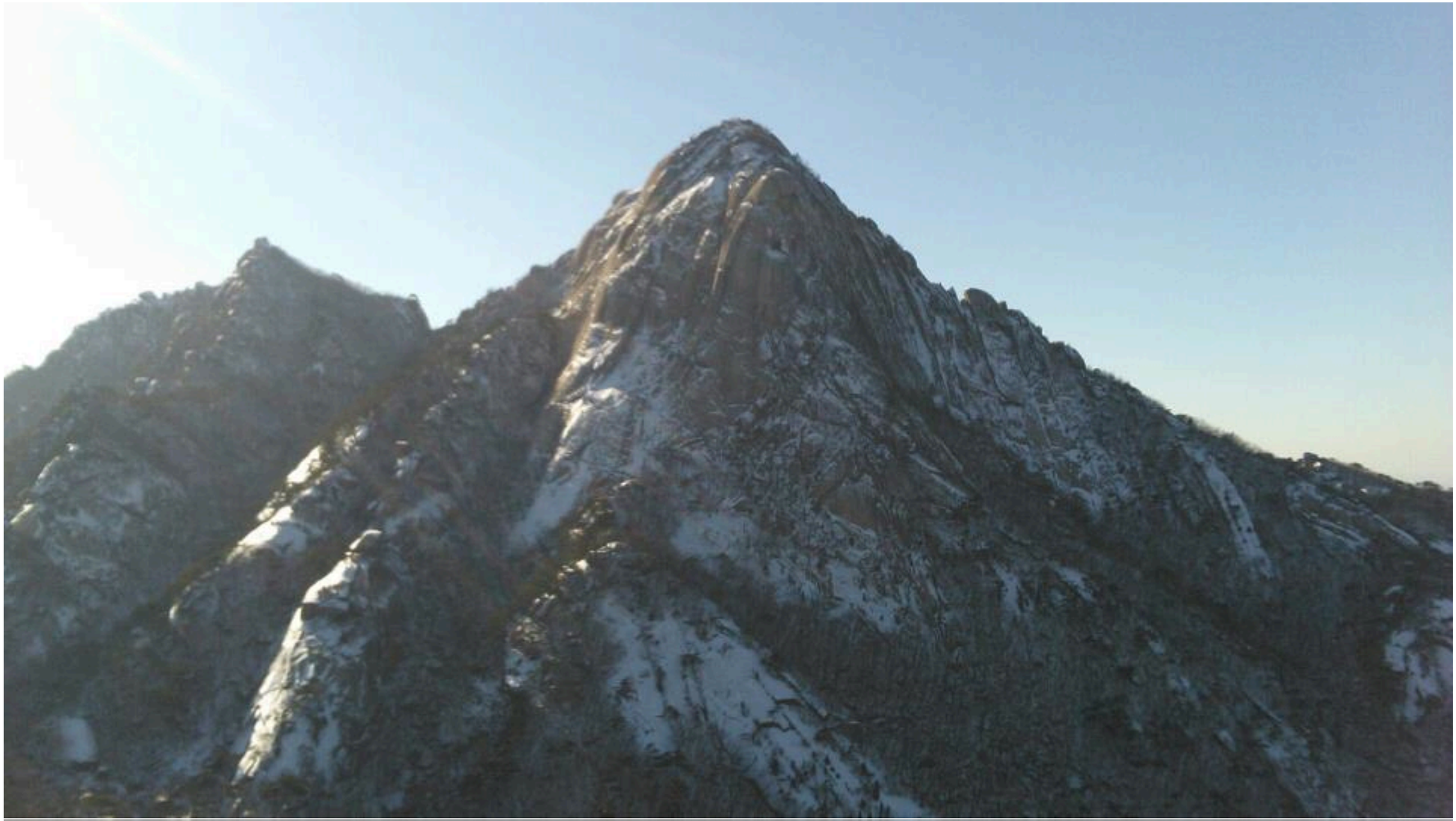


The 12th KIAS protein folding winter school

High1 resort, Jan 21-25, 2013

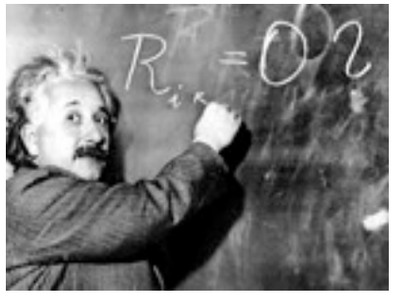
Eighteenth-century science, following the Newtonian revolution, has been characterized as developing the sciences of organized simplicity, 19th century science, via statistical mechanics, as focusing on disorganized complexity, and 20th- and 21st-century science as confronting organized complexity.

Nowhere is this confrontation so stark as in biology.
Nowhere are new conceptual tools so deeply needed.

“Origin of Order” Stuart Kaufman

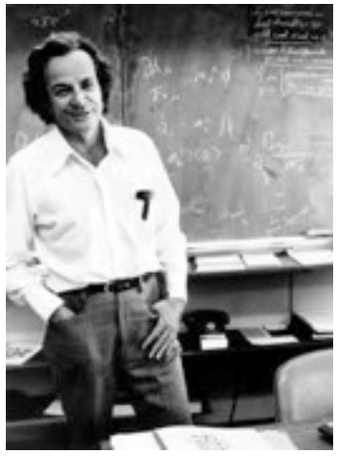












The goal of this lecture: get a global view of Biology

Outline

- Biology by numbers
- Genome
- Gene expression, genetic switch
- Foldings and functions of RNA

1. Biology by numbers

Physical Biology of the Cell by R. Philips et al.

http://bionumbers.hms.harvard.edu

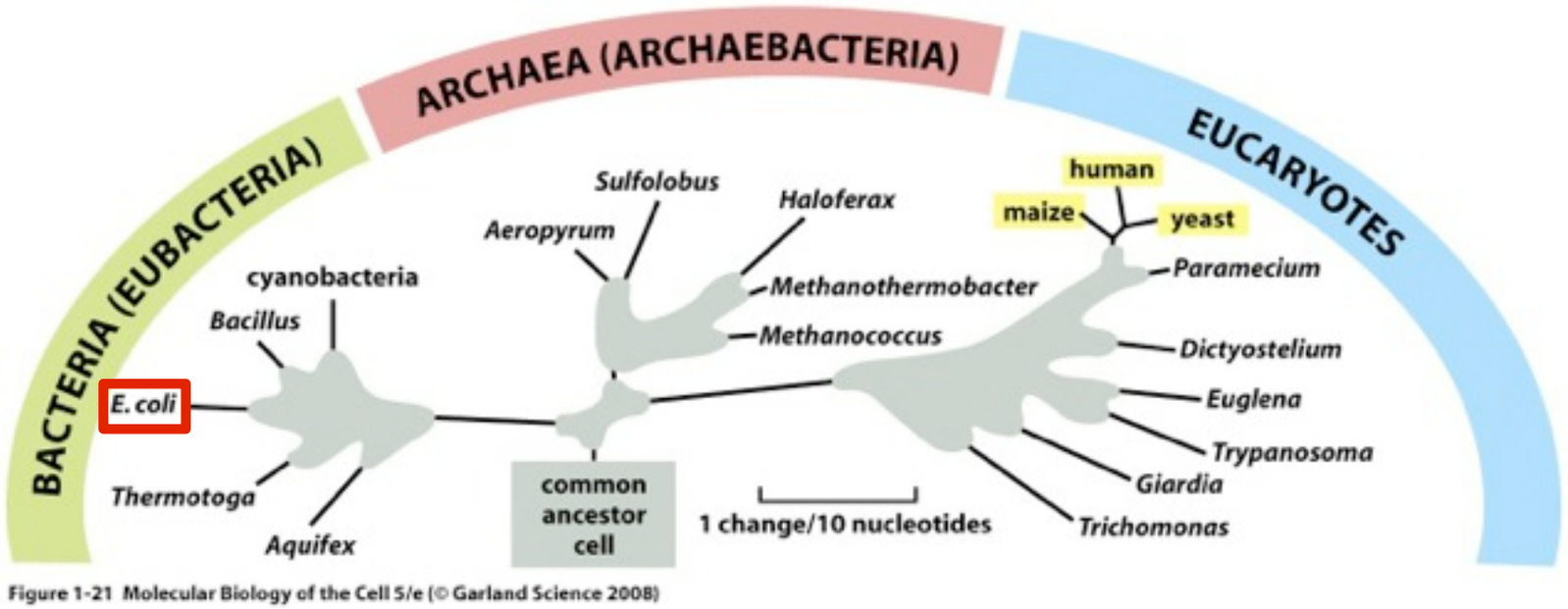
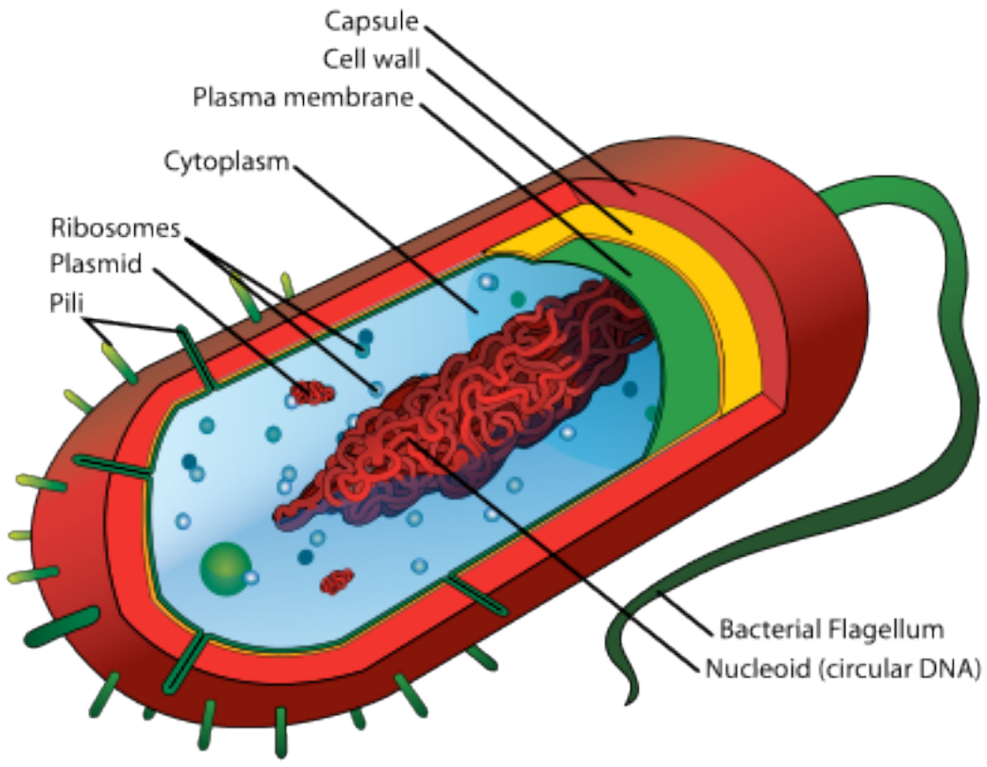
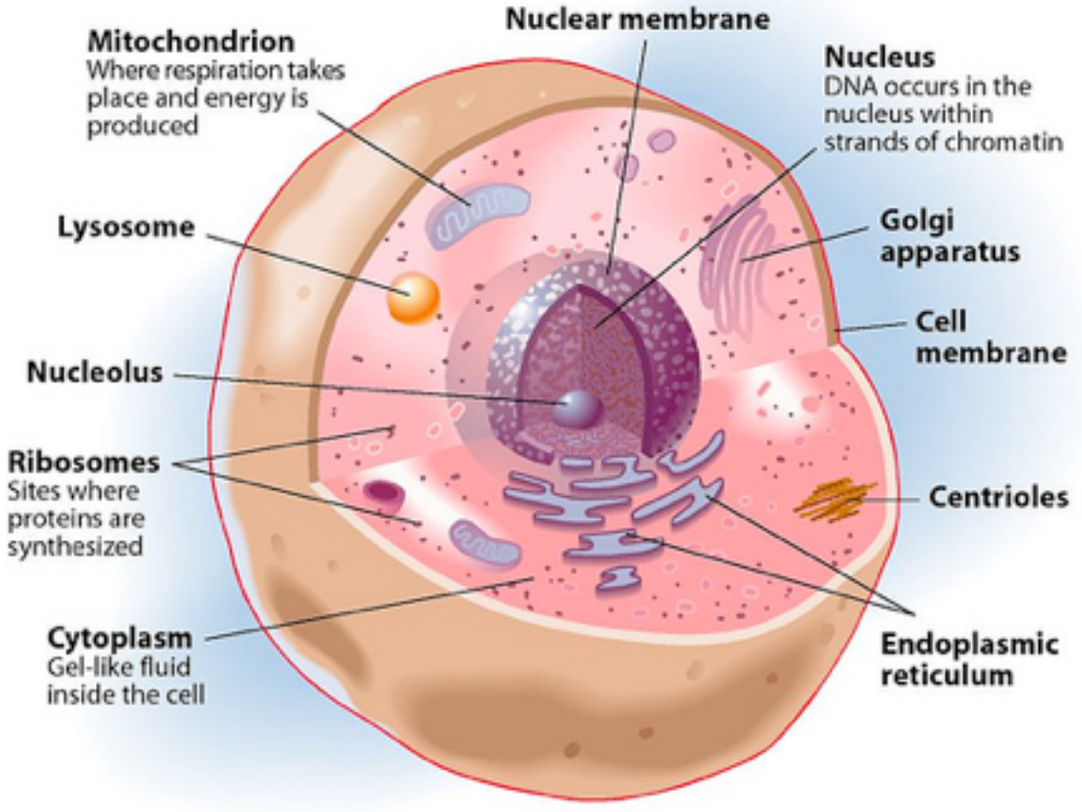


Figure 1-21 Molecular Biology of the Cell 5/e (© Garland Science 2008)



bacteria
(prokaryote)



eukaryote

$$V_{E.Coli} \approx 1\mu m^3 = 1fL$$

$$m_{E.Coli} \approx 1pg$$

$$\rho_{E.Coli} \approx 1g/mL \approx \rho_{water}$$

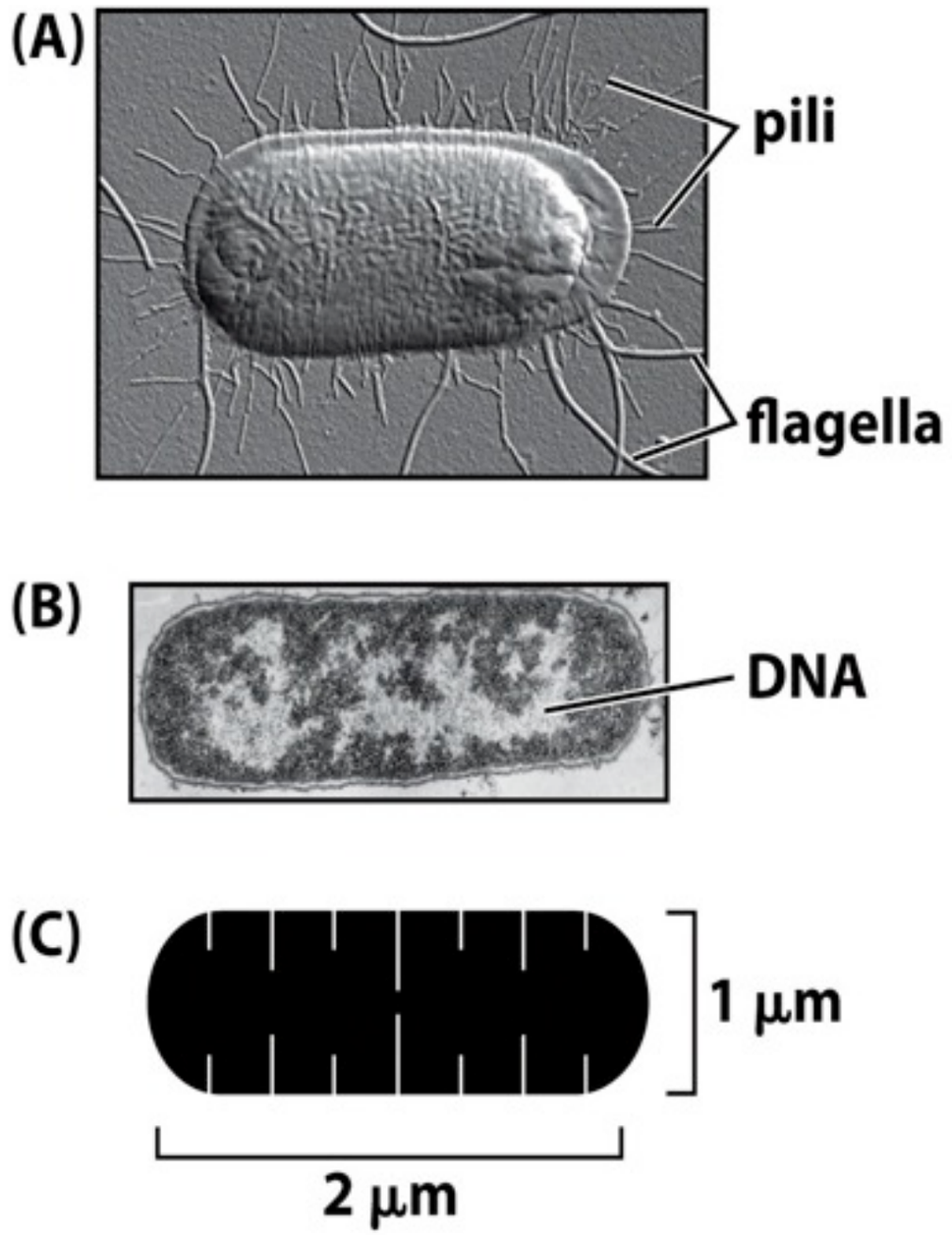


Figure 2.1 Physical Biology of the Cell (© Garland Science 2009)

$$1\,molecule/1fL = \frac{1mol}{6\times10^{23}\times10^{-15}L} \approx 2nM$$

Water is 70 % of the cell mass $(m_{E.Coli} = 1pg)$

Dry mass of the cell (30% of 1pg) = 0.3 pg

Half of the dry mass (=0.15 pg) = proteins

1 Da = mass of a hydrogen atom = $\frac{1}{6\times10^{23}}g = 1.6\times10^{-24}\,g$

1 amino acid = 100 Da

average protein size (E. coli) = 300 a.a. $\rightarrow$ 30,000 Da

$$N_{prot} = \frac{0.15\,pg}{30{,}000\,Da} = \frac{0.15\,pg}{30{,}000\times1.6\times10^{-24}\,g} \approx 3\times10^{6}$$

cf. average copy number of proteins: 167

$$\begin{cases} \frac{1}{3}N_{prot} \rightarrow \text{membrane proteins} \\ \frac{2}{3}N_{prot} \rightarrow \boxed{\text{cytoplasmic proteins}} \end{cases}$$

$m_{water} = 0.7\,pg$

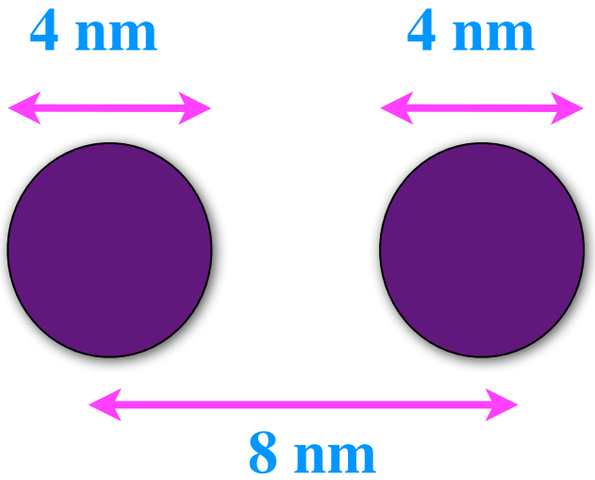
$\rightarrow$

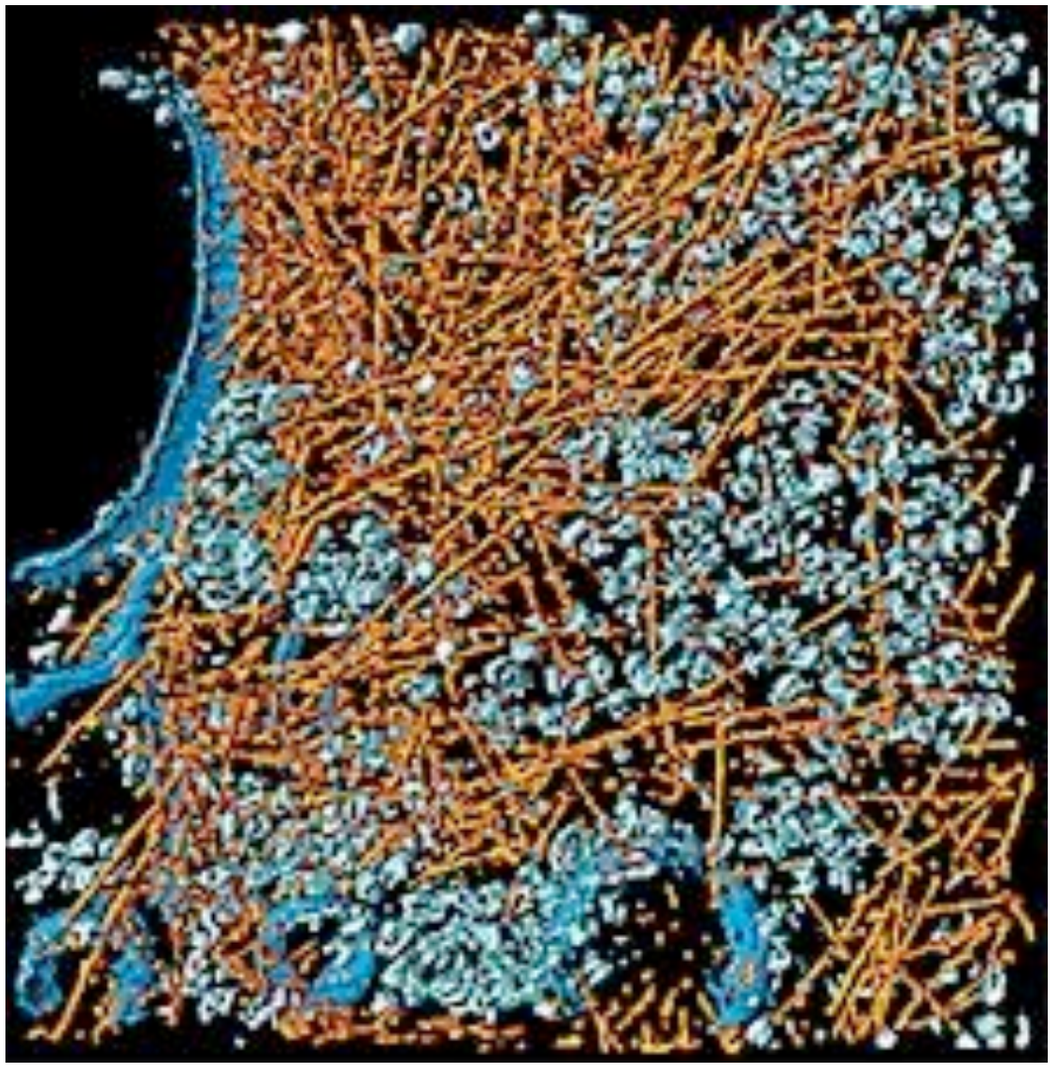
$$N_{water} = \frac{0.7\,pg}{18\,Da\times(1.6\times10^{-24}\,g/Da)} \approx 2\times10^{10}$$

There are 2×10^6 proteins in cytoplasm

$$c_{prot} = 2 \times 10^6 \, / 1 \mu m^3$$

$$d_{prot-prot} = c^{-1/3} = \left(\frac{\left(10^3 nm\right)^3}{2 \times 10^6} \right)^{1/3} = \left(500 nm\right)^{1/3} \approx 8 nm$$



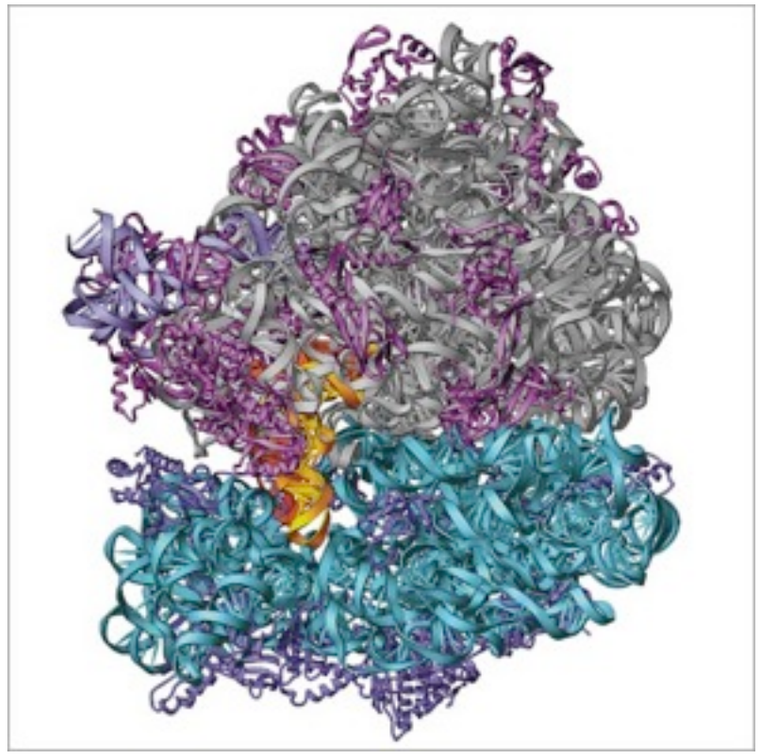


RIBOSOME

20% of the protein complement of a cell = ribosomal proteins

Ribosome (70S) = Large subunit + small subunit
Large subunit (50S) = 23S r-RNA + r-proteins
Small subunit (30S) = 16S r-RNA + r-proteins

S : Svedberg constant (sedimentation constant)
- A heavier particle sediments faster in the centrifugation, thus have a larger S value.



ribosome (2.5MDa) $\begin{cases} r-RNA: & 2/3 \text{ of the mass} \\ r-protein: & 1/3 \text{ of the mass} \end{cases}$

$$m_{r-protein} = 830{,}000\,Da$$
$$M_{r-protein} = 20\% \times M_{protein}$$

$$\therefore N_{ribosome} = \frac{M_{r-prot}}{m_{r-prot}} = \frac{0.2 \times 0.15\,pg}{830{,}000\,Da} = 20{,}000$$

$$\approx 19{,}000$$

yeast cell : model system to study a single eukaryote cell

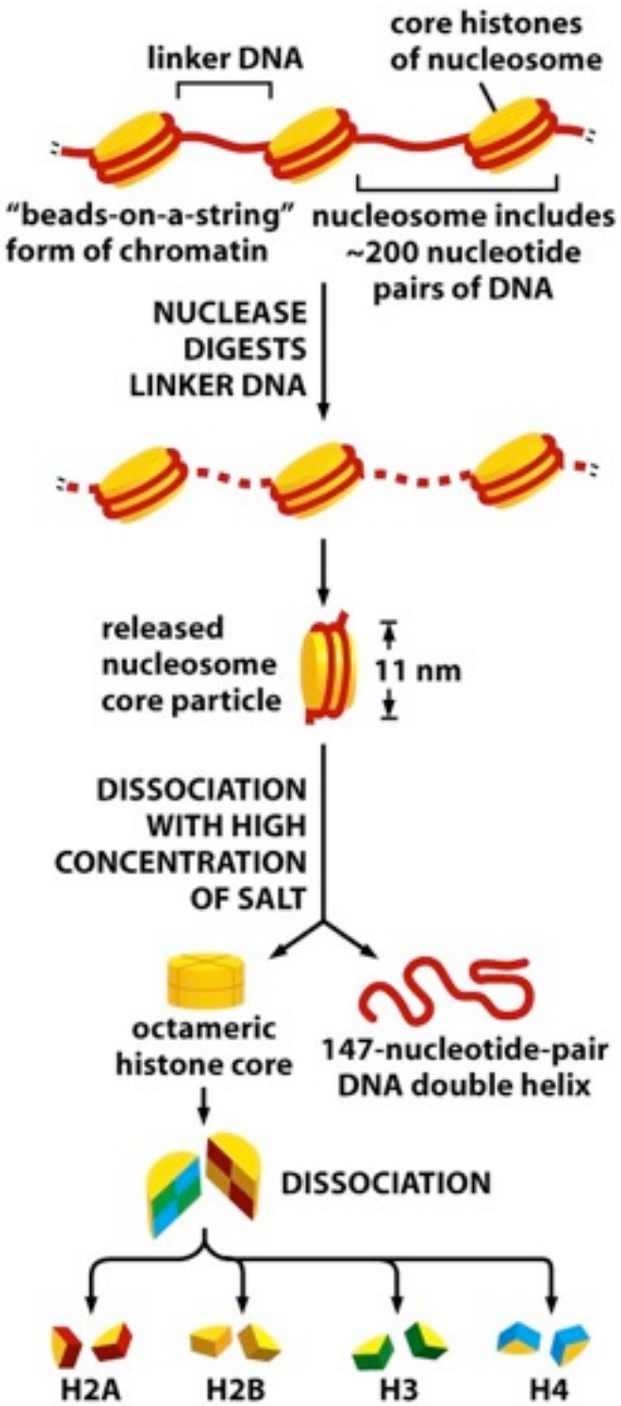


Figure 4-23 Molecular Biology of the Cell 5/e (© Garland Science 2008)

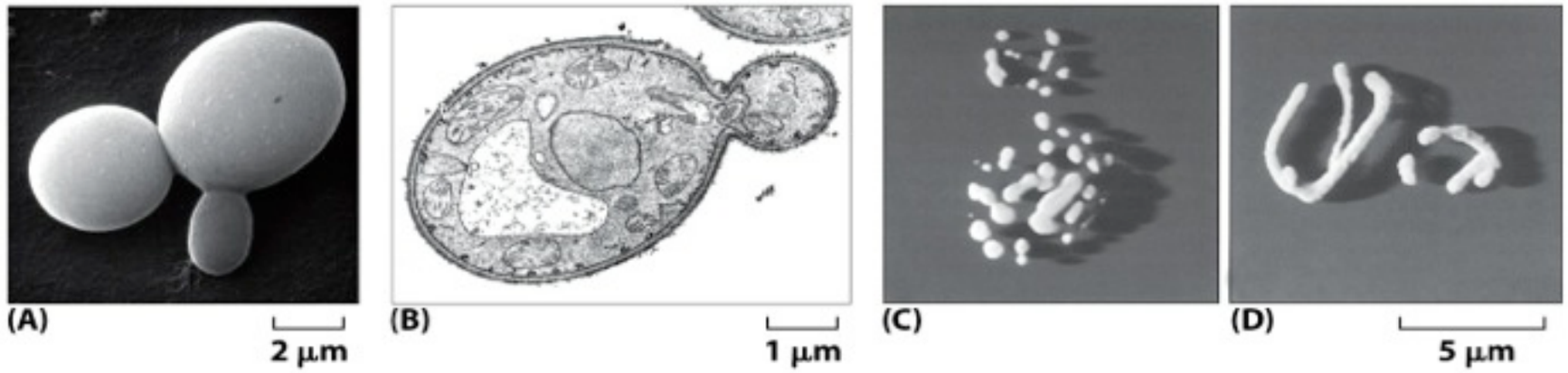


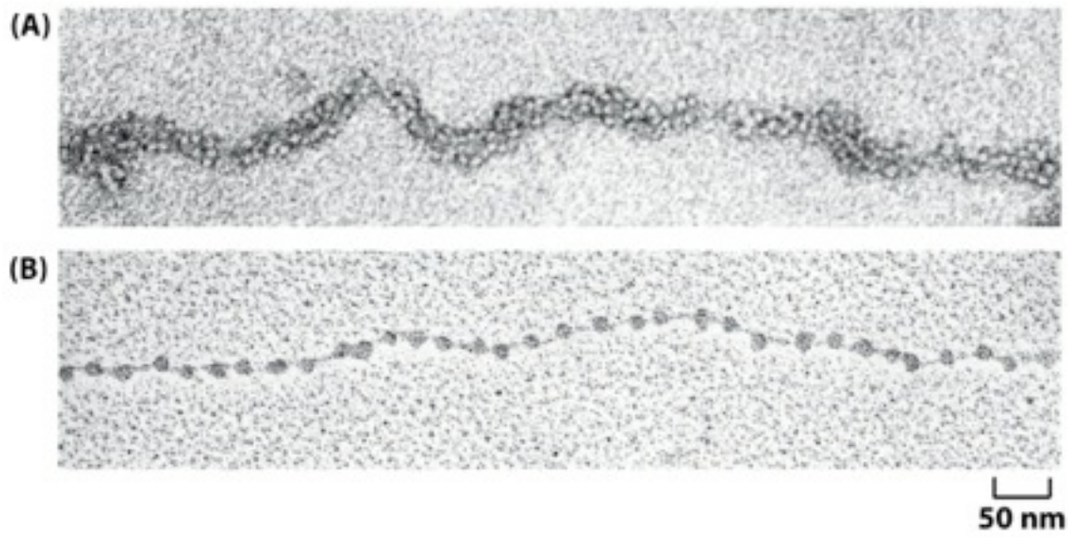
Figure 2.9 Physical Biology of the Cell (© Garland Science 2009)

$$L_{yeast} \sim 5\mu m \qquad V_{yeast} = \frac{4}{3}\pi\left(2.5\mu m\right)^3 \approx 60\mu m^3 \approx 60V_{E.Coli}$$

$$N_{protein}^{Yeast} \approx 60 \times N_{protein}^{E.Coli}$$

$$N_{lipid}^{Yeast} \approx \frac{2 \times 0.5 \times \left(80\mu m^2\right)}{0.5nm^2} \approx 2 \times 10^8$$

$$N_{genome} \sim 1.2 \times 10^7 bp \qquad N_{nucleosome} \simeq \frac{1.2 \times 10^7 bp}{200bp\ /\ nucleosome} = 60{,}000$$



Cells

Important model organisms in Biology

- Bacteriophage: Gene structure, gene regulation
- *E. coli* : Genetic network (lac operon), flagella
- Yeast (*S. cerevisiae*) :
  - Cell cycle, genetics, cell biology
  - quick and easy to grow
- Flies (*Drosophila melanogaster*)
  - Easy to grow, rapid generations
  - Easy mutation induction, many observable mutation
  - Giant chromosome in salivary glands that can be examined under a light microscope
  - Molecular genetics, Population genetics, Developmental biology
- *C. elegans*
  - Excellent model of the genetic control of development and physiology
  - The first multicellular organism whose genome was completely sequenced.
- Mice
  - Genome size and organization is very similar to human's
  - Small, easy to breed, ideal for lab study.
  - Transgenic, knock-out mouse.
- Human
  - cf) HeLa (**He**nrietta **La**cks) cell
    - A cell type in an immortal cell line used in scientific research
    - Oldest and most commonly used human cell line.
    - Incredibly hardy, which makes them useful for medical research
    - The line derived from cervical cancer cells of a patient taken on 1951
    - No aging, no programmed cell death (apoptosis)
- and others (Giant squid) ....

