for understanding self-assembly.

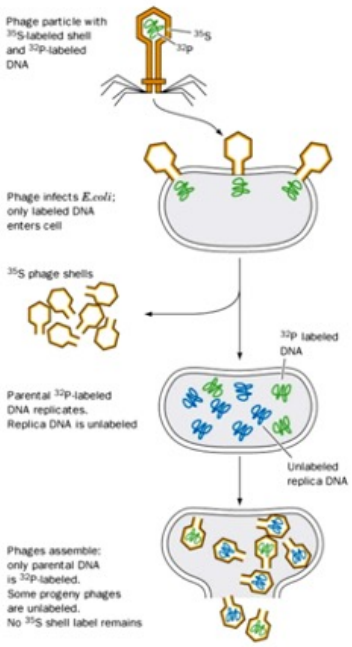
Reminder from earlier lectures…

Genes are made from DNA: Bacteriophage T2



Bacteriophages infecting a bacterium

Hershey-Chase experiment showed that only nucleic acids entered the bacterium



Voet Biochemistry 3e
© 2004 John Wiley & Sons, Inc.

Phages were also used to show that genes are made of DNA (see A Turberfield lectures, Hershey-Chase experiment).

Life cycle of a phage

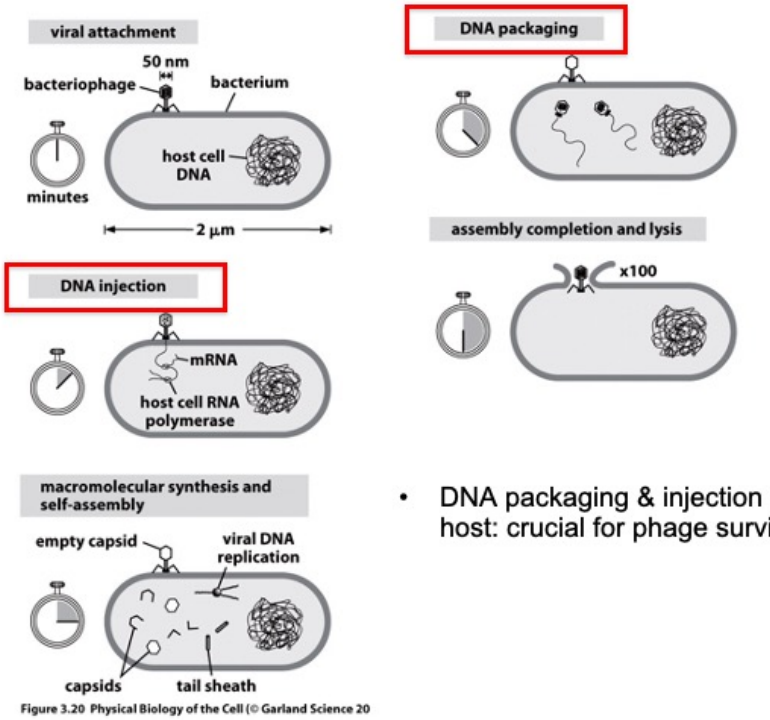
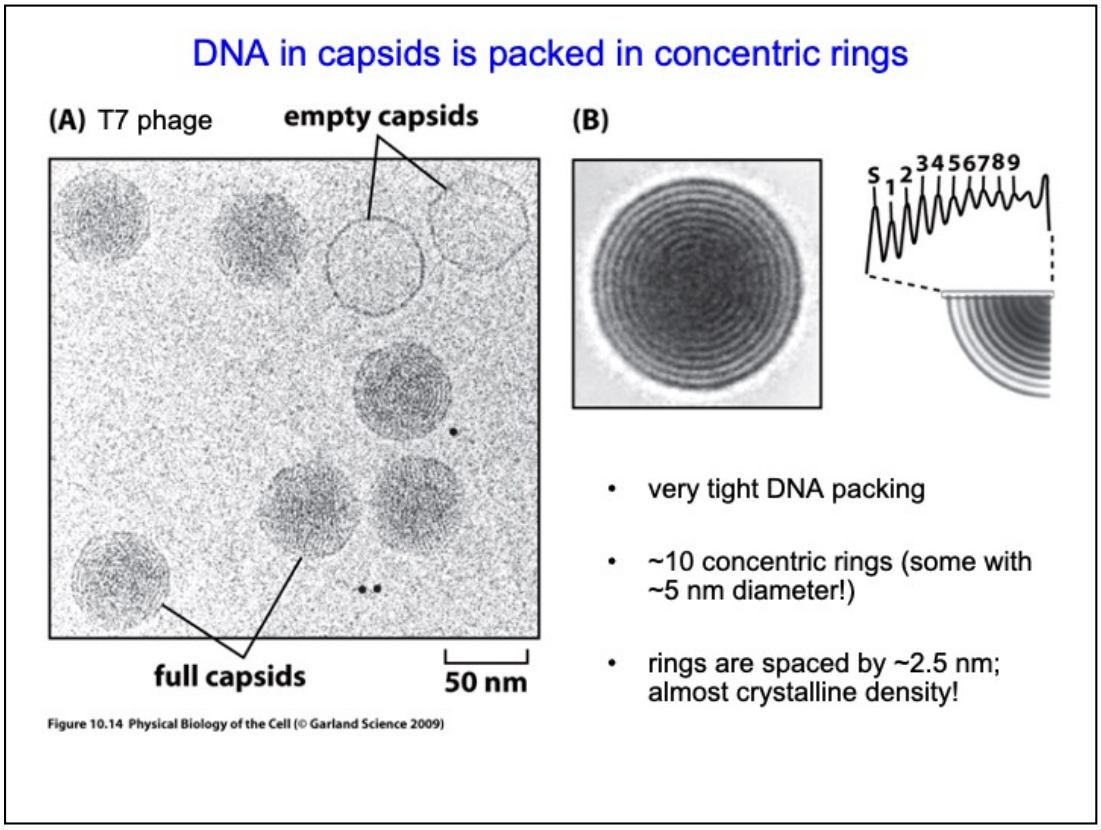


Figure 3.20 Physical Biology of the Cell (© Garland Science 20

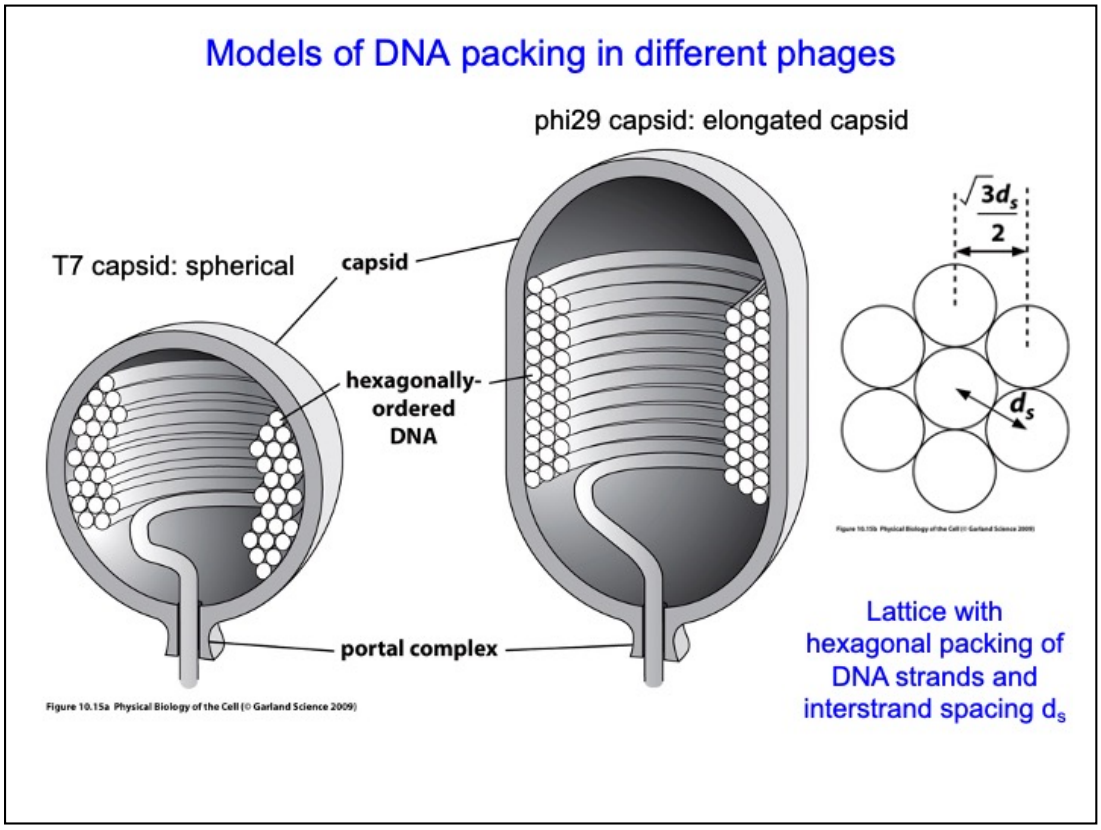
- DNA packaging & injection into host: crucial for phage survival



DNA diameter: 2.4 nm. Inner diameter for T7 phage: 55 nm. Contains T7 genome of 40 kb DNA. Note that the capsids are too small to be visualized with light microscopy.

The projection shown is along the packaging axis of the DNA (i.e., along the tail fibre). The projections represent the average of many images.

Data give rise to the “coaxial spool” (or inverse spool, since the wrapping of the strands is physically done from the outer ring to the inner one) model for the organization of the DNA within phage capsids.



Both phages have an icosahedral symmetry.

DNA is packed in an “inverse spool” configuration that features concentric rings packed from the outside of the capsid towards the center.

How does spacing scale with length of genome packaged?

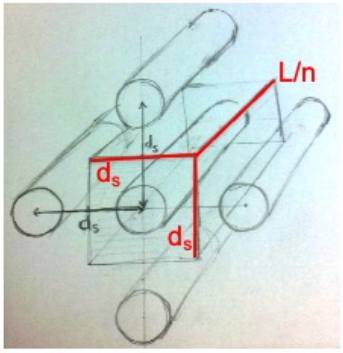
- Consider regular packing with spacing d_s:
- Volume occupied by DNA of length L:

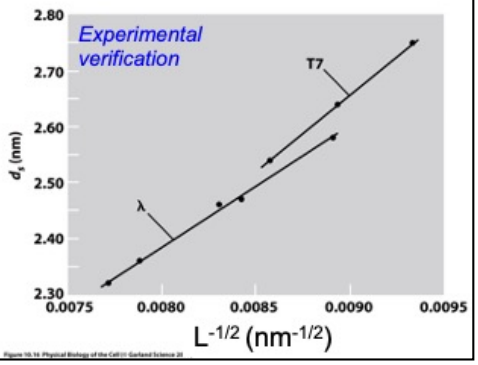
$V_{DNA} \sim d_s^2 L$

Since most of capsid fills with DNA:

$V_{capsid} \sim d_s^2 L$, therefore $d_s = (V_{capsid}/L)^{1/2}$

DNA spacing decreases as more DNA is packed in the capsid

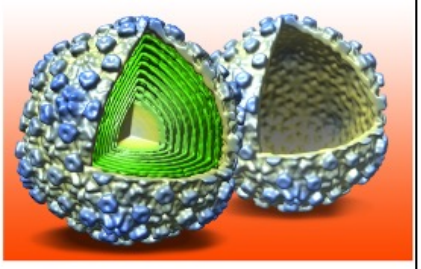




If the contour length is split into *n* segments of length L/n, then each segment occupies an elementary volume of d_s x d_s x L/n (see schematic at top right), and since there are *n* such DNA segments, V_{total}: $d_s^2 L$

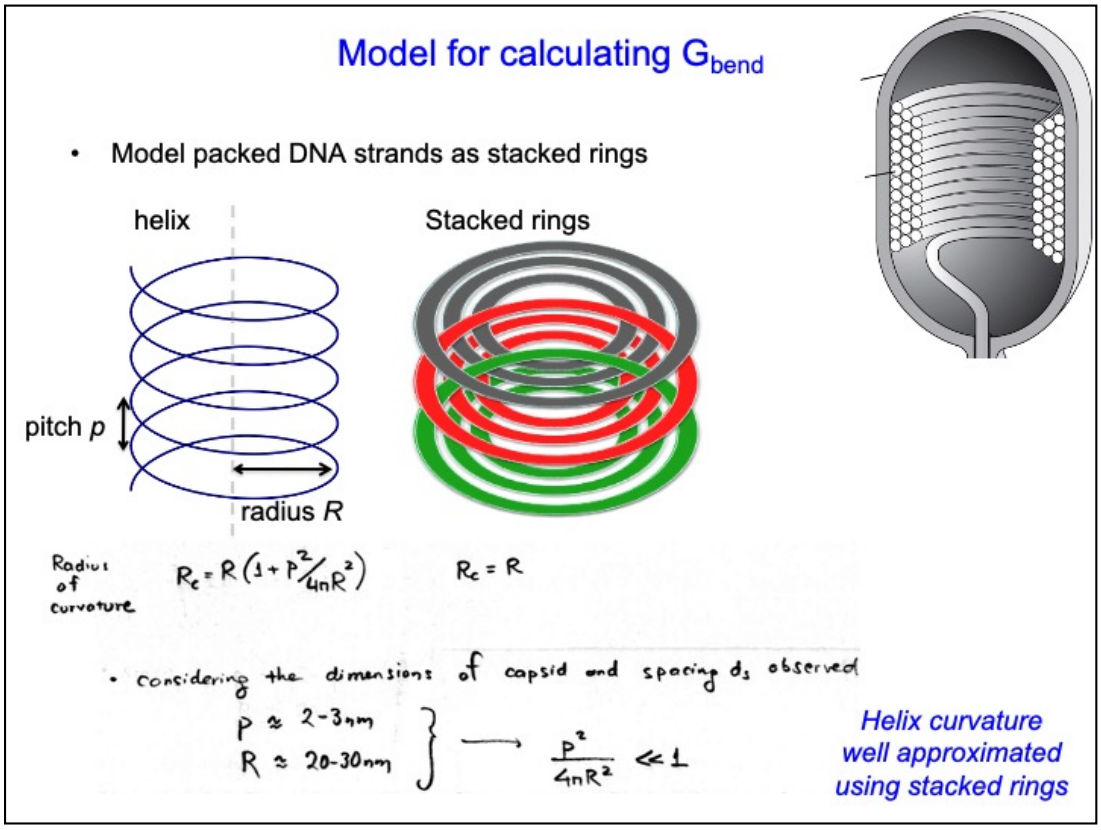
DNA packaging by phages: the phi29 case

- Fact: phages pack DNA into capsids to crystalline density. **How? Why?**
- Case study:
- φ29; 6.6 μm dsDNA in a 42 x 54 nm capsid



- Forces opposing packaging:
  - bending DNA segments shorter that persistence length (G_{bend})
  - electrostatic repulsion between negatively charged DNA (G_{charge})
  - entropic cost of reducing conformational space
- Packing not spontaneous! Done by an ATP-based motor
- Get expression for (G_{bend} + G_{charge}) as function of contour length L, capsid size R_{out}, and spacing d_s

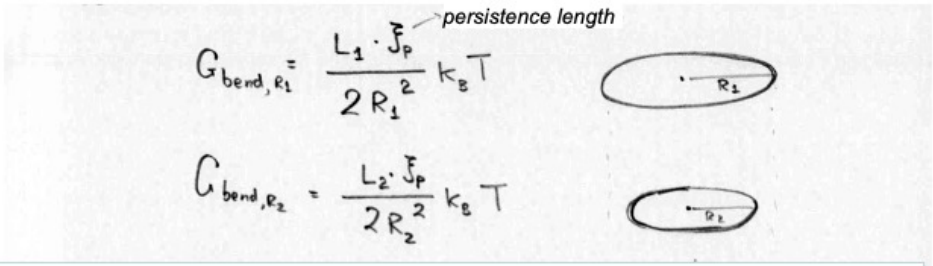
The entropic term is considered to be 10-fold smaller that the other two, so it will be ignored for this analysis.



Most of the curvature due to bending seen in the helix is due to the helix rotation (as expressed by R) and not due to the helix rise (as expessed by *p*).

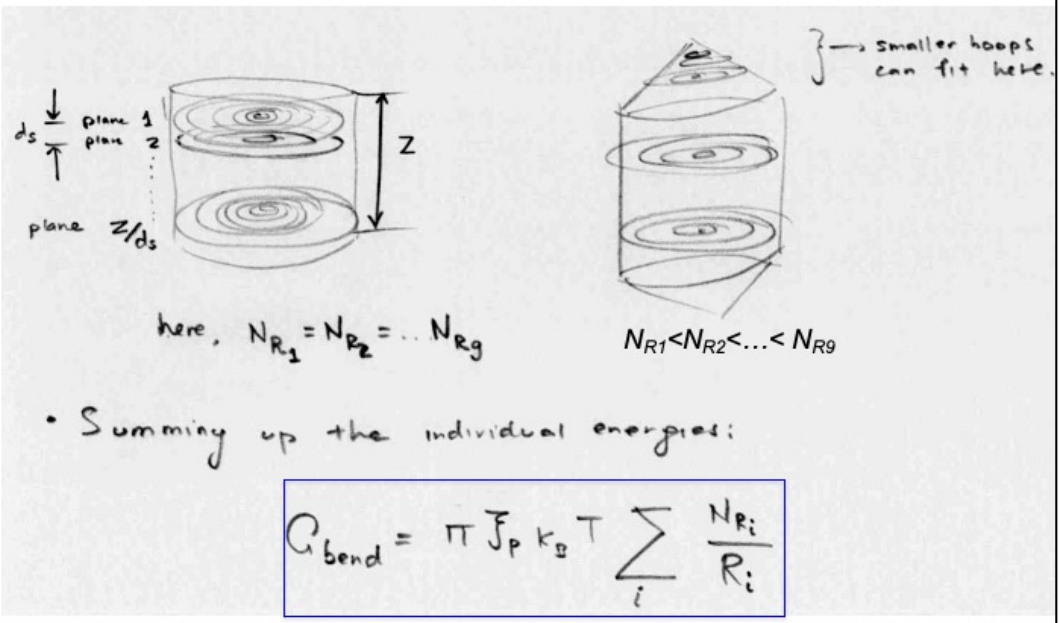
Write G_{bend} for sum of stacked rings

- Consider elastic energy for rings of different radii, R_i:



Write G_{bend} for sum of stacked rings

- Number of rings with radius R_i depends on capsid geometry:



In this example, R_1 corresponds to the outermost ring, and R_9 to the innermost.

G_{bend} for a packed cylinder

- Convert discrete sum into integral (accounting for column spacing *radially*)

$$G_{bend} = \frac{2\pi \bar{J}_p k_B T}{\sqrt{3}\, d_s} \int_{R_{in}}^{R_{out}} \frac{N_{R'}}{R'}\, dR'$$

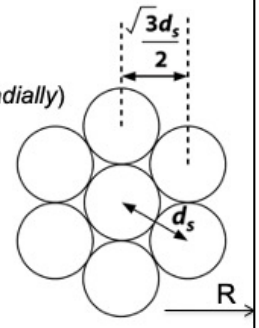
- Similarly for contour length L: $L = \frac{2}{\sqrt{3}\, d_s} \int_{R_{in}}^{R_{out}} 2\pi R' N_{R'}\, dR'$

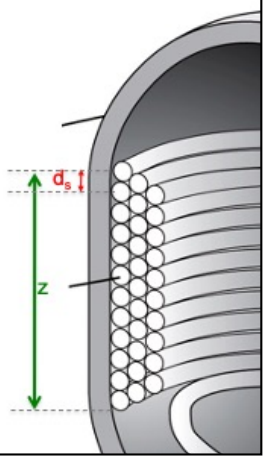
- For a simple cylinder (simplest geometry):

$$N'_R = \frac{z}{d_s}$$

- G_{bend} becomes:

$$G_{bend} = \frac{2\pi \bar{J}_p k_B T z}{\sqrt{3}\, d_s \cdot d_s} \int_{R_{in}}^{R_{out}} \frac{dR'}{R'} = \frac{2\pi \bar{J}_p k_B T z}{\sqrt{3}\, d_s^2} \ln\left(\frac{R_{out}}{R_{in}}\right)$$





Columns of rings are spaced by the projection of d_s along the radial direction.

G_{bend} final expression and related force

- (Problem set) Using the *L* integral, get expression for R_{in} and get G_{bend}…

$$G_{bend} = \frac{\pi \xi_p k_B T z}{\sqrt{3}\, d_s^2} \ln\left(1 - \frac{L\sqrt{3}\, d_s^2}{2\pi z R_{out}^2}\right)$$

… and associated force *f(L):*

$$f(L) = \frac{\xi_p k_B T / 2R_{out}^2}{1 - \sqrt{3}\, d_s^2 L / 2\pi z R_{out}^2}$$

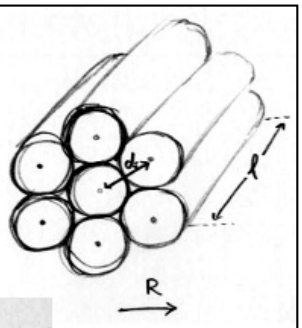
Such equations give bending energy for many capsid geometries

Note that the final expression of the force depends on experimentally accessible properties of the phages and DNA:

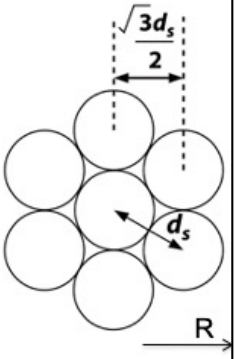
- persistence length ξ_p
- inner diameter of the capsid, which is approximately R_{out} (*not* to be confused with R_{in}, which is the diameter of the innermost DNA ring)
- ring spacing d_s
- cylinder height z, and
- Contour length of phage genome

Electrostatic repulsion term

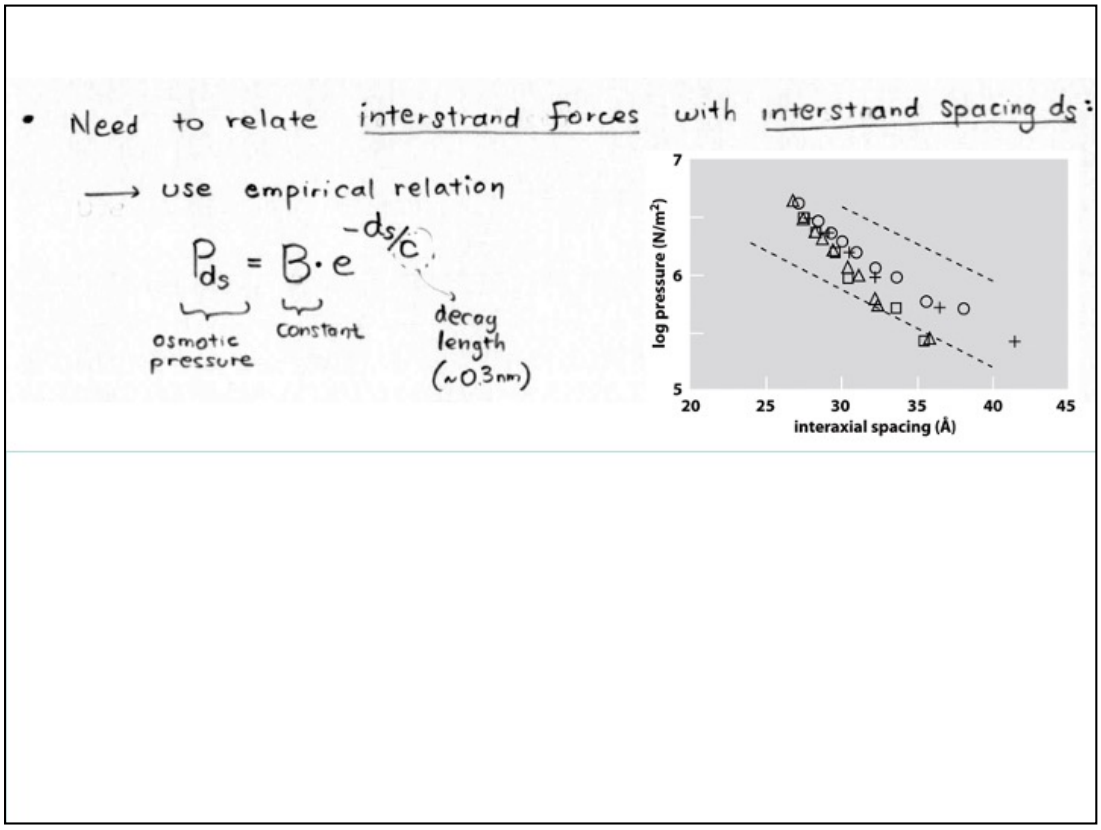
- Model:
- *N* parallel rods of length *l* stacked in a hexagonal packing



- Total volume $V = N \cdot \underbrace{\frac{\sqrt{3}}{2} d_s^2 l}_{\text{volume per strand}}$



The factor 3 arises from the fact that for each rod there are 6 nearest-neighbour interactions but there are counted twice, so we have 3 interactions per pair of strands.



Pressure depends exponentially on d_s.