Bio“polymers”

- Cells are made of biopolymers (DNA, RNA, proteins, cytoskeletal filaments, ...)
- Length : $L = N \times a$
- Size : $R_g \sim N^{\nu}$
- Flexibility (persistence length)

$$\left\langle \vec{u}(s) \cdot \vec{u}(s') \right\rangle = e^{-|s-s'|/l_p}$$

$$\left\langle R^2 \right\rangle = \int_0^L ds \int_0^L ds' e^{-|s-s'|/l_p} = 2Ll_p \left[1 - \frac{l_p}{L}\left(1 - e^{-L/l_p}\right) \right]$$

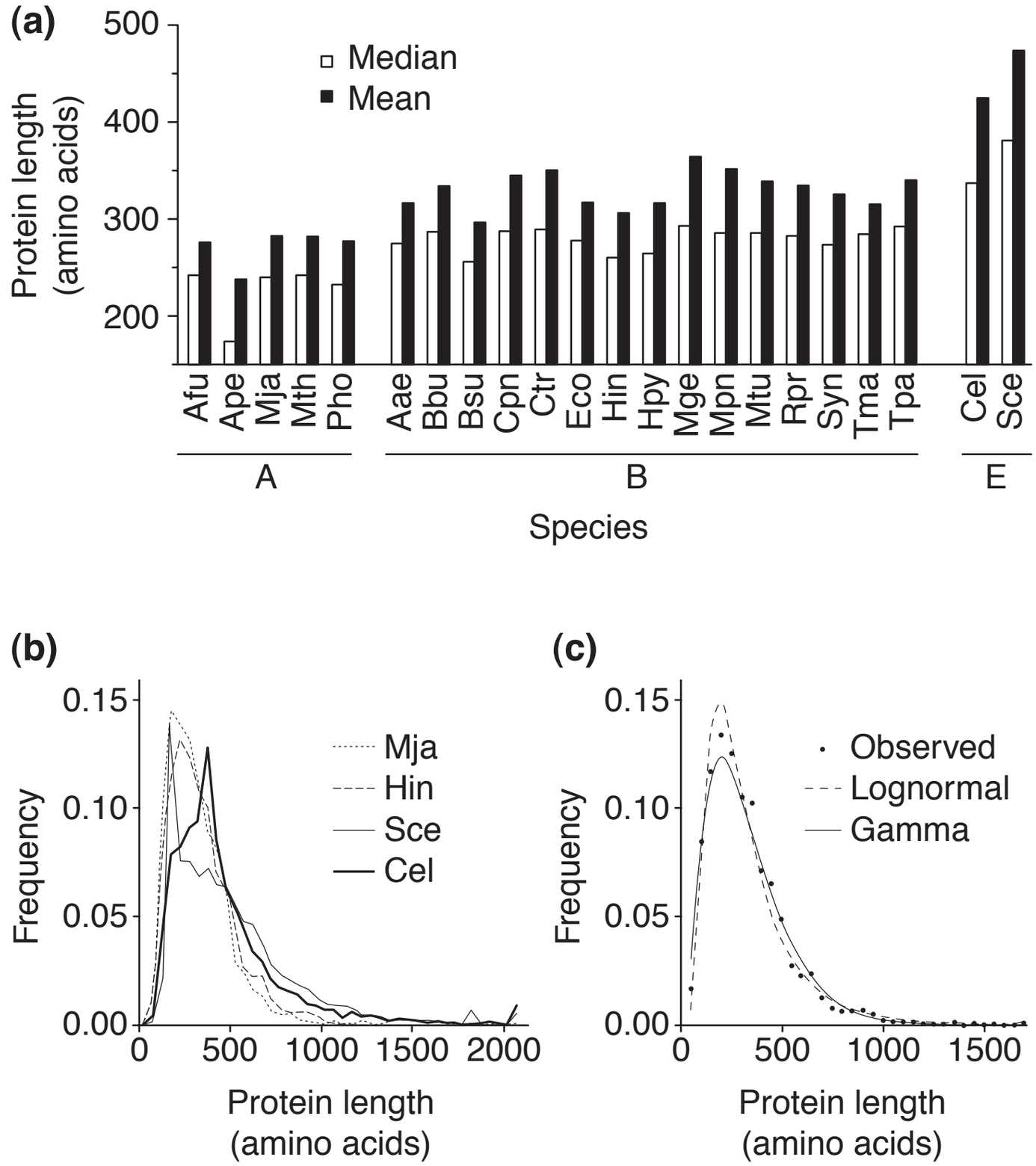
Length (contour length)

$$L = N \times a$$

Proteins

cf. Largest protein size : titin
33423 aa

FIGURE 1. Distributions of protein lengths of 22 completely sequenced genomes



trends in Genetics

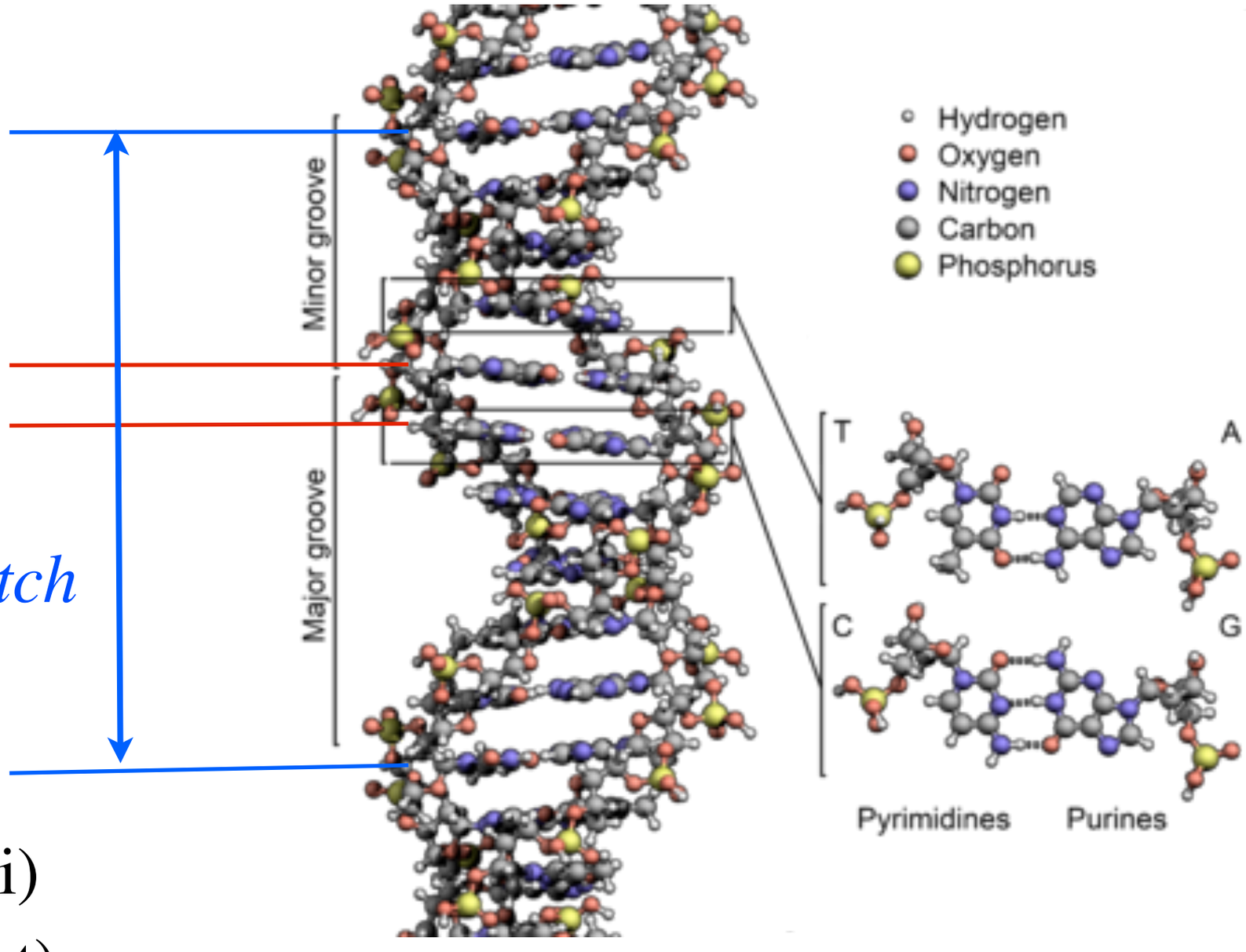
DNA

Length (contour length)

$$L = N \times a$$

$a = 0.34nm$

$10.5bp$ / *pitch*



N=5×10^6 bp (E. coli)

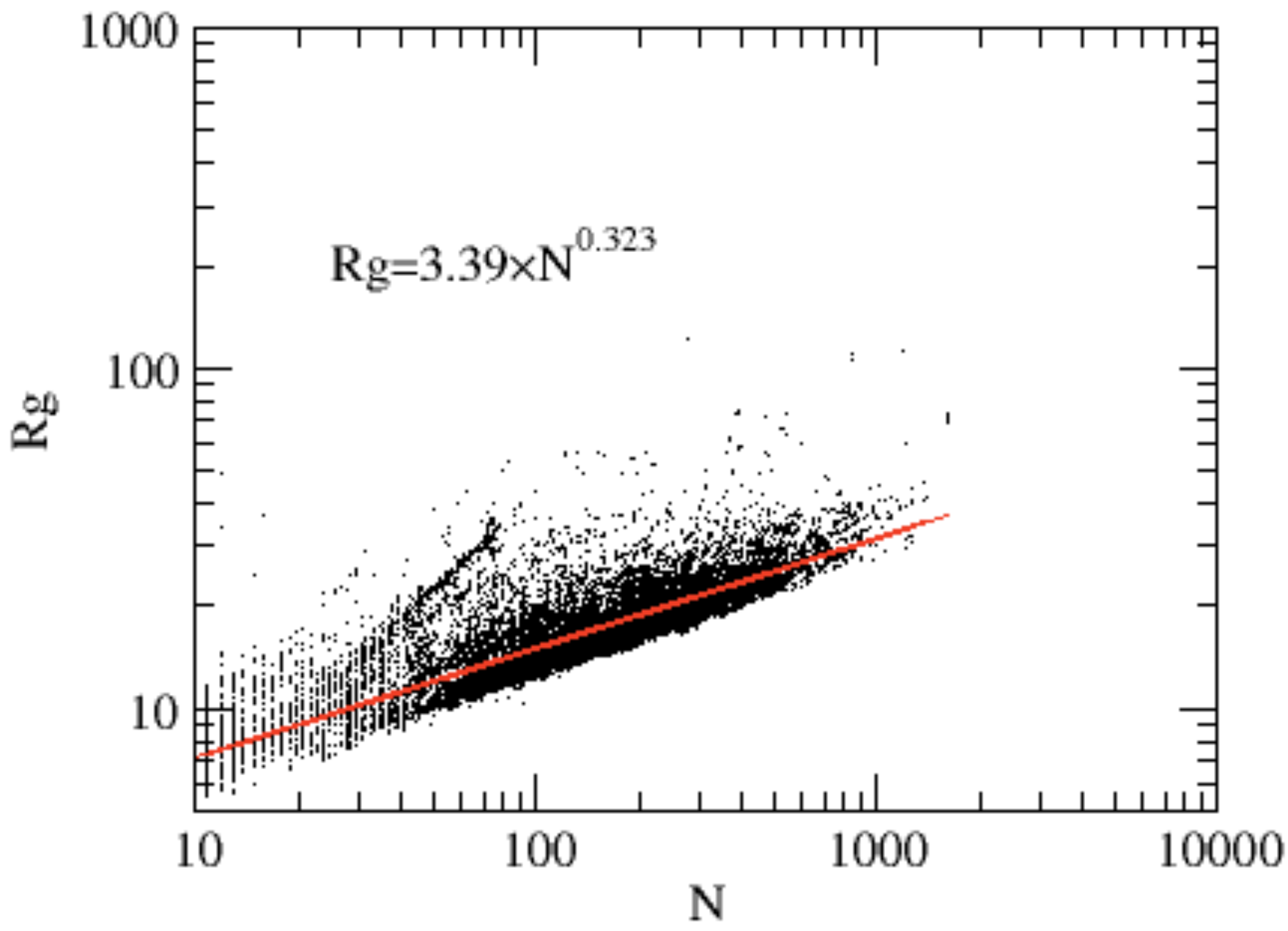
N=1.2×10^7 bp (yeast)

N=3×10^9 bp (human)

Size

$$R_g = aN^{\nu}$$

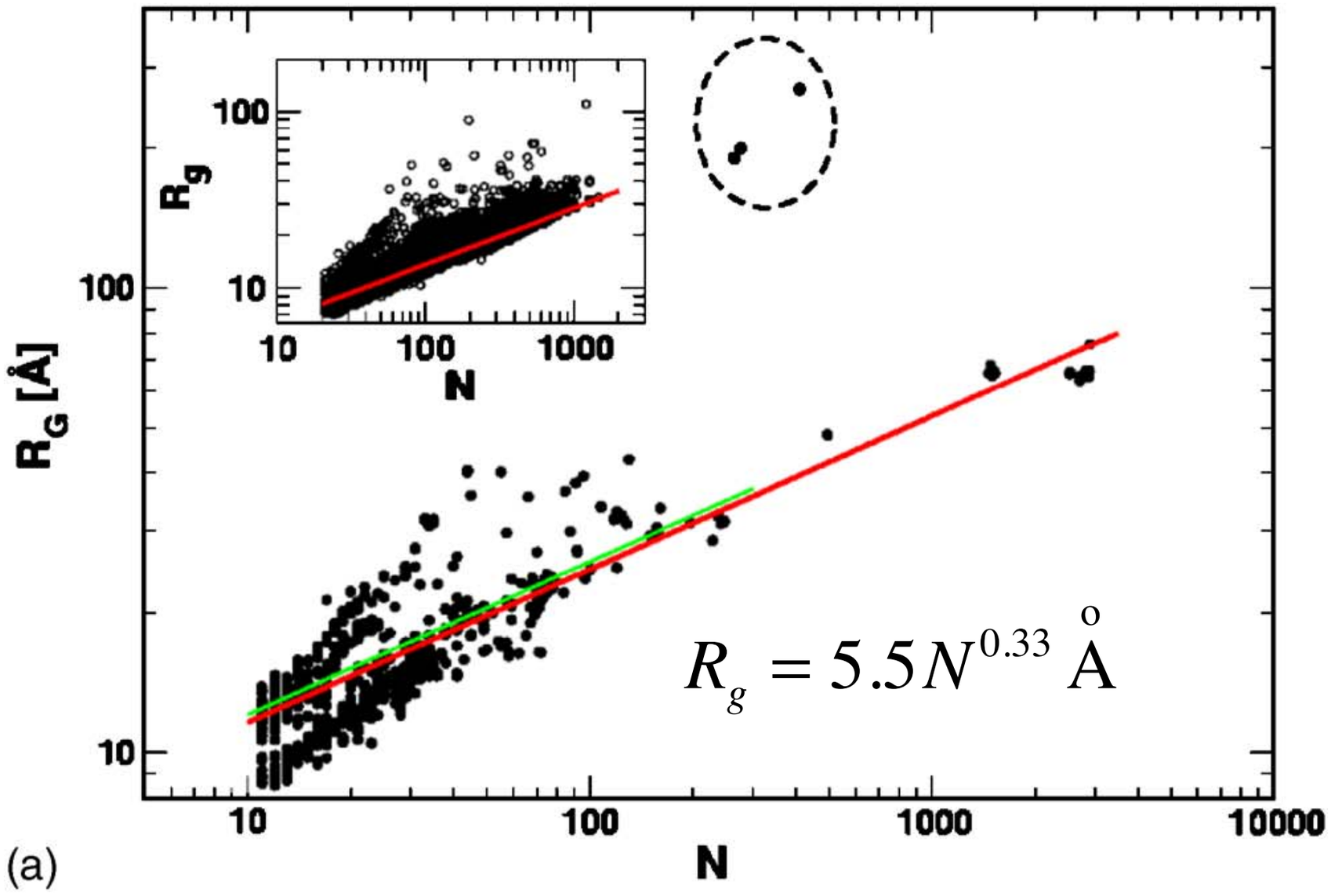
protein



Size

$$R_g = aN^{\nu}$$

RNA



Estimating the rate of biological processes

Replication

800 bp/sec

$$\frac{dN_{bp}}{dt} \approx \frac{N_{bp}}{\tau_{cell}} \approx \frac{5\times10^{6}\,bp}{3000\,\text{sec}} \approx 2000\,bp/\text{sec}$$

(or 1000bp/sec per DNA replication complex)
Biochemical study found 250-1000bp/sec range
Let's take it 800 bp/sec.

Transcription

40 nt/sec

$$40\,nt/\text{sec}/\,RNAP$$

typical length of transcript : $3nt/a.a\times300a.a\approx1{,}000nt$

The time to produce a transcript using an RNAP ~ 25 sec

Protein synthesis

15 aa/sec

$$\frac{dN_{protein}}{dt} \approx \frac{N_{protein}}{\tau_{cell}} \approx \frac{3\times10^{6}\,proteins}{3000\,\text{sec}} \approx 1000\,proteins/\text{sec}$$

$$\rightarrow \frac{(3\times10^{6})\times300a.a.}{3000\,\text{sec}} \approx 3\times10^{5}\,a.a/\text{sec}$$

protein synthesis rate per ribosome
= 3,000,000 proteins/3000 sec /20,000 ribosomes
= 0.05 proteins/sec/ribosome → 15 aa/sec/ribosome

Energy scale relevant for biomolecules

Biomolecules are “marginally” stable; thus their dynamics are susceptible to thermal fluctuations, external noises.

$1k_BT = 4.14$ pN×nm $= 4.14\times10^{-21}$ J $= 0.59$ kcal/mol at 300 K

$1eV \approx 38.7k_BT \sim 40k_BT$

The characteristics of “marginal stability” allow biomolecules to adapt their conformational states into, if any, an alternative form.

However, for some molecules or as the size of molecules grows the thermal fluctuation itself is not sufficient to trigger the conformational transitions within a biologically relevant time scale. In this case, various cofactors, NTPs, or interaction with other molecular machines are used as free energy sources (active process).

Binding Free Energy

$$AB \rightleftharpoons A + B$$

Thermal energy scale (at 300K)

$$k_B T = 4.14 \times 10^{-21} J = 4.14\, pN \cdot nm$$

$$\doteq 0.59 kcal / mol$$

$$\Delta G = k_B T \log K_d = k_B T \log \frac{[A][B]}{[AB]} \approx k_B T \log 10^{-9} = -2.3 \times 9 k_B T \approx -20 k_B T$$

Typical protein-protein dissociation free energy

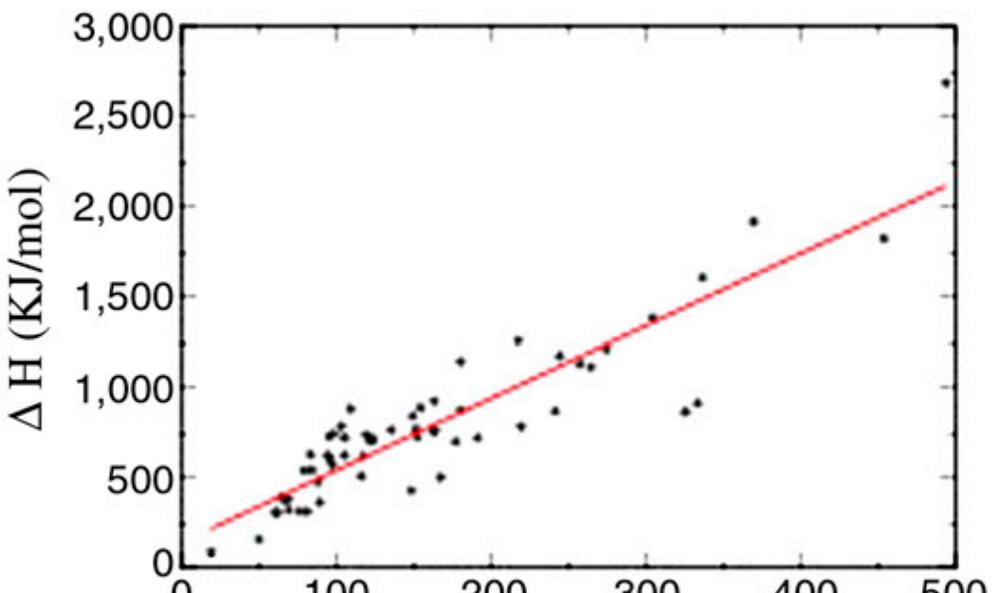
$$\approx 20 k_B T (K_d = 1nM)[1 molecule / cell]$$
$$\approx 14 k_B T (K_d = 1\mu M)[10^3 molecules / cell]$$
$$\approx 7 k_B T (K_d = 1mM)[10^6 molecules / cell]$$

Insulin dimer : 7 kcal/mol $(K_d = 10^{-5} M)$
Trypsin-PTI : 18 kcal/mol $(K_d = 10^{-13} M)$
Haemoglobin αβ dimer : > 11 kcal/mol $(K_d << 10^{-8} M)$

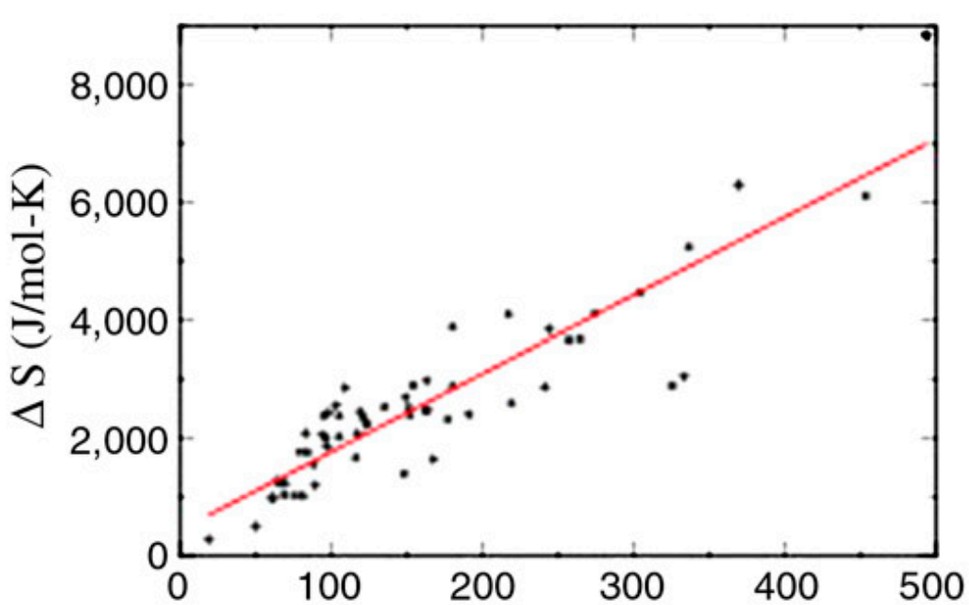
Thermodynamics of protein folding

17876–17882 | PNAS | **November 1, 2011** | vol. 108 | no. 44

Enthalpy of unfolding at 100.5 ^{0}C

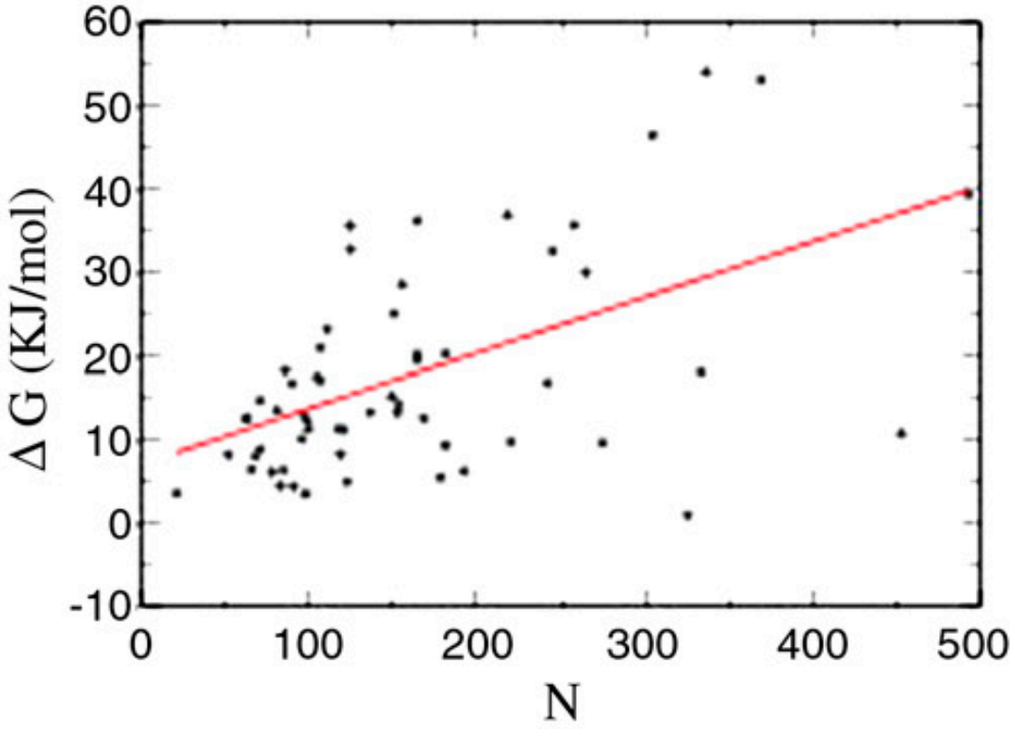


Entropy of unfolding at 120 ^{0}C



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

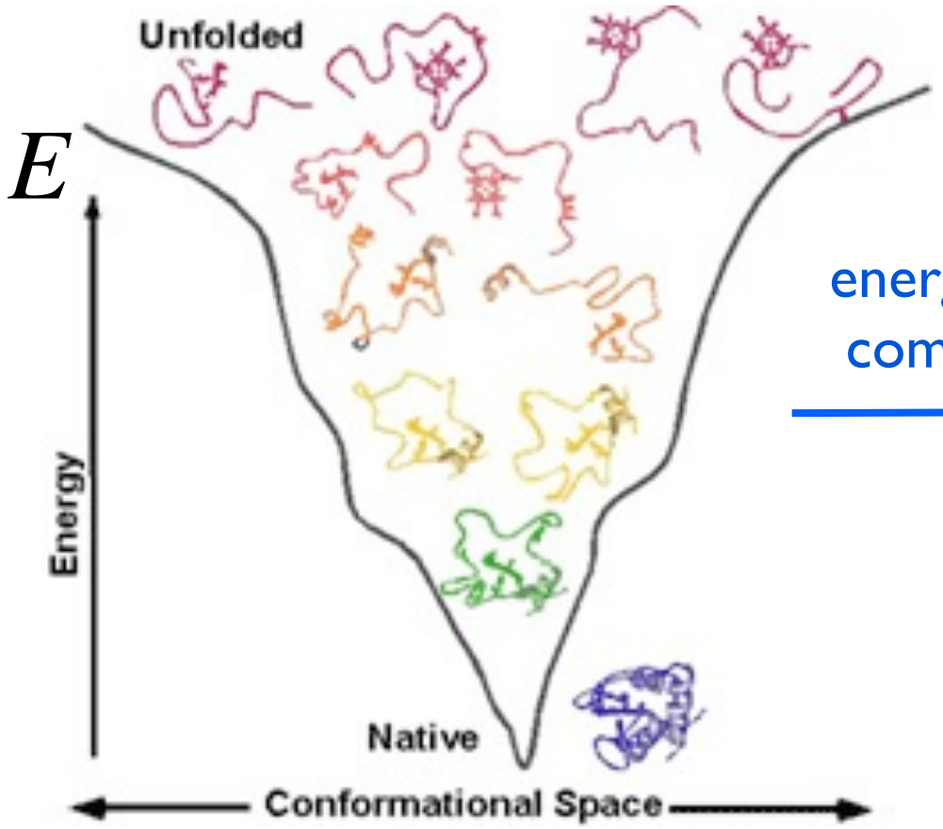
Free energy of folding at 37 ^{0}C



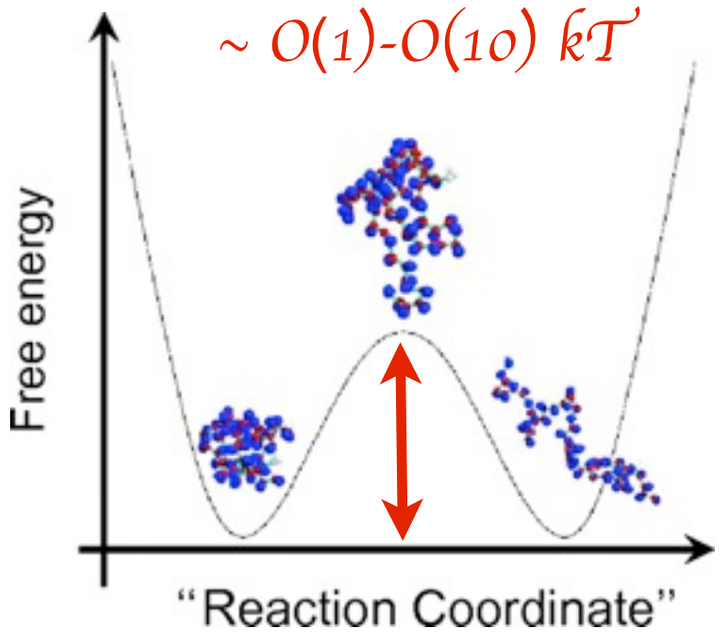
* 1kJ/mol = 0.24 kcal/mol

= 0.4 $k_B T_{300}$

$$F = E - TS$$



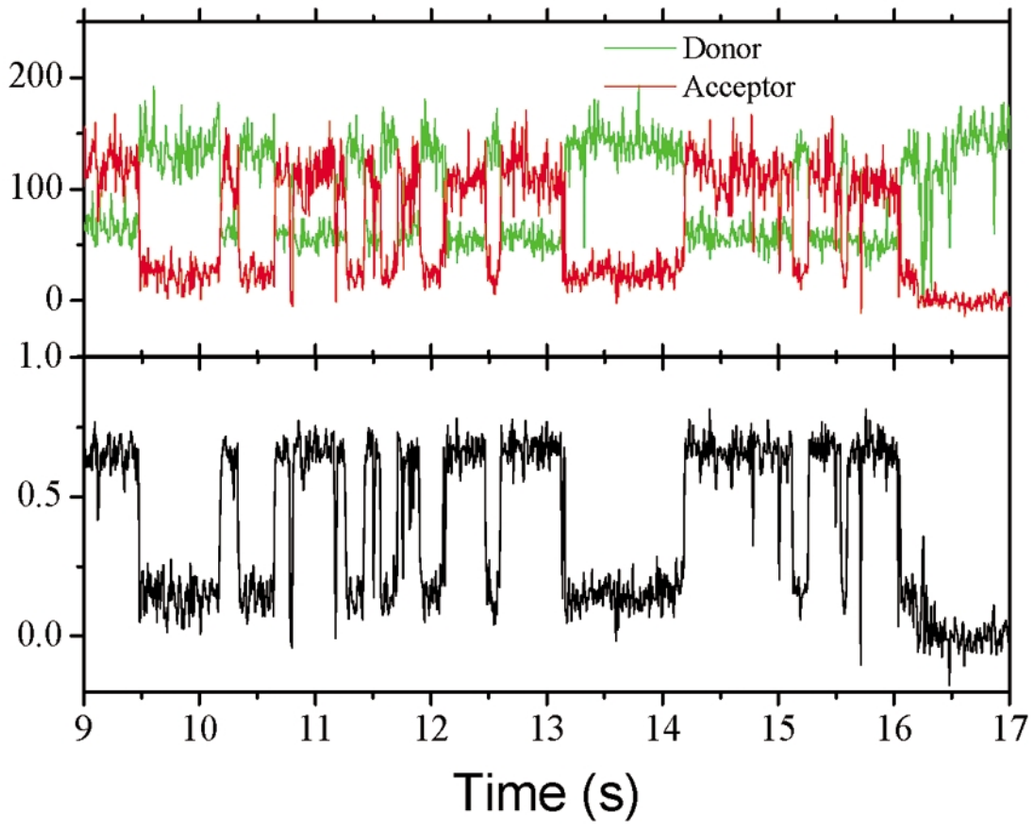
energy-entropy compensation



At transition mid-temperature $T=T_m$

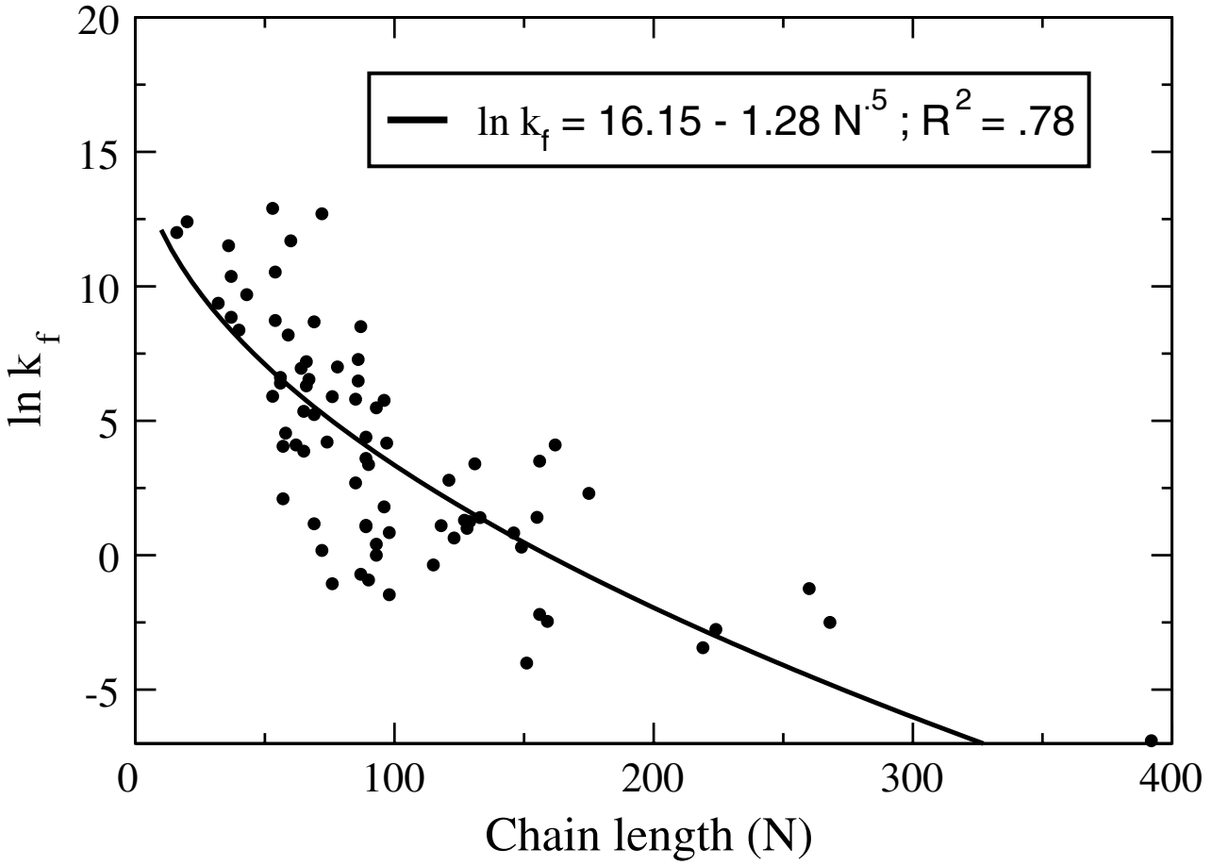
$$S(E) = k \log \Omega(E)$$

Biomolecules are *SMALL*



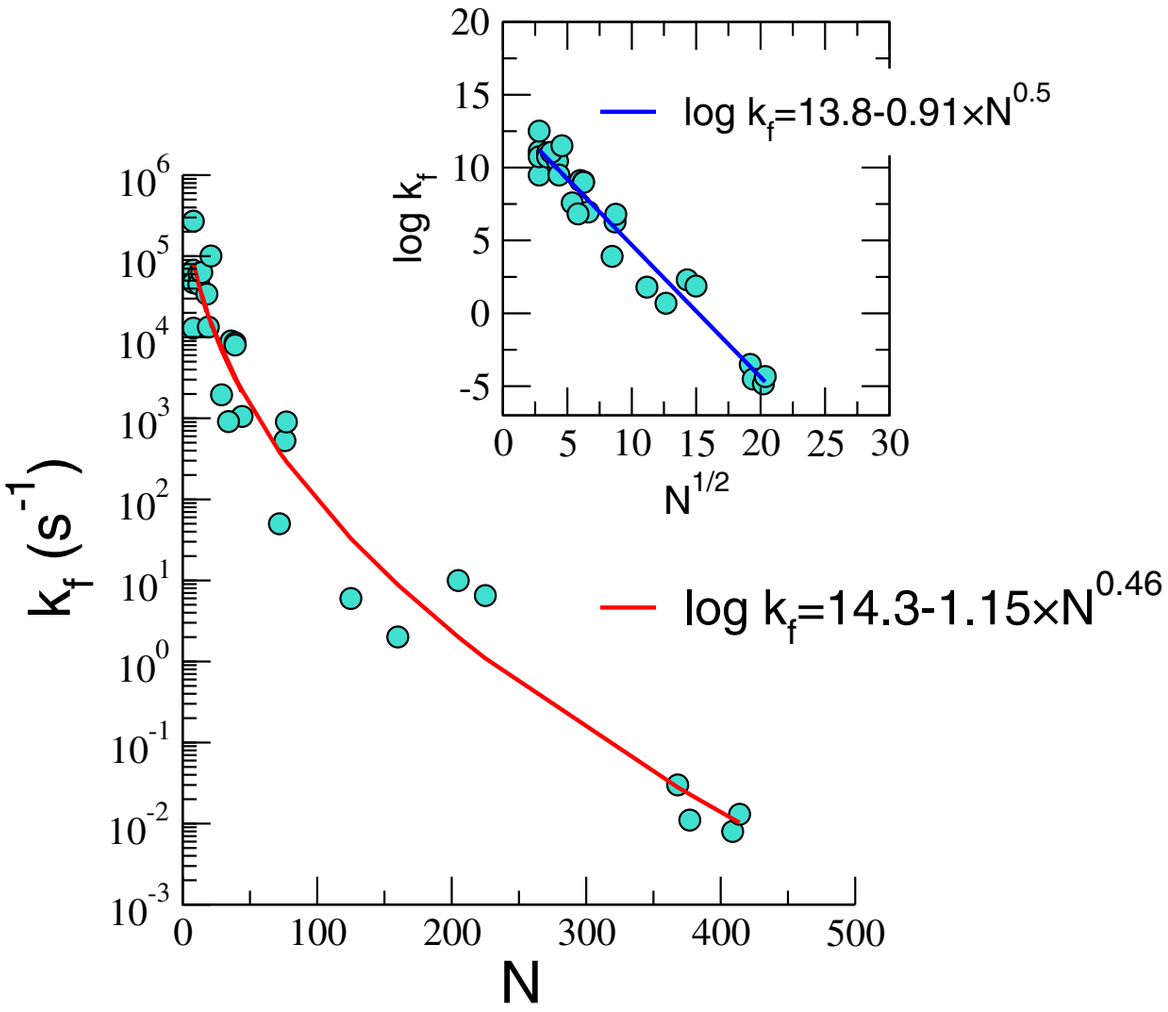
Kinetics of biopolymers

Protein folding time



17876–17882 | PNAS | November 1, 2011 | vol. 108 | no. 44

RNA folding time



Biophysical Journal Volume 102 February 2012 L11–L13

$$k_o^{-1} \simeq 0.1\mu s \qquad k_f = k_o \exp\left(-\alpha N^{\beta}\right) \qquad k_o^{-1} \simeq 1.0\mu s$$

$$\beta \approx 0.5$$

cf. Another Arrhenius-like expression, Eyring equation (or transition state theory (TST)) derived by theoretical chemists H. Eyring and M. Polanyi

$$k_{TST} = \kappa \frac{k_B T}{h} e^{-E^{\dagger}/k_B T}$$

is often used in estimating the height of kinetic barrier. However, it is of note that this equation should only be applied for chemical reactions associated with bond breaking (or formation) dynamics of simple organic (or inorganic) compounds in gas phase. The prefactor $(k_BT/h)^{-1} \simeq 0.2$ psec is the vibrational frequency of a critical molecular bond at the transition state to be ruptured by chemical dynamics. Blind use of this prefactor overestimate kinetic barriers. To describe dynamics in condensed media, including biopolymer folding and biomolecular assembly, it is more appropriate to use Kramers' equation, which correctly predicts the effect of solvent viscosity to the reaction rate.

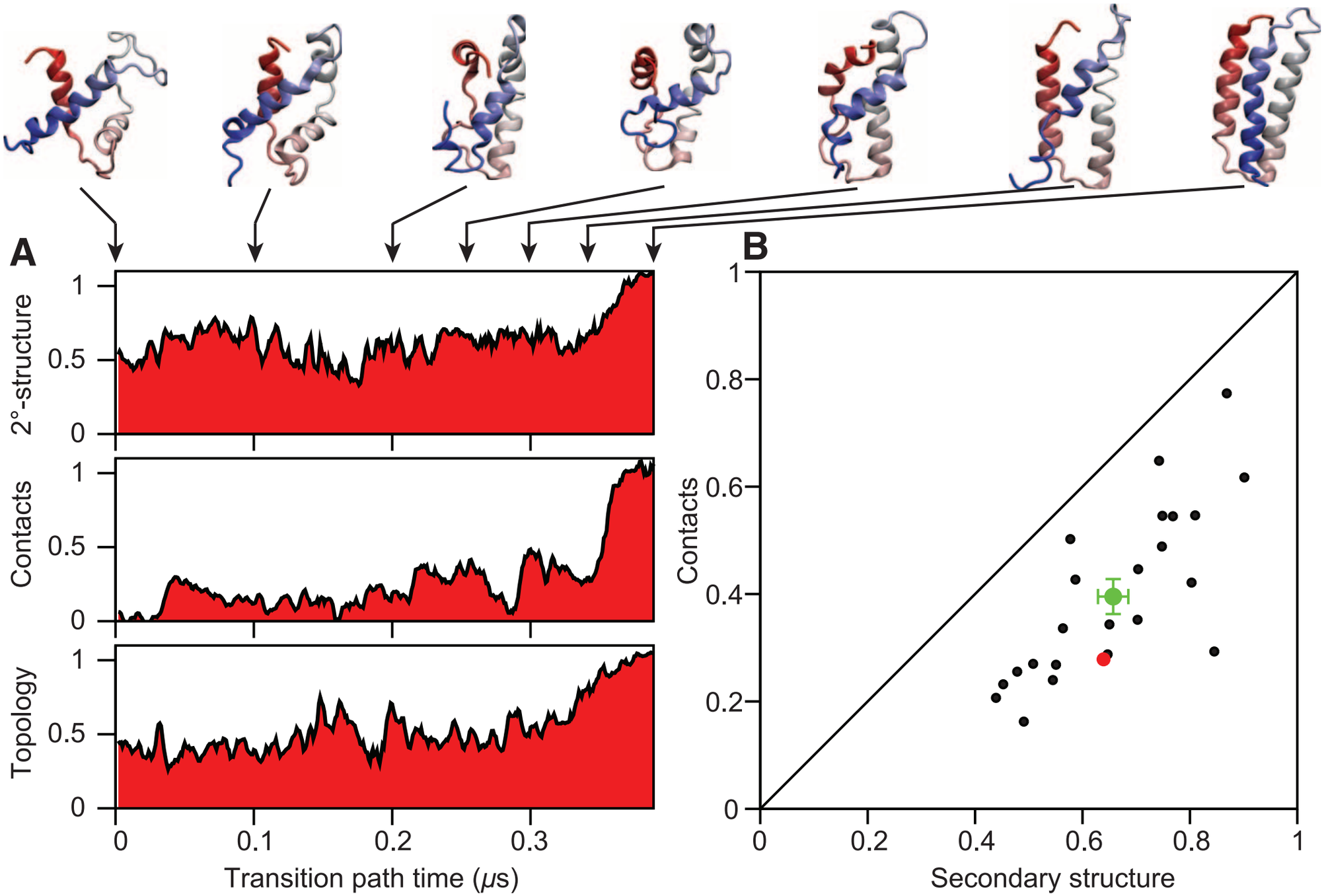
$$k_{KR} = \frac{\omega_0 \omega_b}{2\pi\gamma} e^{-\Delta G^{\dagger}/k_B T}$$

As estimated from the fits to folding time of RNA and proteins, the prefactor of Kramers' rate for RNA and proteins are ~(0.1-1) μsec. These values indeed agree with the recent experimental measurements of **"speed limit" of protein folding**, which were performed using barrierless downhill folders (Gruebele), and computational and experimental reports of **transition path time** by D. E. Shaw and W.A. Eaton and coworkers, respectively.

Transition path time $\tau_0 \sim 1$ μsec is size-independent. Fast and slow folding proteins with vastly different topologies have almost the same value.

How Fast-Folding Proteins Fold

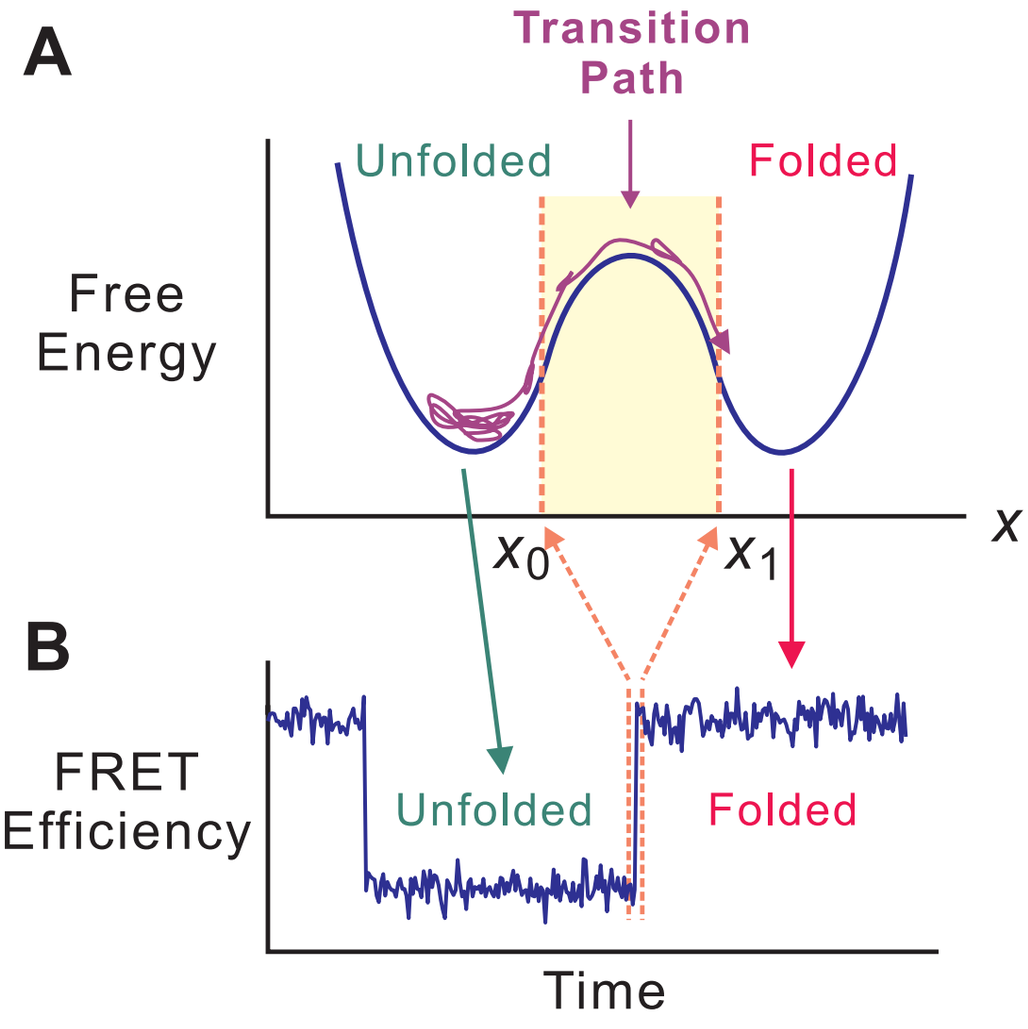
Kresten Lindorff-Larsen,[1]*† Stefano Piana,[1]*† Ron O. Dror,[1] David E. Shaw[1,2]†

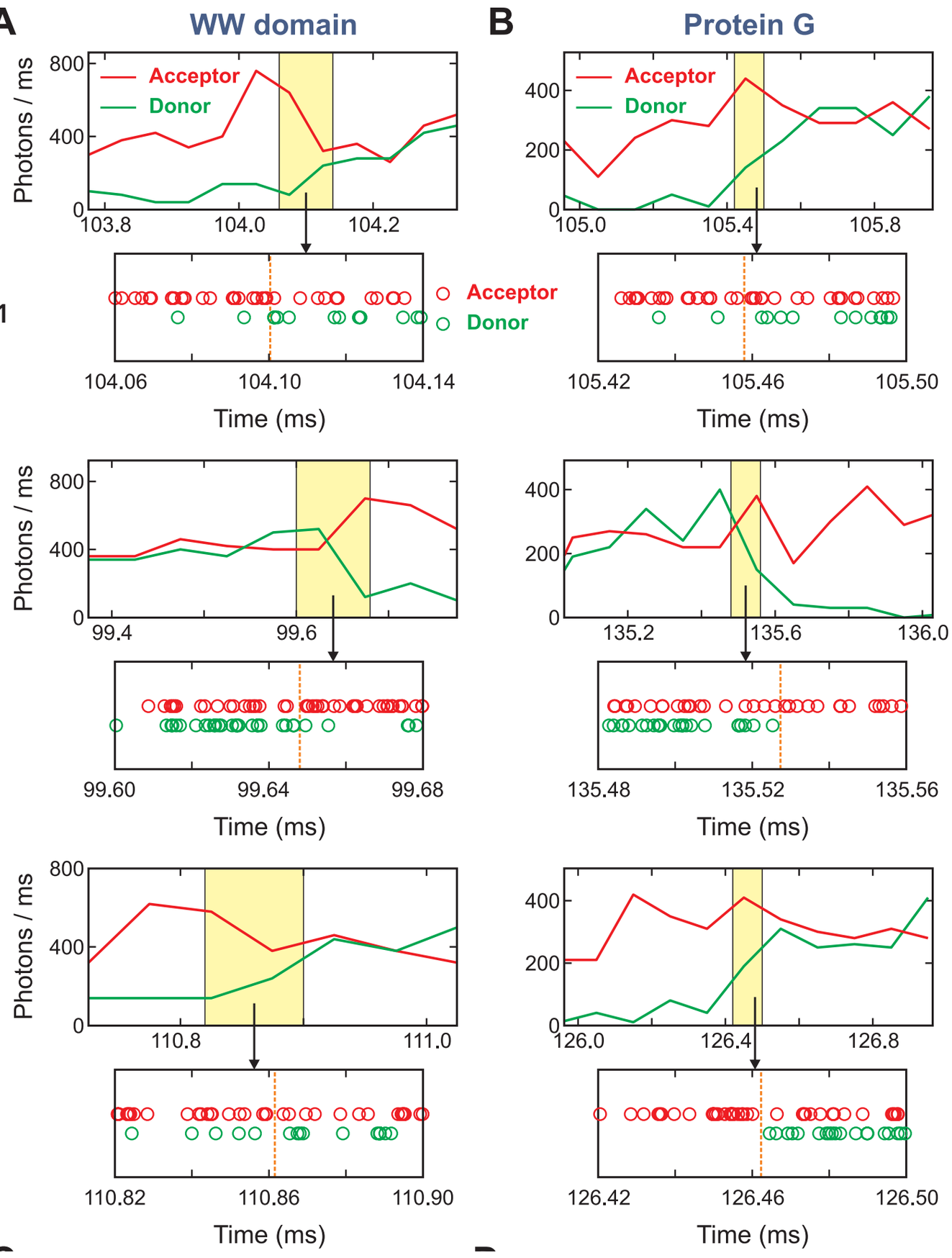


Single-Molecule Fluorescence Experiments Determine Protein Folding Transition Path Times

Hoi Sung Chung,* Kevin McHale, John M. Louis, William A. Eaton*

SCIENCE VOL 335 24 FEBRUARY 2012





2. Genome

A. Chromosome organization inside nucleus

B. Composition of Genome

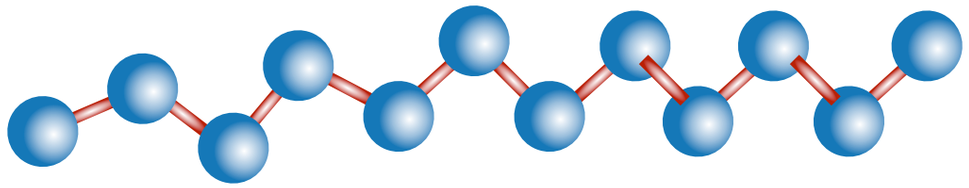
A. Chromosome organization inside nucleus

$L=(3\times10^{9}\,\mathrm{bp})\times(0.34\mathrm{nm/bp})\approx 1\mathrm{m}$

$\rightarrow$

nucleus of 1μm size

Polymer (N>>1)



$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2$$

Gaussian chain (random flight chain)

Freely jointed chain $(k_r \gg 1)$

$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2 + \sum_{i=1}^{N-1} \frac{k_\theta}{2}(\theta_i - \theta_0)^2$$

$$\left(\doteq \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2 - \sum_{i=1}^{N-2} \frac{k_\theta}{2} \hat{r}_{i,i+1} \cdot \hat{r}_{i+1,i+2} \right)$$

Semiflexible chain,
Worm-like chain
$(k_r \gg 1, \theta_0 = \pi)$

$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2 + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \varepsilon \left(\frac{\sigma}{r_{i,j}} \right)^{12}$$

Self-avoiding chain

$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2 + \varepsilon_h \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\left(\frac{\sigma}{r_{i,j}} \right)^{12} - 2\left(\frac{\sigma}{r_{i,j}} \right)^{6} \right] + \varepsilon_l \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\sigma}{r_{i,j}} \right)^{12}$$

attraction (A)

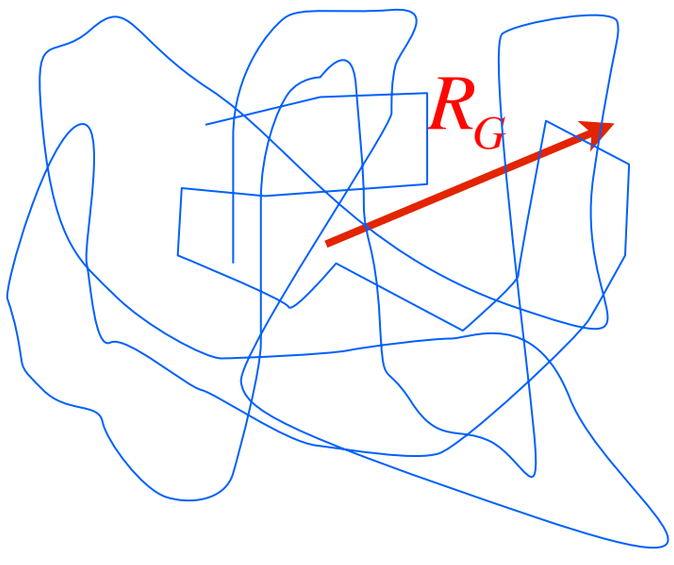
repulsion (R)

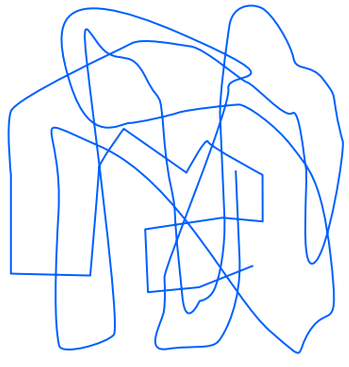
A<<R
Good solvent

A≈R

A>>R
poor solvent

θ





$R_G \sim N^{0.58}$

$R_G \sim N^{1/2}$

$R_G \sim N^{1/3}$

$$\left\langle R^2 \right\rangle = \left\langle \sum_i^N r_i \cdot \sum_j^N r_j \right\rangle = \sum_{i=1}^N \left\langle r_i^2 \right\rangle + \sum_{i \neq j}^N \left\langle r_i \cdot r_j \right\rangle = Na^2$$

$$\therefore R \sim N^{1/2}$$

$$P(R) \sim \exp\left(-\frac{3R^2}{2\left\langle R^2 \right\rangle} \right)$$

$$\left\langle R_g^2 \right\rangle = \frac{1}{N}\sum_{i=1}^N (\vec{R}_i - \vec{R}_C)^2 = \frac{1}{N}\sum_{i=1}^N \left(\vec{R}_i - \frac{1}{N}\sum_{i=1}^N \vec{R}_i \right)^2$$

$$= \frac{1}{2N^2}\sum_{i=1}^N \sum_{j=1}^N (\vec{R}_i - \vec{R}_j)^2 = \frac{1}{N^2}\sum_{i=1}^{N-1} \sum_{j=i+1}^N |i-j|^2 \, a^2 = \frac{Na^2}{6}$$

$$= \left\langle R^2 \right\rangle / 6$$

Size scaling of SAW: de Gennes' argument

$$F = E - TS = \frac{T}{2} c^2 R^d + T \frac{3R^2}{2\left\langle R^2 \right\rangle} \approx \frac{T}{2}\left(\frac{N^2}{R^d} \right) + \frac{3T}{2}\frac{R^2}{Na^2}$$

$$\frac{\partial F}{\partial R} = 0; \quad N^2 R^{-d-1} \sim N^{-1} R$$

$$\therefore R \sim N^{\nu} \quad \text{where} \quad \nu = \frac{3}{d+2} = \begin{cases} 1, & d = 1 \\ 3/4, & d = 2 \\ 3/5, & d = 3 \end{cases}$$

cf. polymer melts

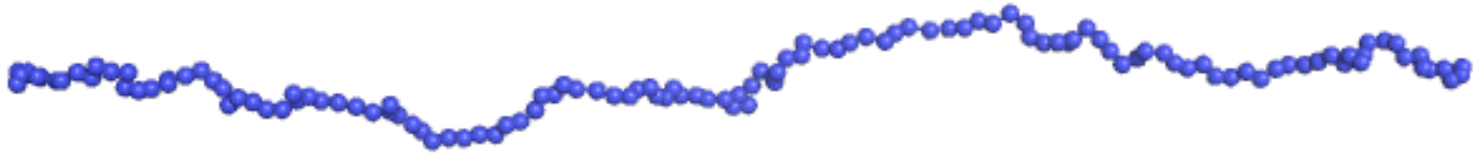
$$\left\langle R_g^2 \right\rangle \sim N$$

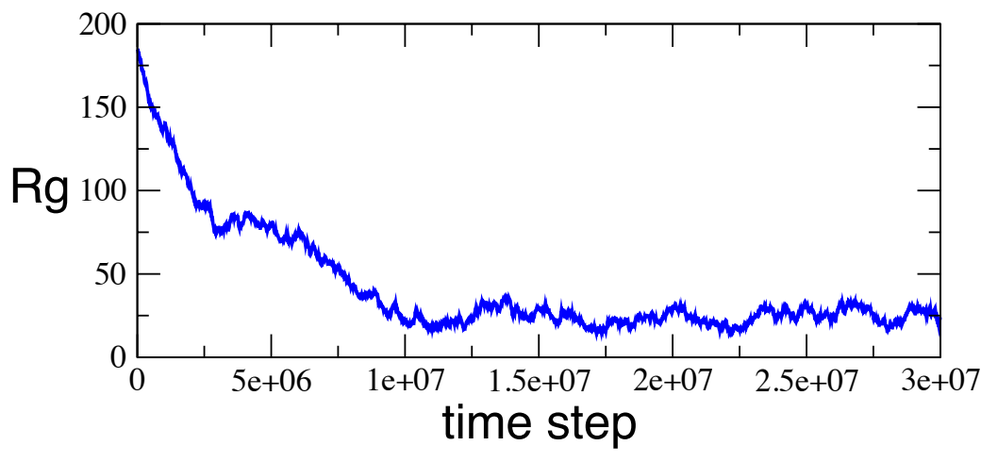
Gaussian statistics

Flexible polymer at θ-condition

$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2$$

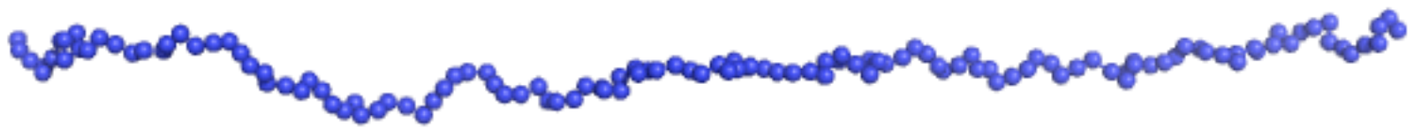
$$\Delta F = \Delta E - T\Delta S$$

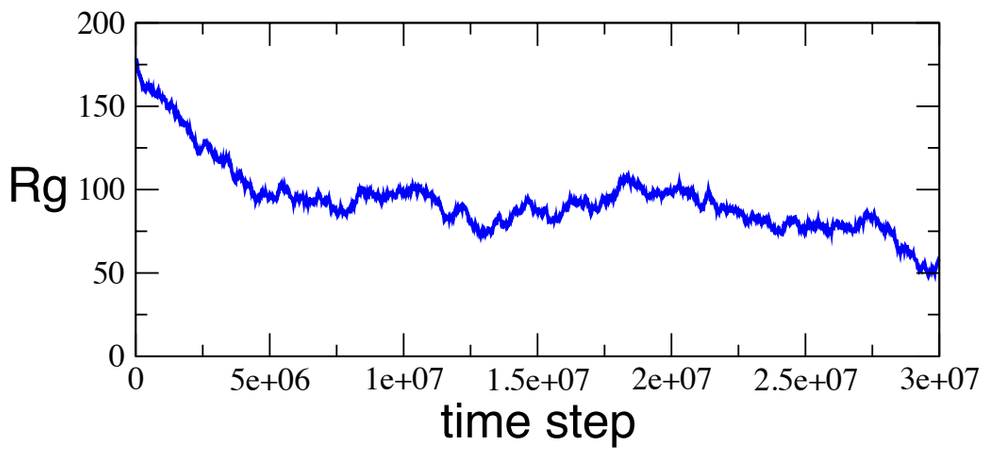




Flexible polymer in good solvent (SAW)

$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2 + \varepsilon_l \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(\frac{\sigma}{r_{i,j}} \right)^{12}$$

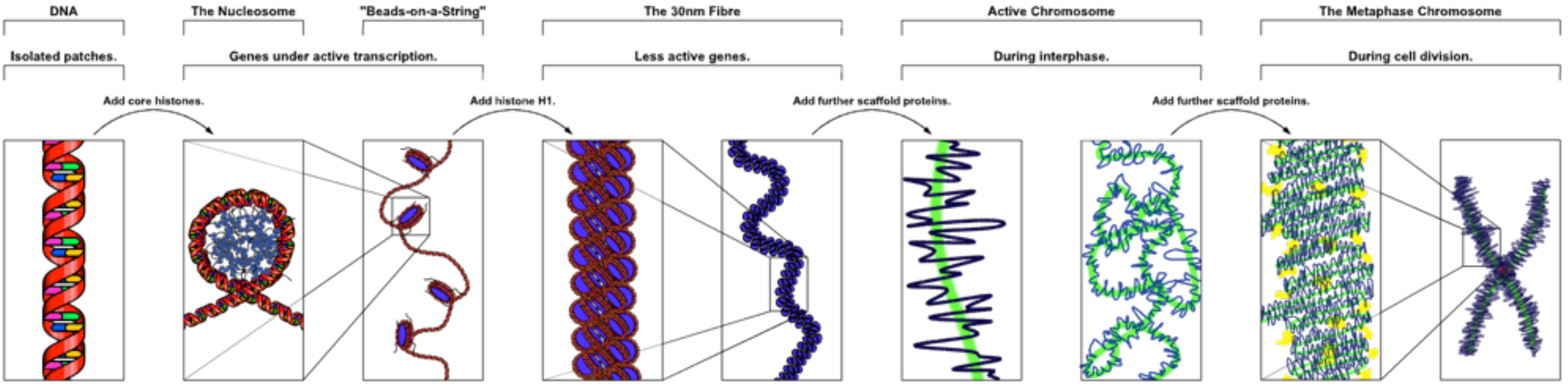


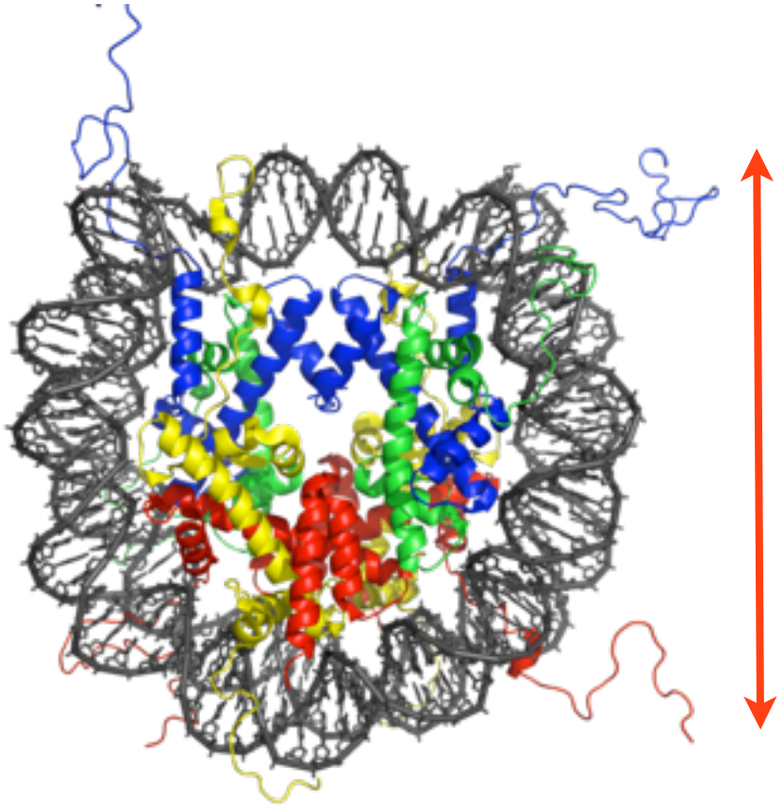


Collapse of flexible polymer in poor solvent

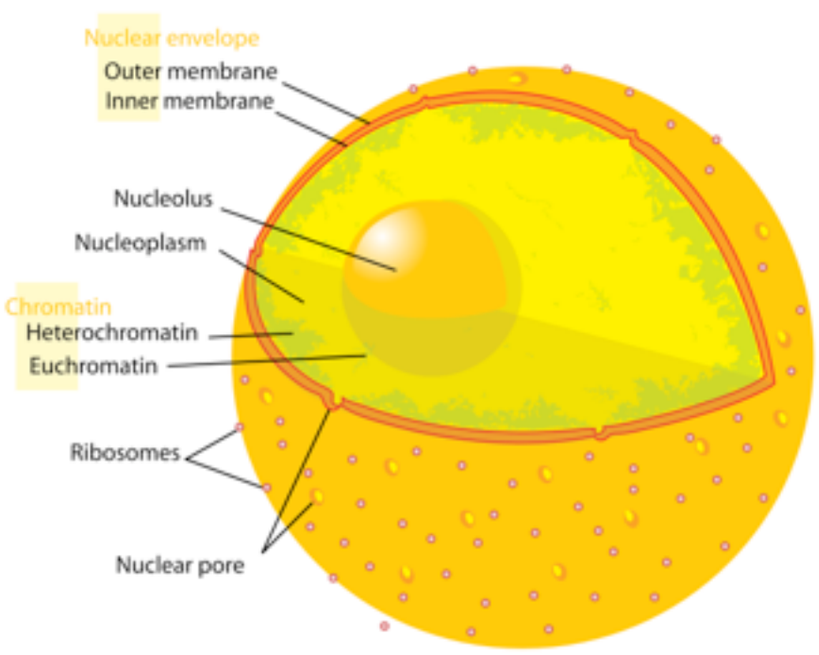
$$U_{\text{int}}(\{\vec{r}_i\}) = \sum_{i=1}^{N} \frac{k_r}{2}(r_{i,i+1} - a)^2 + \varepsilon_h \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[\left(\frac{\sigma}{r_{i,j}} \right)^{12} - 2 \left(\frac{\sigma}{r_{i,j}} \right)^{6} \right]$$