The measurements of spacing d_s as a function of osmotic pressure were done using X-ray diffraction measurements of DNA strands in solution that were compacted into a regular array due to the use of different concentrations of a long chain polymer (that resulted at different values of osmotic pressure p_{ds}).

B values ~ 10^4 pN/nm^2 for experiments in 5-25 mM $MgCl_2$

G_{charge} as a function of d_s

- (Problem set) Solve differential equation below to get v_{ds} and G_{charge}…

$$\frac{dv_{ds}}{dd_s} = \frac{1}{\sqrt{3}} B\, e^{-d_s/c} \cdot d_s$$

v_{ds}

$$G_{Charge} = \sqrt{3}\, B\, (c^2 + c d_s)\, L\, e^{-d_s/c}$$

energy costs depends on L and d_s (also L dependent)

Force as a function of L packaged: theory predictions

for force associated with (G_{bend} + G_{charge}):

$d_s = (V_{capsid}/L)^{1/2}$

- at long L (short d_s):

  - electrostatic term rises sharply

  - bending term rises sharply (high curvature)

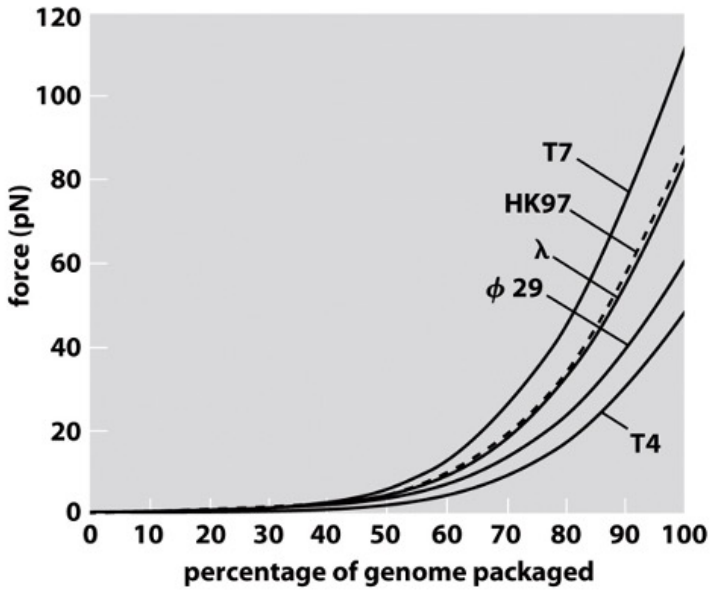


Figure 10.19 Physical Biology of the Cell (© Garland Science 2009)

But can we measure the force during packaging *experimentally*?

Method: tweezers/pipette combo (as with nucleosome)

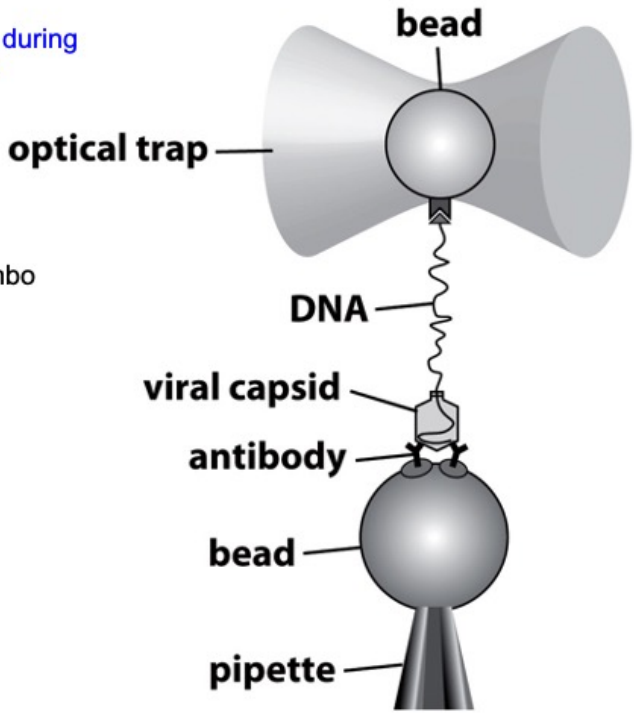
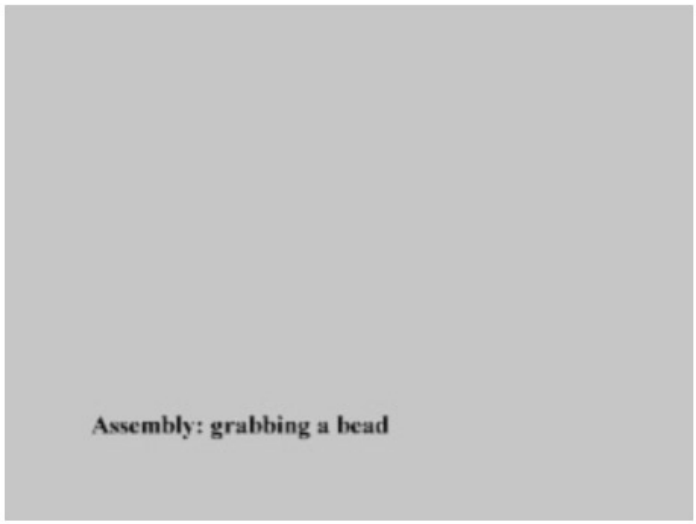
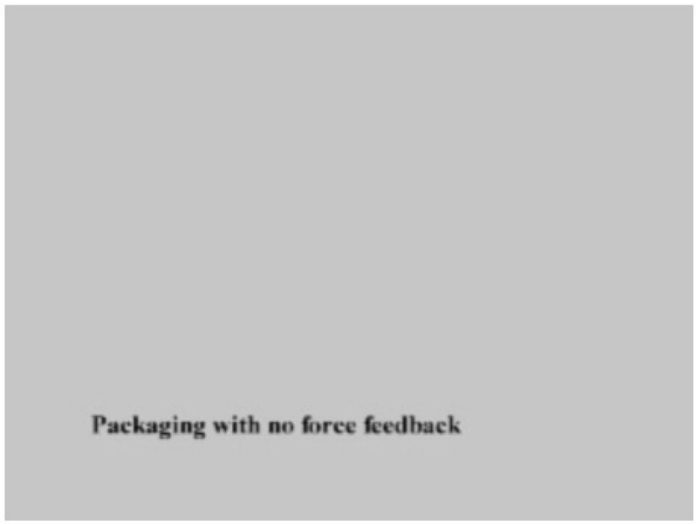


Figure 10.17 Physical Biology of the Cell (© Garland Science 2009)

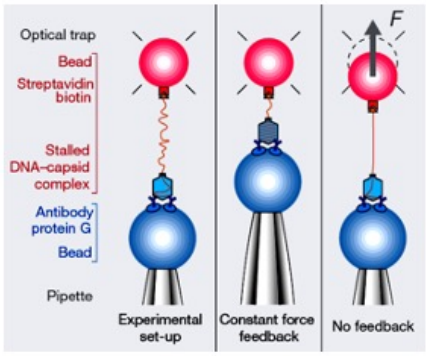

Assembly: grabbing a bead

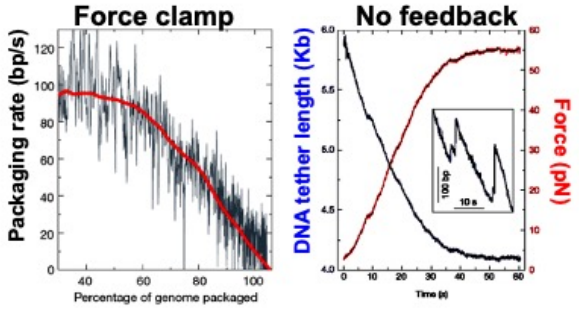

Packaging with no force feedback


Packaging with force feedback

DNA packaging by bacteriophage phi29

- Force-clamp:
  - ~5 min to package 19,000 bp
  - rate drops after 50% DNA packed (opposing F)
- No feedback:
  - Stall force of ~57 pN (strong motor!) *(stall due to bead out of trap…F=kx)*
  - Step per ATP: 2 bp
  - 57 pN * 0.68 nm / 4.1 = 10 kT per ATP (**30% conversion efficiency**)





Smith, D.E. et al. (2001) Nature 413, 748-751

Procedure:

1. stalled, partly prepackaged complexes are attached to small bead through end of unpackaged DNA; this bead is captured by trap
2. Bead with capsid-specific antibody binds capsid; tether is stretched
3. Add ATP and:

a. Run force-clamp experiment (packaging rates)

b. Run no feedback experiment (stall force)

The 2-bp step is an estimate from bulk studies.

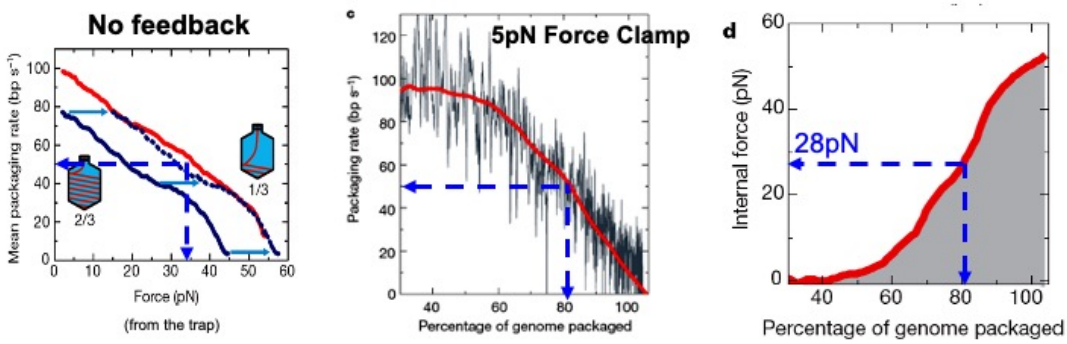
Energy for ATP hydrolysis in the buffer used: ~30kT

Some pauses and slips but otherwise a very processive motor.

For comparison, stall force for kinesin: 6-8 pN, RNAP ~25 pN.

Estimating the internal force and pressure

From F-v curves → relate F for v_{nofeed} to % genome packaged)



- Internal force reaches ~50 pN when all DNA is packaged
- corresponds to a huge internal pressure (6 MPa)

Internal pressure: driving force for DNA injection into host?

Smith, D.E. et al. (2001) Nature 413, 748-751

Note that packaging rates are immediately affected by an opposing force, implying that the rate-limiting step is mechanical in nature.

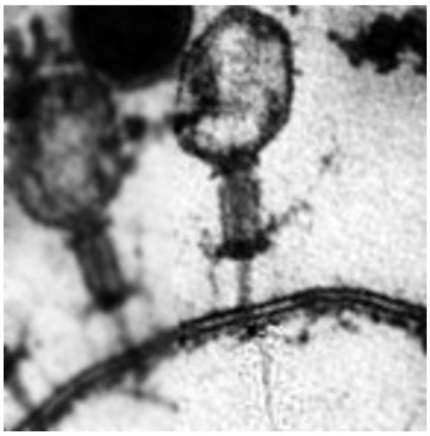
To get internal force:

1. From **F-v experiments** (from trace with no feedback by looking at the rate of tether-length change), one can **relate a certain packaging rate to an opposing force** (coming from the trap).
2. Then, assume that the internal force matches the opposing force that produces the same packaging rate (after subtracting the 5 pN tension applied in the force-clamp experiment).
3. Plot internal force vs. %genome packaged (from force-clamp packaging rate vs. % genome packaged curve)

The phage capsid withstands huge pressures: considering its thickness (~1.6 nm), it has the tensile strength of an aluminium alloy.

The phage phi29 apparently has learned some clever tricks to survive: they use ~20,000 kT (x 3.3=66,000kT, since efficiency in using ATP is 30%) **from the host** during packaging, and use this stored elastic and electrostatic energy for DNA injection to the host!!!

No needle required…

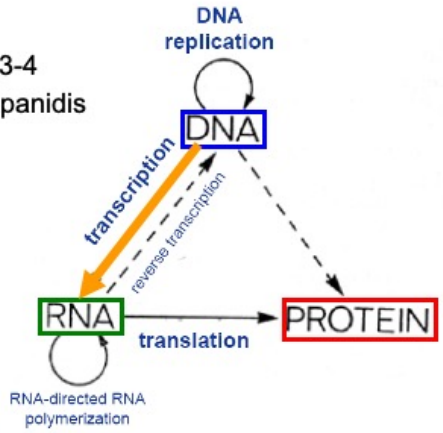


Oxford Physics

Gene transcription

Lectures 3-4
Achilles N Kapanidis

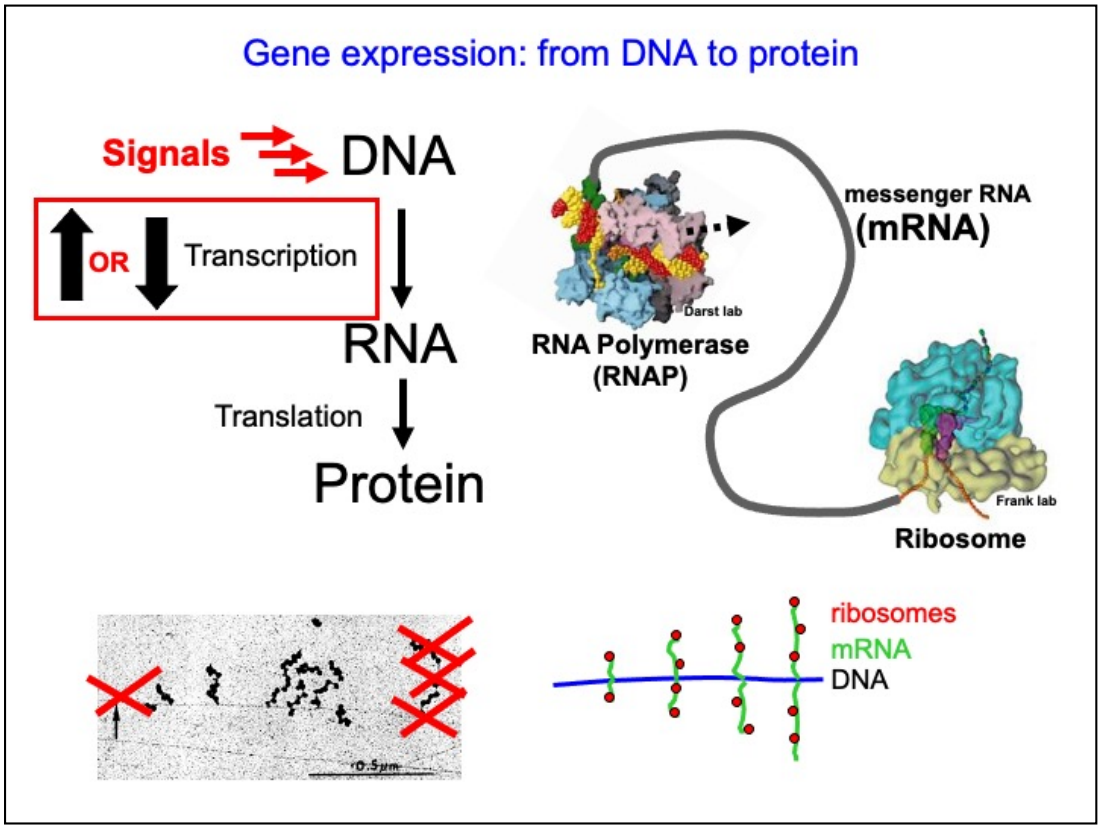
- Alberts, Chapter 6: pg 299-315
- Handout references



Biological Physics Major Option, Jan 2022

Transcription is the first and most important step in gene expression, the path that leads cells from genes to proteins.

Understanding transcription and its control is crucial for understanding how genetic information is converted to the right amount of the right proteins at the right place and at the right time. More on this in Lecture 6 (control of gene expression).

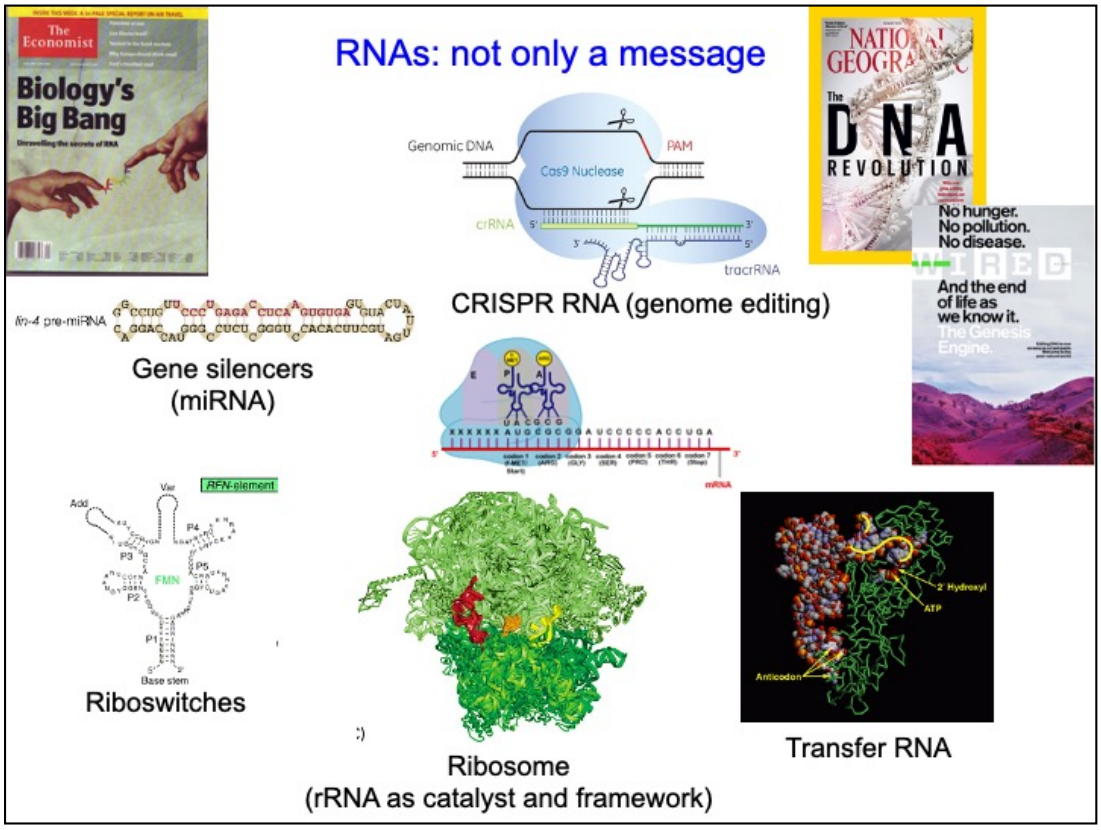


Transcription copies DNA to a messenger molecule, RNA, and it is orchestrated by the multifunctional machine RNAP. Translation uses messenger RNA (mRNA) as a template for making proteins in the ribosomes.

In bacteria, transcription and translation of messenger RNA (mRNA) can occur at the same time, as shown for the electron micrographs known as "Miller spreads" (see bottom left).

Gene expression is much more complex in eukaryotes: transcription and translation are compartmentalised and there are additional processing steps.

Signals that arrive to DNA (from either the cell exterior or interior) increase or decrease the levels of transcription, and thus regulate the levels of proteins. E.g., no transcription, no message, no protein.



Messenger RNA: mRNA

Most of RNA in cells is **ribosomal RNA** (rRNA); only up to 10-15 copies of each mRNA exist per cell. mRNA is a transient species and is continuously made and destroyed (half-life of RNA molecules is in minutes).

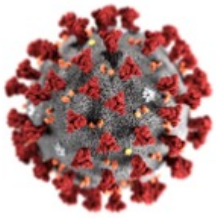
Riboswitches are untranslated parts of a messenger RNA molecule that change their conformation upon binding ("sensing") small biomolecules. Riboswitches can modulate gene expression by increasing the efficiency of transcription termination OR decreasing translation efficiency.

microRNAs (miRNAs) are small double-stranded RNA molecules (less than 30 nt in length) that can repress gene expression of specific genes by blocking the translation of a gene or by destroying the corresponding mRNA. miRNAs also serve as a defence against viruses. They result from processing of longer RNAs. The study of miRNAs belongs to field of gene silencing (aka RNA interference or RNAi) which was featured in the 2006 Nobel in Medicine.

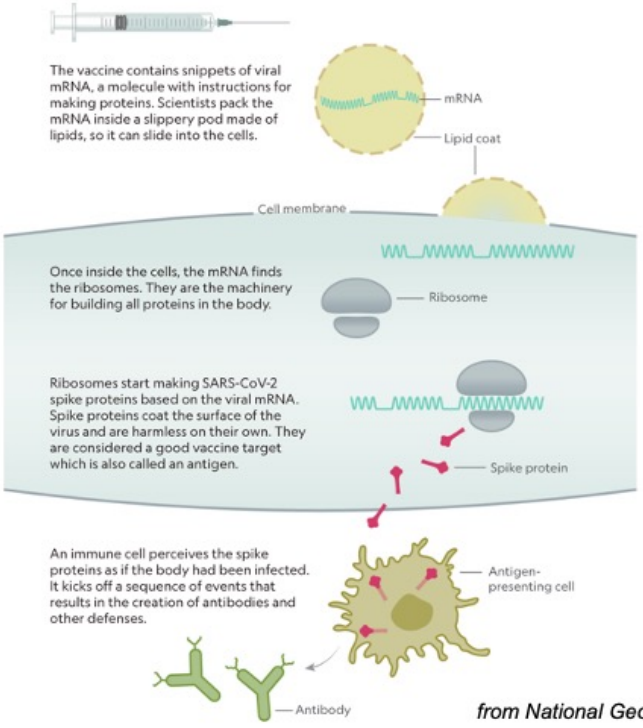
And a new entry: **guide RNAs** for CRISPR gene editing (CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats), a bacterial defence system against foreign DNAs that has been turned into a versatile

tool for genome editing (which can be used for gene therapy) and labelling. CRISPR featured in the 2020 Nobel in Chemistry.

mRNA vaccines: an exit strategy for the covid pandemic



SARS-CoV-2



from National Geographic

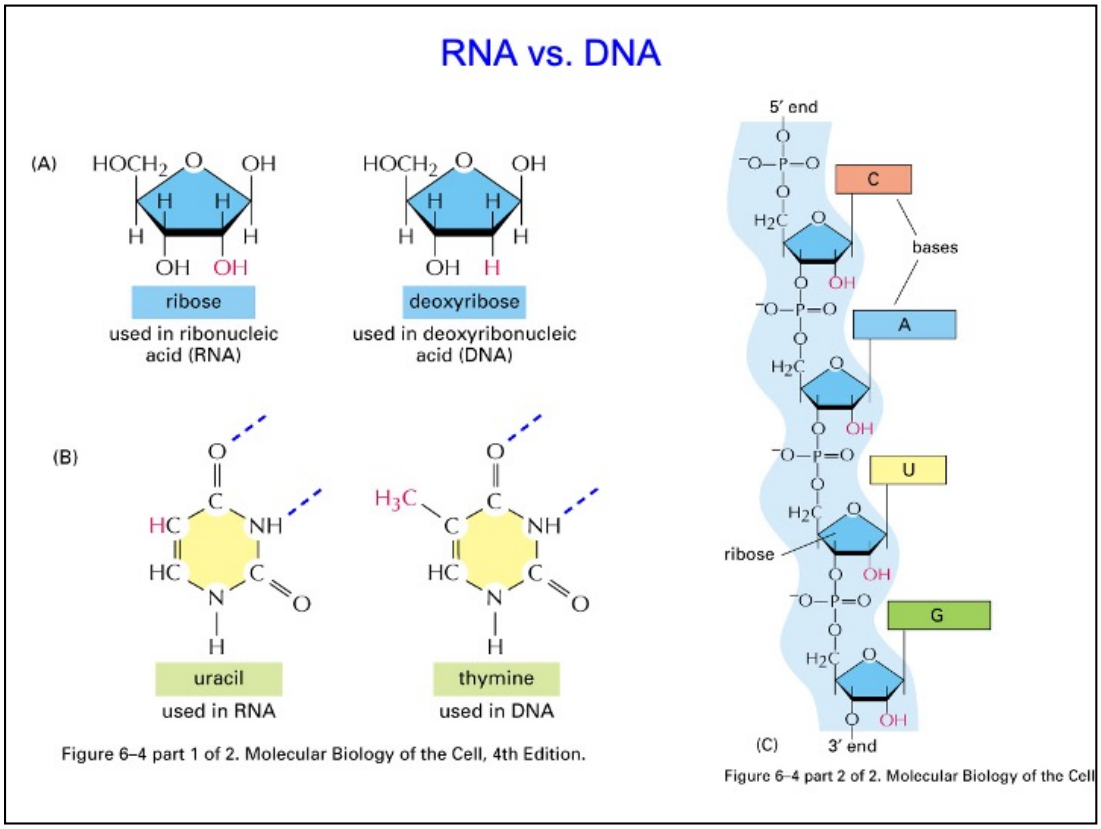


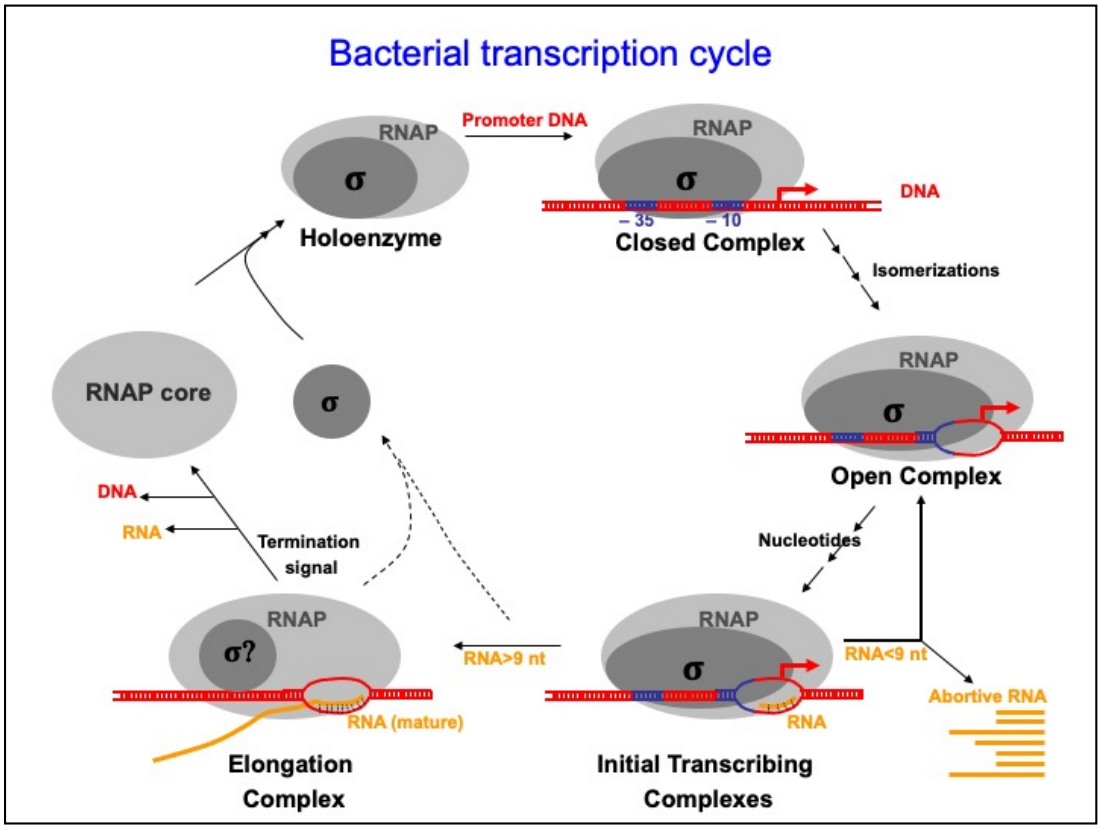
Figure 6–4 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

Figure 6–4 part 2 of 2. Molecular Biology of the Cell

Two main differences:

--RNA has **uracil** (abbreviation: U) as the 4th base (while DNA has thymine, T)

-- RNA has **ribose** as a sugar group (while DNA has deoxyribose). This means that RNA has a hydroxyl group (-OH) at position 2'. This seemingly small difference makes a RNA a much more versatile molecule than DNA, both in terms of structure as well as in function. These capabilities of RNA as linked with the RNA-world hypothesis for evolution, a topic covered by Prof Turberfield.