DNA compaction



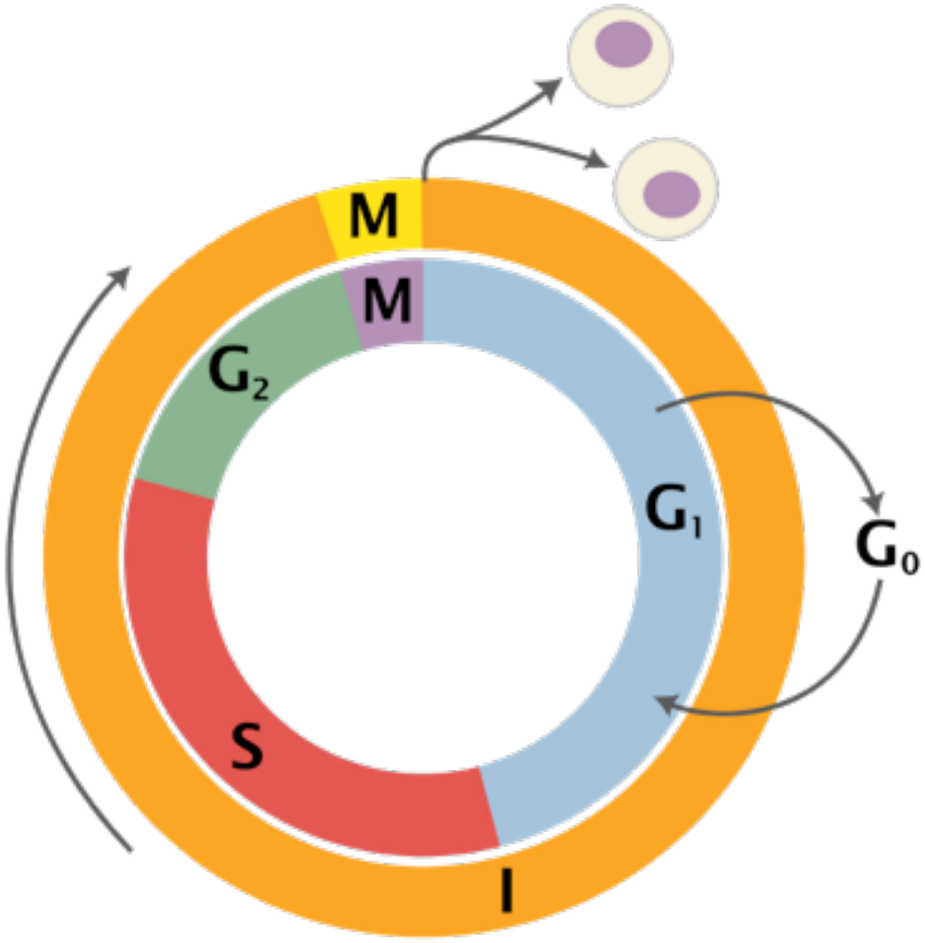


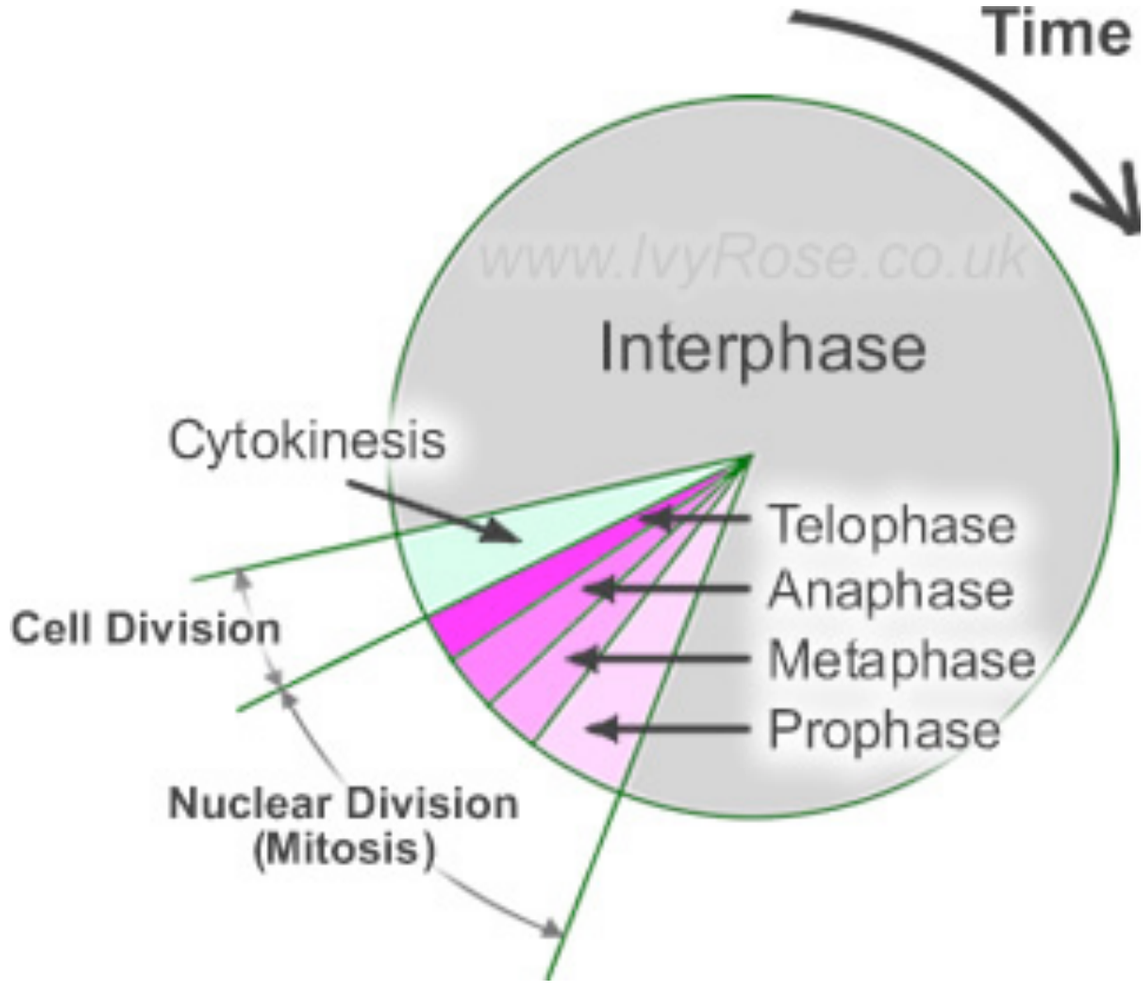
11 nm



Euchromatin : Loosely packed, Transcriptionally active
Heterochromatin : Densely packed, Transcriptionally inactive

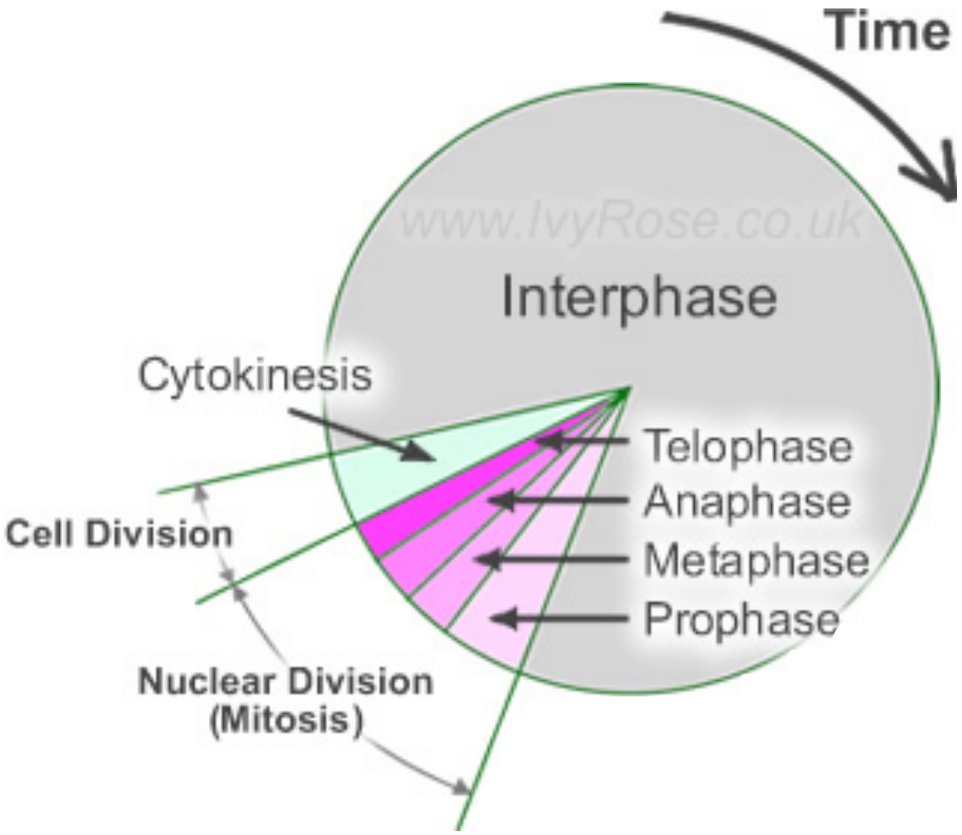
Spatial organization of chromosomes in the interphase?

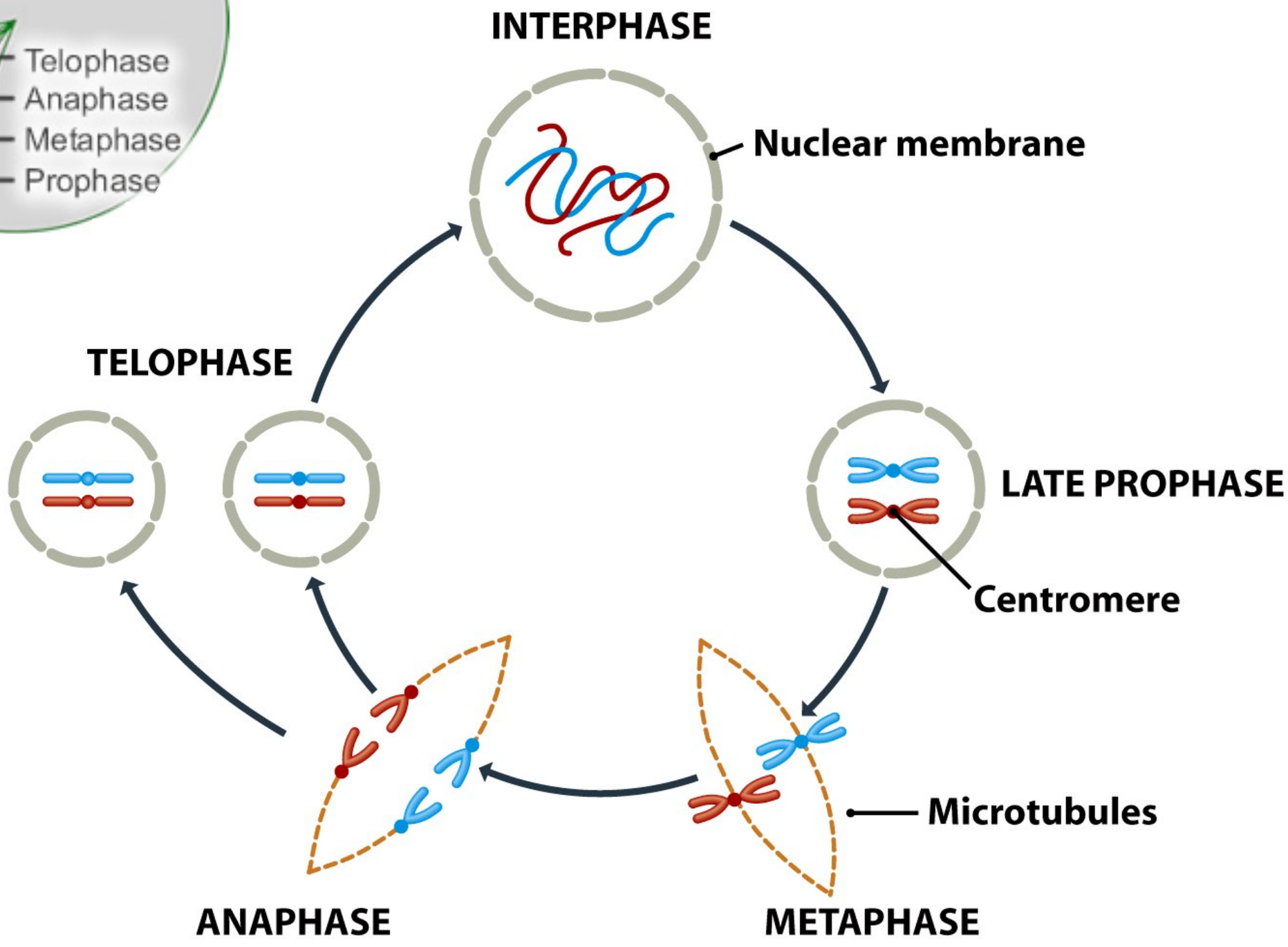




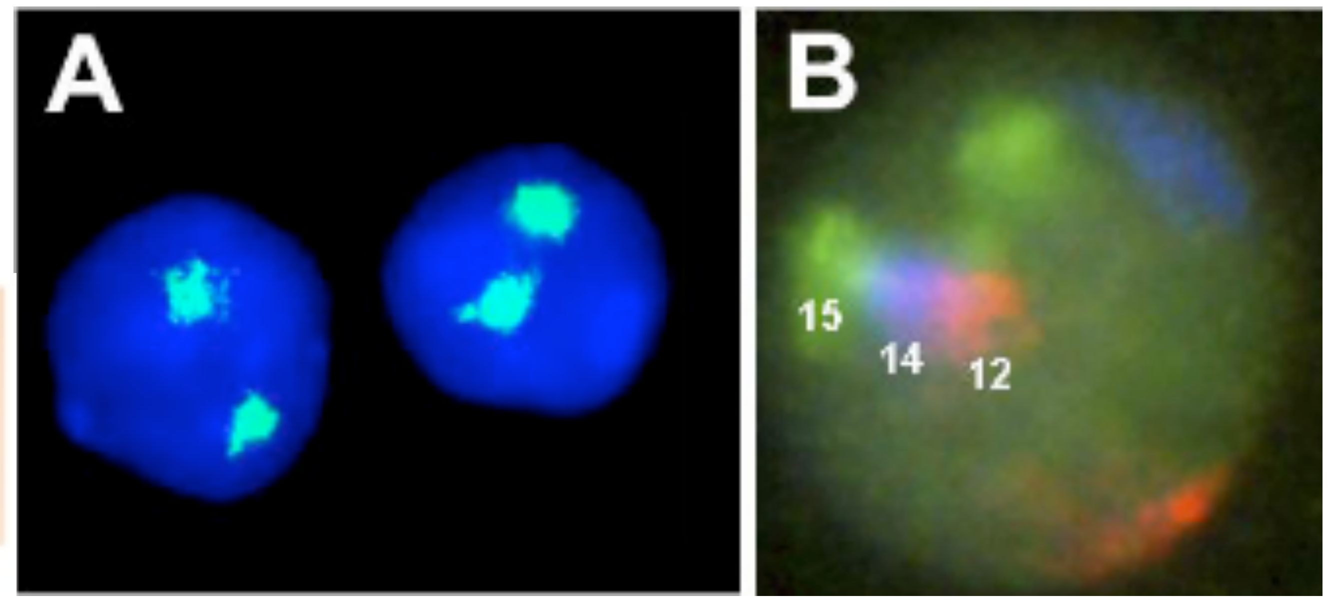
G_0 : a resting phase
G_1 : cells increase in size
S : DNA replication
G_2: cells continue to grow before M phase
M : cells stop growing and divide

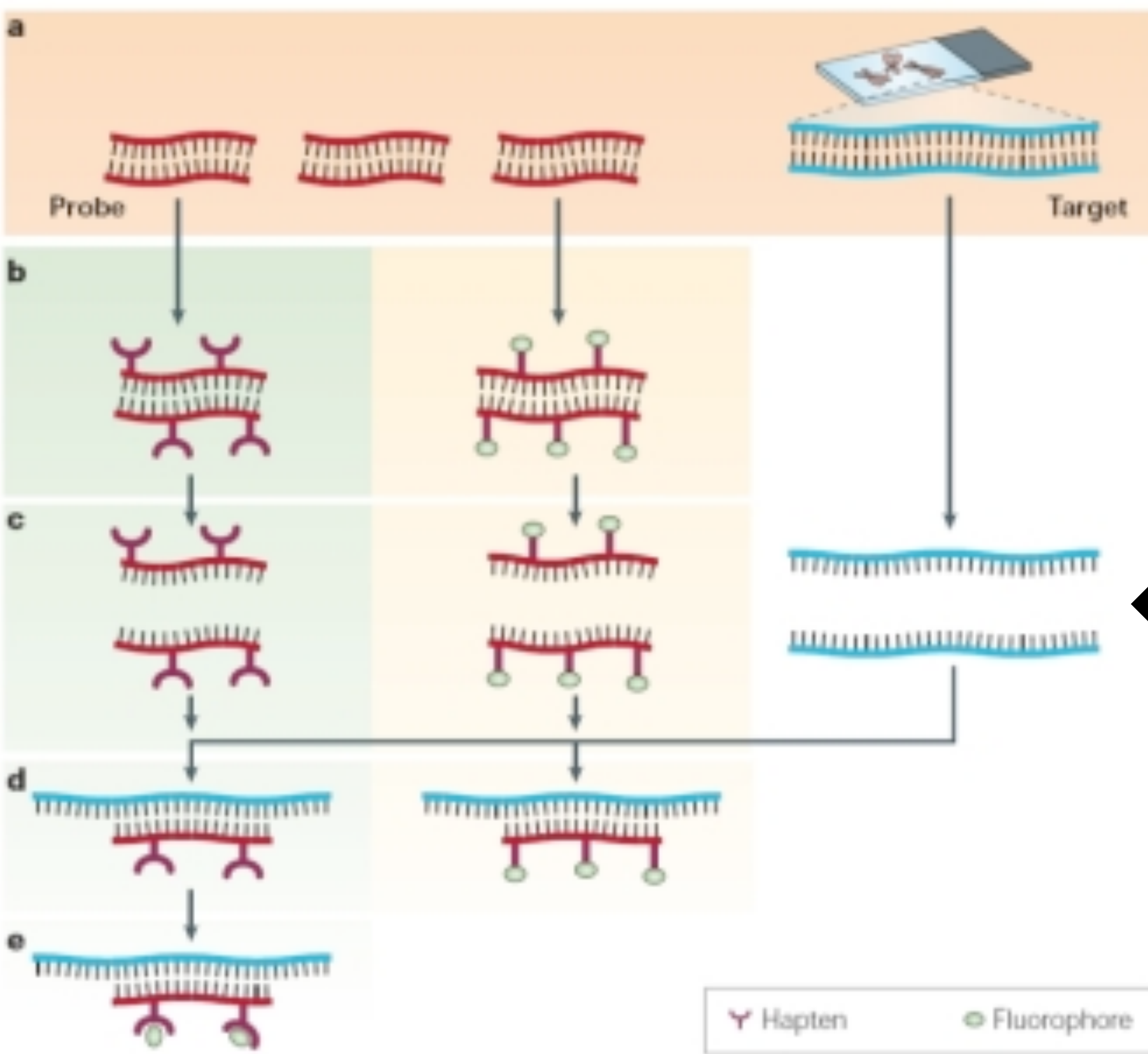
Spatial organization of chromosomes in the interphase?





Visualization of chromosome territories using FISH (fluorescence in situ hybridization)





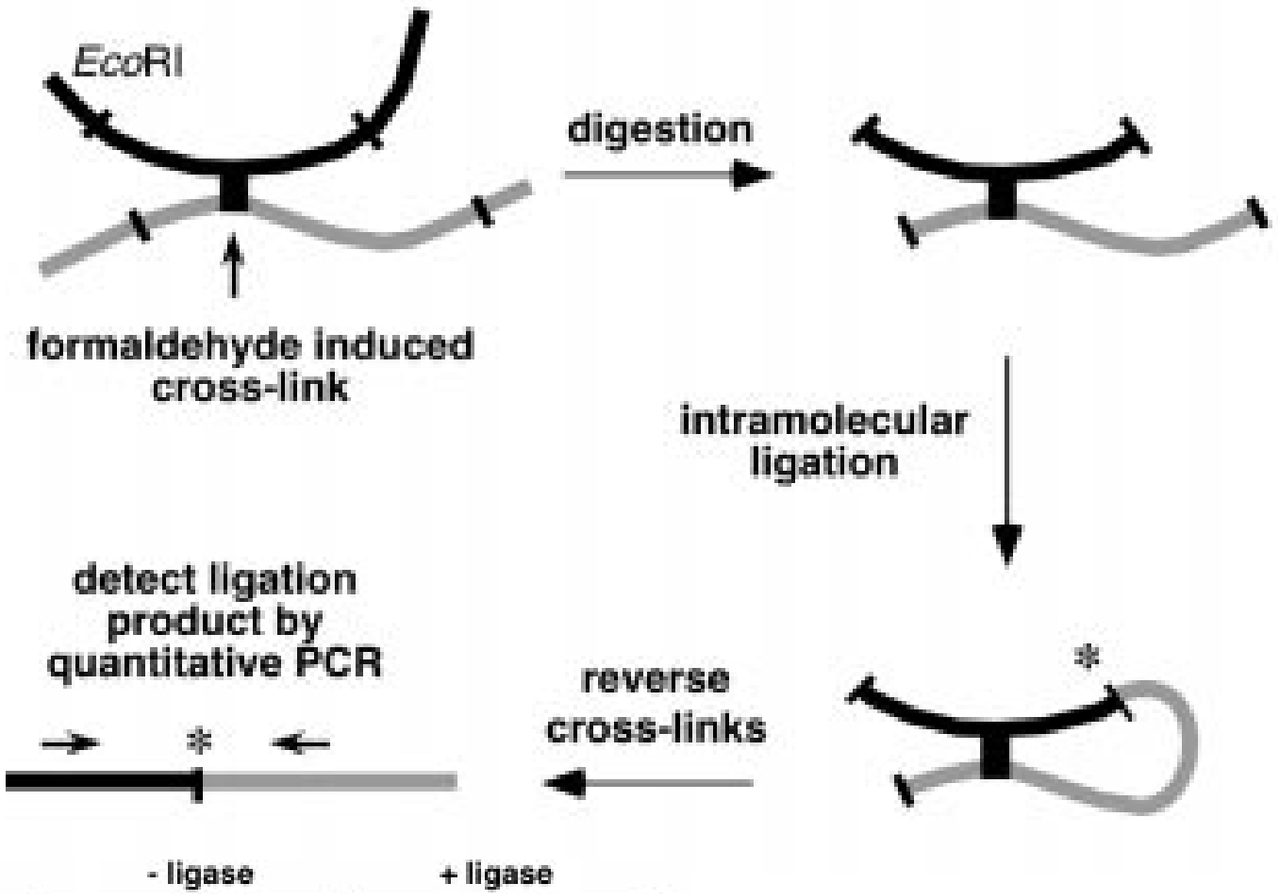
←denaturation

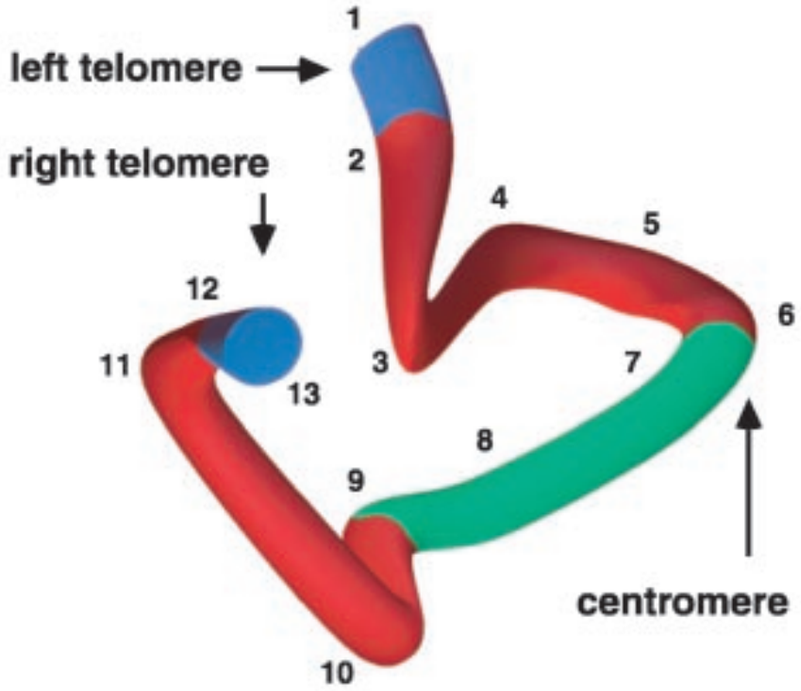
indirect

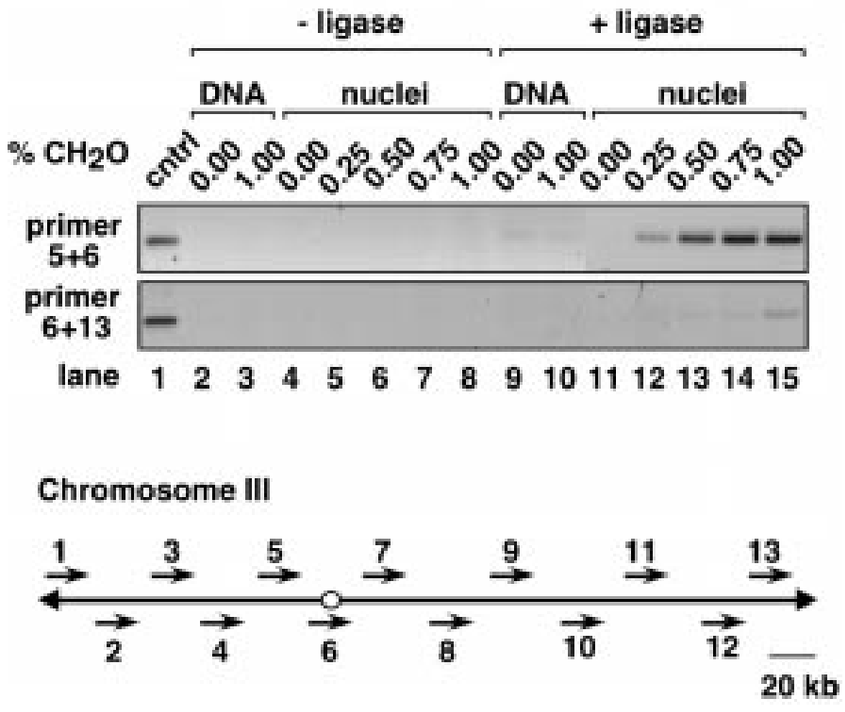
direct

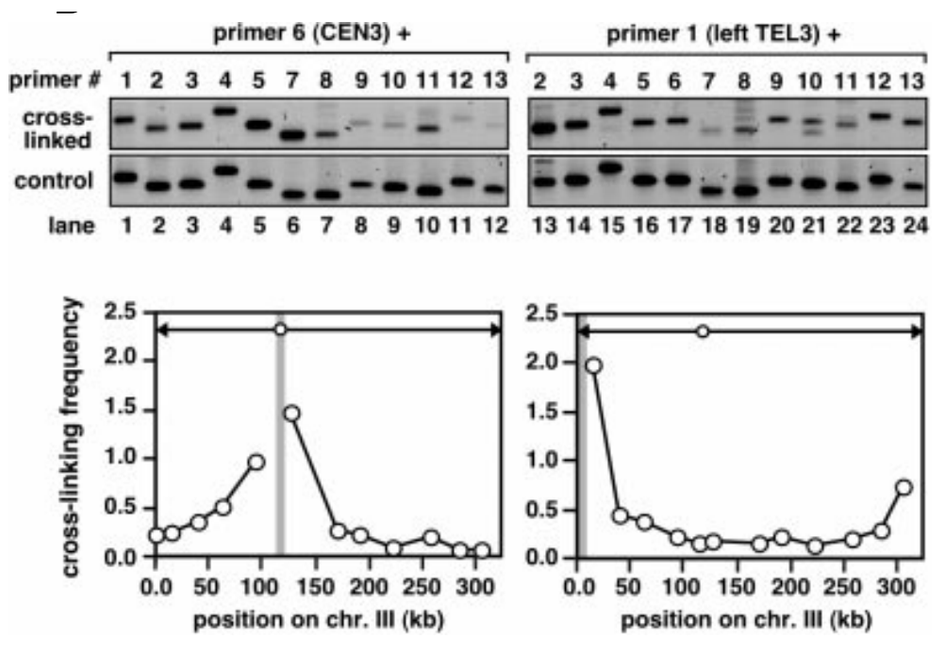
3C (chromosome conformation capture)

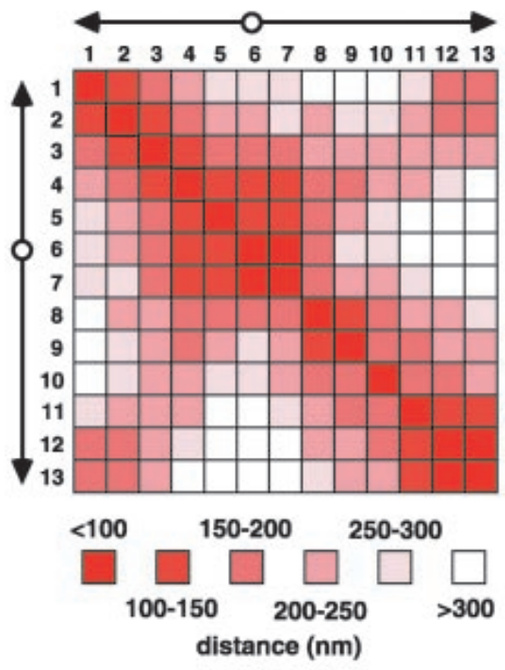
Capturing Chromosome Conformation
Job Dekker *et al.*
Science **295**, 1306 (2002);



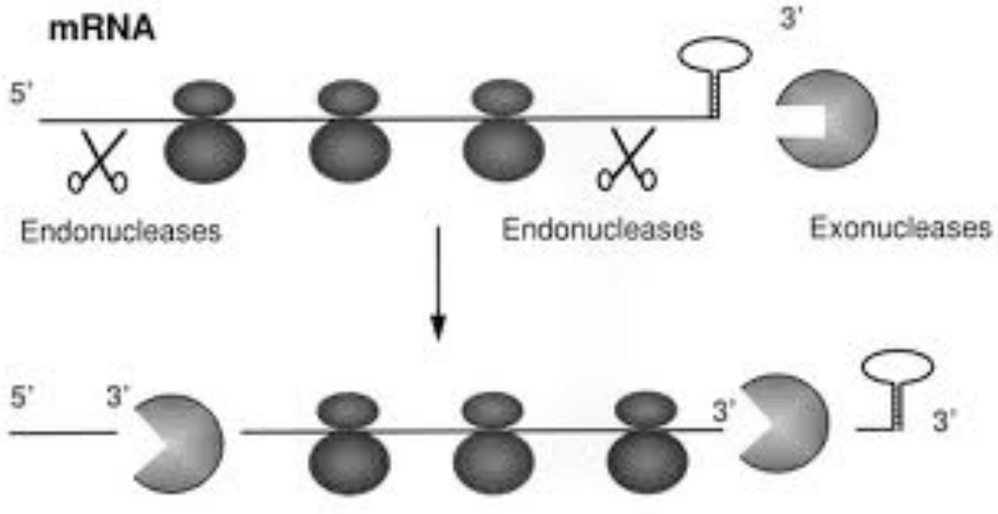




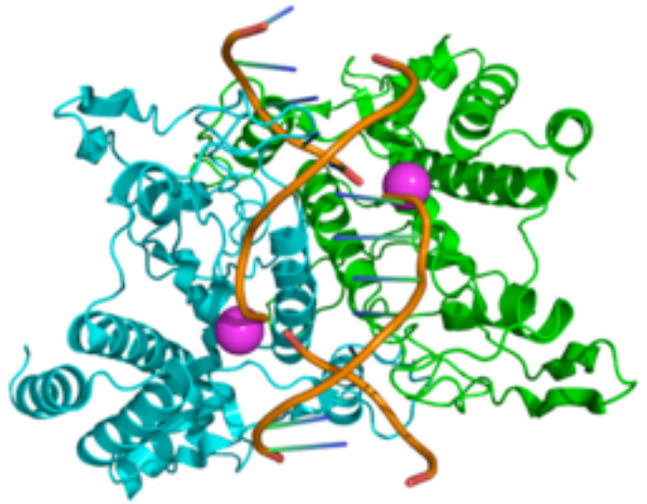




Restriction enzymes



- Type I enzymes cleave at sites remote from recognition site; require both ATP and S-adenosyl-L-methionine to function; multifunctional protein with both restriction and methylase activities.
- Type II enzymes cleave within or at short specific distances from recognition site; most require magnesium; single function (restriction) enzymes independent of methylase.
- Type III enzymes cleave at sites a short distance from recognition site; require ATP (but do not hydrolyse it); S-adenosyl-L-methionine stimulates reaction but is not required; exist as part of a complex with a modification methylase.
- Type IV enzymes target modified DNA, e.g. methylated, hydroxymethylated and glucosyl-hydroxymethylated DNA

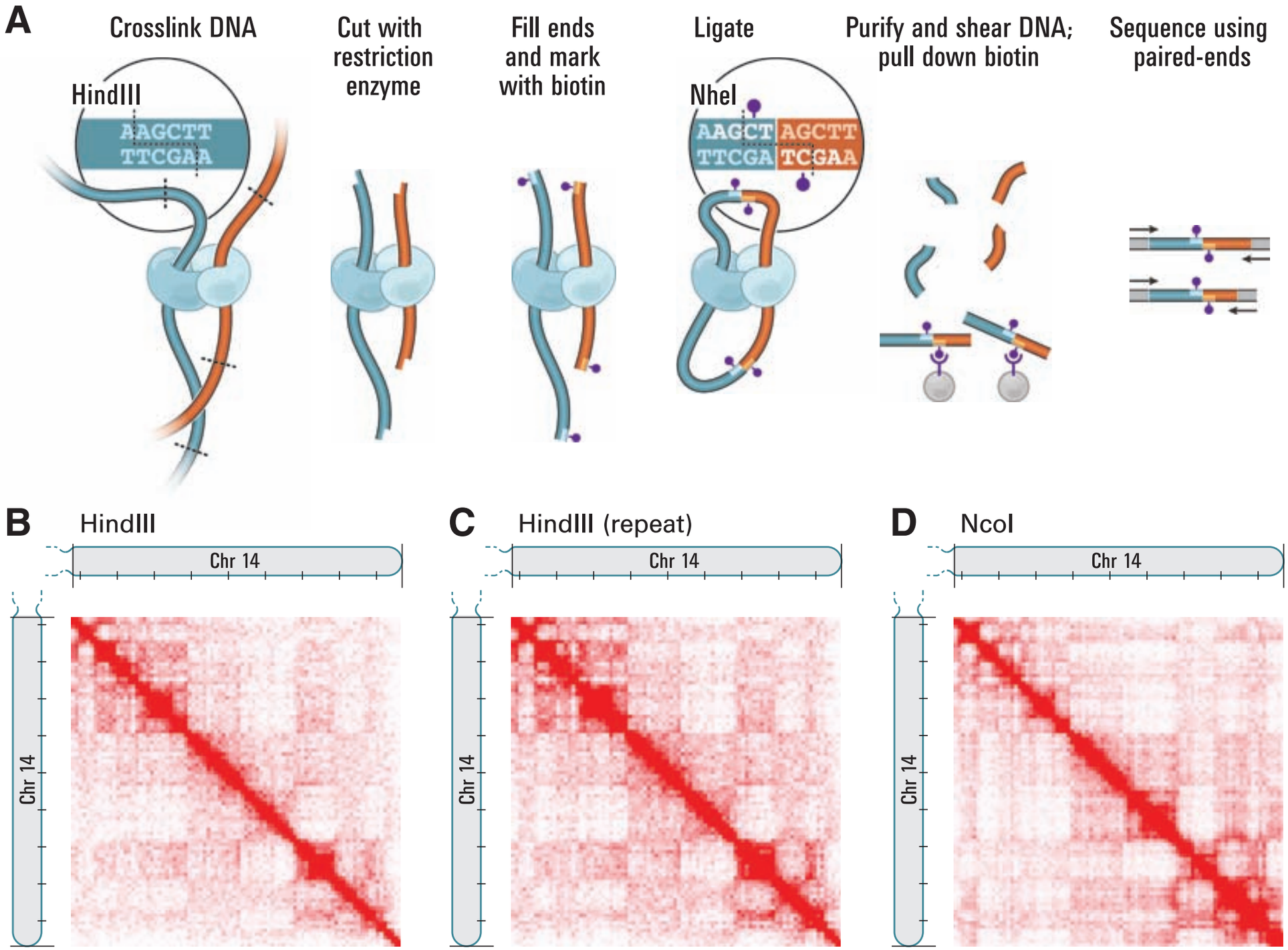


G|A A T T C
C T T A A|G

EcoRI

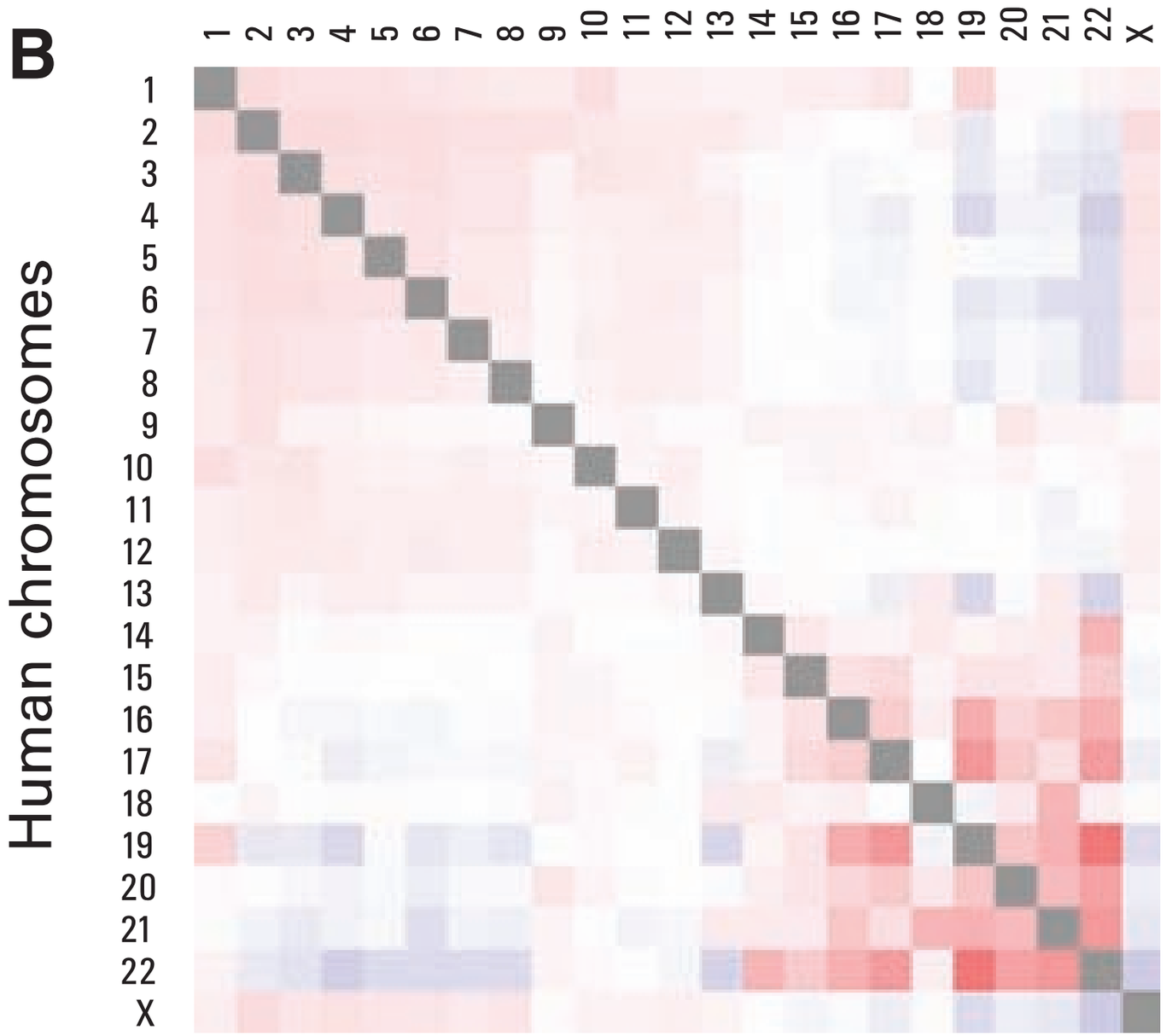
Hi-C method

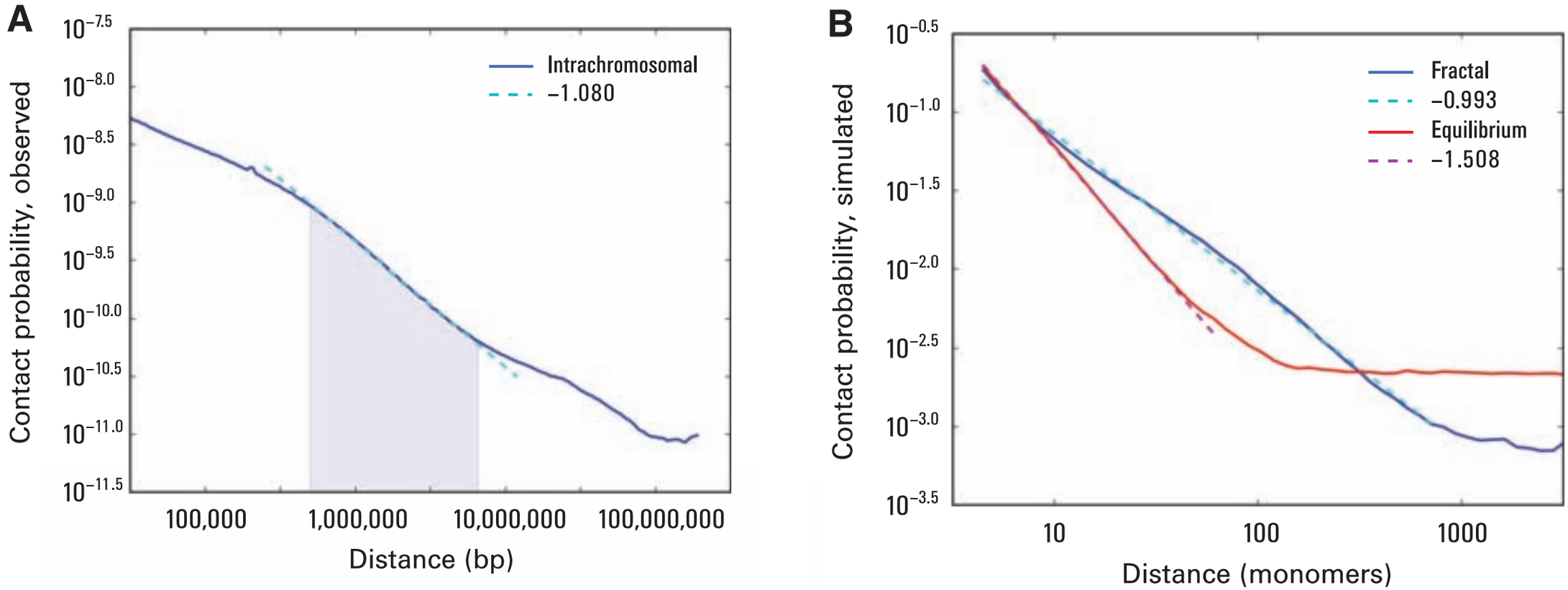
Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome
Erez Lieberman-Aiden *et al.*
Science **326**, 289 (2009);
DOI: 10.1126/science.1181369



(1Mb resolution)

Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome
Erez Lieberman-Aiden *et al.*
Science **326**, 289 (2009);
DOI: 10.1126/science.1181369





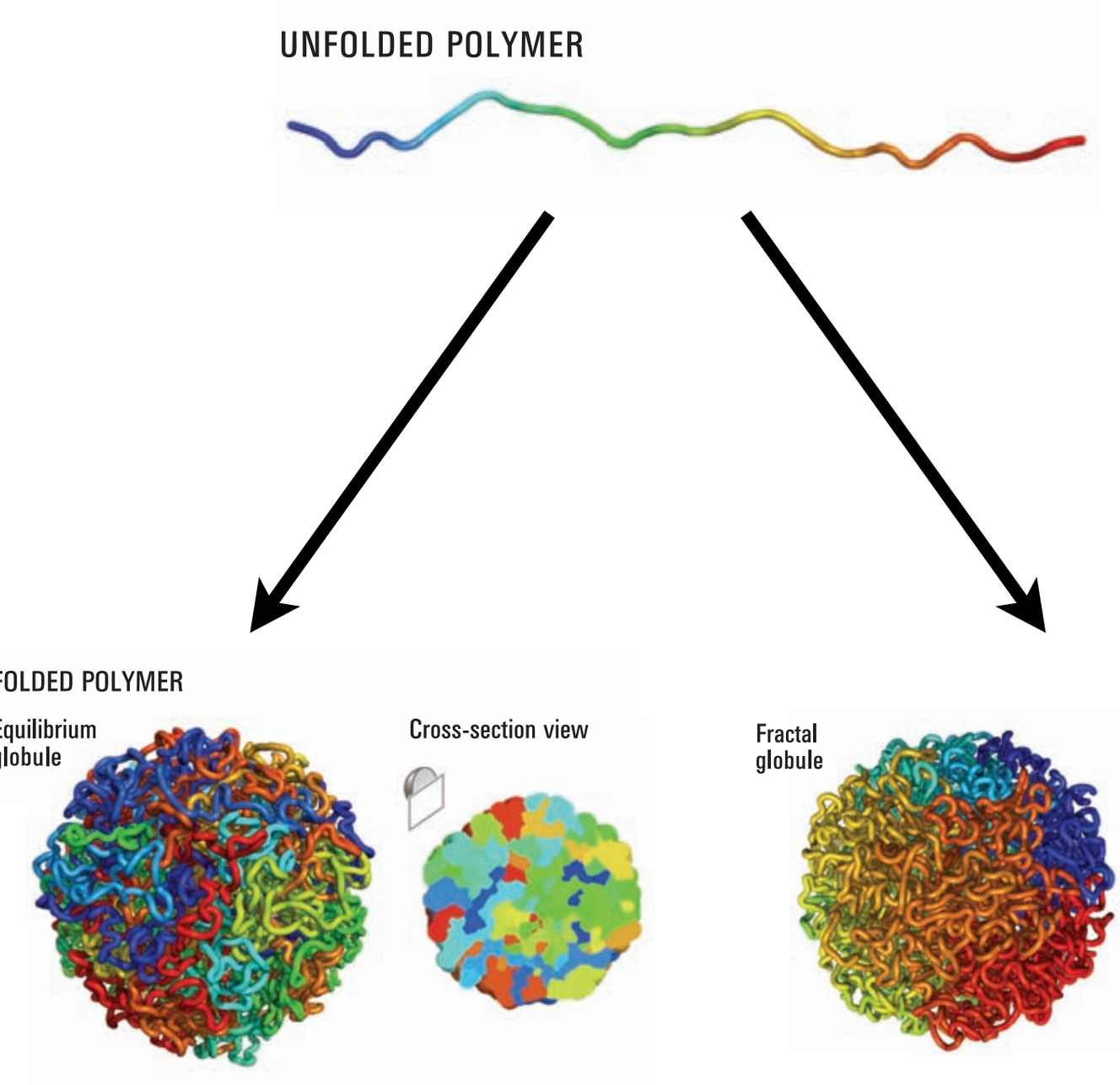
$R_g \sim N^{1/2}$ for polymer melts at an equilibrium, suggesting that each chain in the melts obey statistics of an ideal chain. Therefore, the contact probability [$P(r_1=r_2)$] for a given chain contour should be

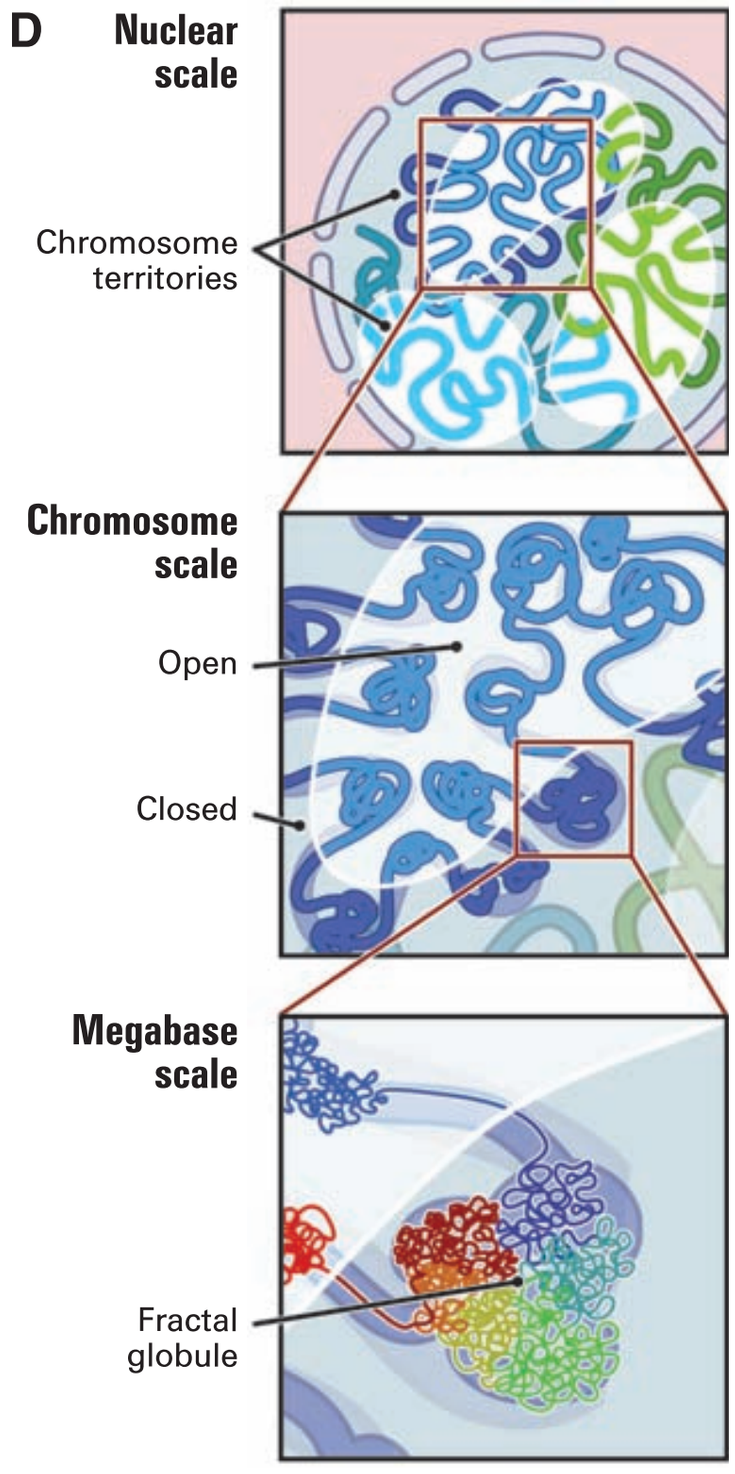
$$P(\vec{r}_1,\vec{r}_2) = \left(\frac{3}{2\pi\langle r_{12}^2\rangle}\right)^{3/2} \exp\left[-\frac{3(\vec{r}_1-\vec{r}_2)^2}{2\langle r_{12}^2\rangle}\right] \rightarrow P(0) \sim s^{-1.5}$$

Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome

Erez Lieberman-Aiden *et al.*

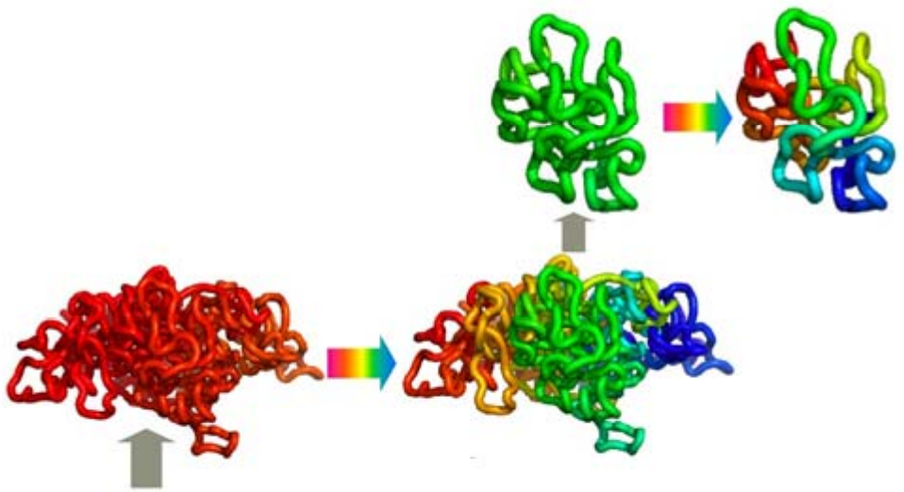
Science **326**, 289 (2009);
DOI: 10.1126/science.1181369

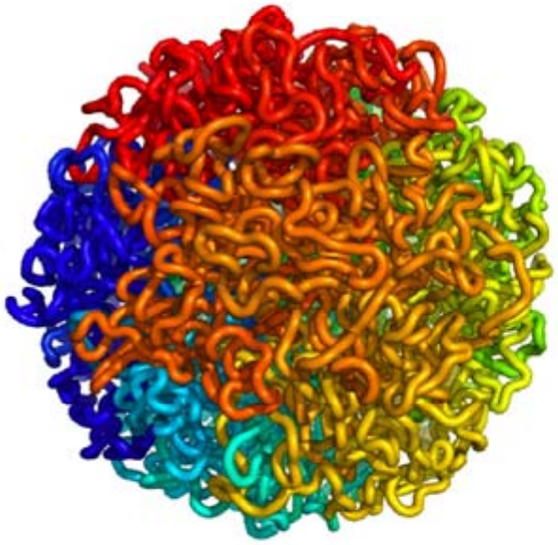


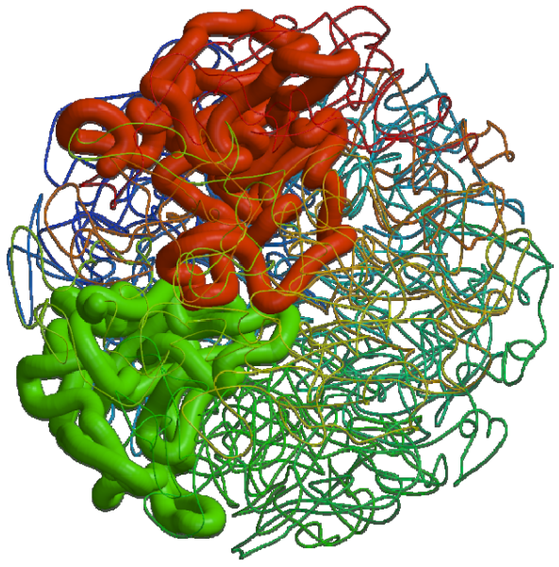


Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome
Erez Lieberman-Aiden *et al.*
Science **326**, 289 (2009);
DOI: 10.1126/science.1181369

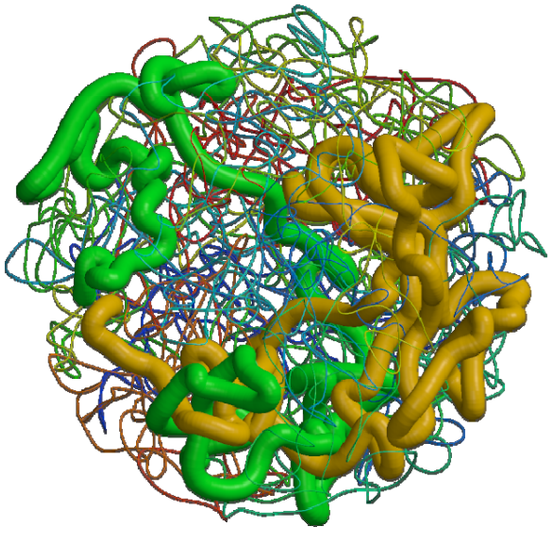
Fig S25. Self-similarity of Fractal Globule







Fractal Globule



Equilibrium Globule

B. Composition of Genome

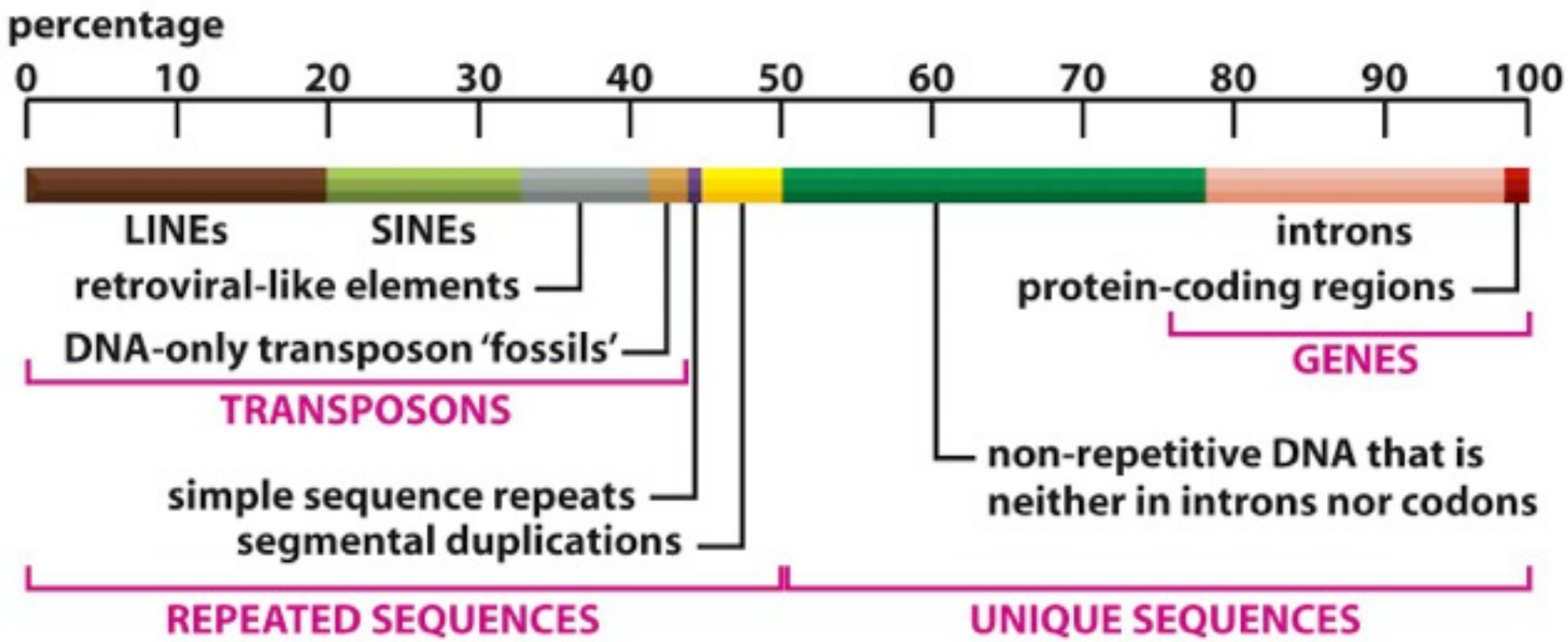
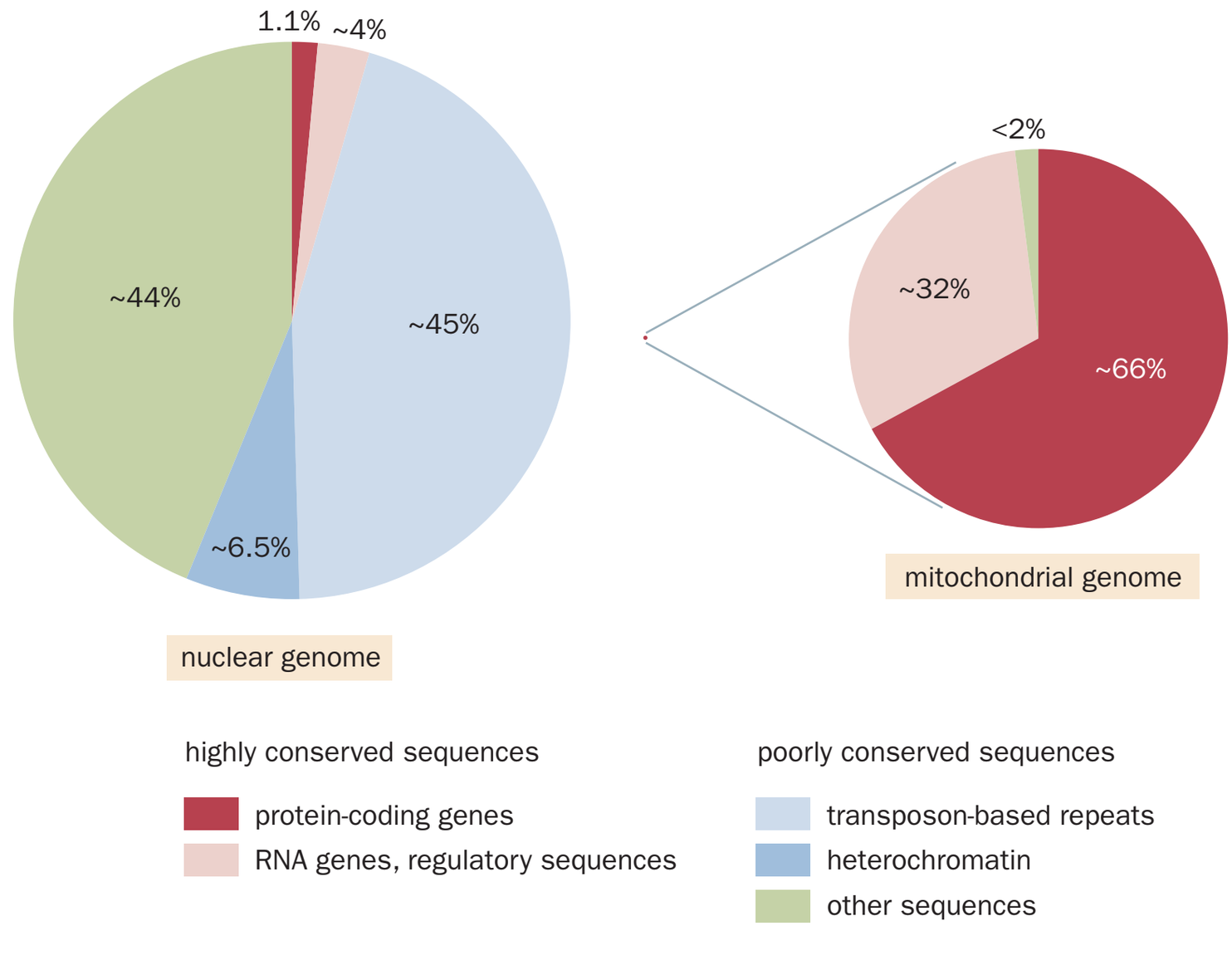


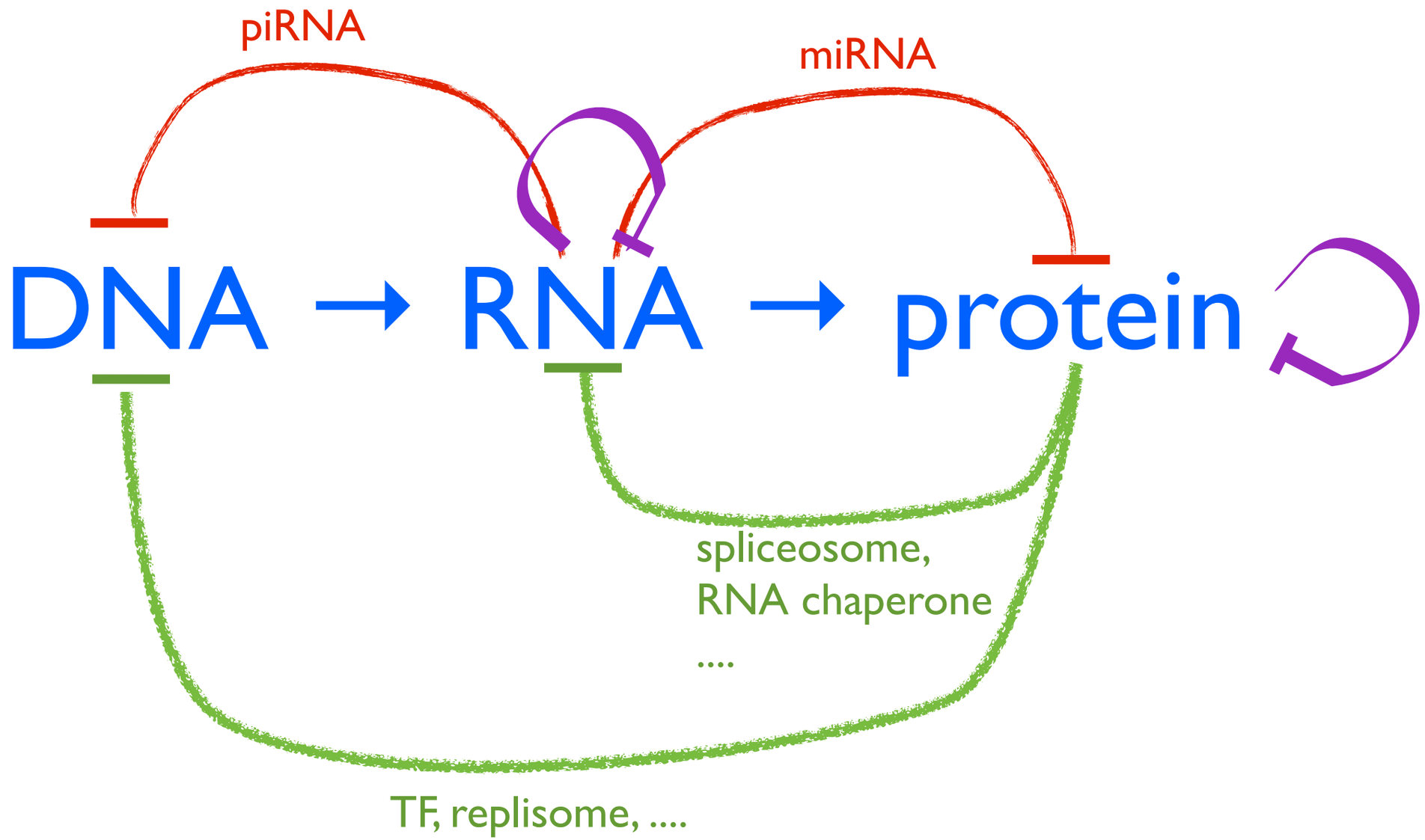
Figure 4-17 Molecular Biology of the Cell 5/e (© Garland Science 2008)



-ome

- Genome: the entirety of an organism's hereditary information, which includes both the genes and the non-coding sequences of the DNA/RNA.
- Transcriptome: the set of all RNA molecules, including mRNA, rRNA, tRNA, and other ncRNA produced in one or a population of cells (connect the genome to gene function).
- Proteome : the entire set of proteins expressed by a genome

Central Dogma



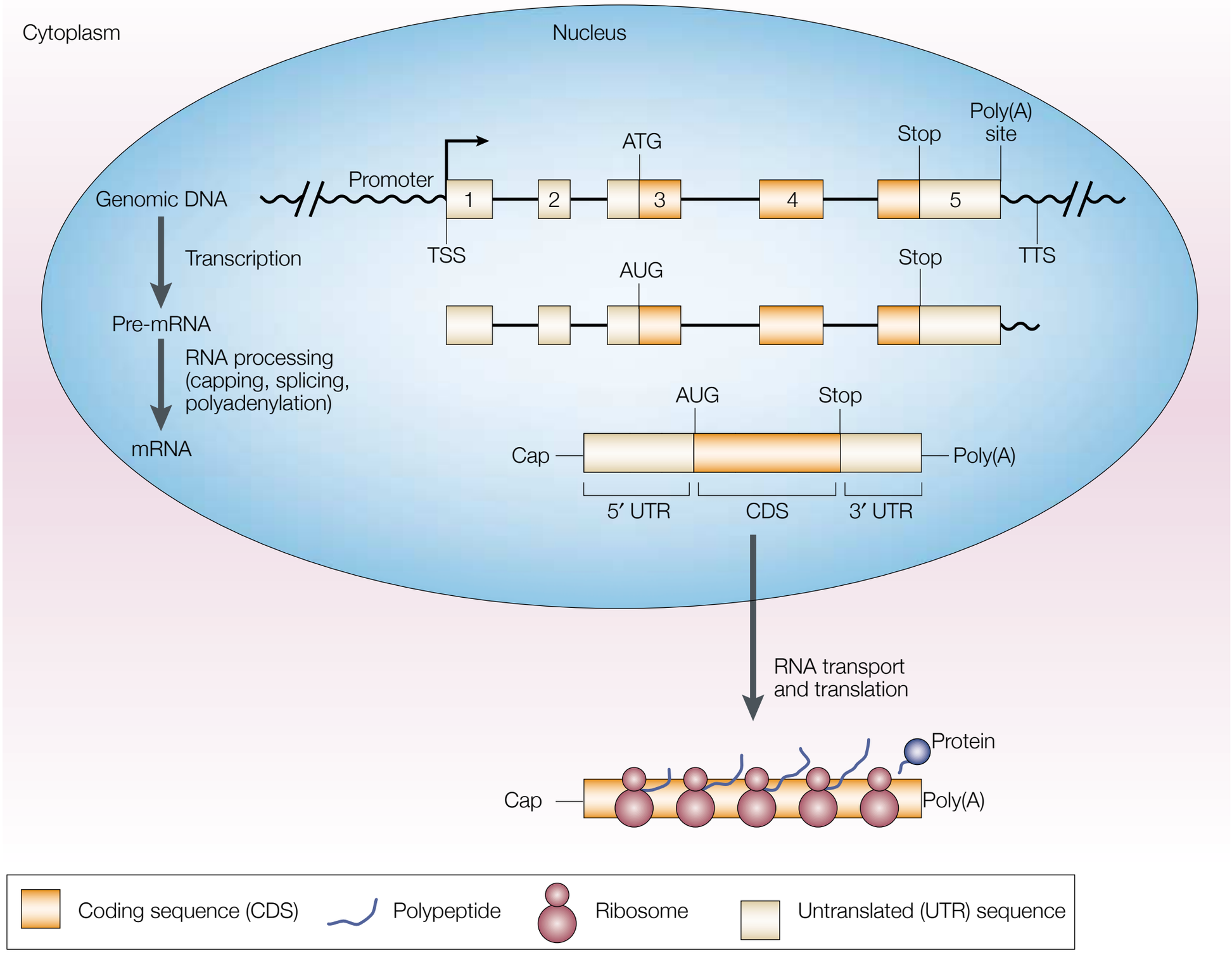
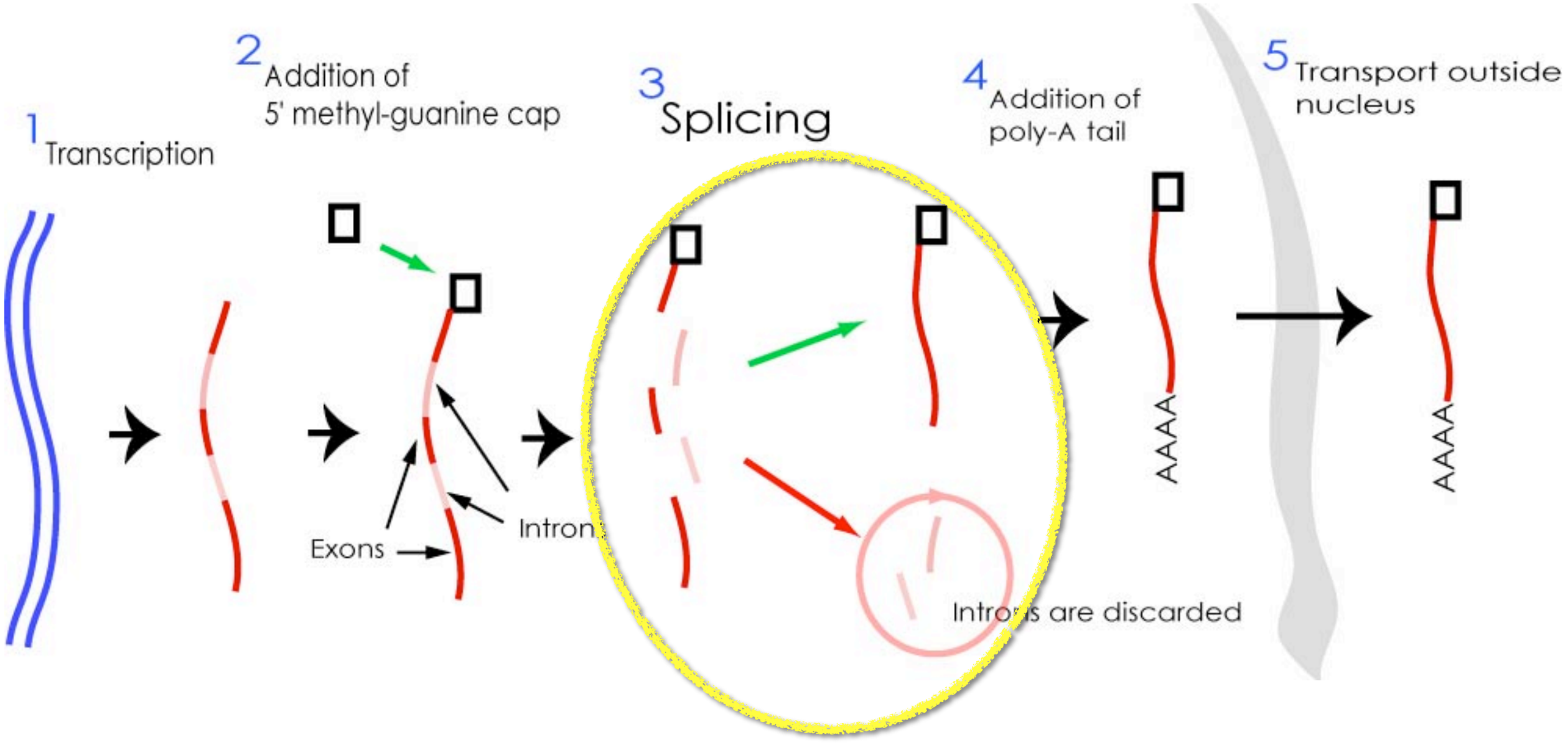
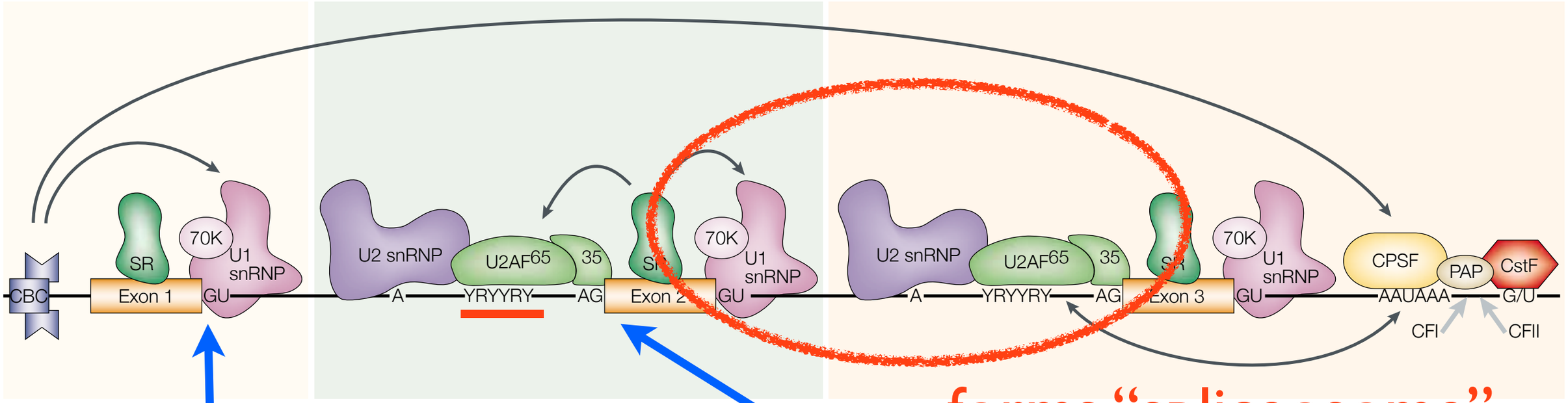


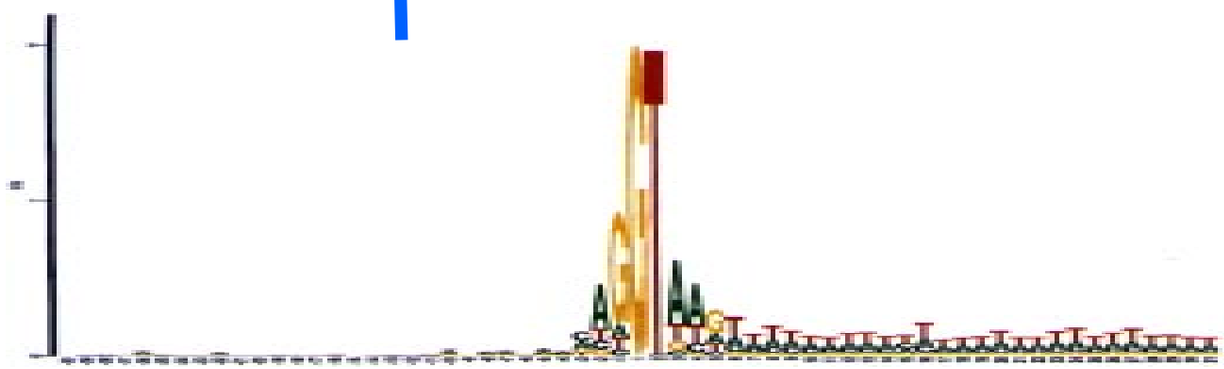
Figure 1 | **The central dogma of gene expression.** In the typical process of eukaryotic gene expression, a gene is transcribed from DNA to pre-mRNA. mRNA is then produced from pre-mRNA by RNA processing, which includes the capping, splicing and polyadenylation of the transcript. It is then transported from the nucleus to the cytoplasm for translation. TSS, transcription start site; TTS, transcription termination site.