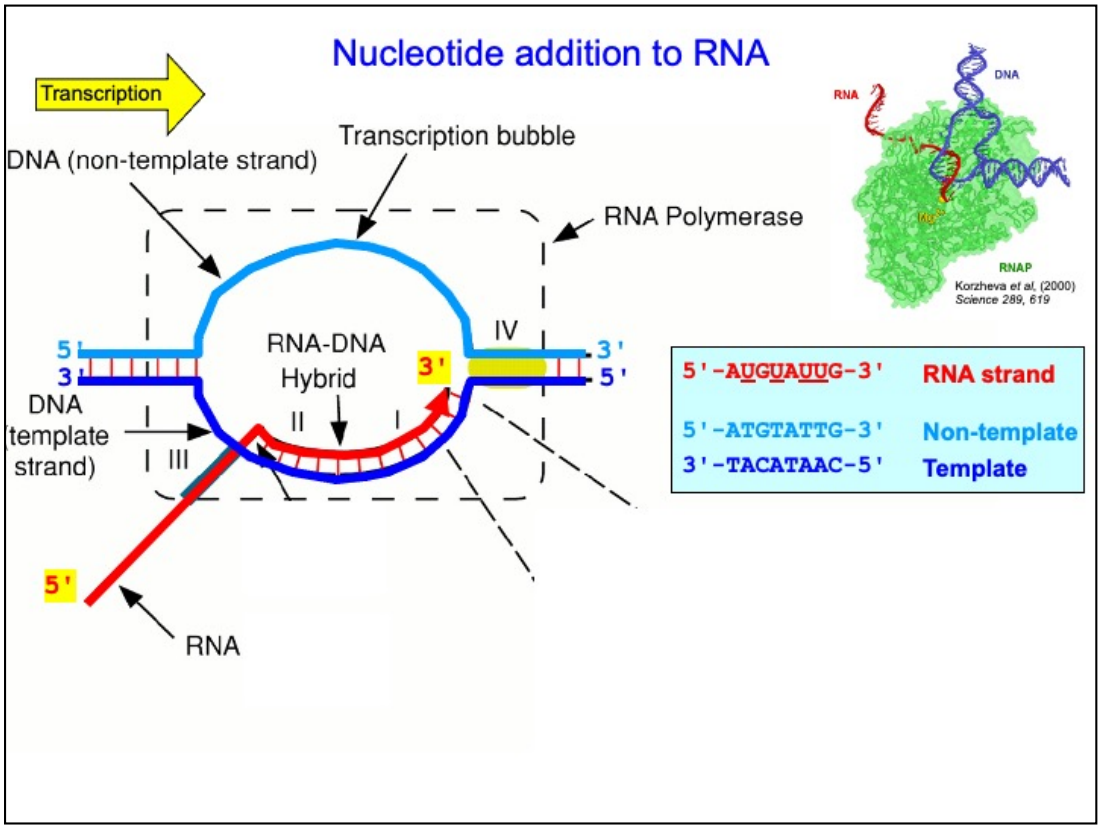
The formation of the RNAP **holoenzyme** (for core RNAP + sigma70) has favourable free energy of ~22 kT.

The ssDNA surrounding the transcription site is known as the "**transcription bubble**". It comprises ~14 nt of ssDNA; remarkably, **no ATP is needed to form the bubble** in most cases in bacteria (however, there are specific cases where ATP is necessary to open the double helix). The formation of the open complex has a favourable free energy of 12-15 kT.

There is structural similarity between bacterial and eukaryotic RNAPs; this allows extension of conclusions drawn for bacterial RNAP to complex eukaryotic systems.

RNAP has several mobile modules that assume different conformations during the transcription cycle. For example, the "jaws" of RNAP close further when the system moves to elongation, increasing the stability of the transcription complex.

Transcription elongation: **20-50 nt/s** (in cells)



As with all nucleic acid polymerases, RNAP polymerises RNA in the 5'→3' direction.

There is a large number of crystal structures and cryoEM structures for bacterial RNAPs and their complexes, as well as for eukaryotic RNAPs and their complexes.

Promoter DNA: pointing where to start

AGGATGTG**TTGACA**ATGTGAGTGCATGTATC**TATAAT**TTACTT**A**ATATGTA

-35 element -10 element

- Promoter DNA: sequences that set where transcription begins
- Recognised by σ (sigma) proteins; most common: σ^{70}
- Consensus by comparing sequences
- Critical steps:
  - search for promoter DNA
  - open complex formation

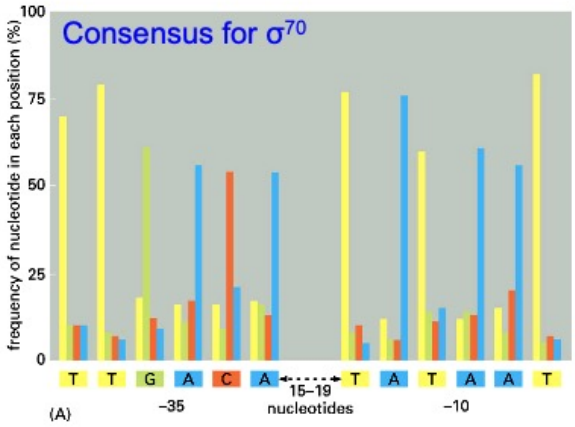


Figure 6–12 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

Sigma proteins (aka "sigma factors") are initiation-specific proteins in bacteria; *E.coli* has 7 sigma factors, each specialising in the transcription of different sets of genes. For example, there is a special sigma (sigma32, the "heat-shock" sigma factor) that promotes expression of proteins involved with counteracting the effects of exposure to high temperatures.

In eukaryotes, the initiation process is very complex, requiring assembly of a large complex with >20 proteins. In this case, ATP is required for opening the double helix.

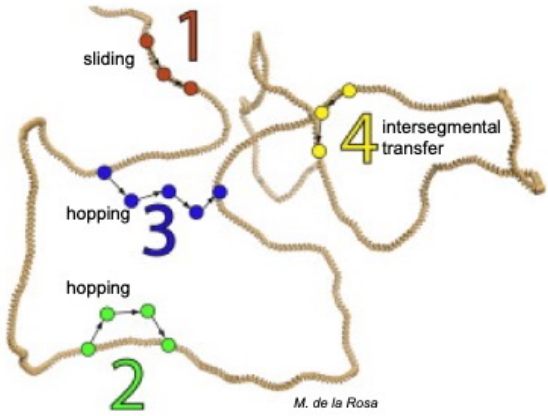
How DNA-binding proteins reach their targets?

- many types of motion relative to DNA

- high selectivity
  - can find single sites of 20 bp within the entire genome

- high speed
  - search rates up to 100x *faster* than diffusion-limited association rate k_a
  - mix of 1D and 3D diffusion

…calculate rate k_a (3D diffusion)

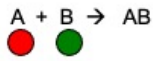


More on the 1D/3D search process in Lecture 6.

Inter-segmental transfer (mode 4 of searching) is available only to proteins that have two separate surfaces for binding DNA. The transfer involves an intermediate where the protein is bound to two separate strands of DNA. Lac repressor (see lecture 6) is the best known example for such a protein.

Diffusion-limited association rate constant

Consider **one** molecule of B colliding with **many** molecules of A:



Reaction determined by rate of molecular collisions:

-- reaction occurs when separation of molecular centres becomes $r_0 = r_A + r_B$

-- reaction determined by diffusion of A molecules to collide with B

Flux of A in 3D (Fick's law): $\mathbf{J}_A = -D_A \nabla C_A$

Flux per unit area (molecules m^{-2} s^{-1})

Assume steady-state conditions: $\partial C_A / \partial t = 0$

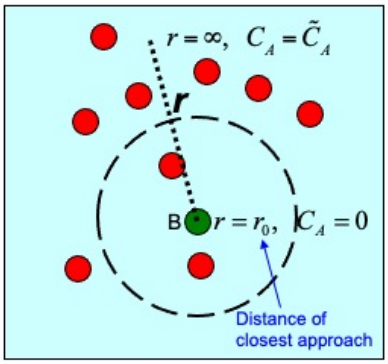
For spherical symmetry: $J_A = -D_A \left(\frac{\partial C_A}{\partial r} \right)$

For a spherical surface at distance *r* from B, total inward flux is:

$$I_A = -area \bullet J_A = 4\pi r^2 D_A \left(\frac{\partial C_A}{\partial r} \right)$$

Flux (molecules s^{-1})

Under steady state conditions:
-- total flux identical for spheres of all radii r
-- total flux is identical to *molecular collisions per sec*



Initially, assume B is stationary.

Units:

C: molar concentration (moles per liter)

D: diffusion coefficients

J: number of molecules per unit area per time

I: number of molecules per time

Standard units for k_a : M^{-1} s^{-1}

N_a: Avogadro's number

Diffusion-limited association rate constant

After integration (problem set):

$$I_A = 4\pi r_o D_A \tilde{C}_A$$

collisions per sec for one B molecule

Accounting for diffusion of B and concentration of B ($\tilde{C}_B$) :

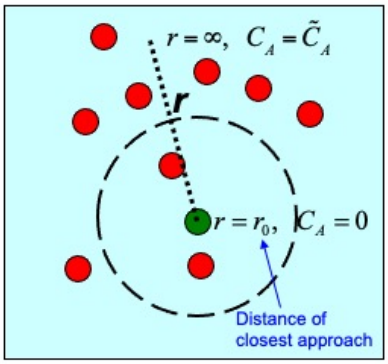
$$I_A \tilde{C}_B = 4\pi r_0 \left(D_A + D_B\right) \tilde{C}_A \tilde{C}_B$$

collisions per sec for many B molecules

(see Cantor and Schimmel textbook, Chapter 16, pg 920)

Convert to rate constant k_a with the appropriate units (problem set):

$$k_a = 4\pi r_0 \left(D_A + D_B\right) N_a 1000$$



C: molar concentration (moles per liter)

D: diffusion coefficients

J: number of molecules per unit area per time

I: number of molecules per time

Standard units for k_a : $M^{-1}\ s^{-1}$

N_a: Avogadro's number

The equation for the diffusion-limited association rate constant is known as Smoluchowski's equation.

Diffusion-limited association rate constants are in the $10^8\ M^{-1}\ s^{-1}$ region, e.g. considering a DNA fragment and a DNA-binding protein.

Kinetics of open complex formation

Simple kinetic scheme for open complex formation:

$$R + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} RP_c \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} RP_o$$

Observed: 1st order kinetics for RP_o formation:

$$[RP_o] = [P_{total}]\left(1 - e^{-k_{obs}t}\right)$$

Assumptions: $[R] >> [P_{total}]$ (true in cells), and steady-state conditions for $[RP_c]$

$$k_{obs} = \frac{k_1[R](k_2 + k_{-2}) + k_{-1}k_{-2}}{k_1[R] + k_{-1} + k_2}$$ (problem set)

Since $k_{-2} << k_2$:

$$k_{obs} = \frac{k_1 k_2 [R]}{k_1[R] + k_{-1} + k_2}$$

$$\tau_{obs} = \frac{1}{k_{obs}} = \frac{1}{k_2} + \frac{k_{-1} + k_2}{k_1 k_2}\frac{1}{[R]}$$

intercept slope

McClure (1980) *PNAS* 77, 5634-8

~3000 RNAP molecules per cell, with approximately half of which are involved in active transcription. These numbers depend on the growth rate and physiology of the cells.

k_{obs}: a first-order rate constant.

RP_c: RNAP-DNA closed complex; RP_o: RNAP-DNA open complex

$k_{-2} << k_2$ applies since the stability of RP_o is very high, with lifetime in tens of minutes to hours.

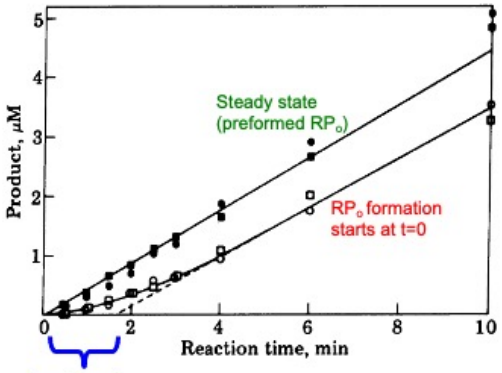
The formation of RP_c occurs fast compared to the isomerisation to RP_o, therefore it can be assumed that the first step reaches equilibrium before the isomerisation step.

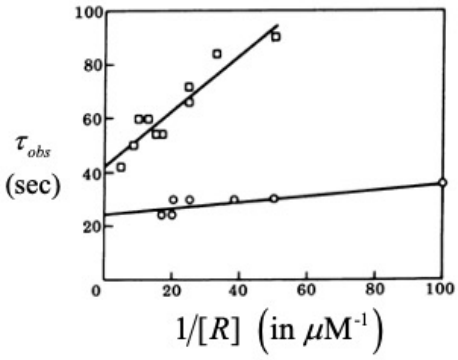
Kinetics of open complex formation

RP_o formation monitored by measuring the first RNA transcript, e.g.:

$$AA + U \rightarrow AAU$$

Assumption: [RP_o] proportional to [AAU]/time





$$\tau_{obs} = \frac{1}{k_{obs}} = \frac{1}{k_2} + \frac{k_{-1} + k_2}{k_1 k_2} \frac{1}{[R]}$$

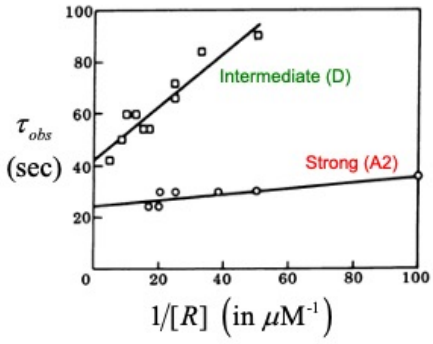
intercept slope

McClure (1980) *PNAS* 77, 5634-8

The RNA product AAU is quantified using gel electrophoresis and autoradiography.

This treatment assumes that the synthesis rate of the first transcript is very fast (completed within a second) compared to the formation of the open complex (completed in tens of seconds or more).

Evaluating promoter strength



$\frac{1}{k_2}$ → Isomerisation constant k_2 for RP_o formation

For rapid equilibrium limit:

$\frac{k_{-1}+k_2}{k_1 k_2} \rightarrow \frac{k_{-1}}{k_1 k_2}$ → Equilibrium constant $K_B = k_1/k_{-1}$ for RP_c formation

- Promoter clearance time (t_{cl}): time from RP_o to empty promoter (through RNA synthesis)
- Minimum t_{cl} ~ 5 sec
- k_2 controls reaction and its dependence on [R]

Table 3. *In vitro* comparison between a strong and a weak T7 promoter

	A2	D
Kinetic constants:		
$k_{on}\left(\frac{k_1 k_2}{k_{-1}}\right)$, M^{-1} sec^{-1}	>9 × 10^6	9.6 × 10^5
k_{off} (k_{-2}), sec^{-1}	1.7 × 10^{-4}	3.3 × 10^{-5}
k_2, sec^{-1}	4 × 10^{-2}	2.4 × 10^{-2}
Apparent equilibrium constants:		
$K_B = K_I$, M^{-1}	>2 × 10^8	4.1 × 10^7
K_{II}	2.4 × 10^2	7.2 × 10^2
K_o, M^{-1}	>5 × 10^{10}	3.0 × 10^{10}

25 sec 40 sec

McClure (1980) *PNAS* 77, 5634-8

Minimum promoter clearance time: need to move by ~80 bp to allow landing of 2nd RNAP. Best case: assume no rate limiting steps after initiation.

-- To synthesize 80 nt with a speed of 20nt/sec: ~ 4 sec.

-- To account that 1st step is slower (~1/20 of normal rate) ~ 1 sec

-- Minimum clearance time: ~5 sec.

-- **Strong promoters**: fast isomerisation step: $1/k_2$ < 10sec, approaching the minimum clearance time.

-- **intermediate-strength promoters**: isomerisation step: 100sec > $1/k_2$ > 10sec, depends on [R]

-- **Weak promoters**: very slow isomerisation step: $1/k_2$ > 100sec. Weak at all [R]

Note: this approach considers only rate-limiting steps up to RP_o formation; rate-limiting steps can exist after this.

nature
International journal of science

LETTER

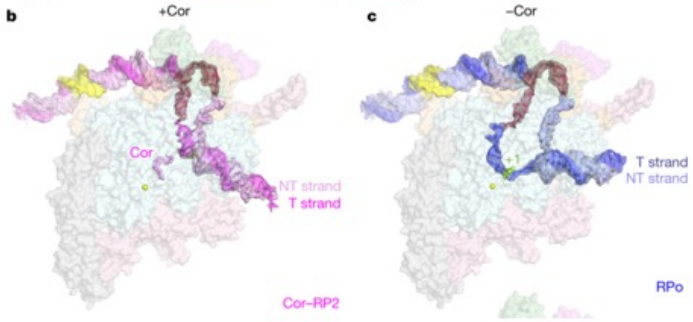
Jan 9th, 2019

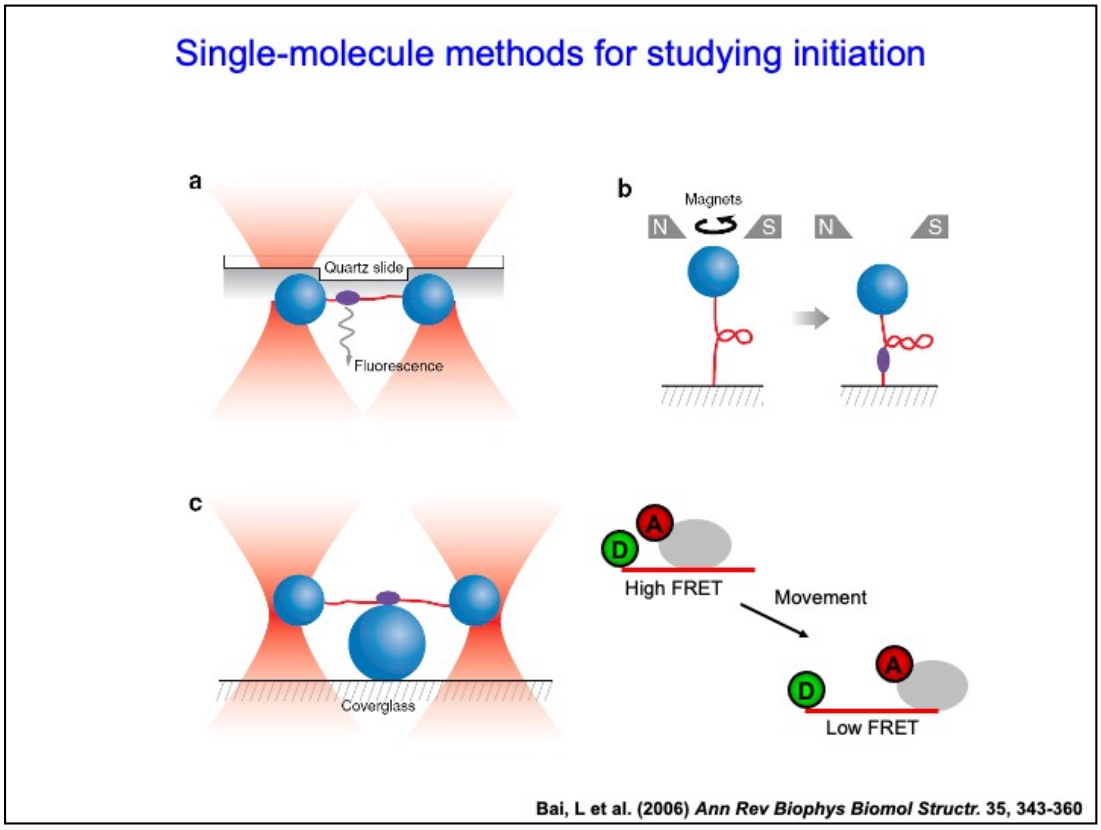
https://doi.org/10.1038/s41586-018-0840-5

Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding

Hande Boyaci[1], James Chen[1], Rolf Jansen[2], Seth A. Darst[1] & Elizabeth A. Campbell[1]*

antibiotic corallopyronin blocks full DNA opening





a: DNA was extended between two beads held by 2 optical traps; the movement of a fluorescent RNAP was monitored by imaging.

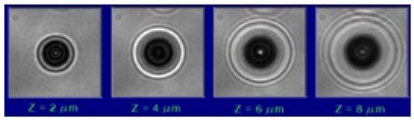
b: see following slides

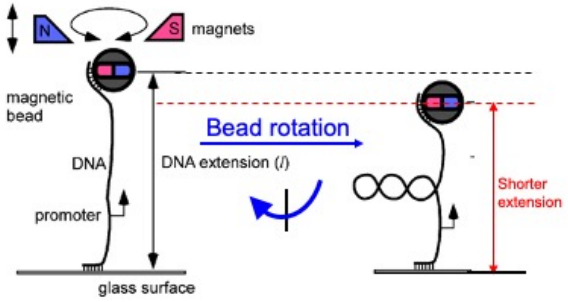
c: DNA was extended between two small beads (held by 2 optical traps); one of the beads was oscillated by moving the trap; since the other bead is coupled to the first through the DNA, it oscillated at the same frequency. Then, the DNA was brought close to a larger bead decorated sparsely with RNAP molecules; when an RNAP molecule bound DNA, it uncoupled the oscillation of the two small beads. Transcription elongation was followed using the movement of the bead that held the DNA segment downstream of the transcription start site.

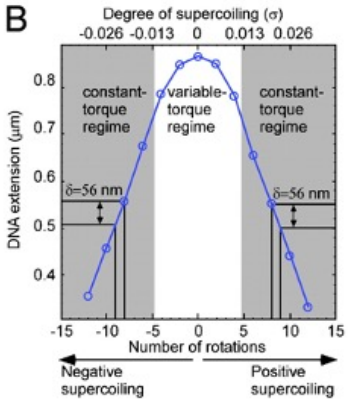
d: a FRET pair is introduced at two specific points in the complex. Movement within the complex changes the separation between the probes, altering FRET efficiency.

Magnetic tweezers for DNA nanomanipulation

- Manipulate dsDNA multiply-linked to magnetic beads
- DNA extension (l) monitored by bead position
- Bead rotation => supercoils (+ or -)







Revyakin et al (2004) *PNAS* 101, 4776

Convention for representing supercoils: negative supercoils are shown on the left of the extended DNA. Turning the bead anticlockwise introduces positive supercoils.

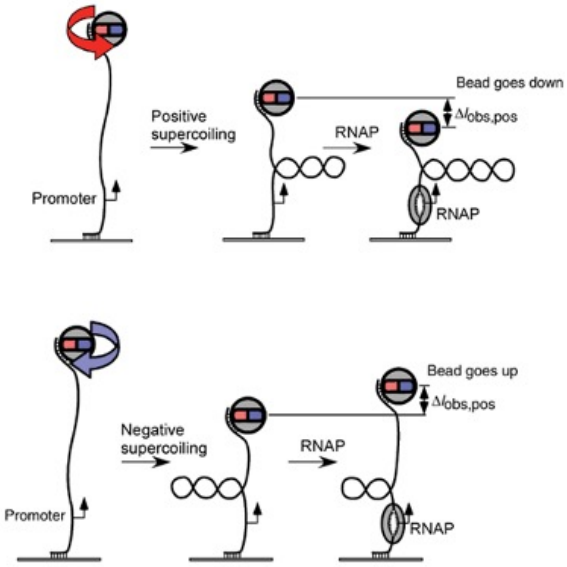
One turn of the magnets in the **constant-torque regime** causes a **~56 nm change in the DNA extension** (e.g., see the difference between 9 and 10 negative supercoils).

The bead position can be identified by characteristic patterns it produces using a simple light microscope with a video camera (see patterns at top right).

The pioneers of this technique (initially developed for studying polymer physics) are Vincent Croquette and David Bensimon (Ecole Normale Superieure, Paris).

Magnetic tweezers for studying processes on DNA

- Proteins bind and change DNA extension
- For torsionally-constrained DNA, linking number Lk is constant:
  Lk = Twist + Writhe
- Unwinding by RNAP (-1 Twist) compensated by positive supercoiling (+1 writhe)



HUGE signal amplification:
+1 turn of unwinding → 56 nm of change in DNA extension

Revyakin et al (2004) *PNAS* 101, 4776

Protein-mediated changes in DNA extension can arise from changes in DNA supercoiling, DNA bending, or DNA wrapping.

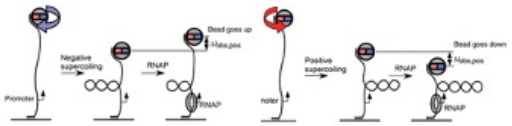
Writhe: number of supercoils

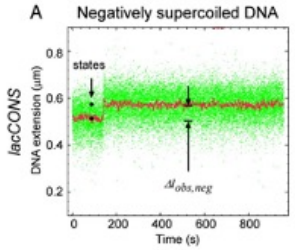
One additional positive supercoil adds to the existing positive supercoils, moving the bead down (top panel).

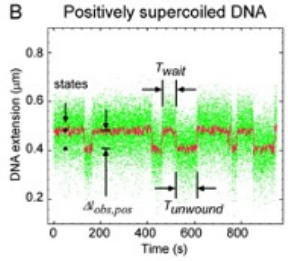
One additional positive supercoil cancels one of the existing negative supercoils, moving the bead up (bottom panel).

Study transcription using magnetic tweezers

- DNA unwinding during RP_o formation
- 4 levels of information:
  - Number of states
  - Changes in DNA extension (*l*)
    - DNA unwinding
    - DNA compaction (bending, wrapping)
  - Waiting time in initial state (T_{wait})
    - Equivalent to τ_{obs}
    - Depends on [R]
    - **Yields K_B and k_2**
  - Waiting time in final state ($T_{unwound}$)
    - Independent of [R]
    - **Yields k_{-2}**







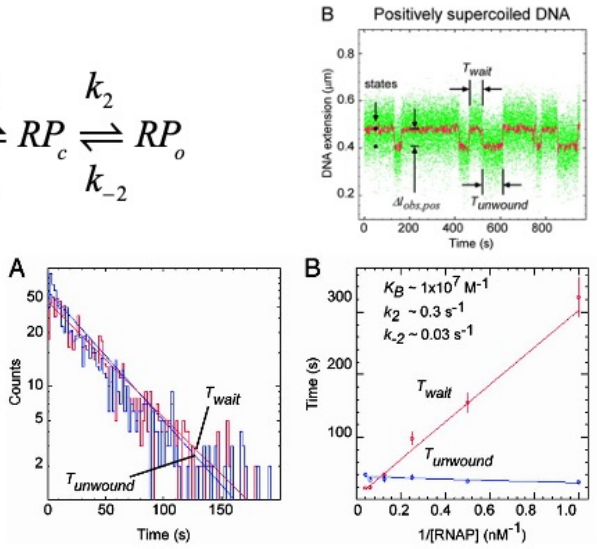
$$R + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} RP_c \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} RP_o$$

Revyakin et al (2004) *PNAS* 101, 477

Kinetics of open complex formation

Plotting T_{wait} vs 1/[R] provides K_B, k_2, k_{-2} to characterise promoter

$$R + P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} RP_c \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} RP_o$$

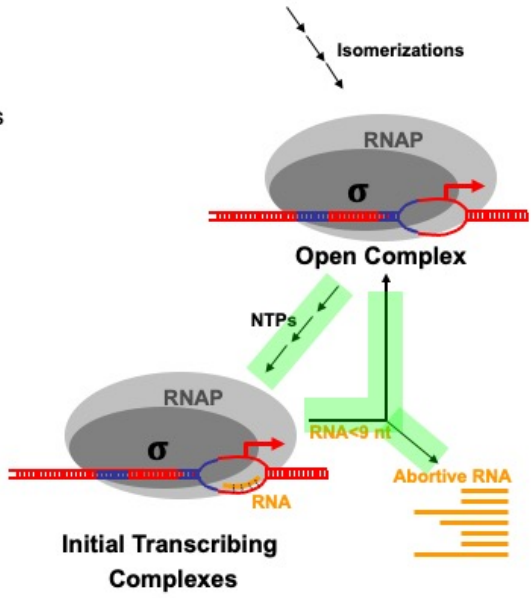


Revyakin et al (2004) *PNAS* 101, 4776

Plotting dwell-time probability histograms can be fit with single-exponential distributions and yield mean values for $T_{unwound}$ and T_{wait}.

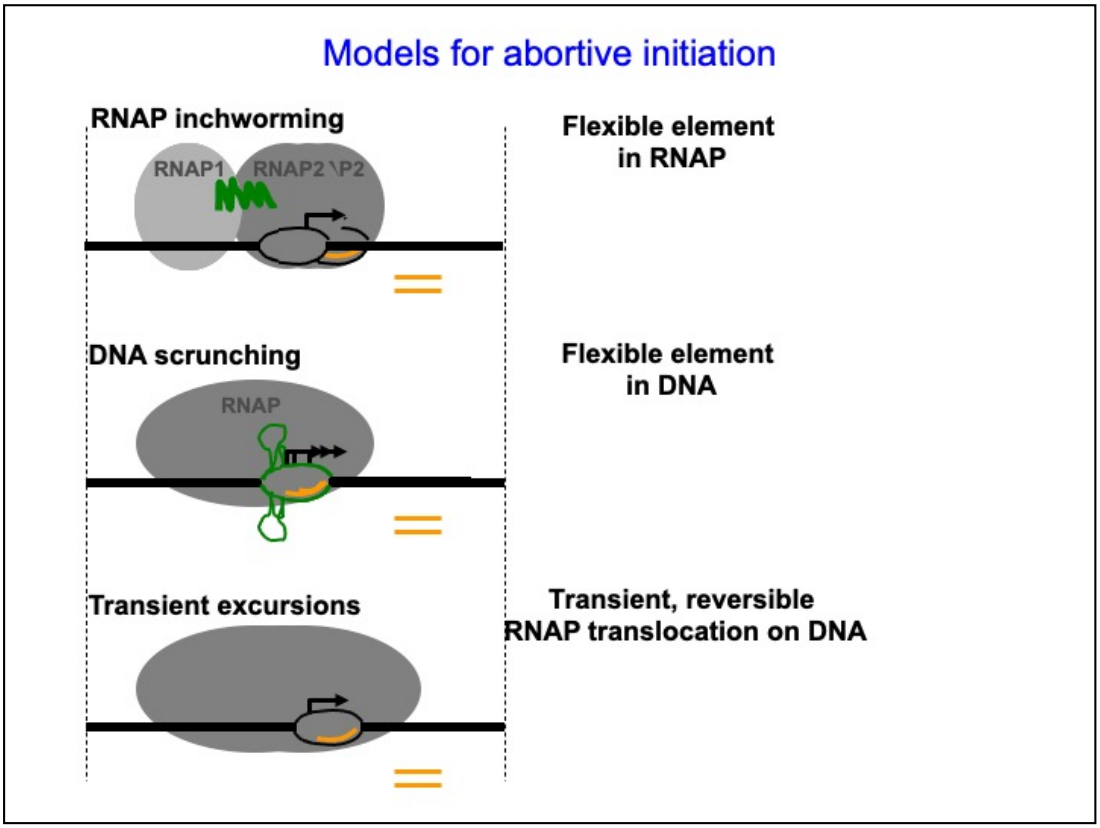
Mechanism of abortive initiation

- First described in 1980
- Important for:
  - regulation (can be rate-limiting step)
  - promoter escape
  - antimicrobial therapy using antibiotics (e.g., rifampicin)



Abortive initiation is a phase of the initial transcription by RNA polymerase.

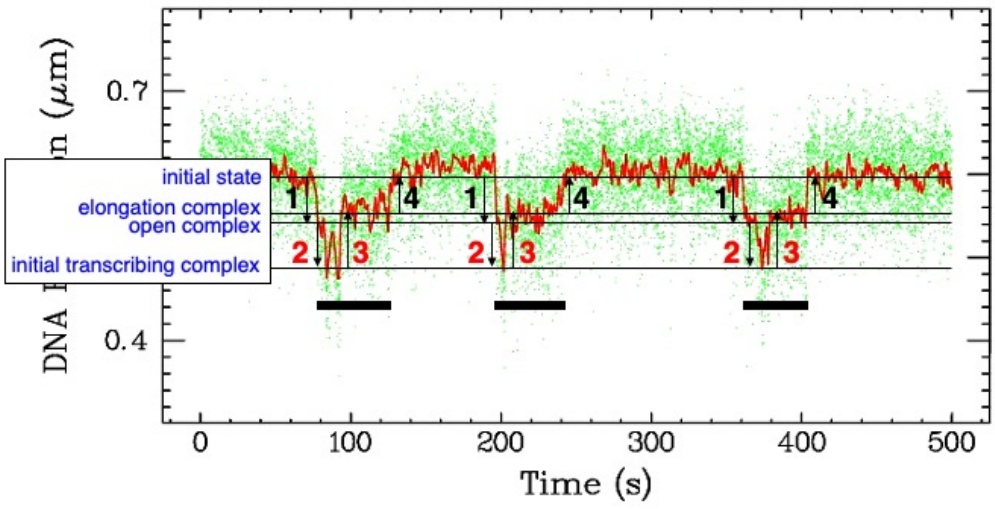
Abortive initiation is blocked by rifampicin, one of the most effective antibiotics vs. tuberculosis, leprosy, staphylococcal infections and other infections.



The distinguishing feature for finding the operating mechanism through magnetic tweezers was the expansion of the bubble during the abortive initiation reaction; only the scrunching model predicted an enlarged bubble.

Additional DNA unwinding in initial transcription

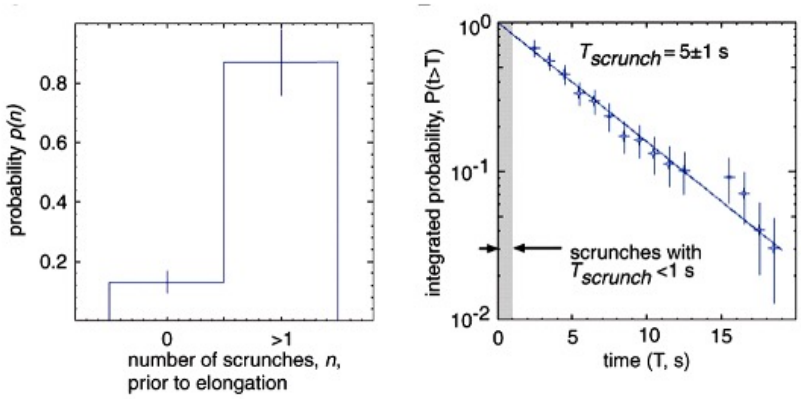
- Magnetic tweezers work; all nucleotides added (productive transcription)
- system cycles between distinct 4 states



Revyakin et al (2006) *Science* 314, 1139-1143

Separate experiments that use conditions that "lock" the system in the abortive initiation mode also show additional unwinding during initial transcription.

Promoter escape also involves DNA scrunching

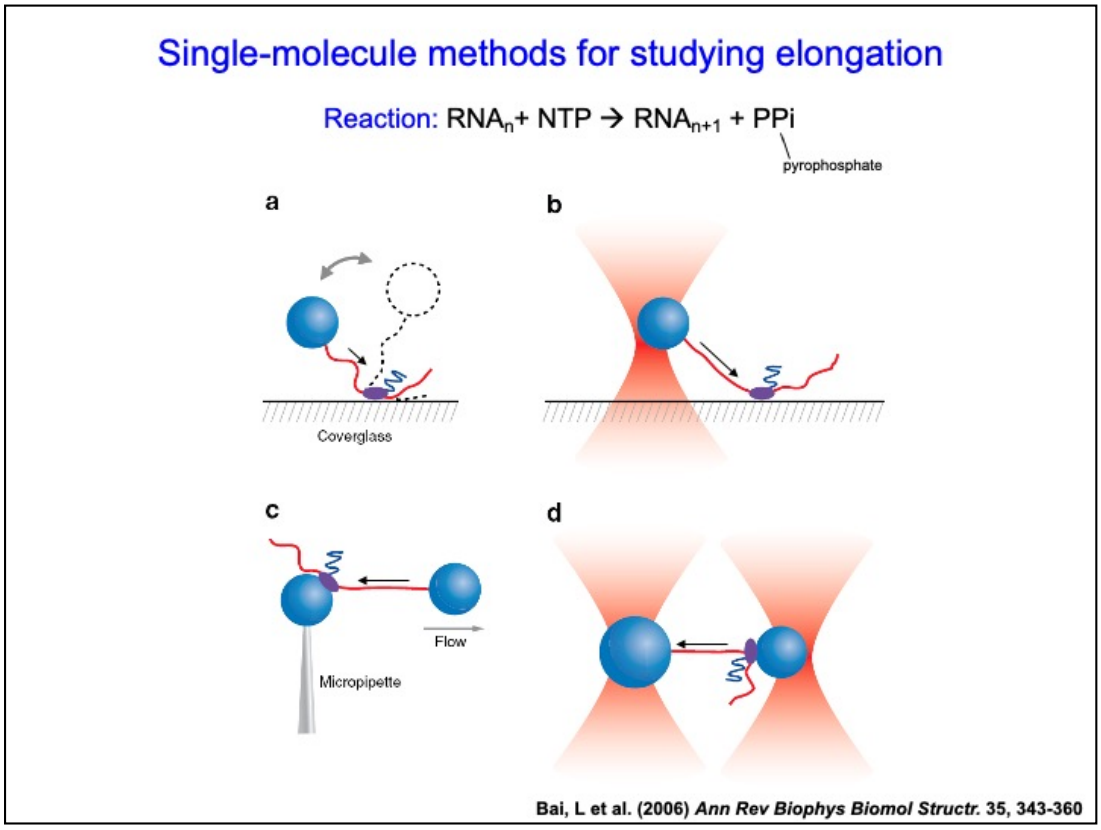


- 85% of promoter escape events have detectable scrunches
- 15% not detectable due to short duration (< 1 sec)
- => 100% of transcription cycles have scrunches

Scrunching is obligatory intermediate for escape

"scrunch and go!"

Revyakin et al (2006) *Science* 314, 1139-1143



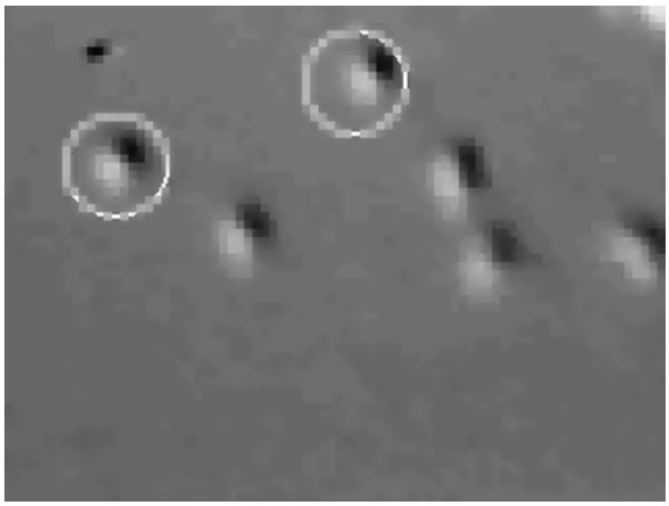
A. Single-molecule elongation using tethered-particle motion: monitored elongation through increased confined Brownian motion of tethered particles, since elongation lengthens DNA-tether and blurs bead image. Transcription rate was 9-17 nt/s (similar to solution, 13-20 nt/s). This work was a tour-de-force for 1991 and open new avenues for the field (although it was lacking resolution and did not detect any pauses during elongation).

B. Optical tweezers approach with a surface-immobilised RNAP pulling a bead out of an optical trap. This experiment was done either at a constant assisting or opposing force (using feedback; this is known as a "force-clamp" experiment) or at an increasing opposing force (without feedback); see L1 lecture on phage DNA packaging.

C. A DNA tethered to a bead is extended by gentle flow; an RNAP molecule (immobilised on a second bead) pulls the DNA and the DNA-tethered bead. Translocation of RNAP is monitored by tracking the position of the DNA-tethered bead.

D. Similar to B but the whole system is now in solution, held by the dual beam optical trap. More in the following slides.

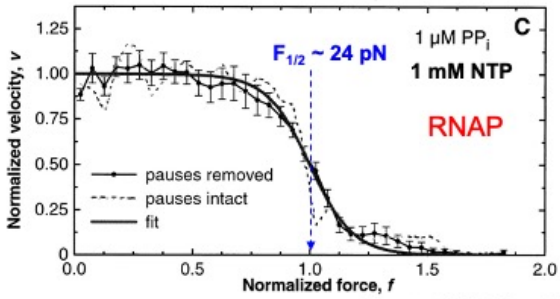
Tethered particle motion (TPM) movie



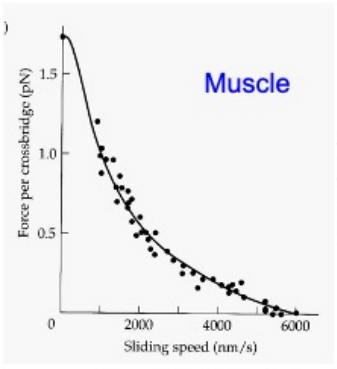
Schafer et al, *Nature* 1991

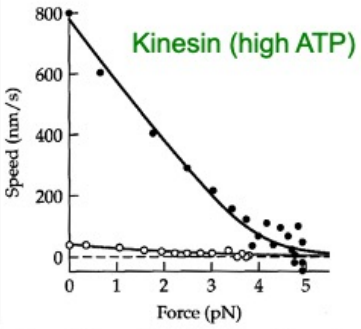
RNAP mechanochemistry: different from kinesin, myosin

- Mechanical step in cycle NOT rate-limiting at low loads
- Rate-limiting step is biochemical
- Brownian ratchet? or power stroke?
- Step size?



Wang et al (1998) *Science* 282, 902-907





Howard, J. (2001). Mechanics of motor proteins and the cytoskeleton. Sinauer Press

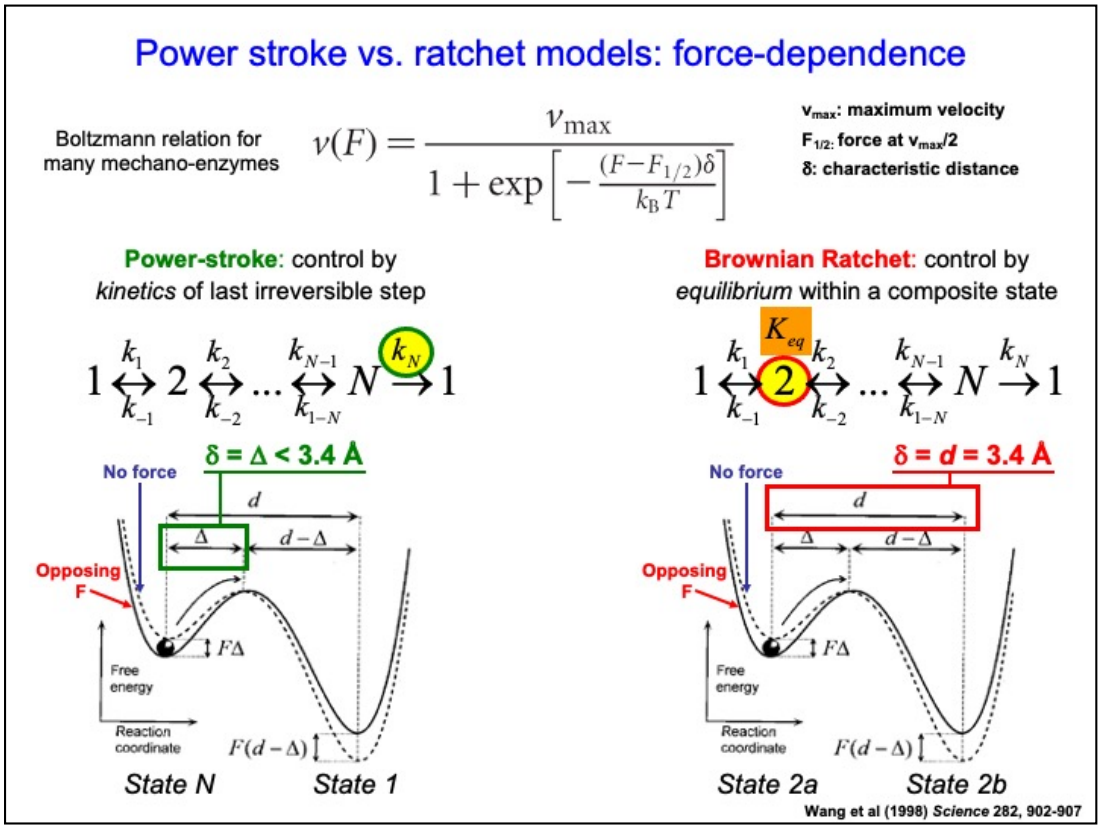
RNAP uses NTP hydrolysis to drive the transcription reaction forward.

Note that the transcription elongation rate is heavily regulated by DNA sequence, RNA secondary structure and elongation-related proteins. This helps control gene expression.

Normalised force f=1.0 is the force at which RNAP reaches 50% of its no-load velocity, V_o.

Normalised velocity is the velocity of RNAP compared to its no-load velocity, V_o.

RNAP is a powerful motor: ~25 pN stall force (compare with kinesin, myosin).



For transcription by RNAP:

- a powerstroke model predicts that there is a conformational change within a flexible RNAP that allows the system to alternate between energetically relaxed and energetically stressed states.

- a Brownian ratchet model is based on thermal fluctuations between states that have different binding affinity for an incoming NTP.

Parameter δ: a characteristic distance that describes the effective distance over which the force exerts its effect. Its characterisation provides access to the mechanical step of the enzymatic cycle.

If $F = F_{1/2}$, then $v(F) = 0.5\ v_{max}$

Power stroke: kinetic reaction control

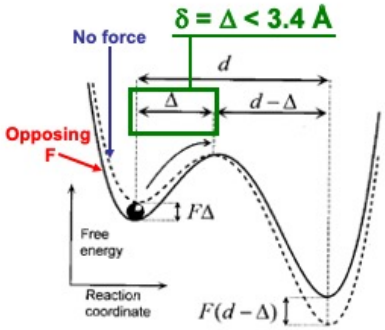
Kinetic scheme for catalytic cycle:

$$1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \ldots \underset{k_{1-N}}{\overset{k_{N-1}}{\rightleftharpoons}} N \xrightarrow{k_N} 1$$

Irreversible step; mechanical transition

Assuming Eyring kinetics for mechanical step:

$$k_N = k_N^0 e^{(F\Delta/k_BT)}$$



Time τ for a complete cycle in the presence of F:

$$\tau = \tau_1 + \tau_2 e^{(-F\Delta/k_BT)}$$

chemistry mechanics

Invert terms, clean up:

$$A = \frac{\tau_2}{\tau_1}$$

$$V(F) = \frac{V_{max}}{1 + Ae^{(-F\Delta/k_BT)}} = \frac{V_0(1+A)}{1 + Ae^{(-F\Delta/k_BT)}}$$

Wang et al (1998) *Science* 282, 902-907

Forces can be either assisting (represented as forces with a positive sign) or opposing (represented as forces with a negative sign).