1 Transcription
2 Addition of 5' methyl-guanine cap
3 Splicing
4 Addition of poly-A tail
5 Transport outside nucleus
Exons
Introns
Introns are discarded
AAAA
AAAA

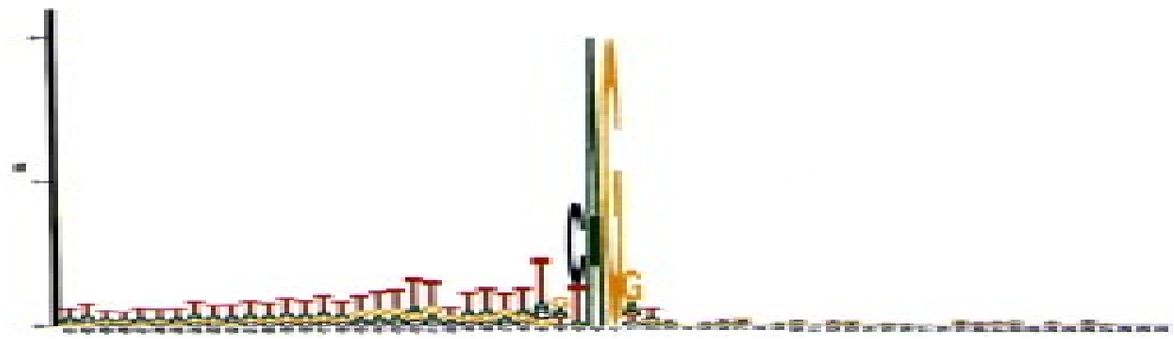
Exon-Intron boundary consensus sequences



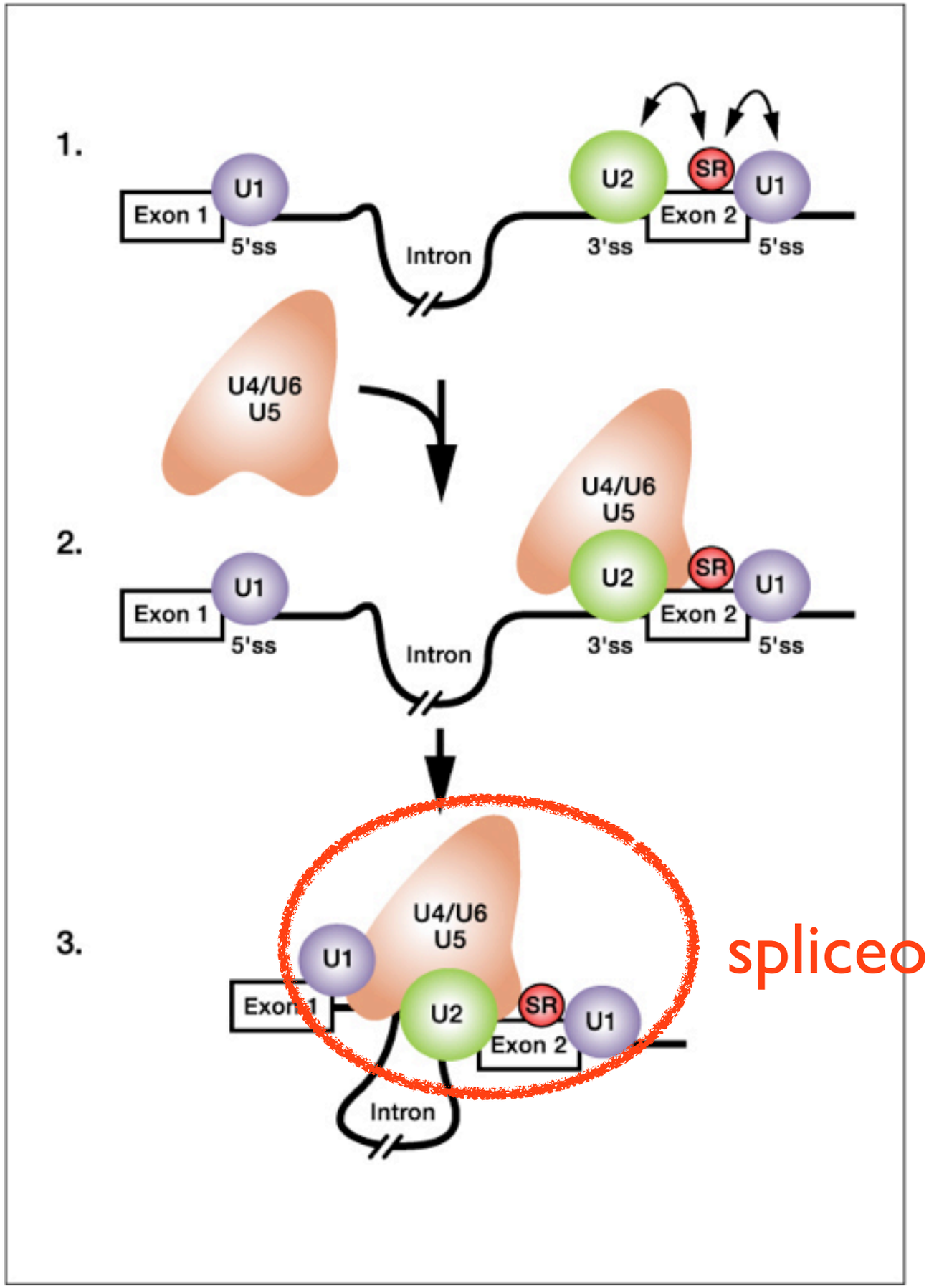
forms "spliceosome"



AG|GTAAGT

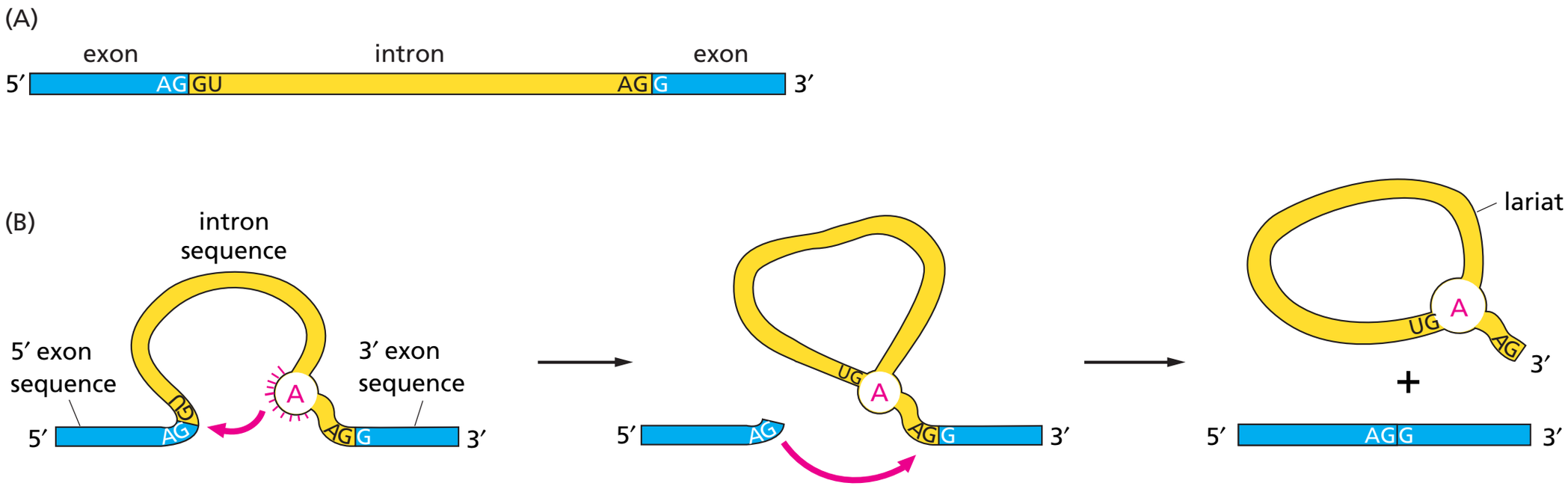


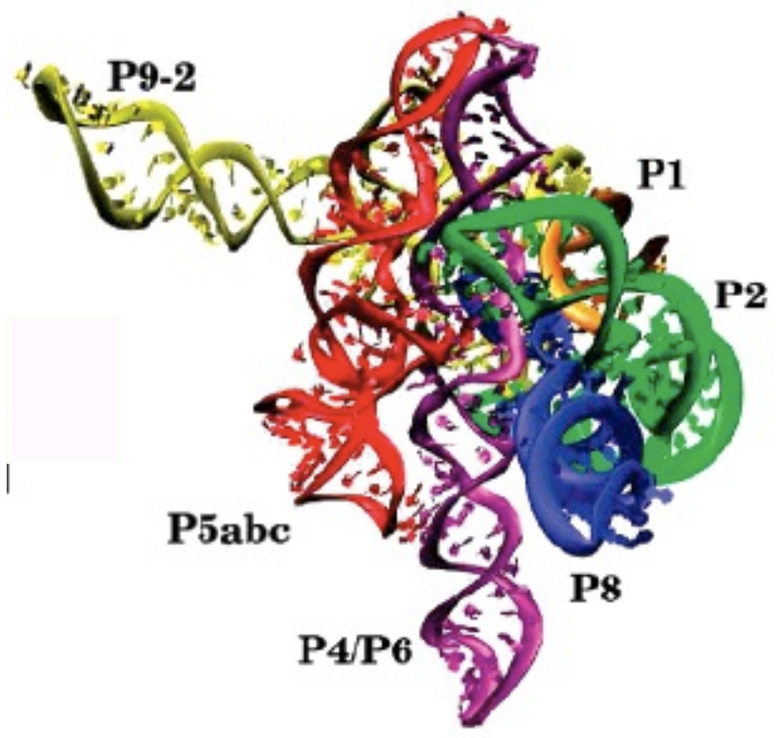
TAYAG|GT



spliceosome (> ribosome)

cf) group I, II intron
Self-splicing ribozyme



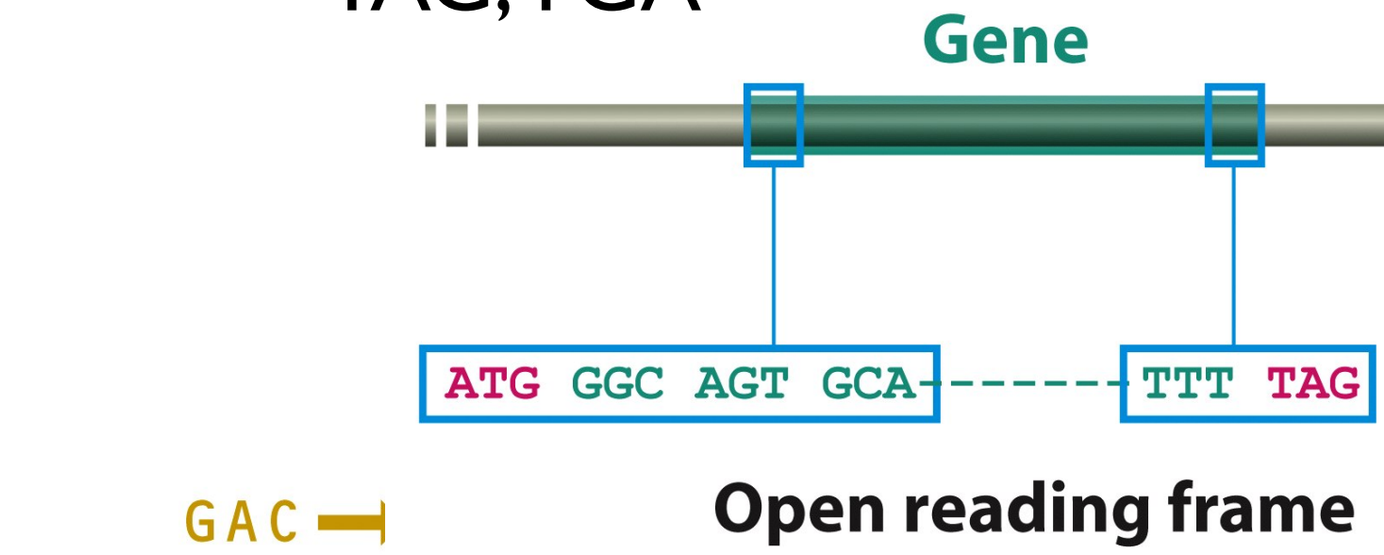


Tetrahymena group I intron ribozyme

Locating the Genes in a Genome Sequence

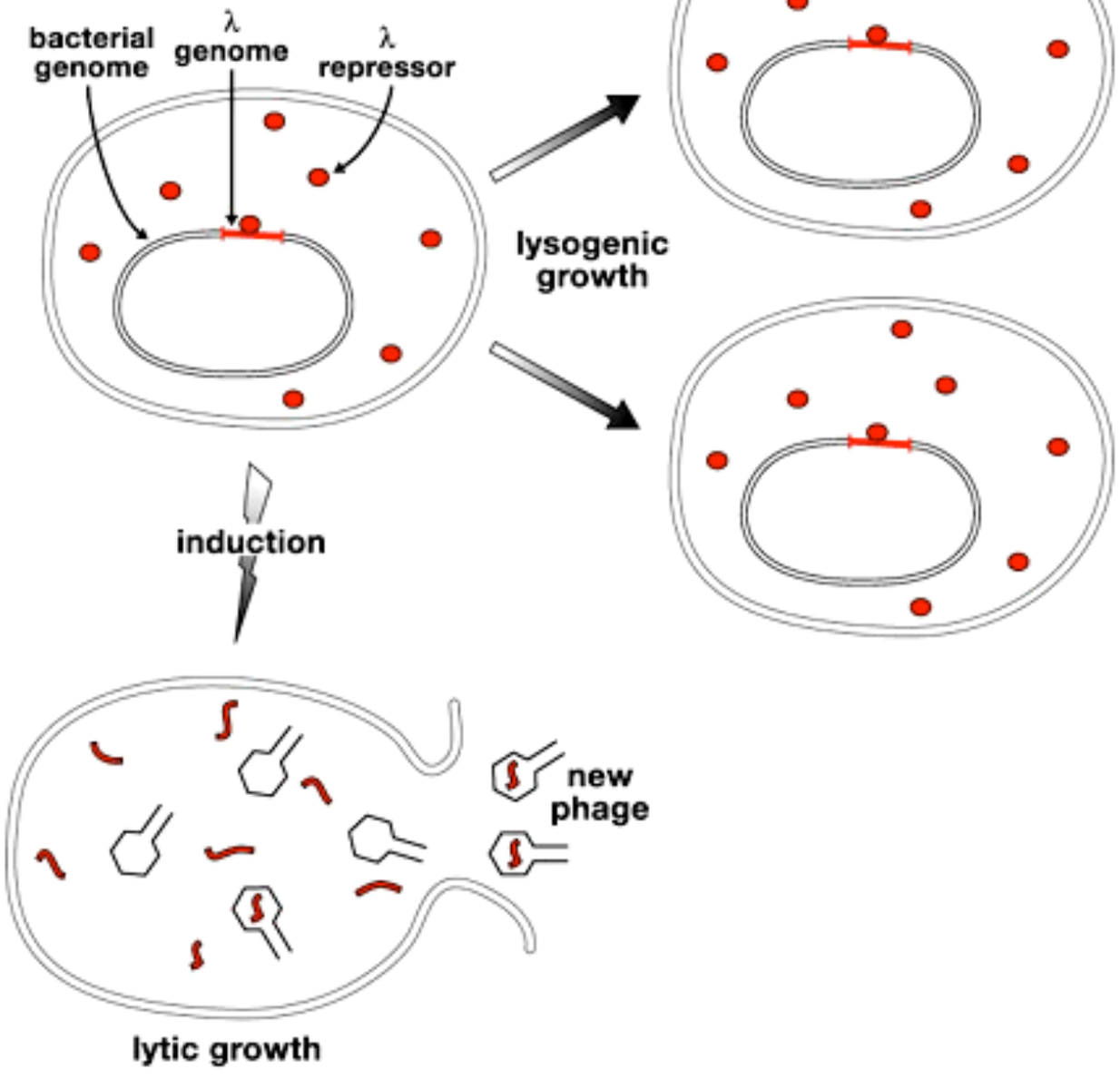
- ORF scanning
  - Open Reading Frames (ORF)

Initiation codon: ATG, Termination codon: TAA, TAG, TGA

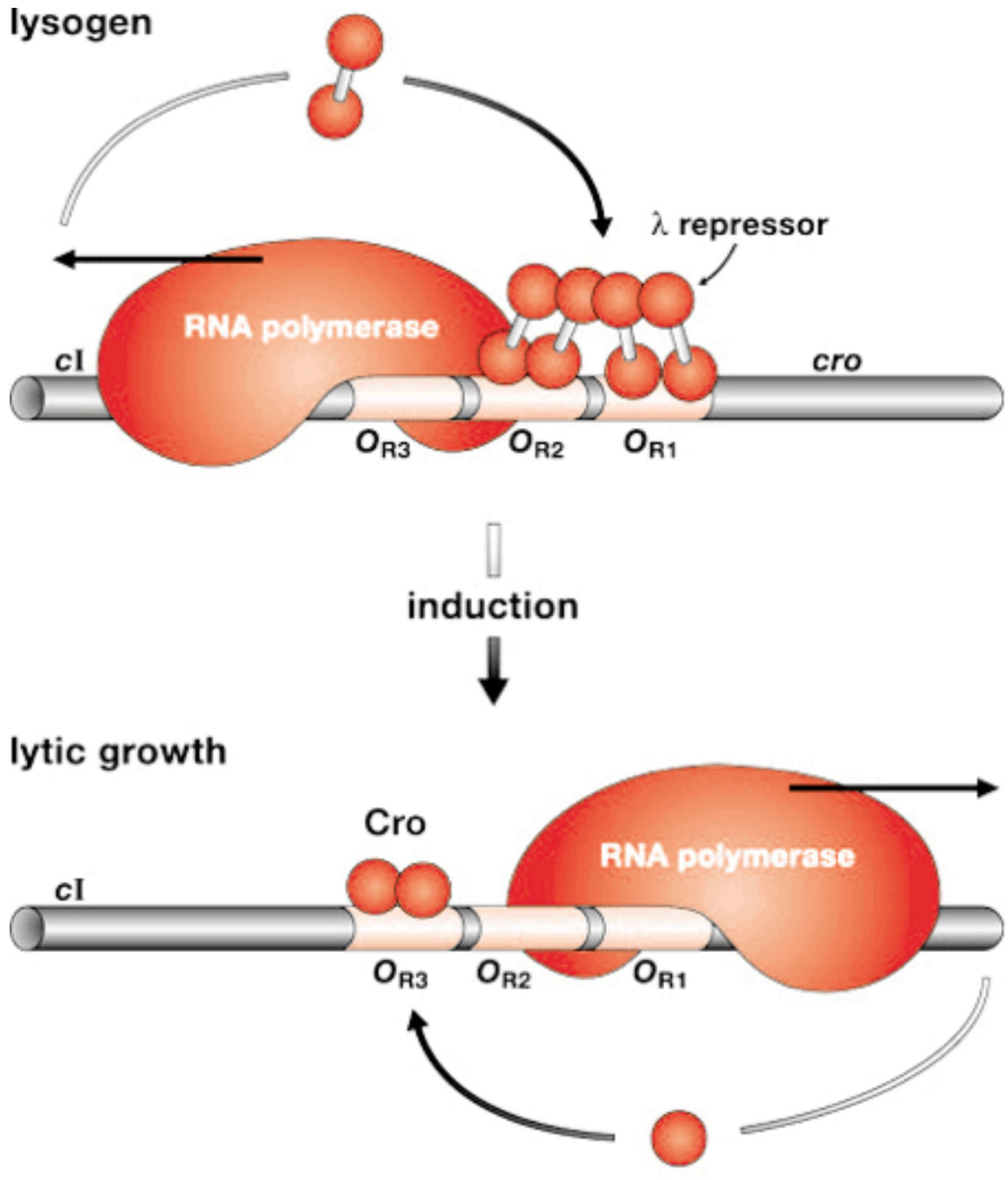


6 possible reading frames

cf) lysogenic vs lytic pathways



Genes and Signals, © 2002 by Cold Spring Harbor Laboratory Press, Chapter 1, Figure 10



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/.)

Locating the Genes in a Genome Sequence (ORF scanning)

- If DNA has a random sequence each termination codon (TAA, TAG, TGA) appear once every 4^3=64 bp;

  GC content is greater than 50 %;

  Frequency of termination codon from random sequence would be 1/200 - 1/100;

  Thus, the shortest length of a putative gene should be ~> 50-100 codons.

- For higher eukaryote, ORFs are not continuous but often split by introns; thus ORF scanning makes use of the followings:

  (1) Codon bias

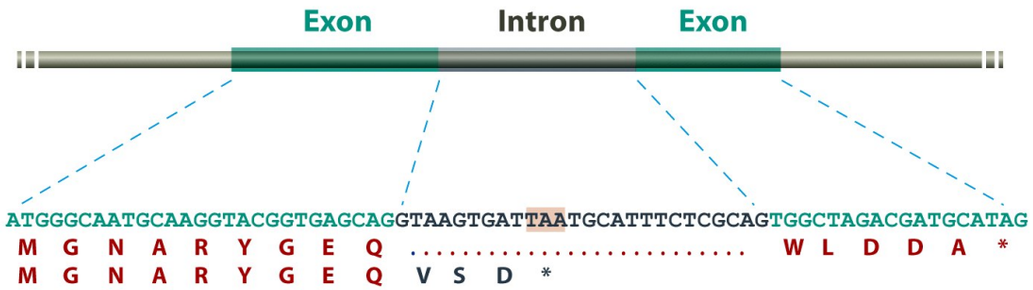
  (2) Exon-Intron boundary consensus sequences
  (introns in prokaryotes are extremely rare)

  (3) Upstream regulatory sequences

  (4) Upstream CpG islands (vertebrate)

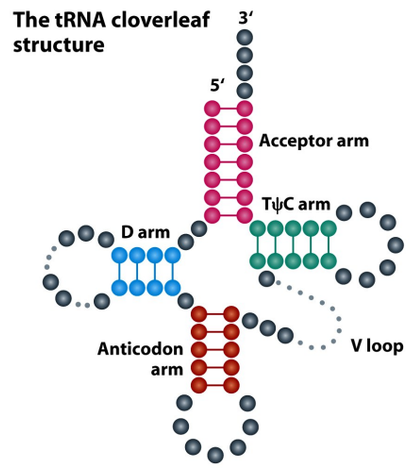
  (5) Comparative genomics (use homology search)

  (6) Databases of cDNA sequences



Locating the Genes for Functional RNAs

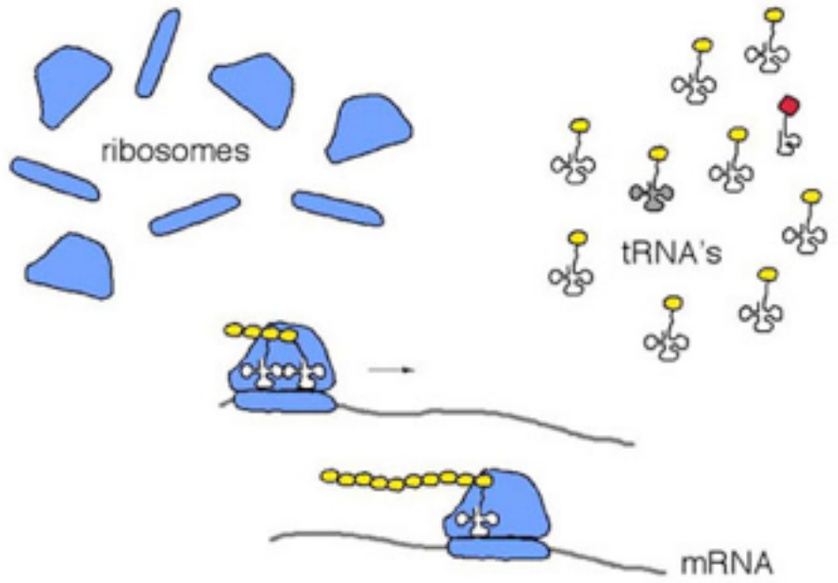
- signatures of intramolecular base pairing
- tRNA, ...



Sequence of one of the *Escherichia coli* tRNA^{leu} genes

5' GCCGAAGTGGCGAAATCGGTAGTCGCAGTTGATTCAAAATCAACCGTAGAAATACGTGCCGGTTCGAGTCCGGCCTTCGGCACCA 3'

Codon bias



CODON USAGE IN *E. COLI* GENES[1]

	Codon	Amino acid[2]	%[3]	Ratio[4]	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio	Codon	Amino acid	%	Ratio	
U	**UUU**	**Phe (F)**	1.9	0.51	**UCU**	**Ser (S)**	1.1	0.19	**UAU**	**Tyr (Y)**	1.6	0.53	**UGU**	**Cys (C)**	0.4	0.43	**U**
	UUC	**Phe (F)**	1.8	0.49	**UCC**	**Ser (S)**	1.0	0.17	**UAC**	**Tyr (Y)**	1.4	0.47	**UGC**	**Cys (C)**	0.6	0.57	**C**
	UUA	**Leu (L)**	1.0	0.11	**UCA**	**Ser (S)**	0.7	0.12	**UAA**	**STOP**	0.2	0.62	**UGA**	**STOP**	0.1	0.30	**A**
	UUG	**Leu (L)**	1.1	0.11	**UCG**	**Ser (S)**	0.8	0.13	**UAG**	**STOP**	0.03	0.09	**UGG**	**Trp (W)**	1.4	1.00	**G**
C	**CUU**	**Leu (L)**	1.0	0.10	**CCU**	**Pro (P)**	0.7	0.16	**CAU**	**His (H)**	1.2	0.52	**CGU**	**Arg (R)**	2.4	0.42	**U**
	CUC	**Leu (L)**	0.9	0.10	**CCC**	**Pro (P)**	0.4	0.10	**CAC**	**His (H)**	1.1	0.48	**CGC**	**Arg (R)**	2.2	0.37	**C**
	CUA	**Leu (L)**	0.3	0.03	**CCA**	**Pro (P)**	0.8	0.20	**CAA**	**Gln (Q)**	1.3	0.31	**CGA**	**Arg (R)**	0.3	0.05	**A**
	CUG	**Leu (L)**	5.2	0.55	**CCG**	**Pro (P)**	2.4	0.55	**CAG**	**Gln (Q)**	2.9	0.69	**CGG**	**Arg (R)**	0.5	0.08	**G**
A	**AUU**	**Ile (I)**	2.7	0.47	**ACU**	**Thr (T)**	1.2	0.21	**AAU**	**Asn (N)**	1.6	0.39	**AGU**	**Ser (S)**	0.7	0.13	**U**
	AUC	**Ile (I)**	2.7	0.46	**ACC**	**Thr (T)**	2.4	0.43	**AAC**	**Asn (N)**	2.6	0.61	**AGC**	**Ser (S)**	1.5	0.27	**C**
	AUA	**Ile (I)**	0.4	0.07	**ACA**	**Thr (T)**	0.1	0.30	**AAA**	**Lys (K)**	3.8	0.76	**AGA**	**Arg (R)**	0.2	0.04	**A**
	AUG	**Met (M)**	2.6	1.00	**ACG**	**Thr (T)**	1.3	0.23	**AAG**	**Lys (K)**	1.2	0.24	**AGG**	**Arg (R)**	0.2	0.03	**G**
G	**GUU**	**Val (V)**	2.0	0.29	**GCU**	**Ala (A)**	1.8	0.19	**GAU**	**Asp (D)**	3.3	0.59	**GGU**	**Gly (G)**	2.8	0.38	**U**
	GUC	**Val (V)**	1.4	0.20	**GCC**	**Ala (A)**	2.3	0.25	**GAC**	**Asp (D)**	2.3	0.41	**GGC**	**Gly (G)**	3.0	0.40	**C**
	GUA	**Val (V)**	1.2	0.17	**GCA**	**Ala (A)**	2.1	0.22	**GAA**	**Glu (E)**	4.4	0.70	**GGA**	**Gly (G)**	0.7	0.09	**A**
	GUG	**Val (V)**	2.4	0.34	**GCG**	**Ala (A)**	3.2	0.34	**GAG**	**Glu (E)**	1.9	0.30	**GGG**	**Gly (G)**	0.9	0.13	**G**
	U				**C**				**A**				**G**				

[1] The data shown in this table is from the Arabidopsis Research Companion on the World Wide Web (//weeds/mgh.harvard.edu). Codon frequencies for many other bacteria can be found at http://morgan.angis.su.oz.au/Angis/Tables.html.

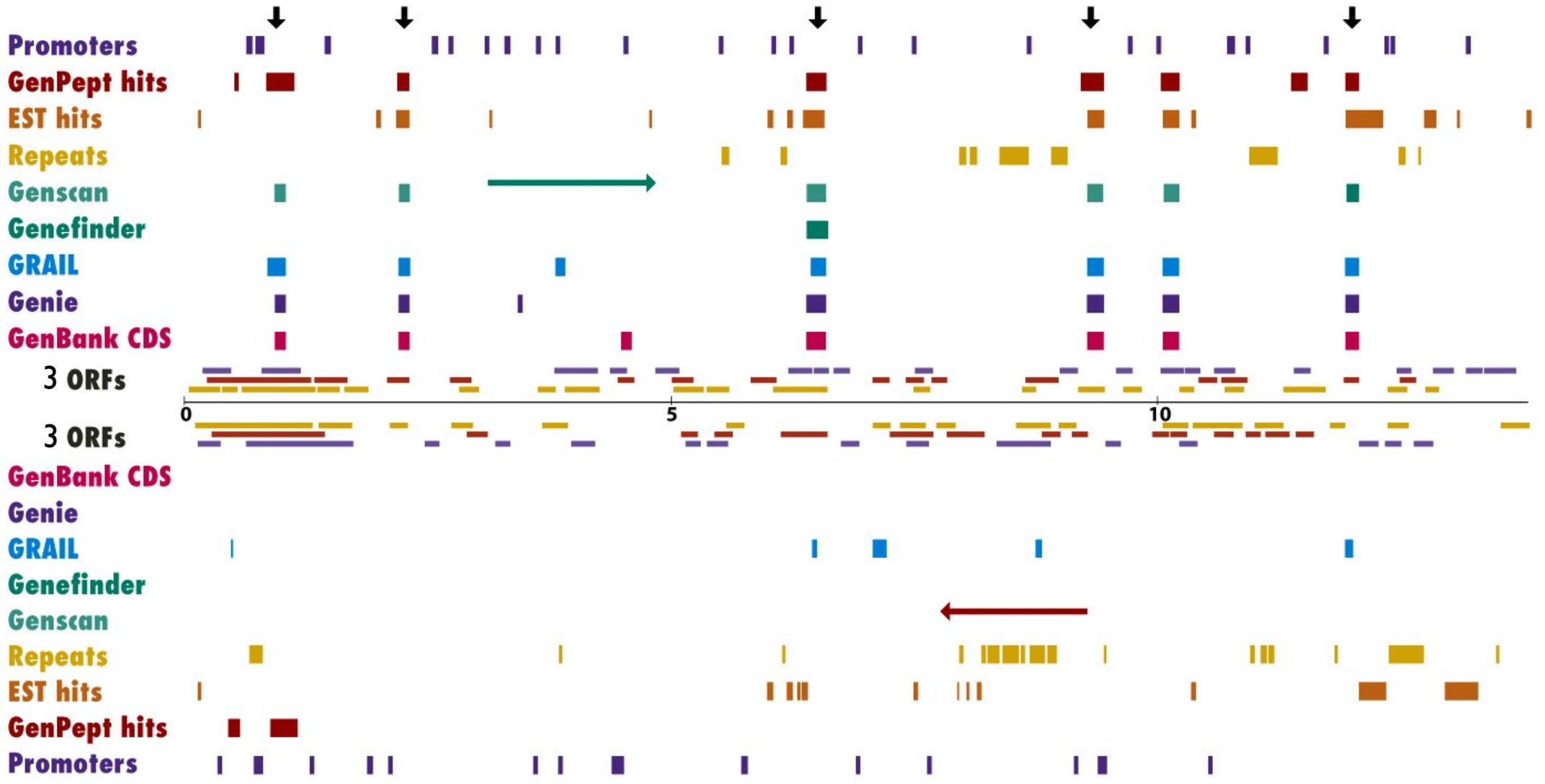
[2] The letter in parenthesis represents the one-letter code for the amino acid.

[3] % represents the average frequency this codon is used per 100 codons.

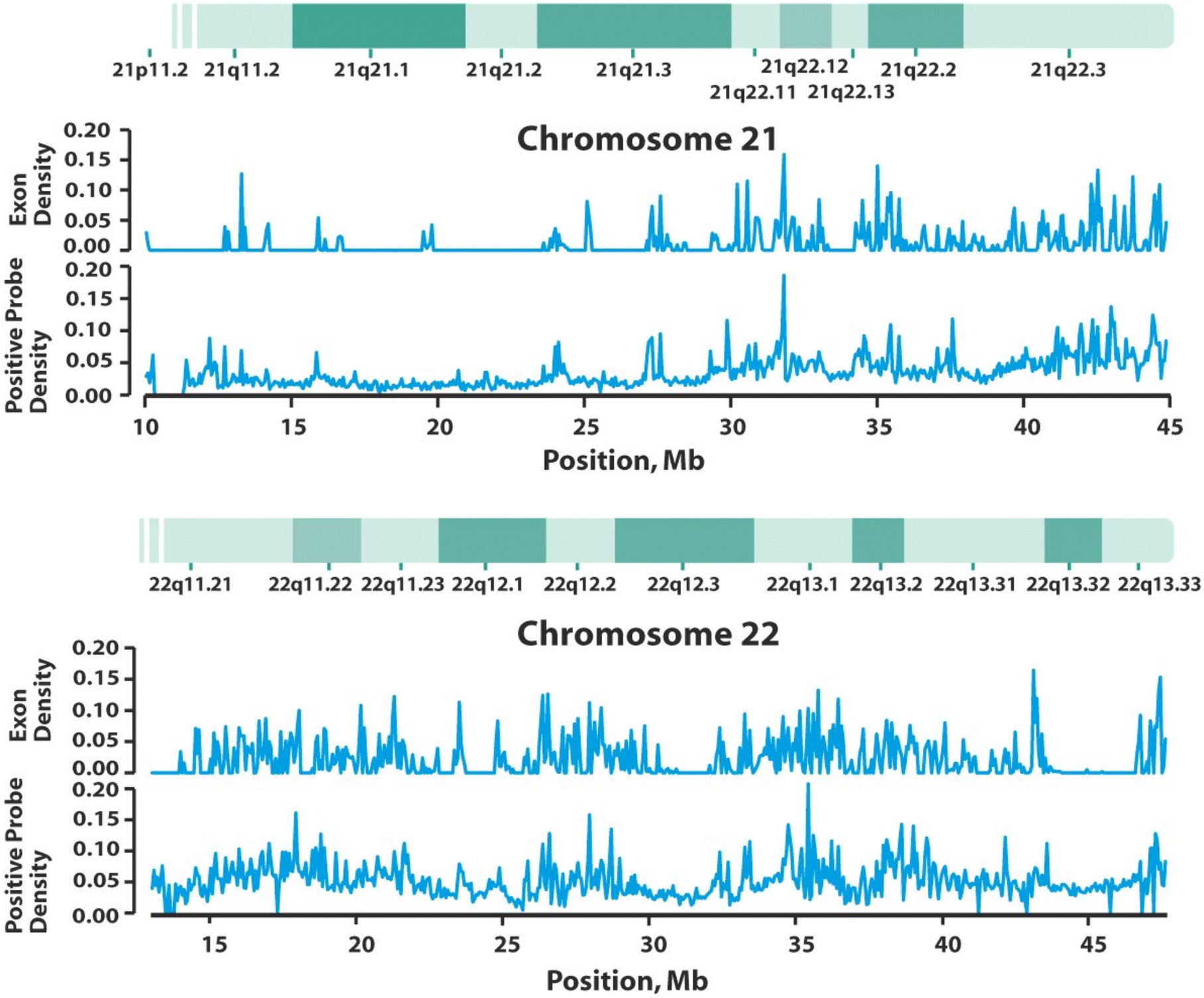
[4] Ratio represents the abundance of that codon relative to all of the codons for that particular amino acid.

Annotation of genome sequences

15 kb segment of the human genome containing a tissue factor gene



Distribution of exon in genome?



Some interesting facts about human genome

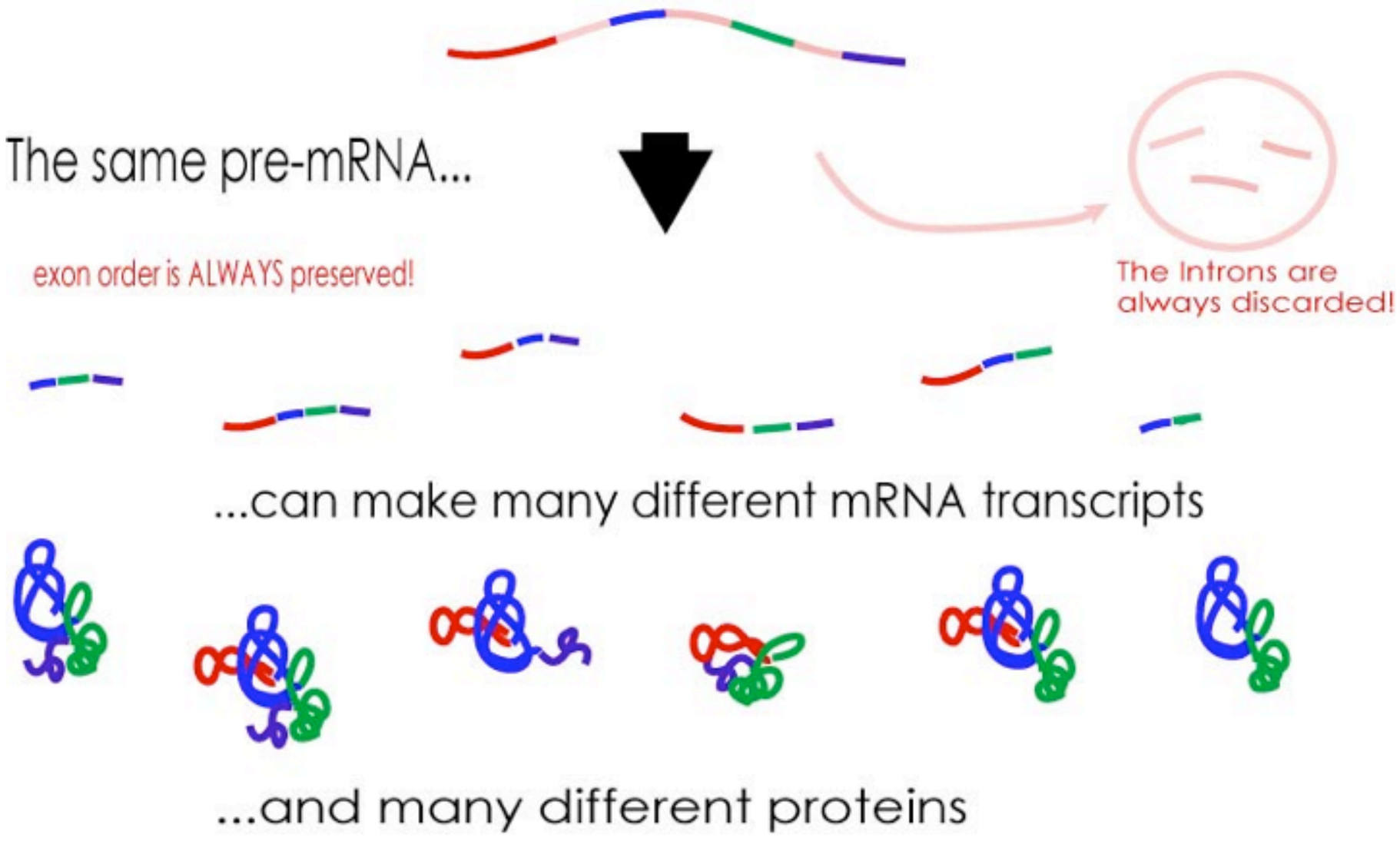
- The smallest protein-coding gene in the human genome is only 500-nt long and has no introns. It encodes a histone protein.
- The largest human gene encodes the protein dystrophin, which is missing or non-functional in the disease muscular dystrophy. This gene is 2.5 million nucleotides in length and it takes over 16 hours to produce a single transcript. However, more than 99 percent of the gene made up of its 79 introns.
- Most of the big differences between human and chimpanzee DNA lie in regions that do not code for genes, according to a new study. Instead, they may contain DNA sequences that control how gene-coding regions are activated and read. - "The differences between chimps and humans are not in our proteins, but in how we use them,"

Sequence variation

- Single Nucleotide Polymorphism (SNP)
- Insertion or deletions of one or more bases
- Repeat length polymorphism, rearrangement

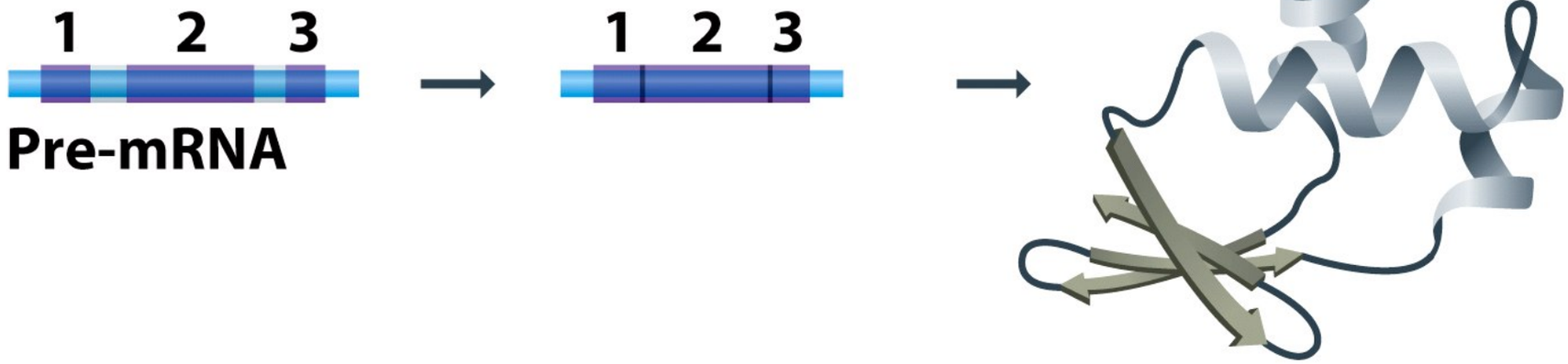
Nonredundant SNP = 1,419,190 in human genome sequence
Average SNP density ≃ one SNP/1.91kb

Alternative Splicing

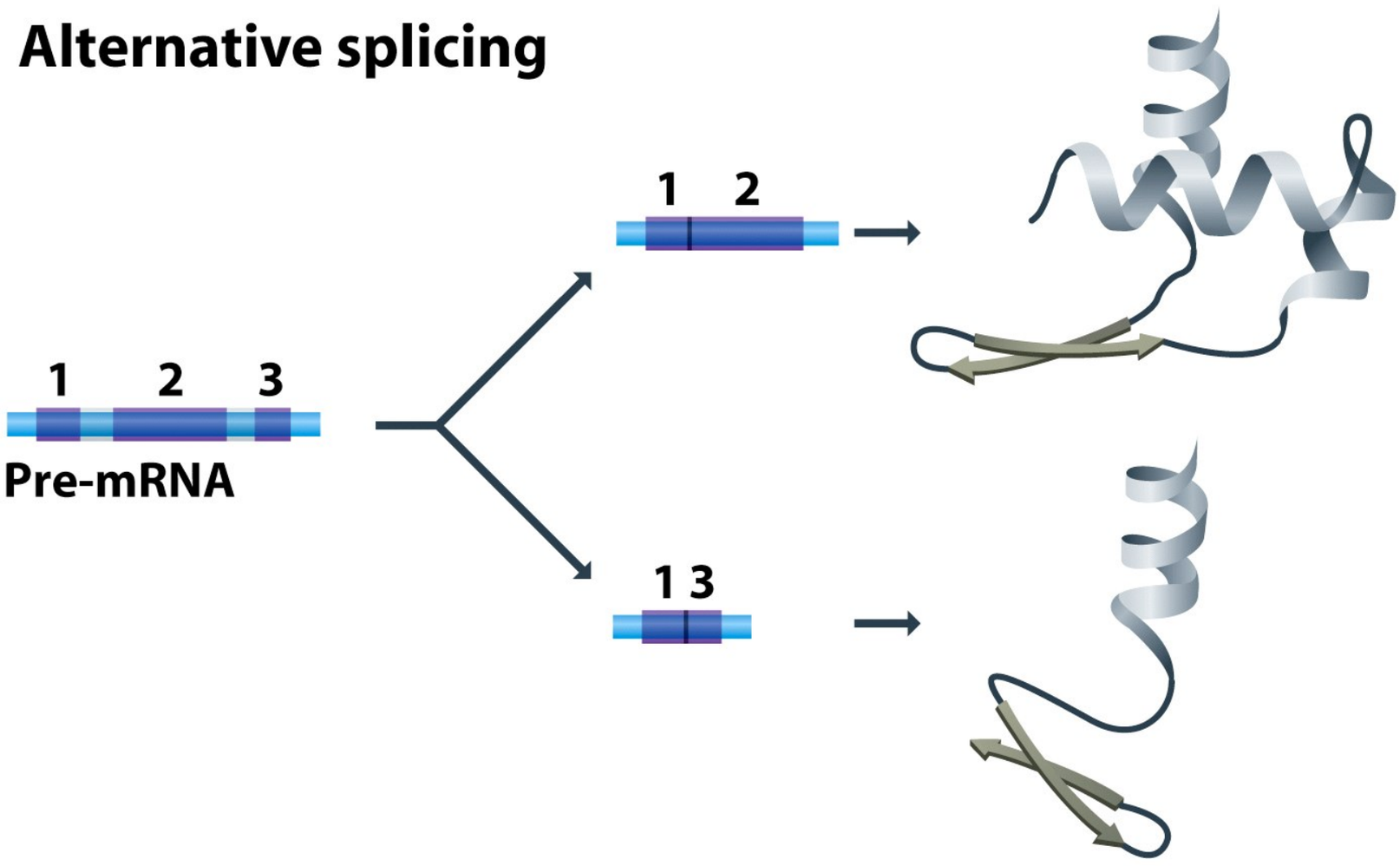


Note that our genome and chimpanzee's are 99 % identical. However, we humans are very different from them. Gene regulation through transcriptome and proteome is extremely important for species variation !

A single splicing pathway



Alternative splicing

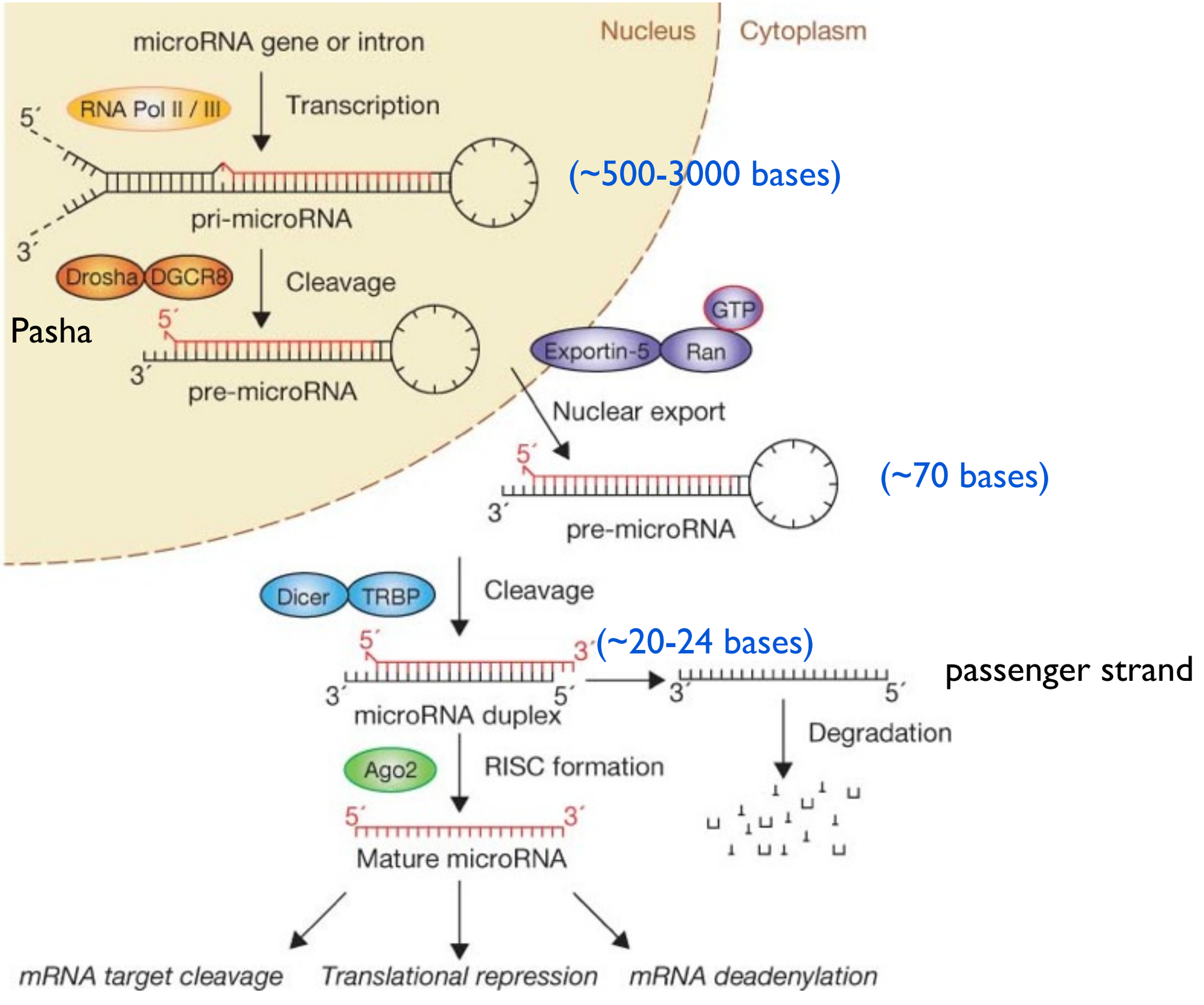


miRNAs

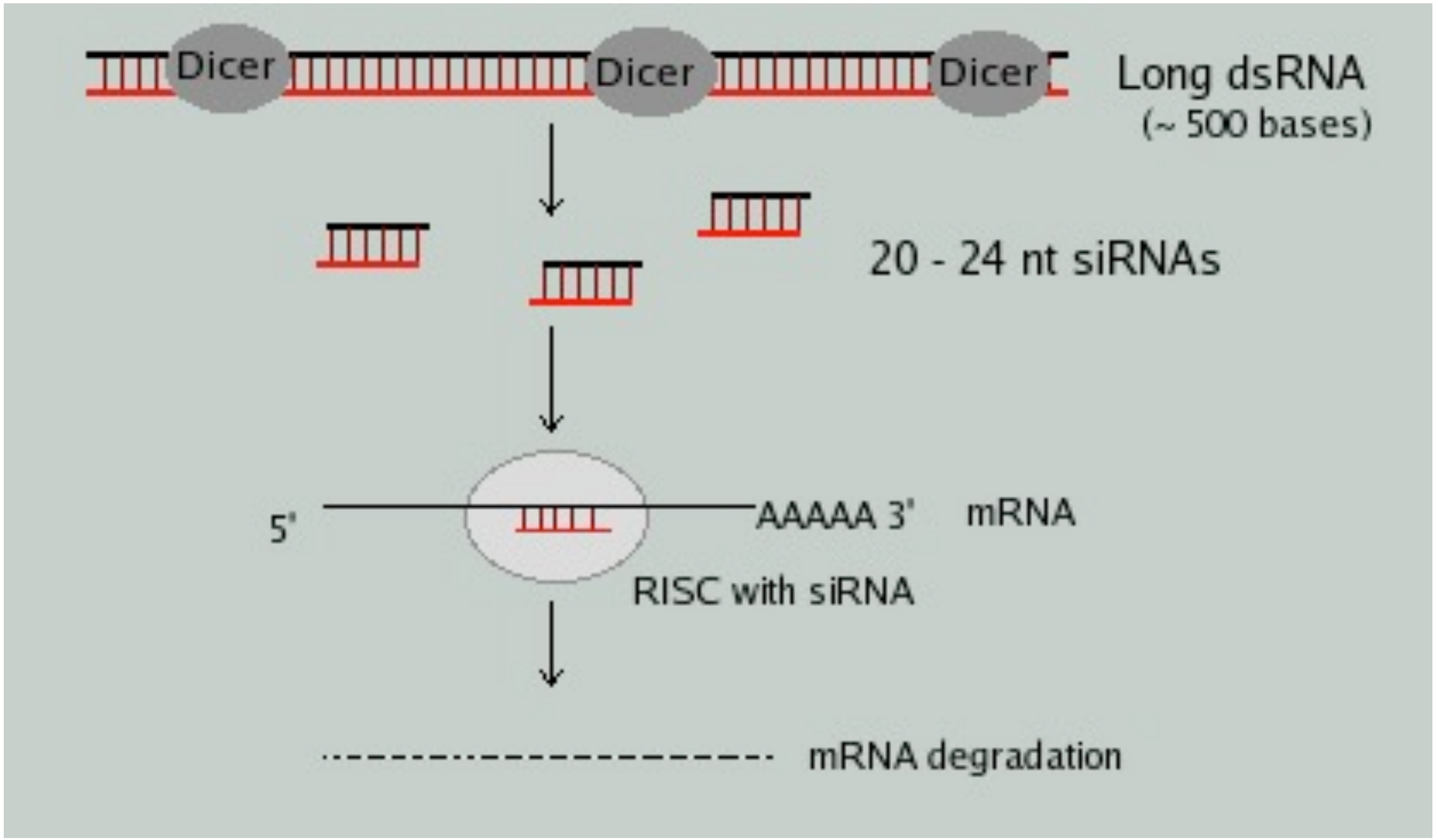
- MicroRNAs are evolutionarily conserved, small (~22 nucleotide) noncoding RNAs that are encoded within the genomes of almost all eukaryotes, from plants to mammals.
- Derive from dedicated genes, exons and introns of coding genes.
- Base pairing to partially complementary sequences in the 3' untranslated regions (3'UTRs) of target mRNAs.
- miRNA mediate mRNA repression by recruiting the miRNA-induced silencing complex (miRISC)
- Computer-assisted estimates predict ~1000 miRNAs in the human genome.
- miRNAs have multiple targets and thus might regulate ~30% of the protein-coding genome.
- miRNAs function in a variety of biological processes, including tissue differentiation and organ development, control of cell proliferation, apoptosis, fat metabolism and insulin secretion.
- miRNAs are more than a fine-tuning agent, should be upgraded to be a partner of TF.

miRNA biogenesis pathways

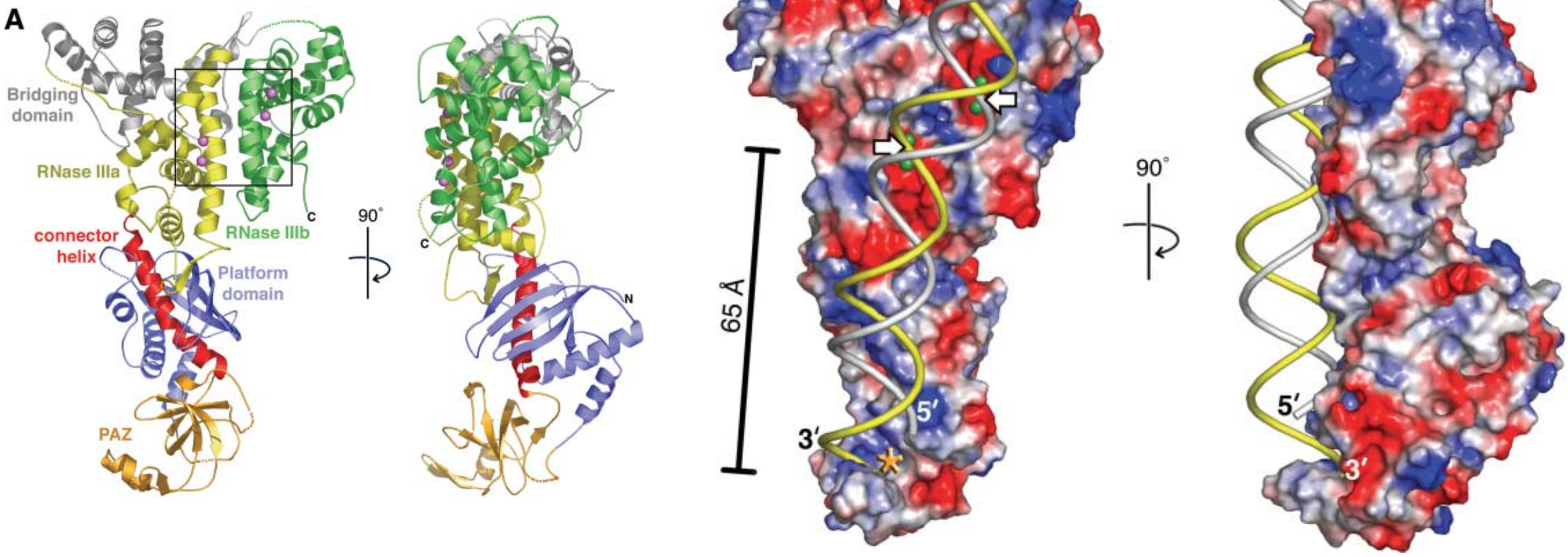
Nature Cell Biology **11**, 228 - 234 (2009)

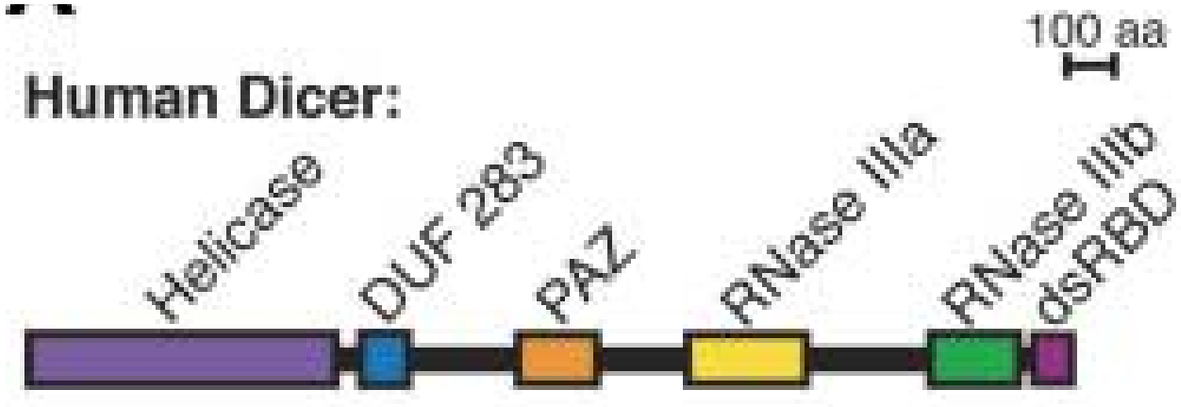


Biogenesis pathway of siRNA (exogeneous)

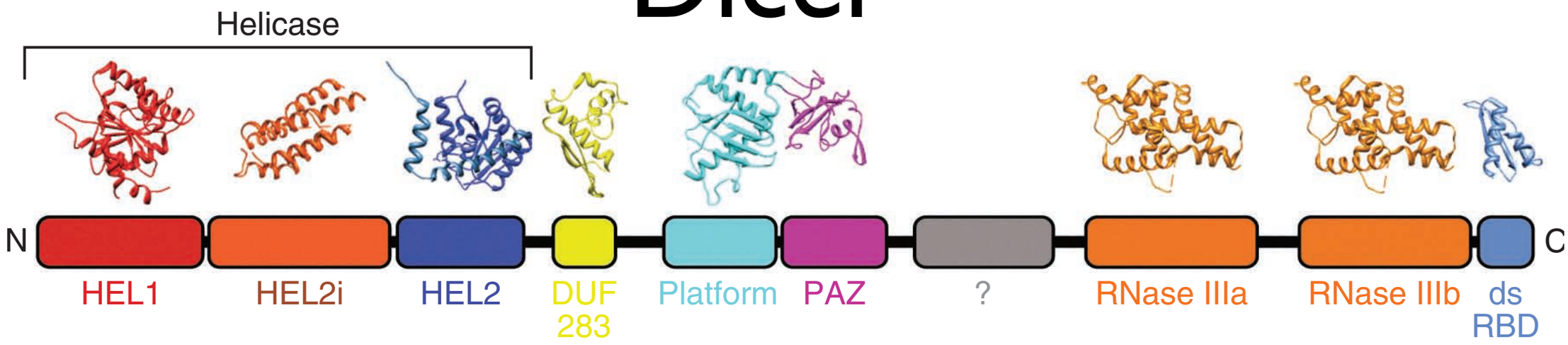


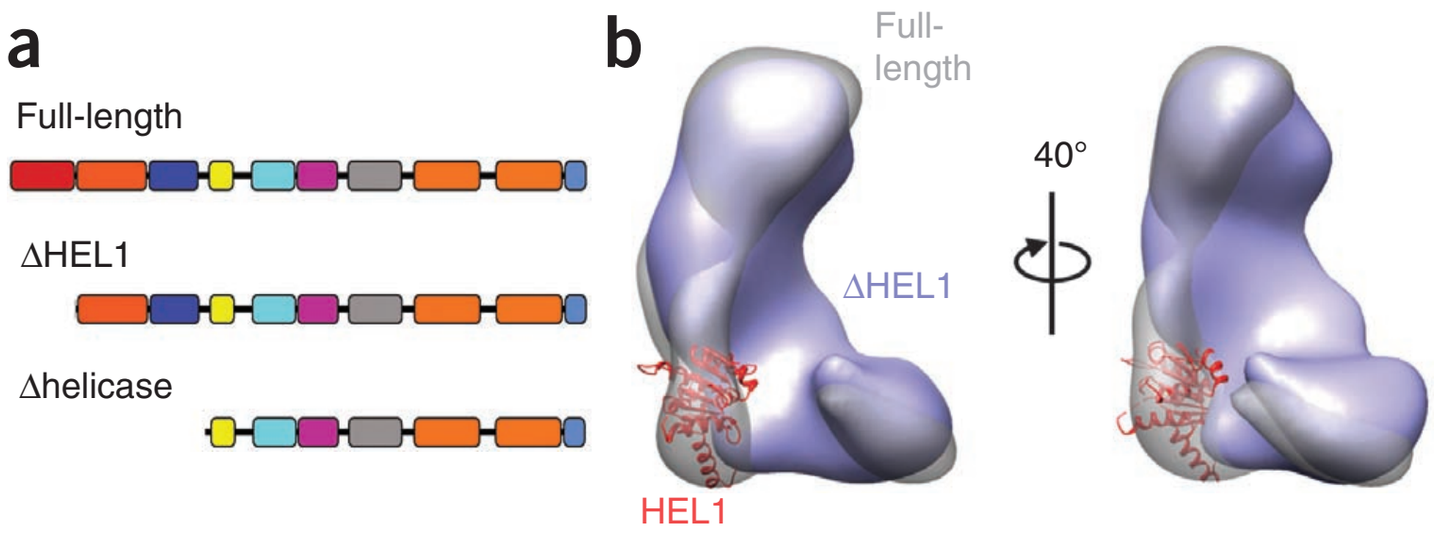
Dicer

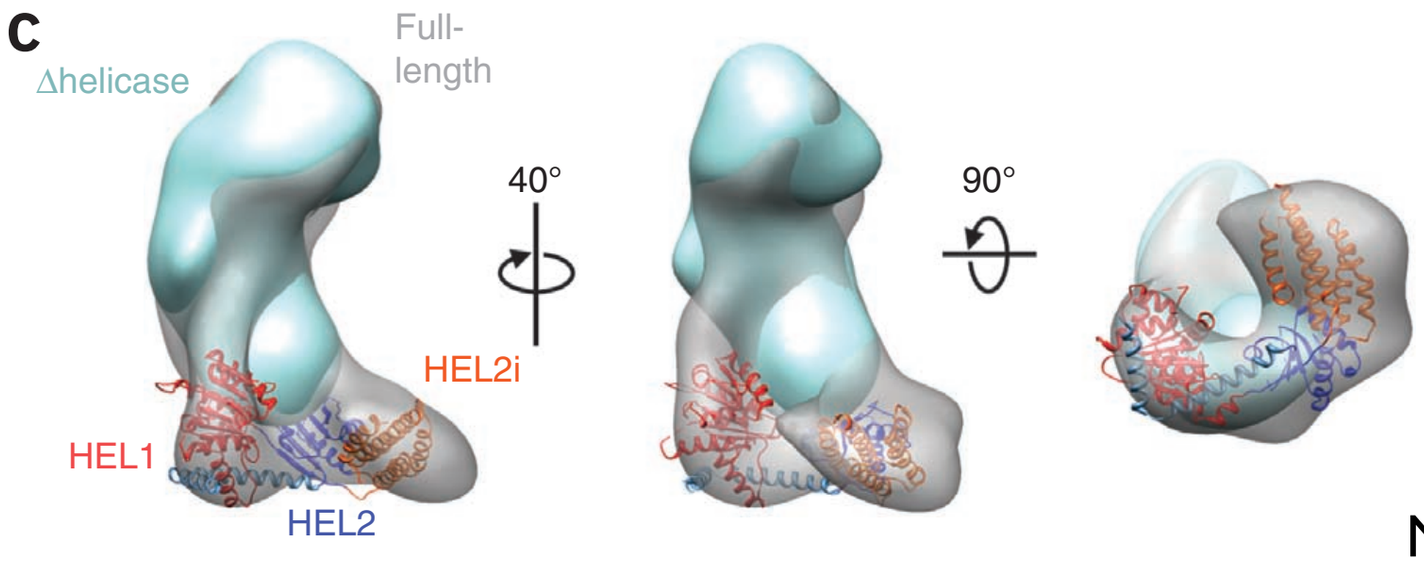

A
Bridging domain
RNase IIIa
RNase IIIb
connector helix
Platform domain
PAZ
C
N
90˚
65 Å
3′
5′
90˚
5′
3′