Parameter $A = \tau_2/\tau_1$, which is small in the case of reactions controlled by biochemical rate-limiting steps (such as transcription at low loads). **Parameter A is force independent constant** .

V_{max}: maximum velocity, reached at high assisting forces. Represents the rate of the reaction when the only rate-limiting step is the biochemical step.

V_o: reaction velocity in the absence of force. It depends on the dwell time in biochemical steps and the dwell time in the mechanical step.

The irreversible step in the power stroke mechanistic model is tied to the PPi release from the complex after NTP hydrolysis. The transition state is between the PPi-bound and PPi-released RNAP states.

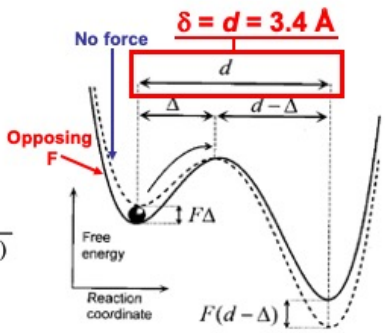
Brownian ratchet: equilibrium reaction control

Catalytic cycle:

$$1 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2 \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \dots \underset{k_{1-N}}{\overset{k_{N-1}}{\rightleftharpoons}} N \xrightarrow{k_N} 1$$

One state is load-dependent composite of 2 sequential substates in rapid equilibrium:

$$\left[\frac{p_+}{p_-}\right]_F = K_{eq,F=0} e^{(Fd/k_BT)}$$



Since reaction rate = constant * [p_+/p_-] , time τ for a complete cycle:

$$\tau = \tau_1 + \frac{1}{A' \cdot \left[\frac{p_+}{p_-}\right]_F} = \tau_1 + \frac{1}{A' \cdot K_{eq,F=0} e^{(Fd/k_BT)}}$$

chemistry mechanics

By cleaning up:

$$V(F) = \frac{V_{\max}}{1 + A' e^{(-Fd/k_BT)}}$$

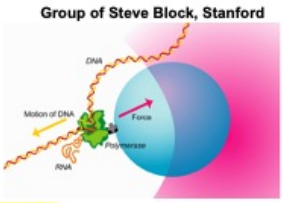
Wang et al (1998) *Science* 282, 902-907

The rate of the overall reaction is proportional to the relative occupancy of the forward vs. rearward states.

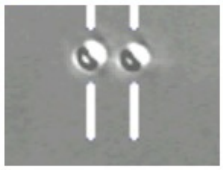
Parameter A' is again a force-independent constant.

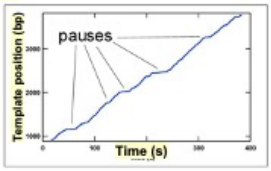
The route to single base-pair resolution

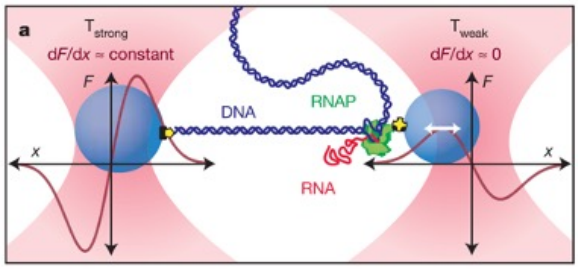
- Better temporal/spatial resolution → single base-pair resolution (1 bp ~ 0.34 nm !)
- Key 1: RNAP levitation
- Key 2: Ultra-stable trap: 1-Å resolution!
  - Helium to improve laser-pointing stability
  - Constant force without feedback



RNAP Dumbbell movie







Neuman et al. (2003) *Cell* 115, 437; Abbondanzieri et al (2003) *Nature* 438, 460

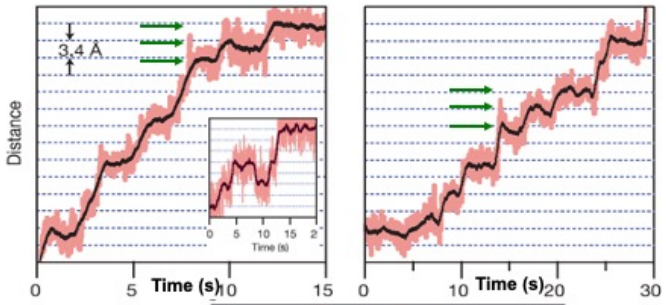
Levitation decouples the system from noise associated with the stage. For levitation, the elongation complex binds bead 1 through RNAP (e.g., T_{weak} in the bottom panel) and binds bead 2 (e.g., T_{strong} in the bottom panel) through DNA. Transcription is monitored through simple monitoring of bead displacement (tether lengthens or shortens and the bead in the weak trap moves).

Helium has a refractive index much closer to 1 than air; this decreases deflection of photons directed to the position detectors (due to density fluctuation of the medium crossed by the laser beams).

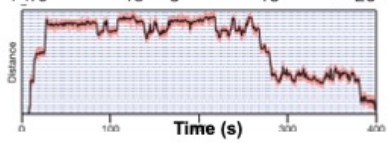
Constant force without feedback is achieved by placing the bead of the weak trap at a region where the force remains constant (a ~50 nm region, ~220 nm from trap centre). In that arrangement, the DNA tether lengthens, and the bead in the weak trap moves right.

Observing RNAP translocation at high resolution

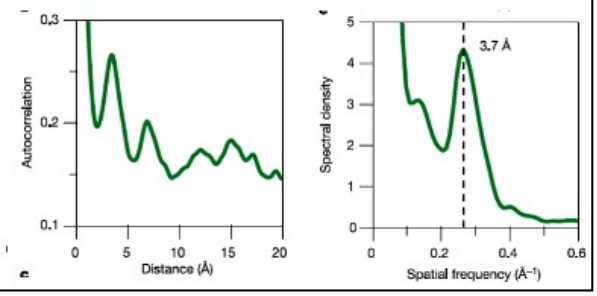
- Staircases of translocation



- Spatial autocorrelation: peaks at ~*n*3.4 Å



- Power spectrum of autocorrelation: → steps of 3.7+/-0.6 Å
- Consistent with the bp step in B-DNA: 3.4+/-0.5 Å
- Step size: 1 bp



Abbondanzieri et al (2003) *Nature* 438, 460

Stepping Staircases: some clear steps seen, but many dwells are absent or skipped (too fast). Backstepping (1-2 bp) and backtracking (>3 bp) are also observed.

Autocorrelation analysis was performed between mean values of RNAP position for extended dwell times.

Models for translocation mechanism

Power stroke coupled to a product (PPi) release

δ: a characteristic distance over which force influences reaction rates

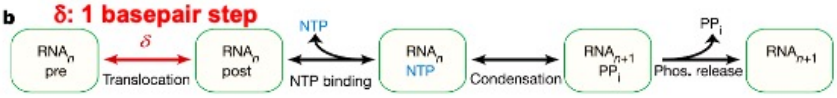
a



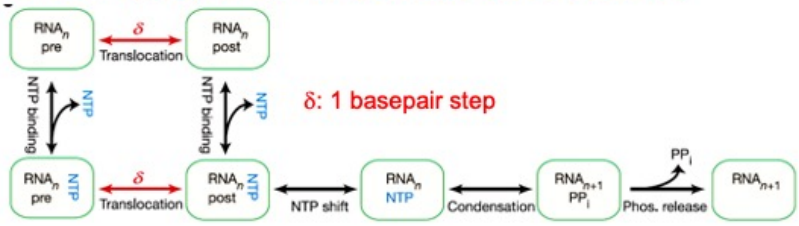
δ: a fraction of a basepair

Brownian ratchet; NTP binds after translocation

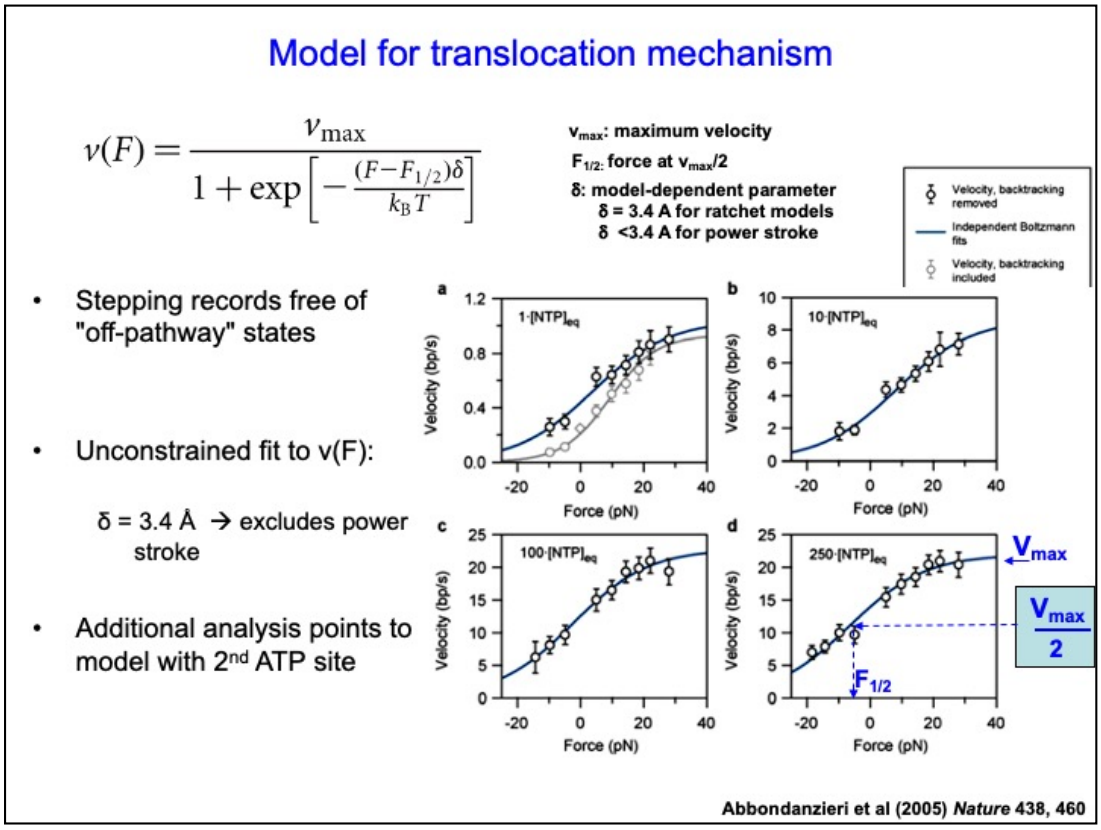
b **δ: 1 basepair step**



Brownian ratchet with 2nd binding site; NTP binds before or after translocation



Abbondanzieri et al (2005) *Nature* 438, 460



Records were cleaned from backtracking/ back-stepping (off-pathway states) that can otherwise preclude F-v analysis.

General characteristics:

High assisting force (positive force), higher transcription velocity.

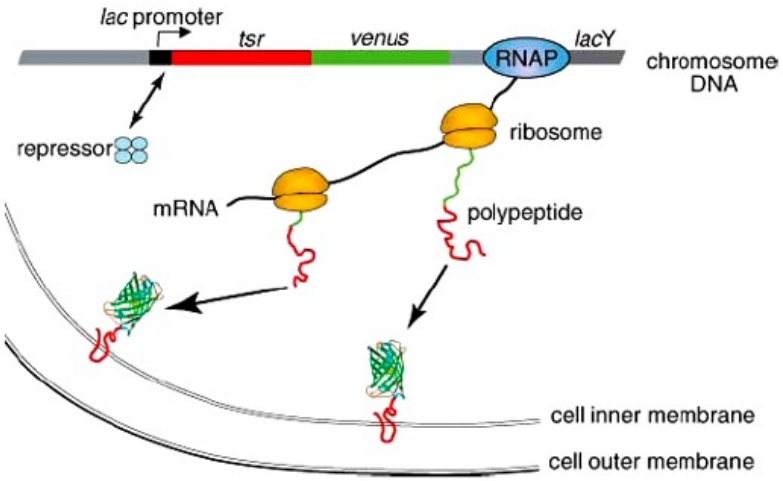
Higher [NTP], higher transcription velocity.

The 2nd ATP site ratchet model predicts better the larger sensitivity to force at high [NTP].

Structural and biochemical evidence for a potential 2nd NTP binding site exists.

Gene expression: single-molecules in single cells

- Observe and analyse single events of gene expression *in vivo*?
- Detection of expression of single GFP molecules (Xie group, Harvard)



Yu et al. (2006) Science 311, 1600-1603

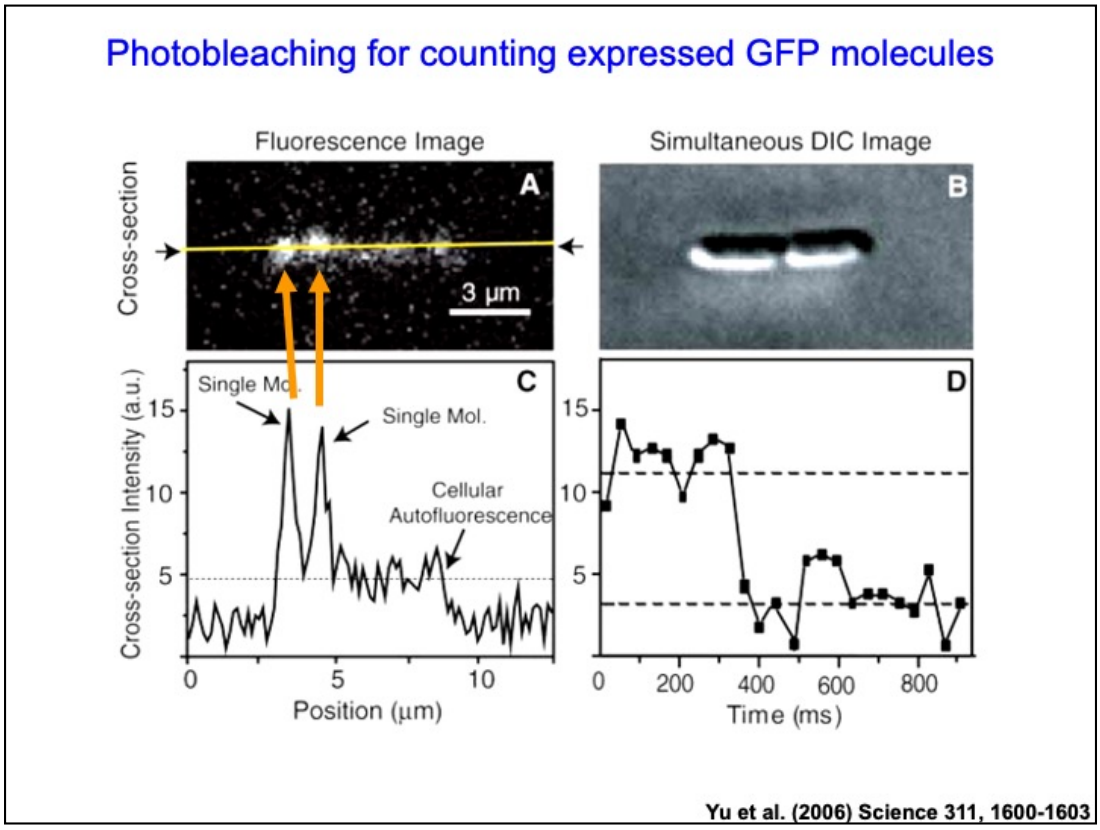
Since most genes exist as a single copy (or, at most, few copies) in cells, gene expression is of stochastic nature. Probing gene expression at the single-molecule level gives access to this stochastic behaviour.

The GFP is fused with a membrane protein to slow its diffusion and allow imaging of single molecules (by increasing the S/N).

Protein synthesis in bacteria at the single-molecule level



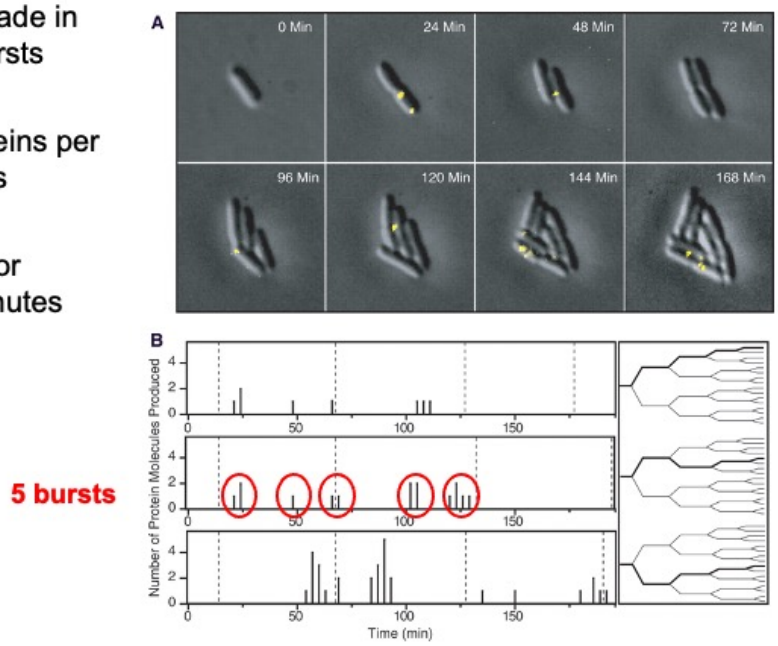
Yu et al. (2006) Science 311, 1600-1603



Single-step photobleaching: signature of a single fluorescent molecule.

Real-time imaging of protein expression

- Proteins made in random bursts
- No. of proteins per burst varies
- Burst last for several minutes



5 bursts

Yu et al. (2006) Science 311, 1600-1603

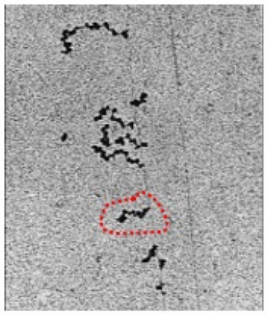
100ms laser exposure per frame (every 3 min), followed by photobleaching of GFP molecules by exposing them for 1.1sec. This short exposure avoids cell damage.

Tracking one protein in a bacterial cell: equivalent to tracking a single person in Wales (~1 in 2.5 million)

Each protein-production burst lasts for 3-15 min.

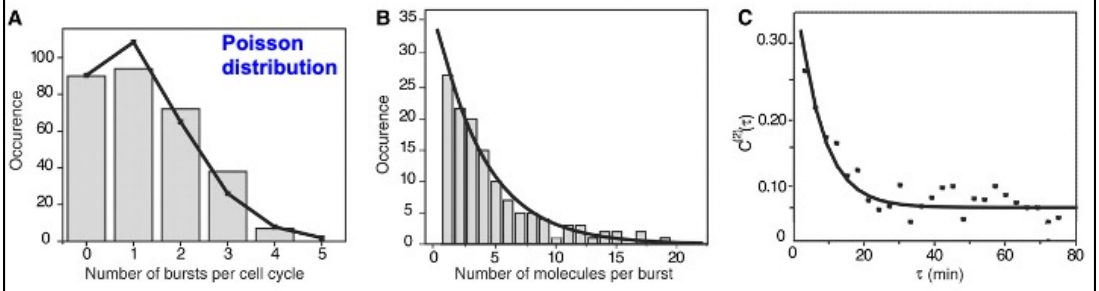
Data analysis of gene expression bursts

- ~1.2 expression events per cycle (random)
- Proteins arise from ~1 copy of RNA
- ~ 4.2 proteins per burst; ~5 per cell cycle
- Temporal spread: ~7 min (GFP folding?)



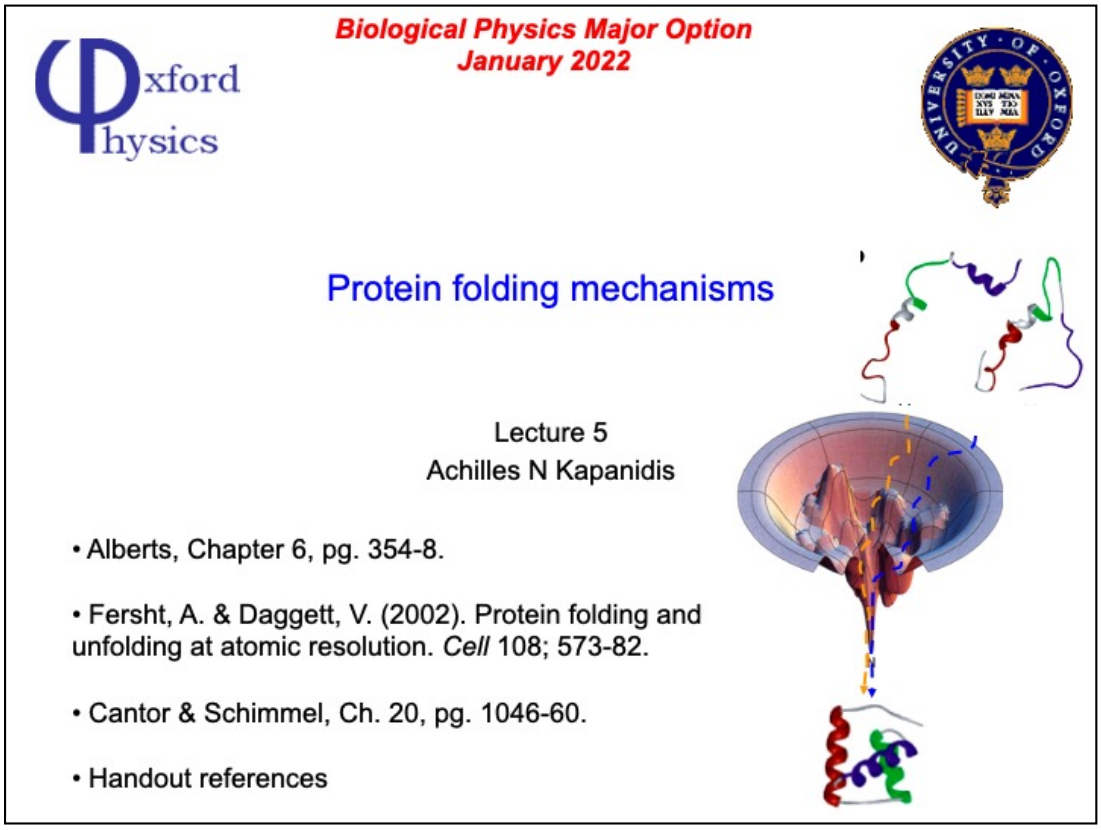
Miller, O. et al (1970). Science 169:392-5

1 RNA → many copies of a protein



Yu et al. (2006) Science 311, 1600-1603

This amazing technical feat opened new avenues for studying stochastic processes and low-copy biomolecules in living cells and at real time.



Importance of protein folding field:

--predict high-resolution 3D structure and dynamics from nucleotide sequence (great for the millions of proteins of unknown structure and function)

--misfolded proteins can form non-functional proteins OR might result in insoluble, fibrous structures that can lead to many human diseases (Alzheimer's, Parkinson's) as well as animal diseases (bovine spongiform encephalopathy, BSE; a.k.a., mad cow disease).

--being able to decipher 3D-structure (and dynamics) from 1D-protein sequence, along with understanding of biomachines and devices, will allow us to engineer machines with novel functions, and building blocks for construction at the nanoscale.

Protein folding revision

- Proteins: self-assembled machines and building blocks

GYHHARAKIILLPAGGYV….

- Protein native 3D structure dictated by amino acid sequence



- *Thermodynamic requirement*: native state is thermodynamically most stable state under natural conditions (Anfinsen 1962)
- *Kinetic requirement*: proteins must follow pathways to fold (random search longer than age of Universe; Levinthal's paradox)

Figure 3–13. Molecular Biology of the Cell, 4th Edition.

After protein synthesis in the ribosome, proteins fold to become functional. Protein structure: coded in the sequence. Folding *in vivo* is helped by molecular chaperones, but the final structure is the same if protein is folded *in vitro.* Proteins are marginally stable, with 5-25 kT of stabilisation energy (worth just a few hydrogen bonds).

Anfinsen (Nobel Chemistry 1972) and his **"thermodynamic hypothesis"** of folding (1962): the native protein conformation occurs because it is the **thermodynamically most stable** in the intracellular environment. This conformation results from the constraints of peptide bonds as modified by other amino acid properties (that is, considering all **intramolecular interactions**). To test his hypothesis, Anfinsen unfolded the protein ribonuclease and observed it returning to its original structure **spontaneously** when he returned the environment to conditions similar to the ones in cells.

Levinthal's paradox (Cyrus Levinthal, 1968): proteins with 124 residues: 10^{50} conformations possible. If new conformation is tested every 10^{-13} s $\Rightarrow$ 10^{30} years to fold correctly (larger than age of the Universe!). But ribonuclease *in vitro* **folds in <1 min** (and some proteins fold in less than a millisecond; the speed limit is currently at 2 microseconds)! Thus, proteins must fold through specific kinetic pathways.

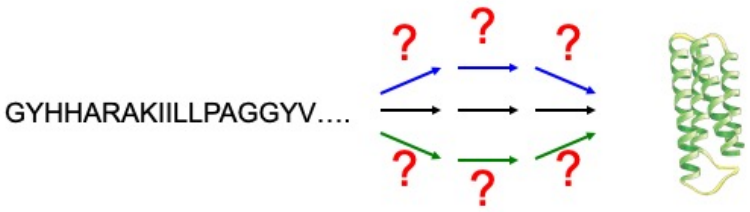
Interesting fact: in the unfolded state, the number of possible configurations is so large, that every molecule in a typical protein sample (with trillions of molecules) could be in a different conformation!

The two protein-folding problems

- Predict 3D structure from sequence

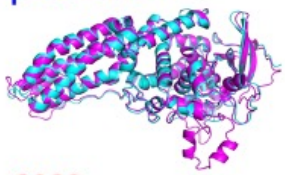


- Pathways of folding



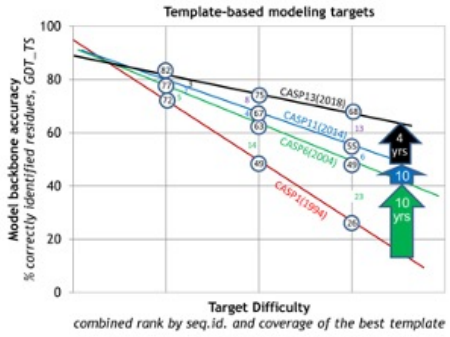
CASP: critical assessment of structure prediction

- competition to objectively assess abilities in protein structure prediction

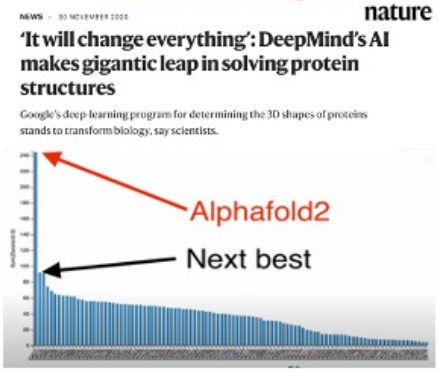


- participants predict unreleased structures (double-blind)

1994-2018:
steady (but slow) improvement



2020:
AI breakthrough



https://deepmind.com/research/case-studies/alphafold

for a webinar on protein folding and CASP, click here

Alpha-fold is based on the multiple sequence-alignment for existing protein sequences. It is based on deep-learning approaches developed initially for language translation.

Resolutions of ~0.2 nm are achieved by Alpha fold; further increase in resolution will help with drug design and understand interactions with other molecules.

Still this advance does not explain the mechanisms that lead to the folded structure.

Department of Physics

UNIVERSITY OF OXFORD

Machine Learning & Physics Seminars

Thursday, 27 January 2022 at 3:00 pm
To receive Zoom room links, send an empty email to
request.zoom.ox.ml.and.physics [AT] gmail [DOT] com

Highly accurate protein structure prediction with AlphaFold

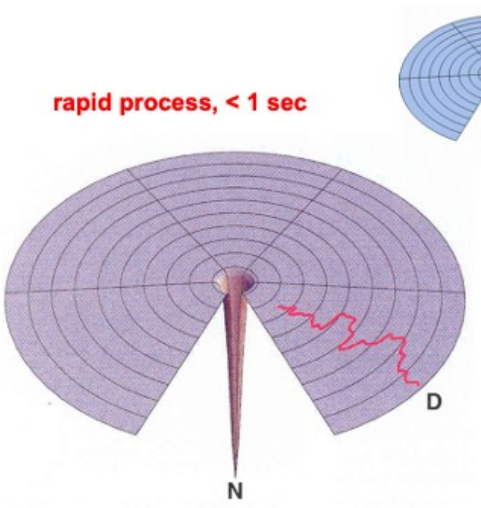
https://zoom.us/j/97744543875

Tim Green
DeepMind

Abstract: Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence — the structure prediction component of the 'protein folding problem' — has been an important open research problem for more than 50 years. AlphaFold, a novel machine learning approach developed at DeepMind, demonstrated accuracy competitive with experimental structures in a majority of cases and greatly outperformed other methods, and has since been recognized as "Method of The Year 2021" by Nature Methods. In this talk I will outline the problem, describe the AlphaFold method, discuss applications in biology, and sketch possible future directions and connections.

Folding landscapes

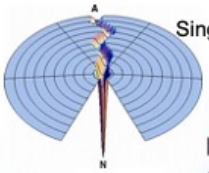
rapid process, < 1 sec

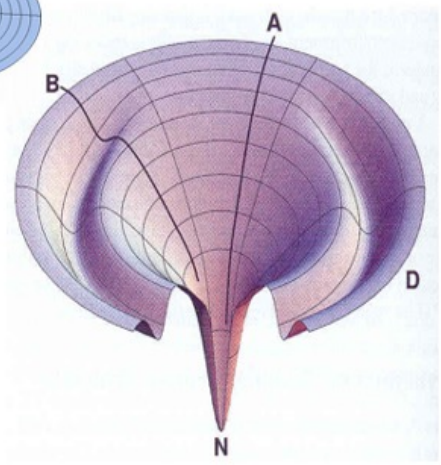


Levinthal 'golf-course'

random search

Single path

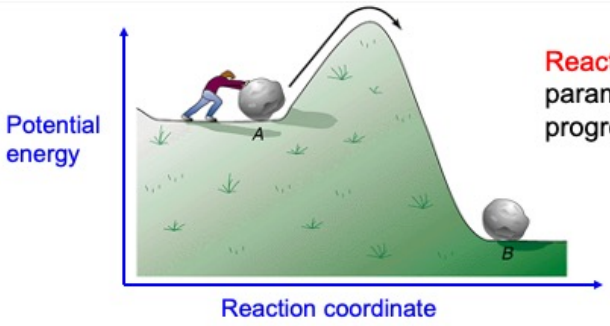




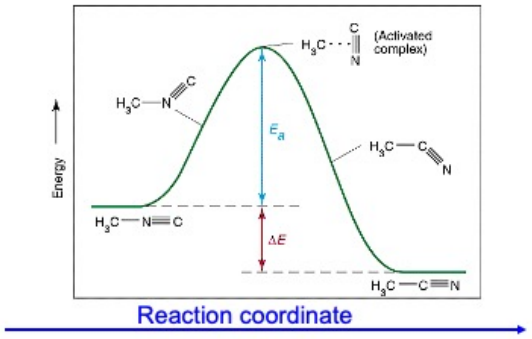
More realistic energy landscape:
'Funnelling' the protein into native state
Many downhill paths

Dill, K.A. & Chan, H.S. (1997) *Nat Str Biol*, 4, 10-19

Potential energy surfaces and reaction coordinate



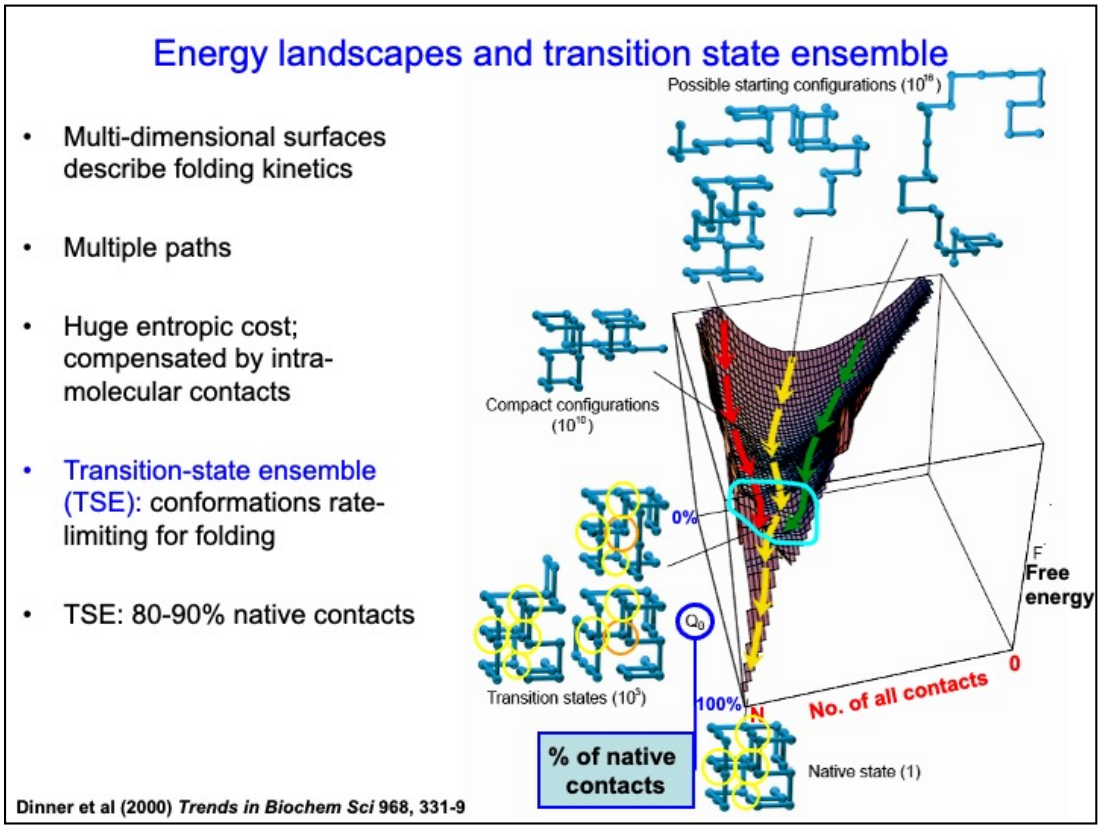
Reaction coordinate: a parameter that measures progress of a reaction



$$k_1 = A \exp\left(-\frac{E_a}{kT}\right)$$

Kinetics of chemical reactions are described by potential energy surfaces (**potential energy** vs **reaction coordinate**). This theory can identify paths for a reaction as well as describe the transition state and any intermediates.

Reaction coordinate: the "distance" along the path of a reaction. In the bottom panel, this refers to the **relative position of the atoms** involved in the reaction.



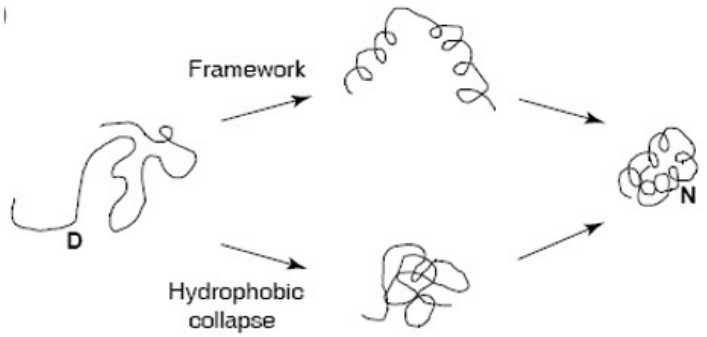
Protein folding has >1000 coordinates, therefore multidimensional surfaces are needed to describe the reaction kinetics.

Multiple pathways are possible toward the native state. The TSE is found at landscape regions through which all molecules must pass to reach native state

Entropy is more important in protein folding than in simple chemical reactions; thus, the protein folding reaction considers free energy (instead of potential energy)

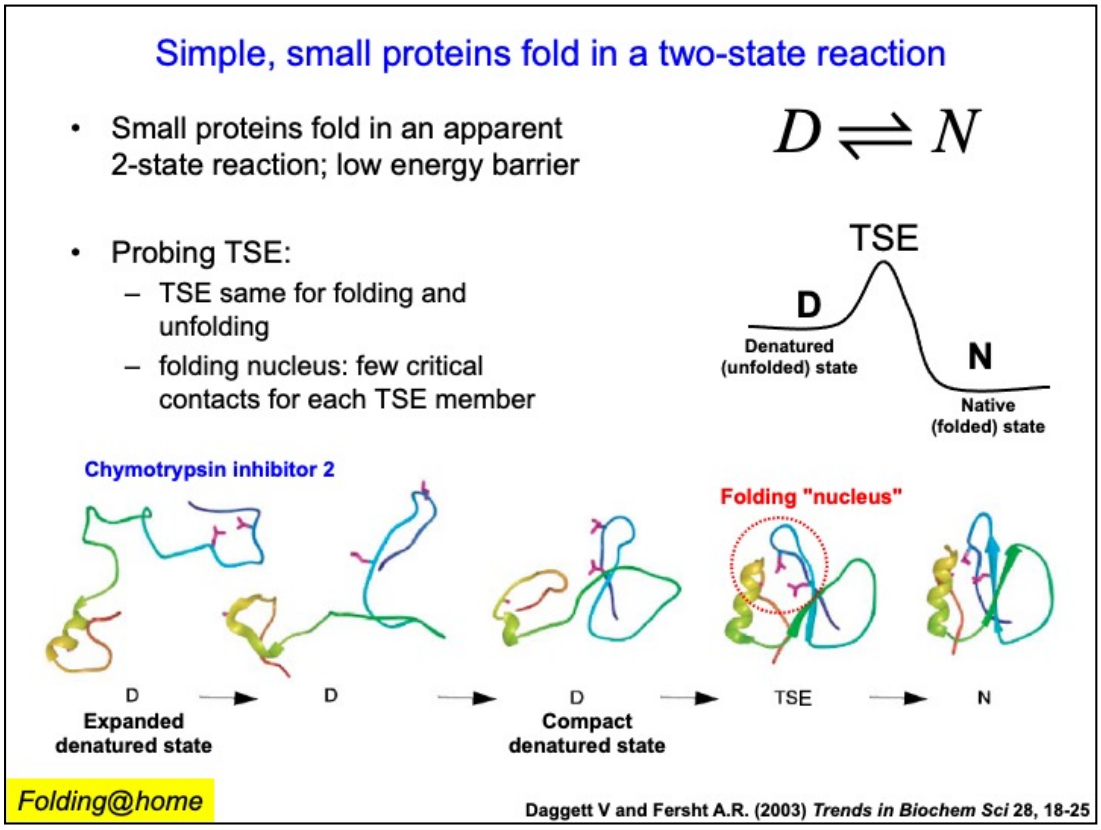
Classical views

- Framework model
  - 2° → 3° structure (similar to native)
  - Hierarchical folding with intermediates
- Hydrophobic collapse
  - Hydrophobic effect compacts protein, narrows conformational search
  - In compact state: non-covalent interactions → folded state



Daggett V and Fersht A.R. (2003) *Trends in Biochem Sci* 28, 18-25

Note that during a general, global hydrophobic collapse, non-covalent interactions lead to a compact state but wrong interactions can also slow process significantly!



The apparent two-state protein folding reaction in CI2 is similar to a small molecule reaction. CI2 folds in 10ms

Since the TS is the same for folding and unfolding (due to principle of microscopic reversibility), the TS can be accessed through unfolding experiments or unfolding MD simulations. These are easier to perform since:

a. the TS is formed early in the unfolding reaction (and, as such, it is within the timescale of the MD simulations)
b. The initial state is the native state, which is the best characterised state in the reaction (as opposed to folding experiments, which start with the least characterised state, the denatured state).

Methods to study protein folding:

- NMR for studying kinetics of formation of folded domains: e.g., see when a certain region of the protein becomes shielded from the solvent.
- Molecular dynamics simulations
- Protein engineering: systematic probing of TS structure by testing energetic costs of single protein mutations; identifies whether mutated amino acid has native contacts in TS or it is denatured in TS. The method requires both equilibrium and kinetic measurements.

Helix → coil transition

- Helix → coil: co-operative transition

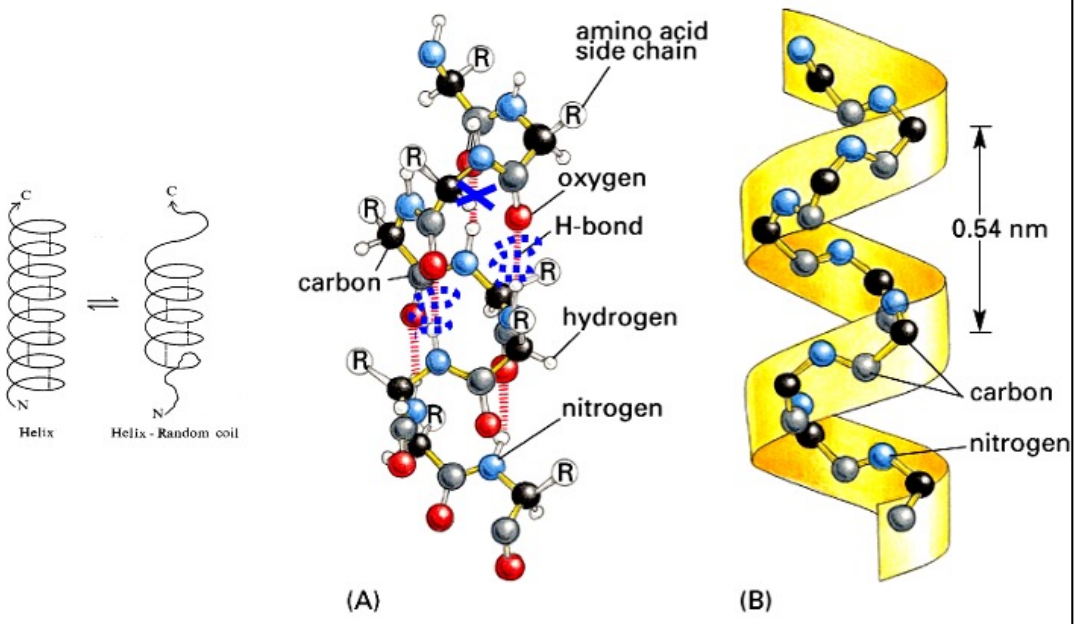


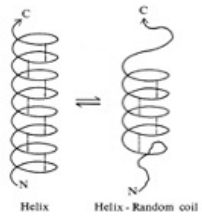
Figure 3–9 part 1 of 2. Molecular Biology of the Cell, 4th Edition.

co-operative transition: polypeptide more susceptible to unfolding after it has been initiated. We can think of it as an "all-or-none" process, or a kind of "phase transition".

Express partition function in terms of equilibrium constants of transitions between states

Helix → coil transition modelling using statistical mechanics

Partition function for various states of molecule



microstates	HHHH	CHHH HCHH HHCH HHHC	CCHH CHCH CHHC HCCH HCHC HHCC	CCCH CCHC CHCC HCCC	CCCC
Stat weight per state	q_0	q_1	q_2	q_3	q_4 (Boltzmann factor)

Partition function $q = q_0 + 4q_1 + 6q_2 + 4q_3 + q_4$

Dividing by q_0 for reference state HHHH: $\Rightarrow q = q_0\left(1 + 4\frac{q_1}{q_0} + 6\frac{q_2}{q_0} + 4\frac{q_3}{q_0} + \frac{q_4}{q_0}\right)$

$\frac{q_1}{q_0} = K_1$ Equilibrium constant for $HHHH \rightleftarrows CHHH$

$$\Rightarrow q = q_0\left(1 + 4K_1 + 6K_2 + 4K_3 + K_4\right)$$

Partition function q: **the sum of the Boltzmann factors of all microstates of a system**.

Assumption: all states with identical number of C residues contribute equally to the partition function; e.g., for the ensemble of states with 1 C residue, the contribution to the partition function is $4q_1$

K_i: probability of any state relative to reference state [HHHH]

Express partition function in terms of equilibrium constants of transitions between states

Helix-coil transition

Setting $E_{HHHH} = 0 \Rightarrow q_0 = \exp(-0/kT) = 1$

$$\Rightarrow q = 1 + 4K_1 + 6K_2 + 4K_3 + K_4$$

Assume $\mathrm{HHHH} \overset{\Delta G^0}{\rightleftarrows} \mathrm{CHHH} \overset{\Delta G^0}{\rightleftarrows} \mathrm{CCHH} \overset{\Delta G^0}{\rightleftarrows} \mathrm{CCCH} \overset{\Delta G^0}{\rightleftarrows} \mathrm{CCCC}$

then

$$K_1 = \exp(-\Delta G^0/kT)$$
$$K_2 = \exp(-2\Delta G^0/kT) = K_1^2$$
$$K_3 = \exp(-3\Delta G^0/kT) = K_1^3$$
$$K_4 = \exp(-4\Delta G^0/kT) = K_1^4$$

Setting $K_1 = s$, we have $q = 1 + 4s + 6s^2 + 4s^3 + s^4 = 1 + \sum_{i=1}^{4} N_i s^i$

For *n* amino acids: $\boxed{q = 1 + \sum_{i=1}^{n} N_i s^i}$

N_i: coefficients of binomial expansion of $(1+s)^n$

s: also known as "stability" parameter.

N_i: coefficients of binomial expansion of $(1+s)^n$

Helix→coil: the zipper model

HHHH → CCCC



Nucleation ($\sigma << 1$) Propagation

HHHH $\xrightarrow{\sigma s}$ CHHH $\xrightarrow{s}$ CCHH $\xrightarrow{s}$ CCCH $\xrightarrow{s}$ CCCC

HHHH $\longrightarrow$ CHHH ~~$\rightarrow$ CHCH~~

number of residues

Nucleation

(Problem set) Propagation

$$q = 1 + \sum_{i=1}^{n} N_i' \sigma s^i \quad \xrightarrow{\text{(Problem set)}} \quad q = 1 + \frac{\sigma s^2}{(s-1)^2}\left[s^n + n s^{-1} - (n+1)\right]$$

Probability of having *i coil* amino acids in a chain of *n* residues:

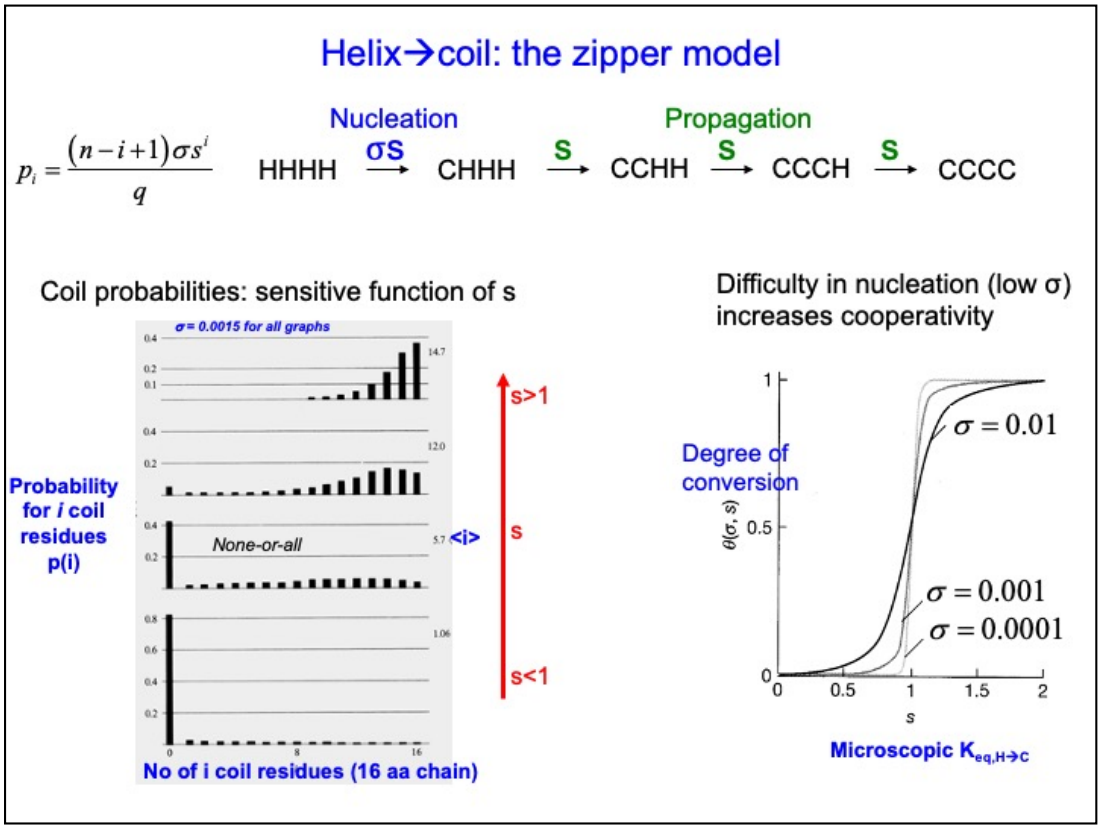
$$p_i = \frac{(n-i+1)\sigma s^i}{q}$$

Mean number of coil residues: $\langle i \rangle = \sum_{i=1}^{n} i p_i$

Degree of coil conversion: $\vartheta = \langle i \rangle / n$

The zipper model allows a h→c transition **only if for h residues adjacent to coil residues;** this way, non-contiguous coil-only sections are not allowed. This is a limitation of the model, which is addressed by a more advanced model, the Zimm-Bragg model (treatment similar to the 1-D Ising model used to describe ferromagnetism).

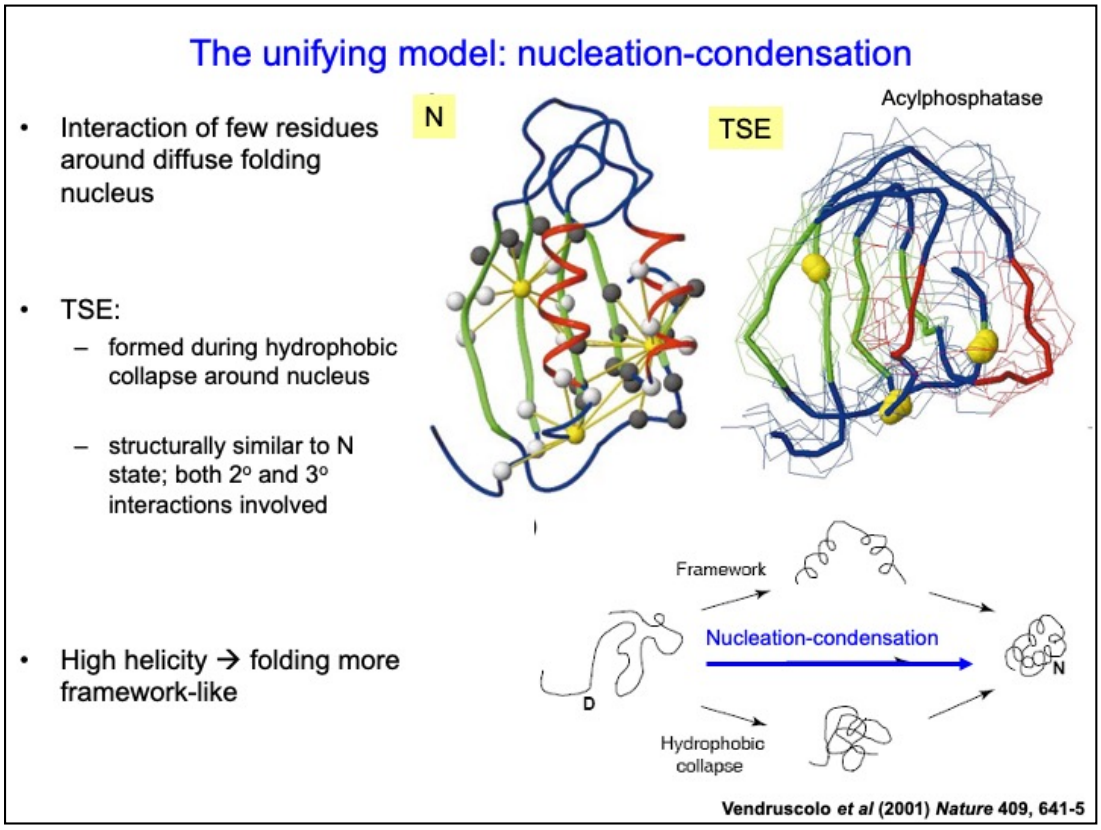
Note that the N_i' coefficients are different from the N_i coefficients of the previous model (you will address this in the problem set).



s>1: mostly coil

s<1: mostly helix

Small σ values will increase the heterogeneity of the graph with <i> = 5.7, disfavouring conformations with short helical lengths.

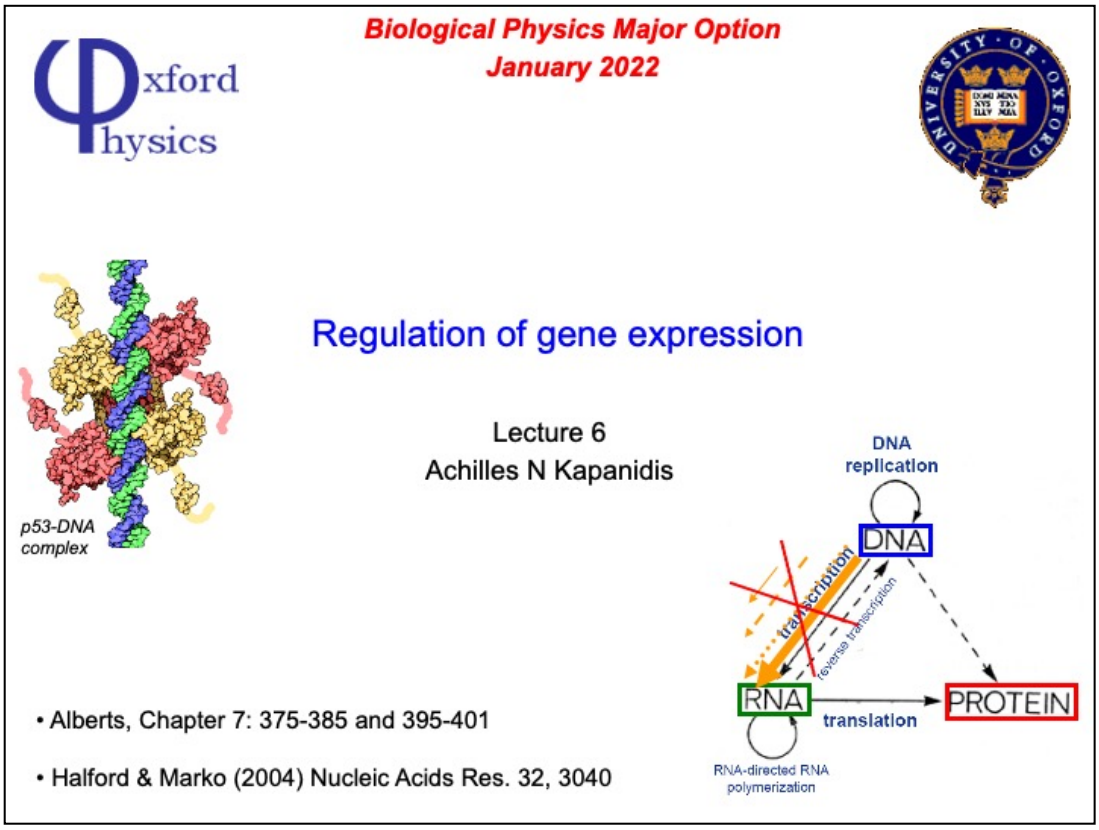


The hydrophobic collapse is global, and is driven by the hydrophobic effect (i.e., the burying of hydrophobic amino acid side chains in the protein interior to avoid their exposure to the aqueous solution).

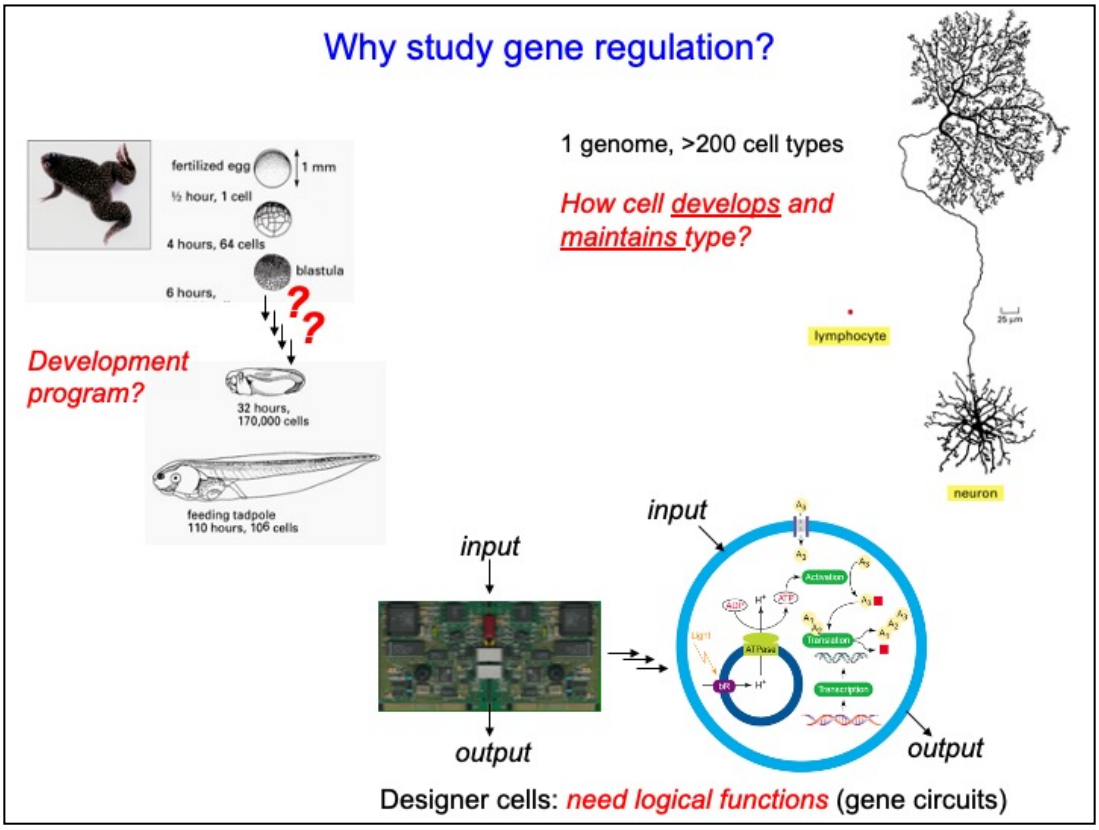
No proteins that formed *solely* due to by hydrophobic collapse (condensation) have been identified yet.

Along folding, 2° structure is stabilized by 3° interactions

Native structure forms late in folding (after TS)



p53 (**protein 53)**, is a transcription factor that act as a tumour suppressor protein (which means that it keep eukaryotic cells from growing in an uncontrolled manner). p53 has been described as "the guardian of the genome". p53 is mutated in ~50% of human cancers.



Understanding gene regulation is important for many fields:

--Organism development, cell differentiation (embryonic stem cells)

--Cell function and response to environment

--Therapeutics: gene therapy with RNA interference, antibiotics

--Cancer biology and treatment: understanding molecular mechanisms of gene deregulation in cancer can lead to better diagnostics and treatment

--Bionanotechnology and synthetic biology: artificial cells

Basis of regulation: genes expressed at different levels

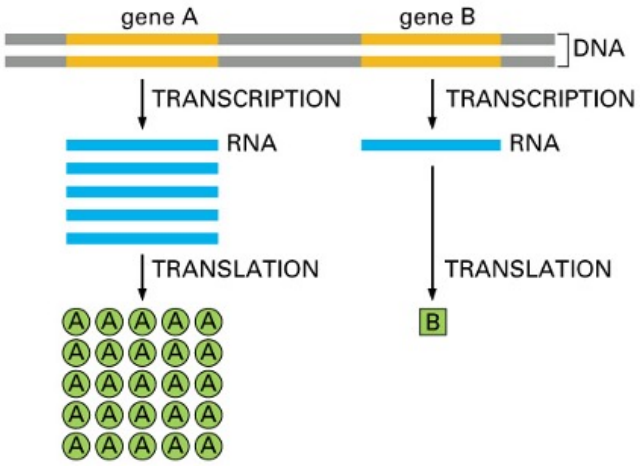


Figure 6–3. Molecular Biology of the Cell, 4th Edition.

Gene expression controlled at many points

Most regulation occurs early, at the level of transcription.

Repressors: Transcription inhibition

- simplest regulation: repression of transcription
- Repressors: DNA-binding proteins; block binding of RNAP to promoter DNA
- Repressor DNA-binding controlled by other molecules

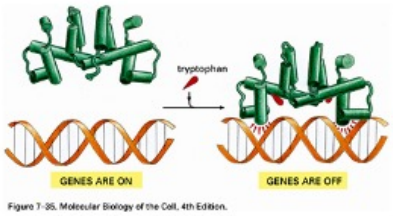
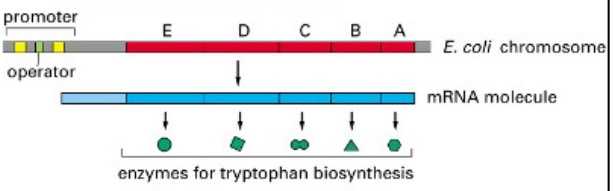
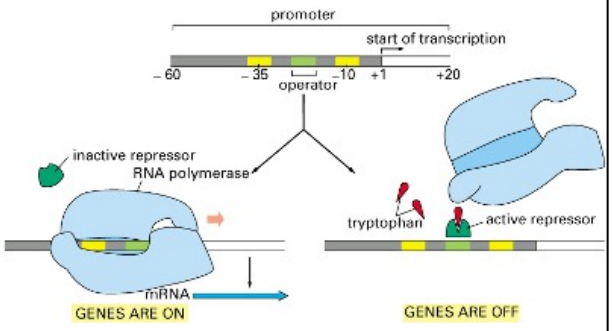
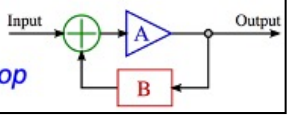


Figure 7-35, Molecular Biology of the Cell, 4th Edition.

Example: Trp repressor







a negative feedback loop

A repressor protein blocks the promoter region, preventing RNAP from landing to DNA and forming an open complex

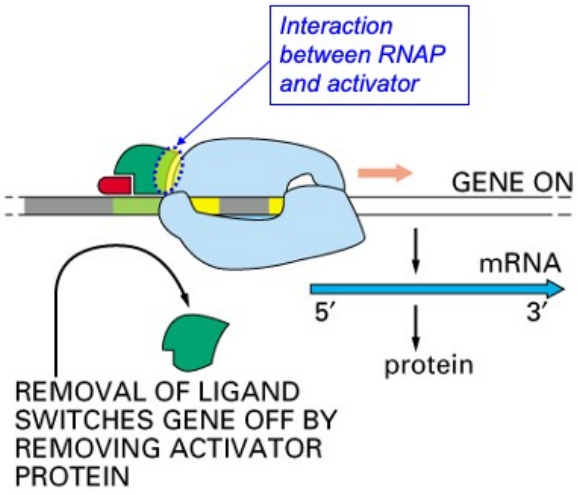
Repressors bind to DNA regions known as "operators".

Repressor activity can be **controlled by other molecules** (i.e., by small molecules such as nucleotides and amino acids, by proteases, by metals etc). This is the key property that allow repressors to sense the environment and act as molecular switches and modulators.

In a **negative feedback loop**, the output of a system opposes changes to the input of the system, attenuating the changes that are produced by the system.

Activators: Transcription acceleration

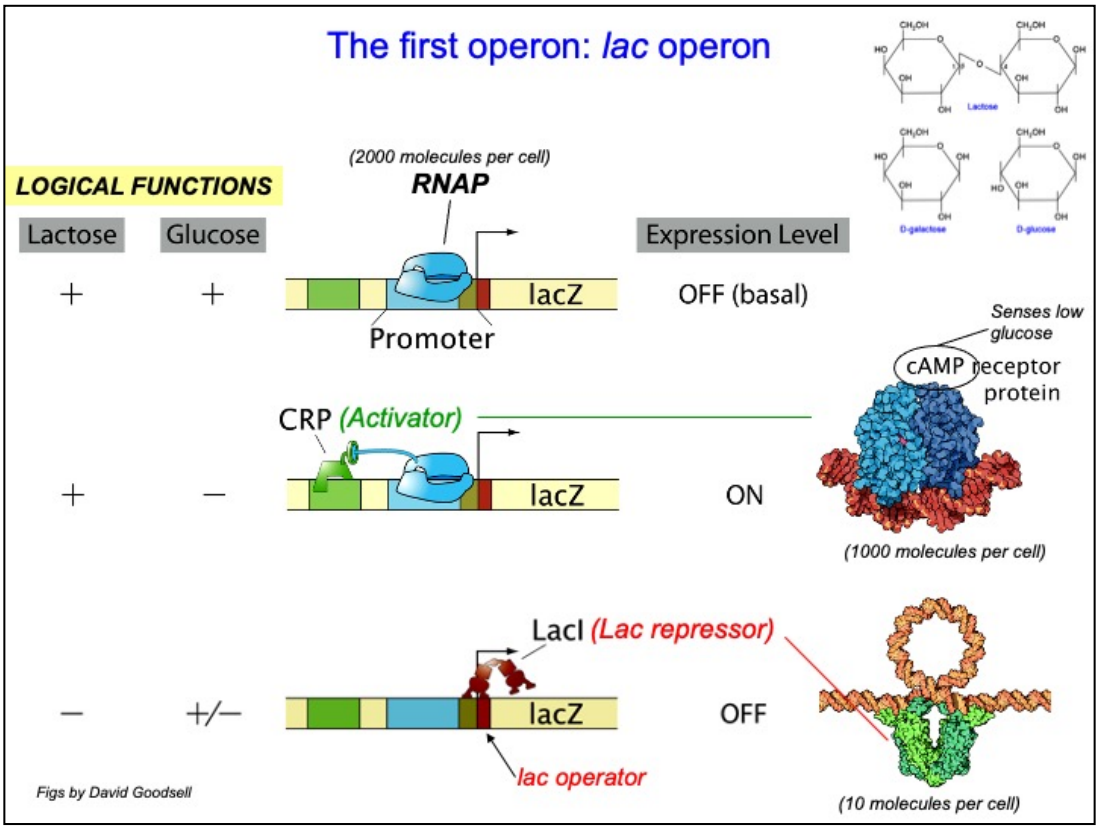
- Positive control: activators of transcription
- Activators: DNA-binding proteins; enhance RNAP binding to promoter DNA ("molecular velcro")
- Activator DNA-binding controlled by other molecules



An activator protein **enhances the rate of the formation of an RNAP-DNA open complex**, increasing the number of initiation events per unit time. This is done primarily through direct interactions with RNAP.

Activators bind to DNA regions upstream of the transcription start site; some can act from thousands of base pairs away.

As with repressors, activator activity can be controlled by other molecules (i.e., by small molecules such as nucleotides and amino acids, by proteases, by metals etc). This is the key property that allow activators to sense the environment and act as molecular switches and modulators. Control regions are usually controlled by a set of repressor and activators.



The concept of the operon was first described by Francois Jacob and Jacques Monod in the early 60s (Nobel in Medicine 1965)

Lactose, glucose: sugars that are broken down by bacteria for energy production.

The CRP protein (also known as CAP, catabolite activator protein) stands for **cyclic-AMP receptor protein**, because is activated by cyclic AMP (cAMP), a small molecule that senses the low concentration of glucose. When the cAMP levels are high, CRP-cAMP binds to DNA and activates transcription by enhancing the formation of the RNAP-DNA open complex.

The **lacI** protein (I stands for inducibility) is know as ***lac* repressor**; it binds an operator region just downstream of the transcription start site, blocking landing of RNAP on the promoter. Upon binding lactose, it releases the promoter so it can used for making proteins for the consumption of lactose, at normal or accelerated levels (depends on the glucose availability).

Thermodynamics of protein-DNA interactions

Protein-DNA interactions: bi-molecular associations

association rate constant ("on-rate") $M^{-1}s^{-1}$
Upper limit set by diffusion (see L2)

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} AB \qquad \frac{d[AB]}{dt} = k_1[A][B] - k_{-1}[AB]$$

dissociation rate constant ("off-rate") s^{-1}

In equilibrium, $\frac{d[AB]}{dt} = 0$, therefore:

$$\frac{[AB]}{[A][B]} = \frac{k_1}{k_{-1}} = \mathrm{K}_a$$ Equilibrium binding constant (M^{-1})

$$\Delta G^o = -k_B T \ln \mathrm{K}_a$$

$$\frac{[A][B]}{[AB]} = \frac{k_{-1}}{k_1} = \mathrm{K}_d$$ Equilibrium dissociation constant (M)

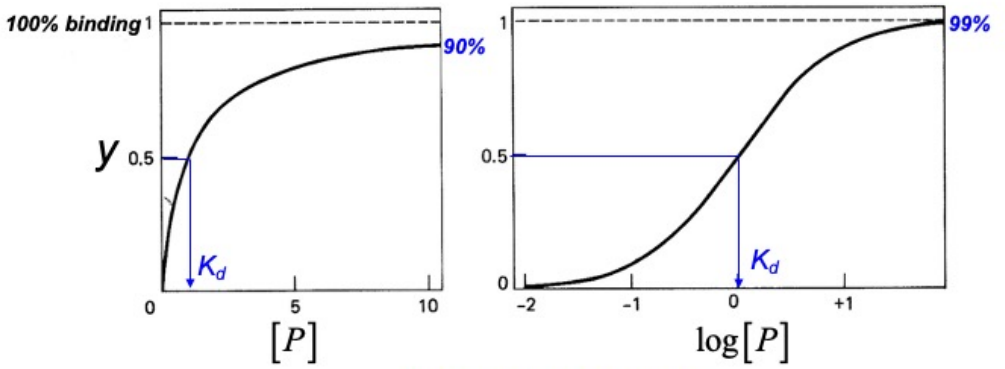
See also Dr Berry's lectures.

Measuring affinity of protein-DNA interactions

Extent of binding: $y = \frac{\text{bound DNA}}{\text{total DNA}} = \frac{[PA]}{[A]+[PA]} = \frac{K_a[P][A]}{[A]+K_a[P][A]} = \frac{K_a[P]}{1+K_a[P]}$

P: protein
A: DNA

Dividing right term with K_a: $y = \frac{[P]}{K_d + [P]}$ Hyperbolic relation of y and [P]



[P] expressed in K_d units

Creighton, *Proteins 1st edition*, Fig 8-1

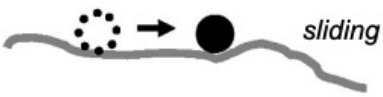
Here we assume that protein is in excess, as in the cell (one or few copies of DNA, multiple copies of protein).

K_d is the concentration of protein that results in 50% occupancy of the DNA sites in question.

Protein-DNA binding curves are collected by keeping the concentration of DNA at low levels, at values much lower that the anticipated K_d (e.g., if the K_d is 10 nM, [DNA] is kept at 0.1-1 nM). This way, one can assume that $[P]_{total} \sim [P]$, where [P] is the concentration of free protein. Then, the concentration of the protein is increased (by addition of protein) and the accumulation of protein-DNA complex (i.e., the concentration [PA]) is monitored using a biophysical or biochemical observable.

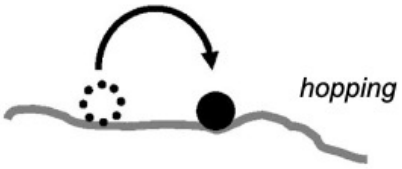
How DNA-binding proteins find their target so fast?

First example: *lac* repressor – *lac* operator binding (Riggs *et al*, 1970)

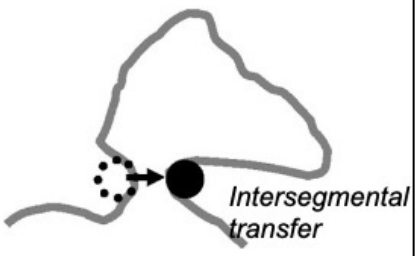


Diffusion-limited association rate constant k_a
10^8 $M^{-1}s^{-1}$

Observed k_a:
$>10^{10}$ $M^{-1}s^{-1}$



"Facilitated diffusion"



Simplified treatment in "How DNA-binding proteins find their targets". Halford & Marko (2004) *Nucleic Acids Res*. 32, 3040.

Target location by 3D diffusion

Assume:

- binding site of size a on DNA fixed in space
- protein at distance r from binding site

Simplified diffusion: $\langle x^2 \rangle = Dt$

3D space between protein and target is divided in 3D-pixels (voxels) of width a: total of $(r/a)^3$ voxels

Considering diffusion:

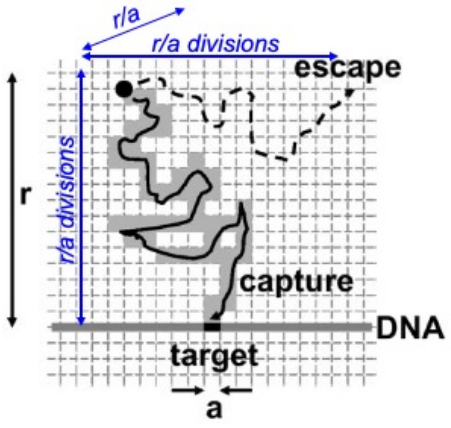
Time per voxel: $t_{voxel} = a^2 / D$

Time to diffuse for distance r: $t_{target} = r^2 / D$

voxels visited: $t_{target} / t_{voxel} = (r/a)^2$

Fraction of total voxels visited: $\frac{(r/a)^2}{(r/a)^3} = \frac{a}{r}$

This is the probability of binding the target site.



This treatment ignores some proportionality constants; these constants either cancel out during calculations or do not change the qualitative description of the process.

D: diffusion coefficient for 3D diffusion. For lacR, D~4x10^{-7} cm^{-2}/s

Sliding length: concept and control

P: probability of protein dissociation from DNA during sliding for a single bp step.

Probability of retention on DNA after N steps: $(1-P)^N = e^{N\ln(1-P)}$

Steps to dissociation (retention probability becomes 1/e): $N = -1/\ln(1-P)$

Since for $P<<1$, we approximate $\ln(1-P) = -P$, and: $\boxed{N = 1/P}$

Sliding length: DNA length explored by 1D diffusion for N steps:

$l_{sl} = \sqrt{N} = 1/\sqrt{P}$ in basepairs $\Rightarrow$

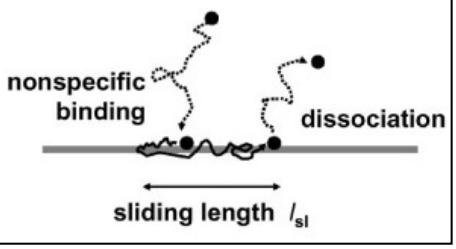
$l_{sl} = 0.34/\sqrt{P}$ in nm

Time per sliding length (1D diffusion): $\tau_{sl} = l_{sl}^2 / D_1$

Rate for dissociation: $k_{off} = 1/\tau_{sl} = D_1 / l_{sl}^2$

and $\boxed{l_{sl} = \sqrt{D_1/k_{off}}}$

k_{off} can be increased by increasing salt concentration.



Simplest way of thinking about how sliding facilitates diffusion: it effectively increases the target size a (the reaction radius; see derivation of diffusion-limited association rate constant).

The protein does not need to land exactly on the binding site: binding with 100 bp or so and 1D sliding can lead to faster target location

D_1: diffusion coefficient for 1D sliding on DNA. It was recently measured to be ~2×10^{-10} cm^2/s (Wang *et al* (2006), PRL 97, 48302).

Targeting radius

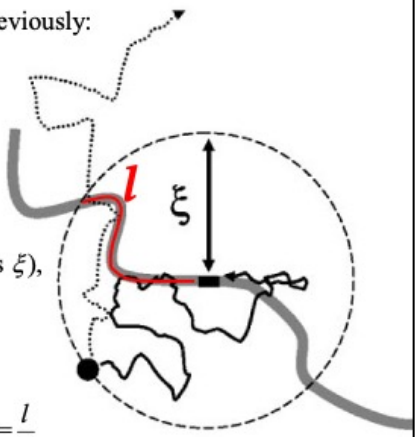
Targeting radius ξ: starting distance from a specific target for which capture probability is 50%.

Assume non-specific binding occurs at a site of size b. As previously:

3D space inside targeting radius : $(\xi / b)^3$ voxels

voxels visited during 3D random walk: $(\xi / b)^2$

Fraction of total voxels visited: $= \frac{b}{\xi}$



A DNA segment of contour length l (corresponding to radius ξ),

contains $\frac{l}{b}$ sites for non-specific binding.

Number of non-specific interactions through 3D walk: $\frac{b}{\xi} \cdot \frac{l}{b} = \frac{l}{\xi}$

Number of non-specific interactions (bind-and-slide) to find target site: $n = \frac{l}{l_{sl}}$

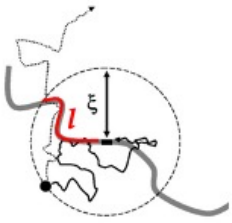
Therefore, for finding the target with 50% probability: $\boxed{\xi = l_{sl}}$

The simple final result does not depend on the exact conformation of the DNA.

Note that binding probability is different from capture probability. Binding is the initial event; 1D sliding then leads to capture of the target.

Search time within targeting radius and within a cell

Total time = time for 3D search within targeting region (radius ξ)
+ time for sliding the entire contour l:

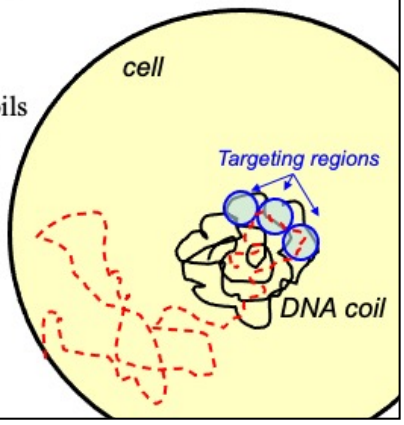


$$\tau_\xi = \frac{\xi^2}{D} + \frac{l}{l_{sl}}\tau_{sl} = \frac{l_{sl}^2}{D} + \frac{l}{l_{sl}}\frac{l_{sl}^2}{D_1} = \frac{l_{sl}^2}{D} + \frac{l\, l_{sl}}{D_1}$$

Since $l >> l_{sl}$ for in vivo situations and $D >> D_1$, the sliding term dominates:

$$\tau_\xi \approx \frac{l\, l_{sl}}{D_1}$$

Using similar arguments for a cell containg large DNA coils
(each containing several targeting regions with radius ξ):



Total search time:

$$\tau_{search} = \frac{V}{D l_{sl}} + \frac{L l_{sl}}{D_1}$$

Consider:

1. 3D diffusion to reach the DNA coil
2. Coil exploration by visiting targeting regions; in targeting regions, the sliding term dominates.
3. 3D diffusion time within coil regions without DNA is small versus diffusion out of the coil, and thus it is ignored.

For the assumptions and simple calculations, check Halford & Marko (2004) *Nucleic Acids Res*. 32, 3040.

Search time for site location and association rate

Total search time: $\tau_{search} = \frac{V}{Dl_{sl}} + \frac{Ll_{sl}}{D_1}$

By converting search time into rate per unit protein concentration:

$$k_a = \left(\frac{1}{Dl_{sl}} + \frac{Ll_{sl}}{D_1 V} \right)^{-1}$$

An approximate diffusion limited $k_a = Da$. Factoring this out gives:

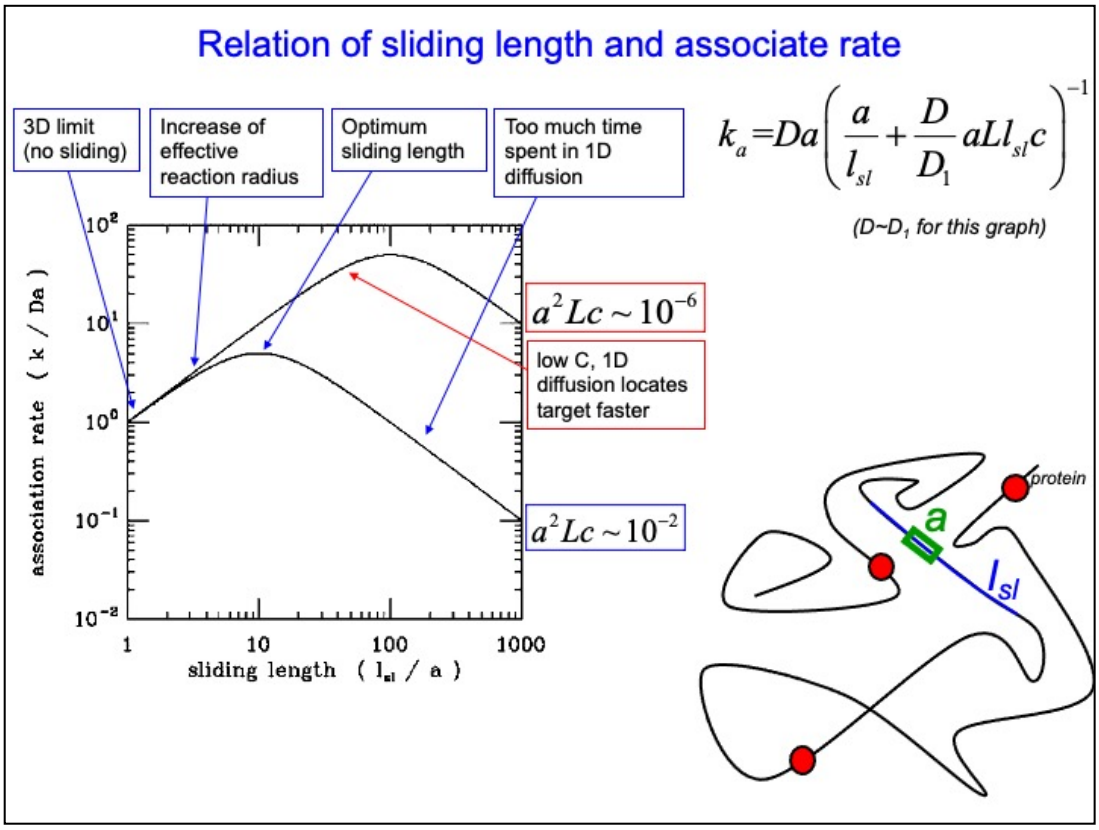
$$k_a = Da \left(\frac{a}{l_{sl}} + \frac{D}{D_1 V} aLl_{sl} \right)^{-1}$$

and considering 1/V = DNA concentration c in a cell of volume V:

$$k_a = Da \left(\frac{a}{l_{sl}} + \frac{D}{D_1} aLl_{sl} c \right)^{-1}$$

Again, for simplification the diffusion-limited association rate constant is k_a=Da (where a is the reaction radius, see lecture 3) instead of 4πDa.

This does not affect the conclusions about the acceleration of target search by 1D diffusion.



Key factor 1: the non-specific sequences flanking the specific site along with sliding essentially increase the reaction radius: a non-specific interaction 50-100 bp from the specific site can lead to formation of a specific complex through 1D sliding.

Key factor 2: long sliding lengths lead to extended and fruitless explorations of DNA away from the binding site. Dissociation and 3D diffusion minimise such unproductive searches.

$a^2Lc \sim 10^{-2}$: "high-concentration" conditions

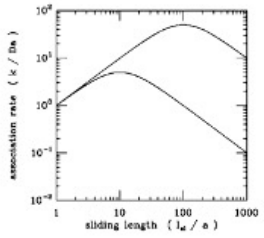
$a^2Lc \sim 10^{-6}$: "low-concentration" conditions

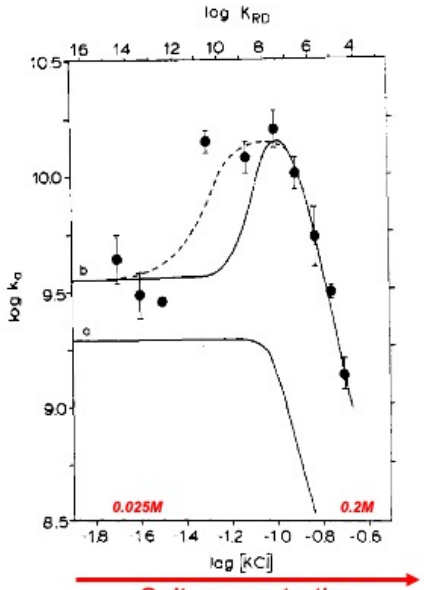
a~ 1nm (a few base pairs)

L_{sl} optimum: ~100 bp (~30nm)

For simplicity, it is assumed $D=D_1$ in this case (in reality, D_1 is much smaller)

Target location: theory vs. experiment





Winter *et al* (1981)
Biochemistry 20, 6961

Compare with experiment:

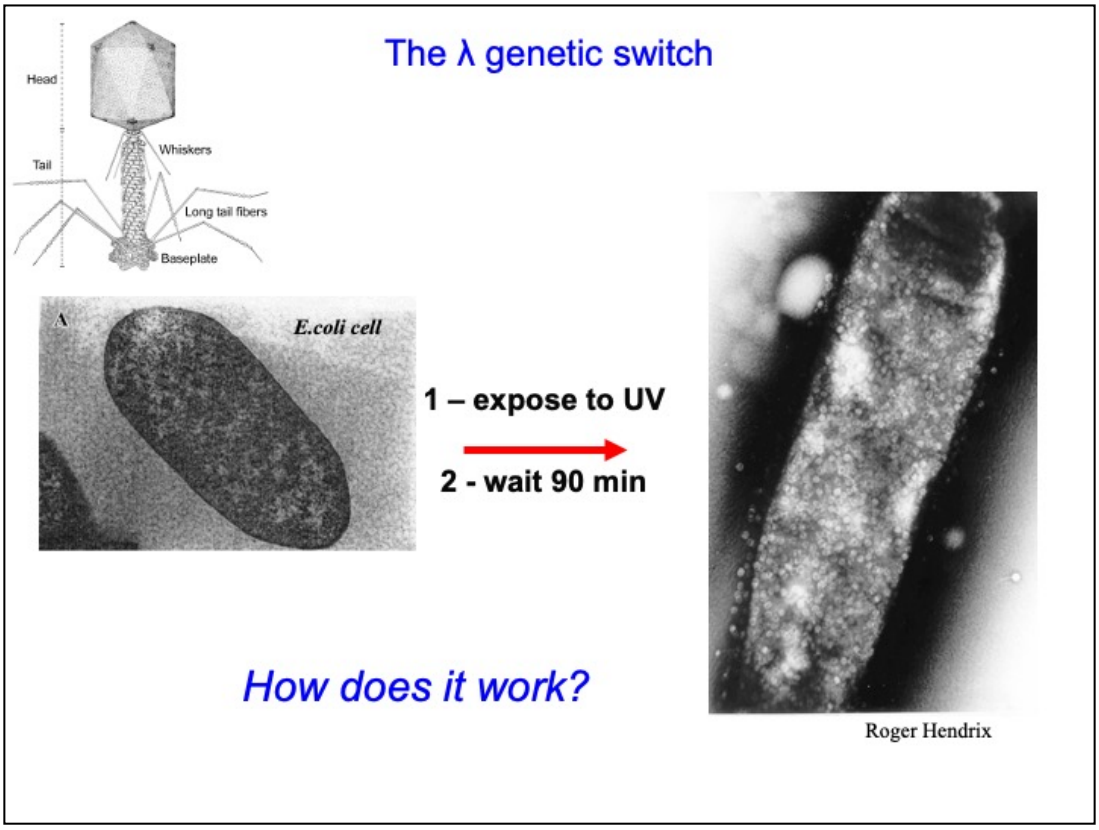
it works!!!

$$l_{sl} = \sqrt{D_1 / k_{off}}$$

k_{off} increases with [salt]

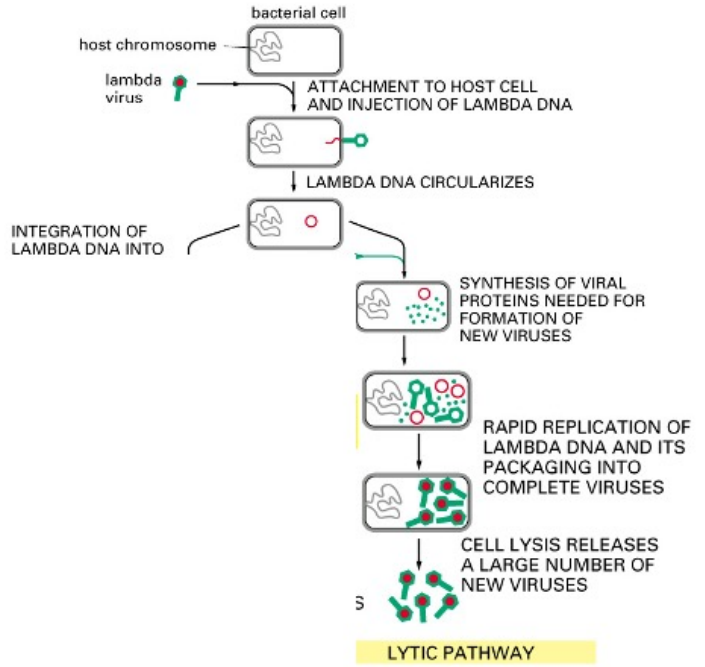
Salt concentration →

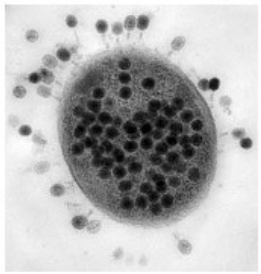
← *Sliding length*



These experiments were done by André Lwoff (Institut Pasteur, Nobel Medicine 1965) in the late 40s.

The life cycle of λ phage



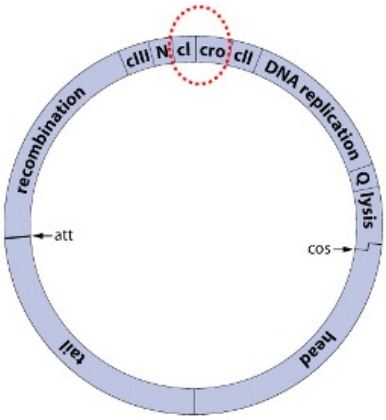


Roger Hendrix

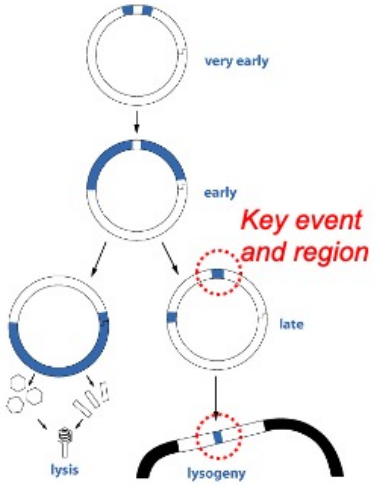
Figure 5–81 part 2 of 2. Molecular Biology of the Cell, 4th Edition.

What is the molecular nature of the switch between lysis and lysogeny?

The λ genome and gene-expression programme



A Genetic Switch, 3rd edition, 2004
© Cold Spring Harbor Laboratory Press
Chapter 3, Figure 1



A Genetic Switch, 3rd edition, 2004
© Cold Spring Harbor Laboratory Press
Chapter 3, Figure 3

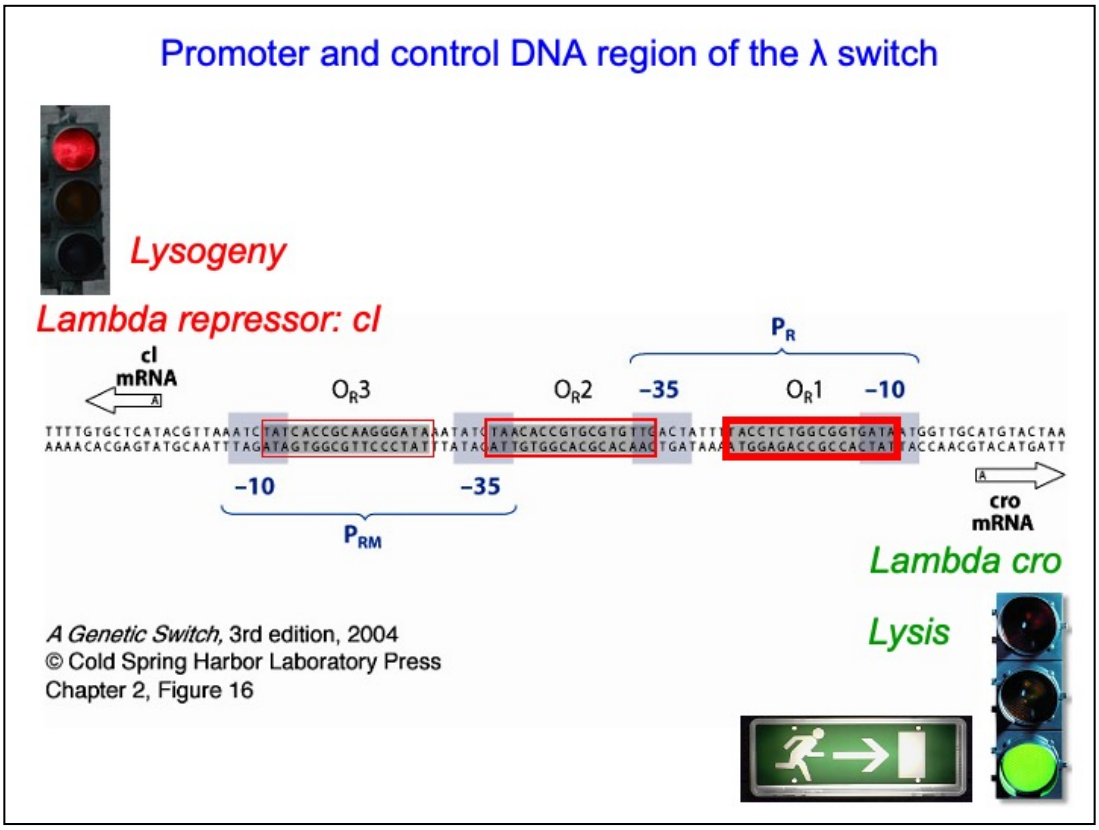
For the lytic pathway:

1—expression of proteins antiterminator N and cro

2—protein N helps to transcribe recombination and DNA replication genes

3—"parts" expressed and assembled

4--Lysis



Lambda repressor can bind to 3 operator regions with the preference of $O_R1 > O_R2 > O_R3$.

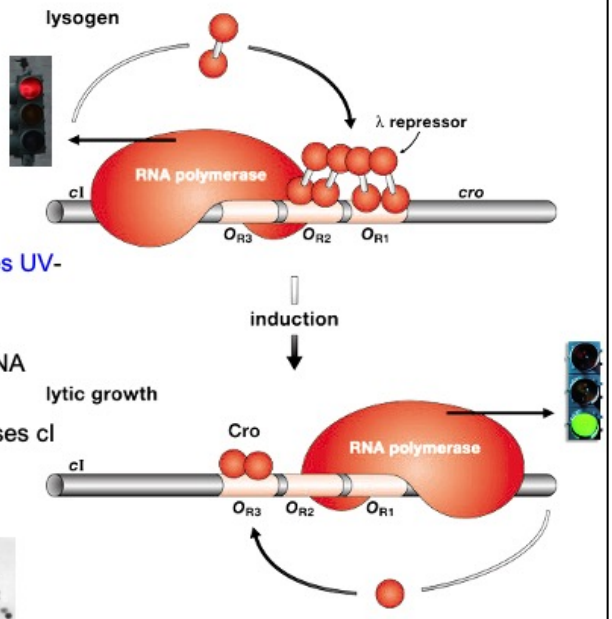
When the repressor binds at O_R1, it stops transcription of the protein lambda cro.

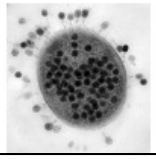
When the repressor at O_R2, it activates transcription of its own gene.

When the levels of Lambda repressor are high, then it also fills the O_R3 site, which shuts down the expression of its own gene.

Flipping the switch

- In lysogeny (prophage state):
  - λ repressor (cI) activates its own expression
  - all other genes shut down
- Upon induction:
  - λ repressor cleaved by RecA (senses UV-induced DNA damage)
  - cleaved λ repressors cannot bind DNA
  - λ cro activates its own gene, represses cI
  - expression, assembly of λ parts
  - lysis, "escape"





Genes and Signals, © 2002 by Cold Spring Harbor Laboratory Press, Chapter 1, Figure 13. (Modified, with permission, from Blackwell Science http://www.blacksci.co.uk/.)

RecA cleaves repressor monomers; these cannot dimerise efficiently, and cannot bind DNA tightly any longer.

Synthetic regulatory networks

- Can we design networks with complex temporal patterns of expression?
- Example: repressilator

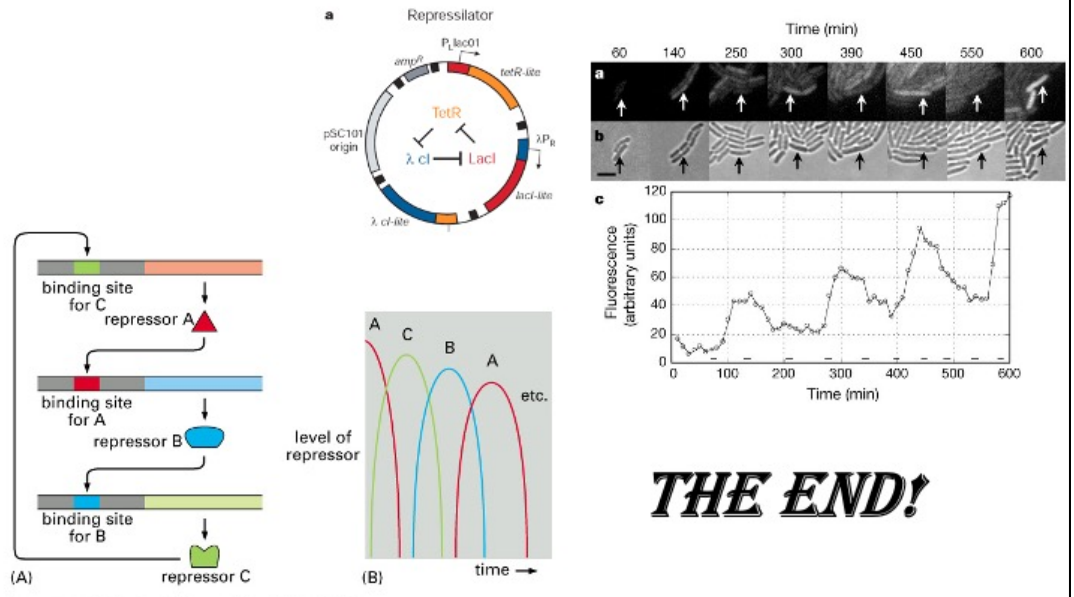


Figure 7–69. Molecular Biology of the Cell, 4th Edition.

M. B. Elowitz, and S. Leibler (2000) *Nature 403:335–338*

The last slide gives you a flavour of gene-regulation topics (feedback loops, oscillating systems etc) that will be covered in Dr Ard Louis lectures.

Best understanding of gene regulation: build artificial networks and predict their behaviour! This is one of the active areas of research in the new fields of Systems Biology and Synthetic Biology (more on this by Ard Louis).

Repressilator: a gene regulatory network that leads to regular oscillation of expressed genes