100 aa
Human Dicer:
Helicase
DUF 283
PAZ
RNase IIIa
RNase IIIb
dsRBD

Dicer

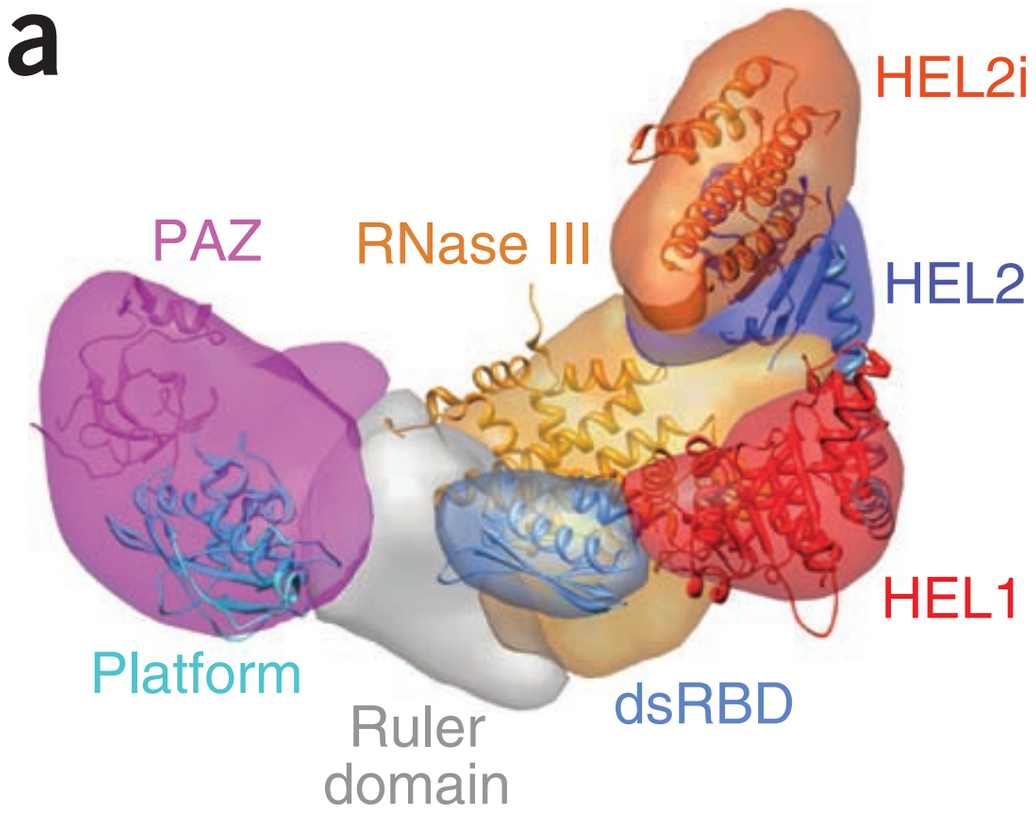


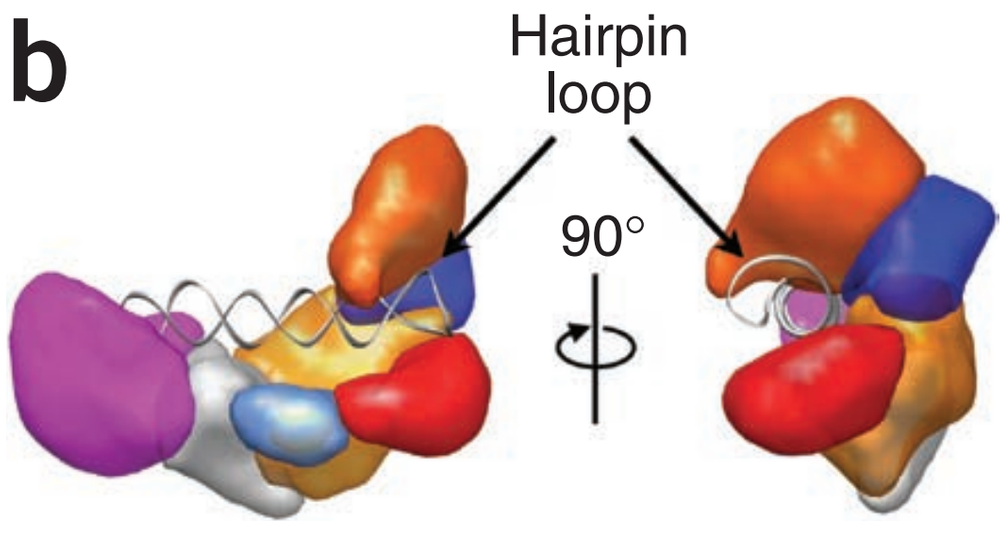


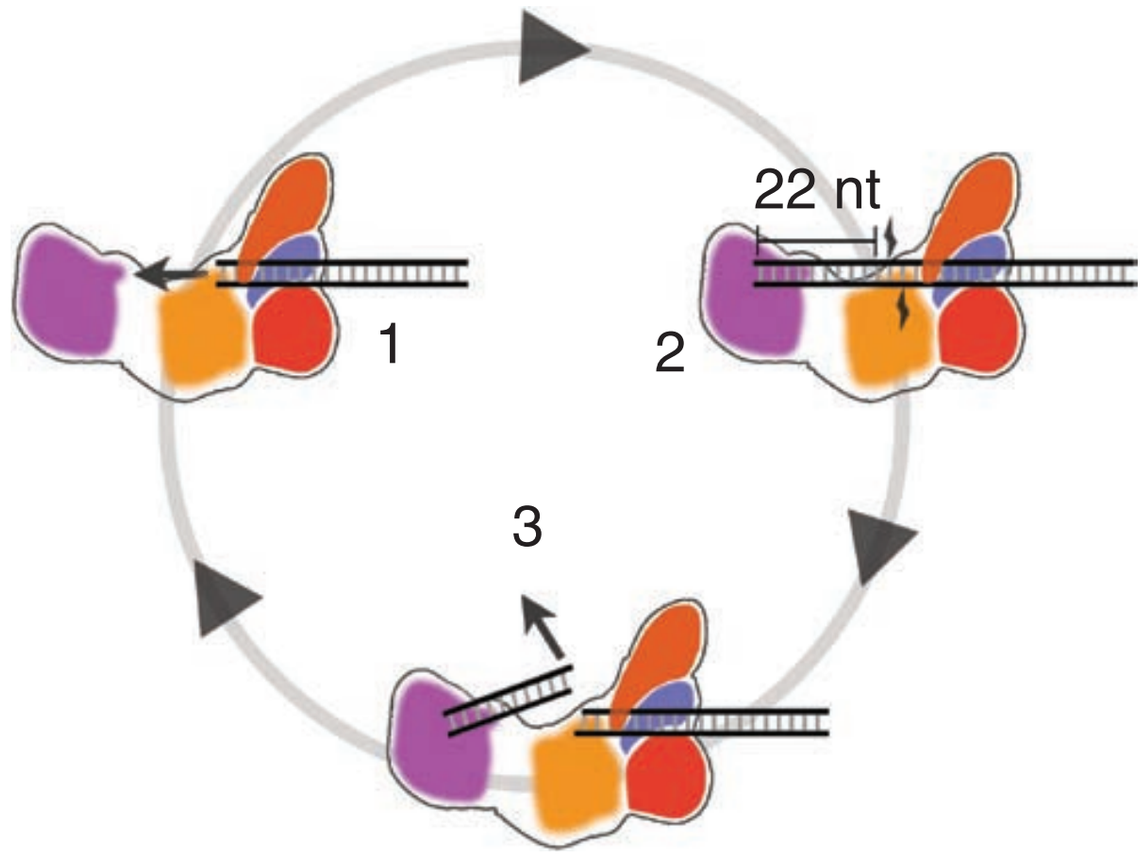


NSMB 19: 436 (2012)

Dicer



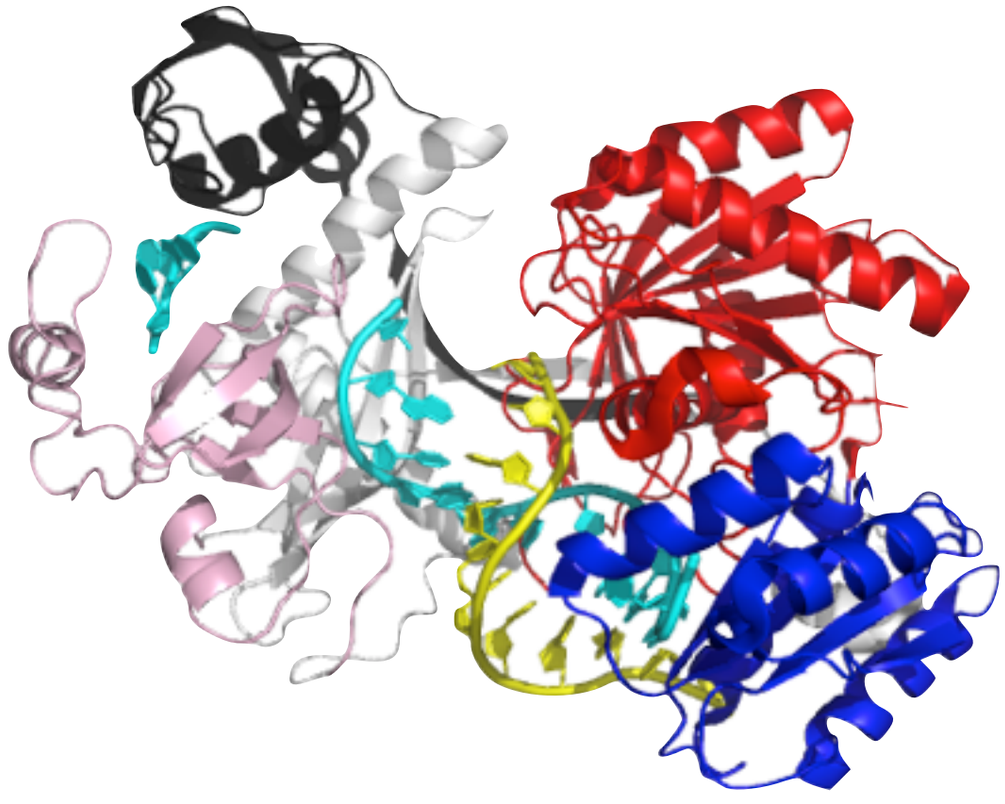




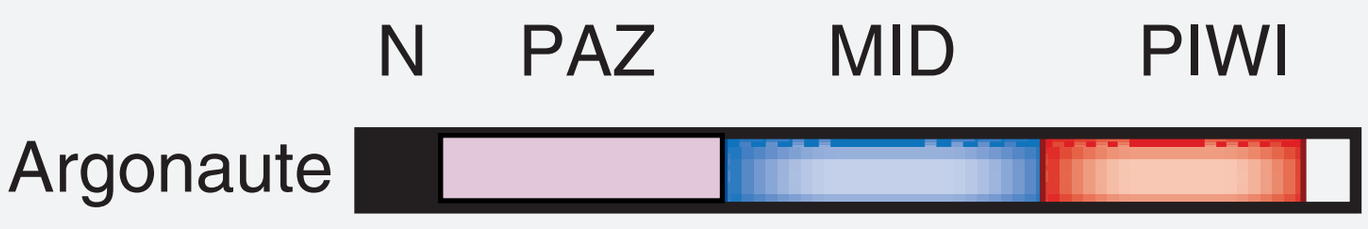
NSMB 19: 436 (2012)

Argonaute (AGO) proteins

- Bilobal architecture consisting of four evolutionarily conserved domains:
  - N-terminal and Piwi-Argonaute-PAZ domains

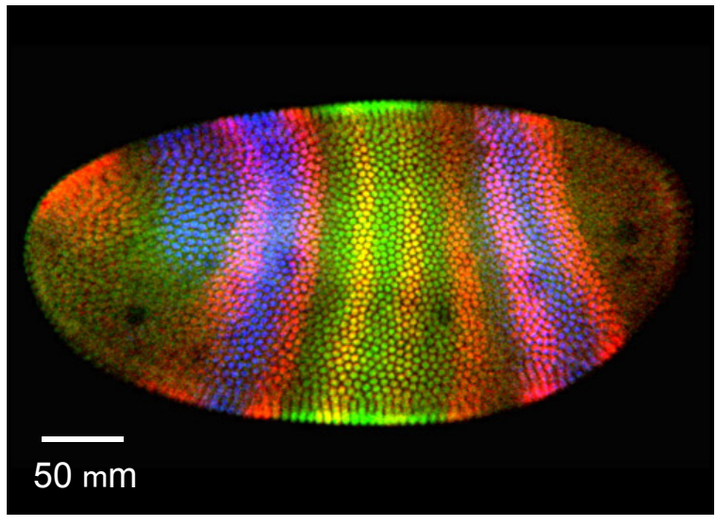


RISC

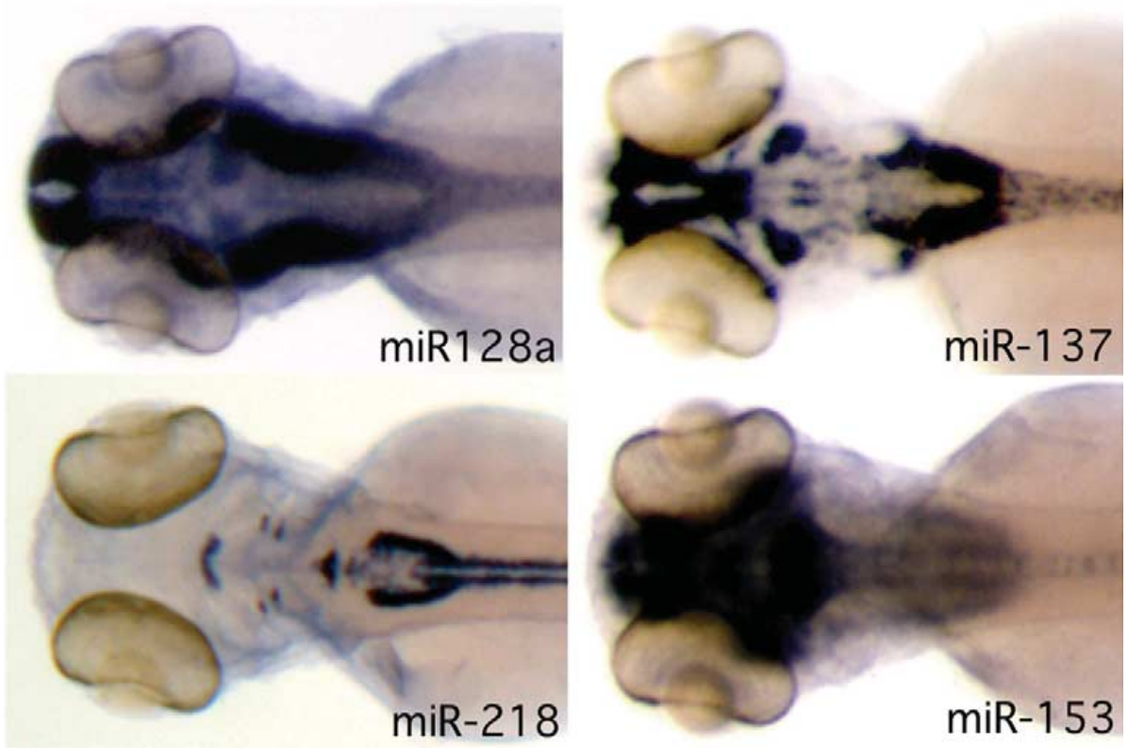


Regulatory role of microRNAs

Drosophila embryo



zebrafish embryos



http://www.mirbase.org/index.shtml

Stem-loop sequence hsa-mir-17

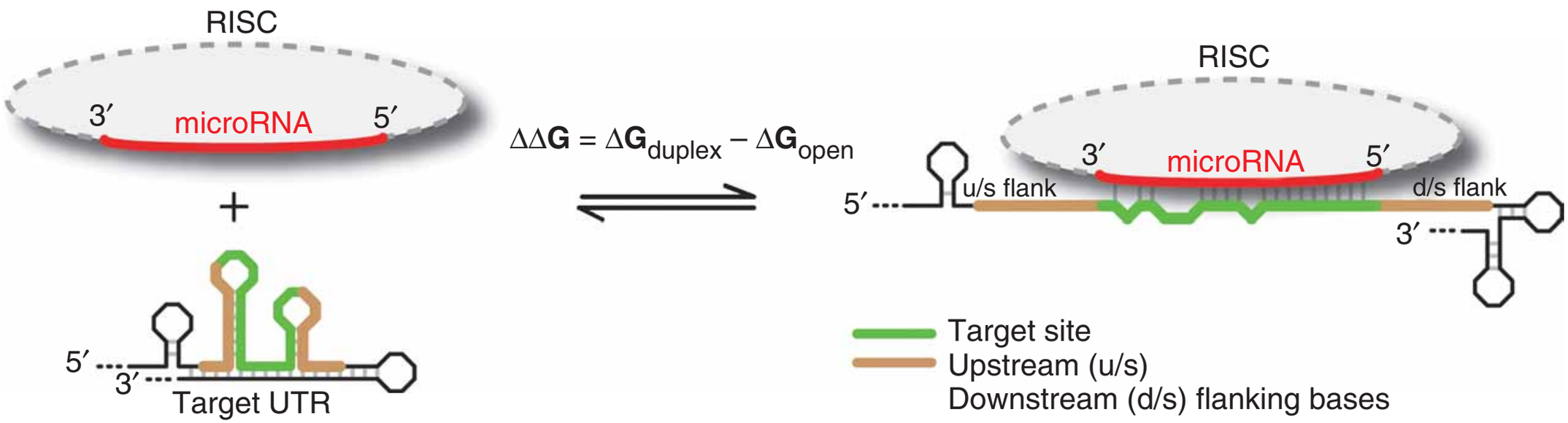
Accession	MI0000071
Symbol	HGNC:MIR17
Description	Homo sapiens miR-17 stem-loop
Gene family	MIPF0000001; mir-17
Community annotation	This text is a summary paragraph taken from the Wikipedia entry entitled mir-17_microRNA_precursor_family. miRBase a The miR-17 microRNA precursor family are a group of related small non-coding RNA family, includes miR-20, miR-91, and miR-103. miRNAs are transcribed as ~70 nucle product. In this case the mature sequence comes from the 3' arm of the precursor. T mRNA. A screen of 17 miRNAs that have been predicted to regulate a number of brea these patients were noncarriers of BRCA1 or BRCA2 mutations, lending the possibility Show Wikipedia entry View @ Wikipedia Edit Wikipedia entry
Stem-loop	ga -ca a g g - au guca auaaugu aagugcuu ca ugcag uag ug a cagu uauuacg uucacgga gu acguc auc ac g gg aug a g - u gu Get sequence
Deep sequencing	35854 reads, 48 experiments reads GUCAGAAUAAUGUCAAAGUGCUUACAGUGCAGGUAGUGAUAUGUGCAUCUACUGCAGUGAAGGCACUUGUAGCAUUAUGGUGAC

Mature sequence hsa-miR-17-5p

Accession	MIMAT0000070
Previous IDs	hsa-miR-17-5p;hsa-miR-17
Sequence	14 - caaagugcuuacagugcagguag - 36 Get sequence
Deep sequencing	31906 reads, 42 experiments
Evidence	experimental; cloned [2,5-8], Northern [4]
Validated targets	TARBASE: hsa-miR-17-5p
Predicted targets	DIANA-MICROT: hsa-miR-17-5p MICRORNA.ORG: hsa-miR-17-5p MIRDB: hsa-miR-17-5p RNA22-HSA: hsa-miR-17-5p TARGETMINER: hsa-miR-17-5p TARGETSCAN-VERT: hsa-miR-17 PICTAR-VERT: hsa-miR-17-3p PICTAR-VERT: hsa-miR-17-5p

Mature sequence hsa-miR-17-3p

Accession	MIMAT0000071
Previous IDs	hsa-miR-17-3p;hsa-miR-17*
Sequence	51 - acugcagugaaggcacuuguag - 72 Get sequence
Deep sequencing	3947 reads, 33 experiments
Evidence	experimental; cloned [1,5,7-8], Northern [1]
Validated targets	TARBASE: hsa-miR-17-3p
Predicted targets	DIANA-MICROT: hsa-miR-17-3p MICRORNA.ORG: hsa-miR-17-3p MIRDB: hsa-miR-17-3p RNA22-HSA: hsa-miR-17-3p TARGETMINER: hsa-miR-17-3p PICTAR-VERT: hsa-miR-17-3p PICTAR-VERT: hsa-miR-17-5p



PITA (by Eran Segal)

http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html

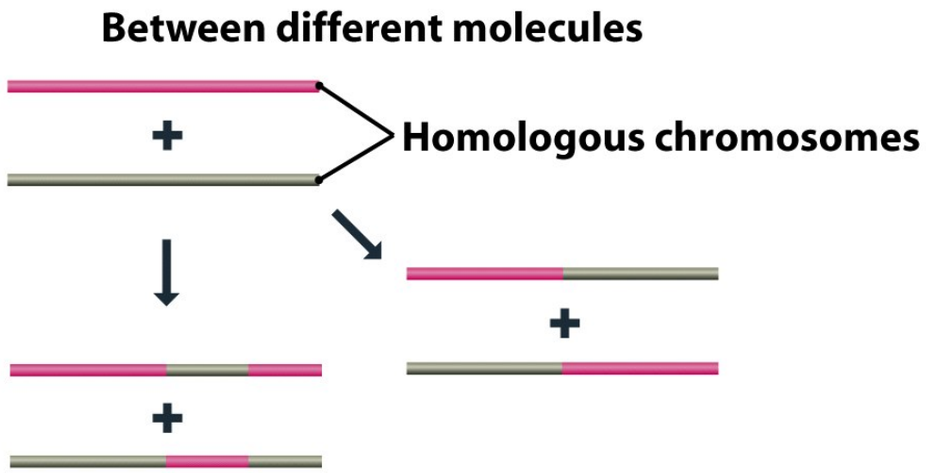
cf. seven base-pair rule

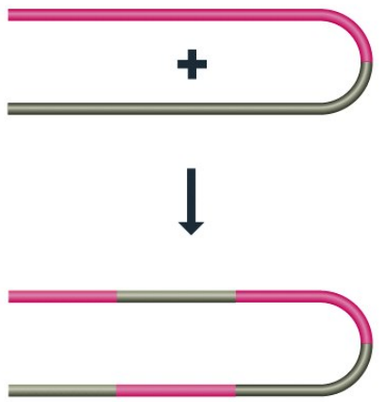
piRNA (Piwi interacting RNA)

- The largest class of small ncRNA molecules expressed in animal cells.
- forms RNA-protein complexes through interactions with piwi proteins
- , which leads to epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germ line cells.
- Distinct from microRNA or siRNA in size (26-31 nt)
- has highly complex and heterogeneous population revealed by recent DNA sequencing.
- Biogenesis of piRNAs not fully understood yet.

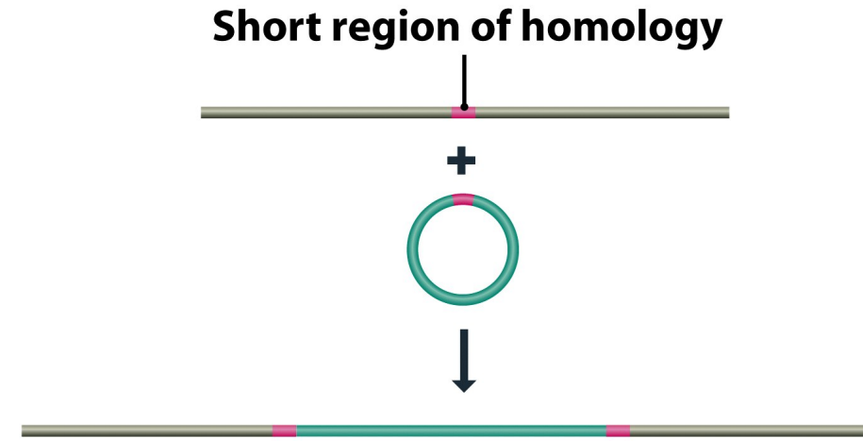
Reorganization of Genome

Homologous recombination

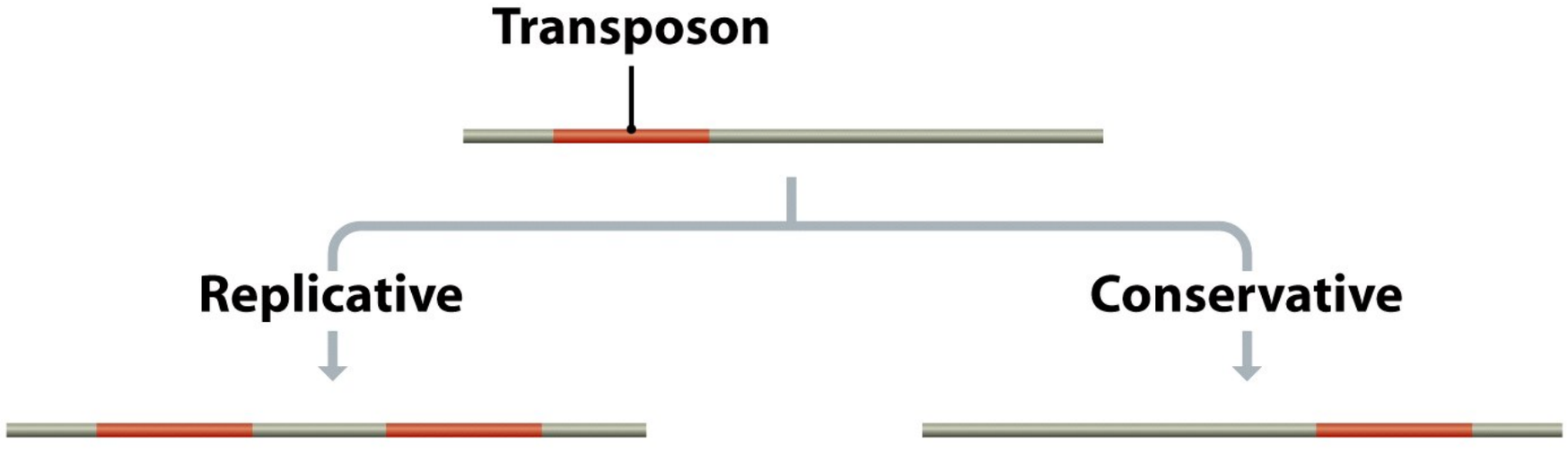




Site-specific recombination



Transposition

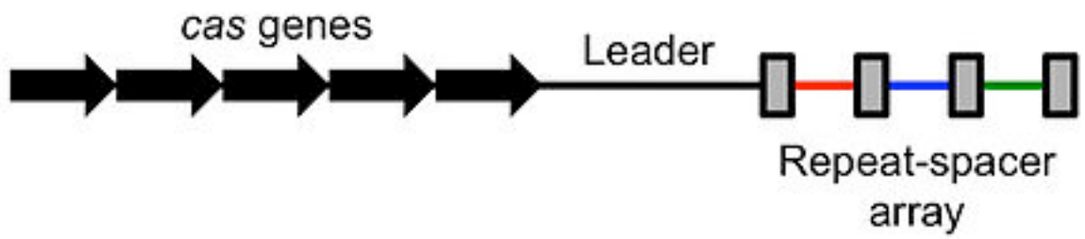


minisatellite : 10-60 bp around core sequence(--GGGCAGGANG--)
microsatellite : 2-13 nt long short tandem repeats

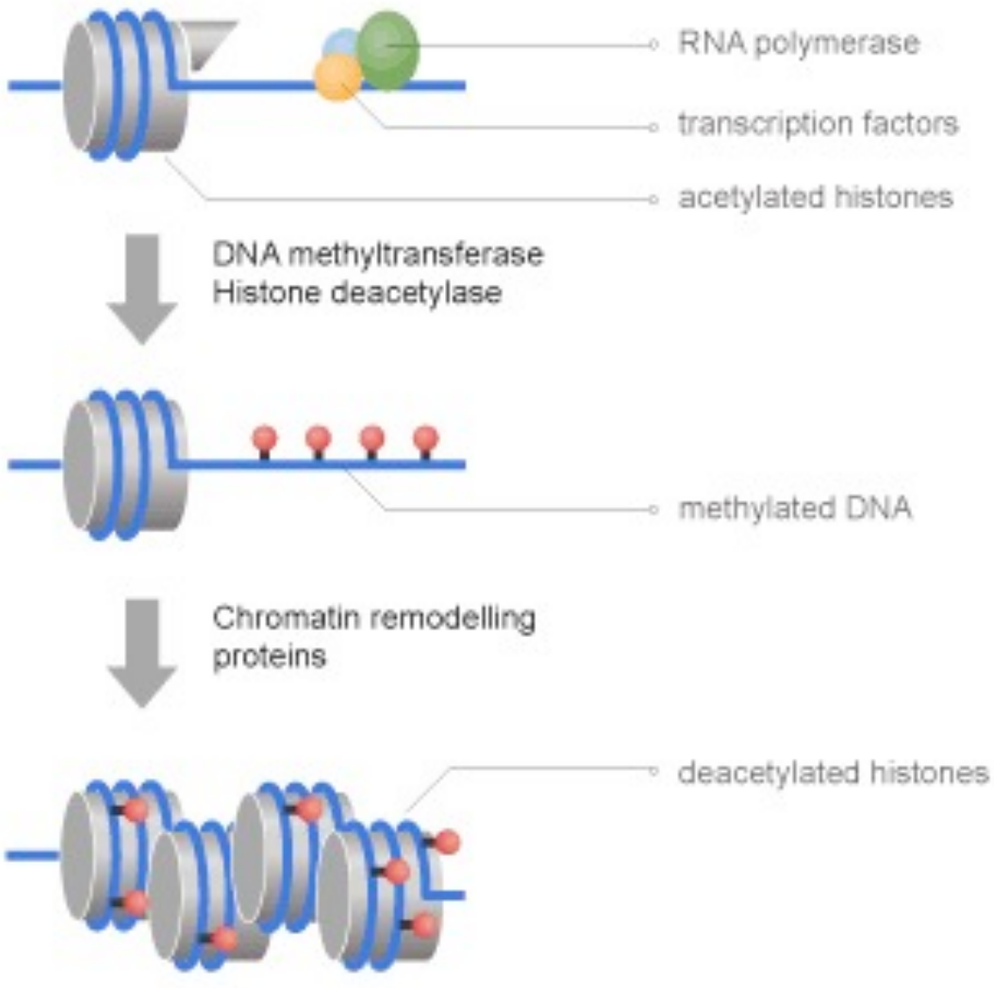
CRISPRs (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats)

Short sequence tags from invading genetic elements are actively incorporated into the host's CRISPR locus to be transcribed and processed into a set of small RNAs that guide the destruction of foreign genetic material.

- Immune system of bacteria and archaea against predators
- Confers resistance to exogenous genetic elements (plasmids, phages).
- Short segments of foreign DNA (spacers) incorporated into genome between CRISPR repeats and serve as a 'memory' of past exposures.
- CRISPR spacers are used to recognize and silence exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

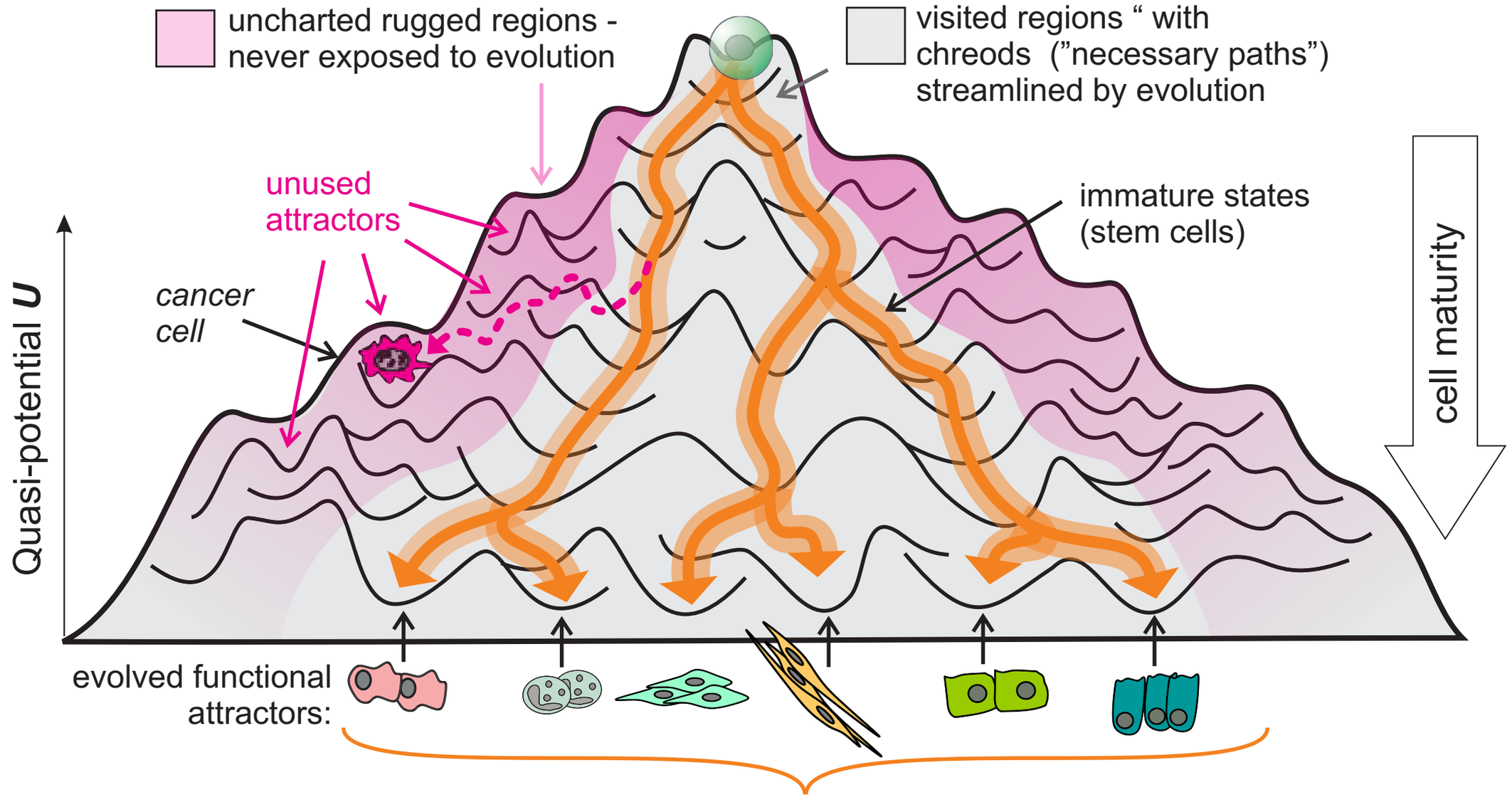


Epigenetics



- Chemical modifications of the DNA (methylation) or the histones (deacetylation) alter the chromatin structure without changing the DNA sequence
- Inappropriate gene silencing (or activation), brought about by epigenetic modification, cause a number of human diseases including cancers.

Epigenetic landscape



Decision-making problem of cells

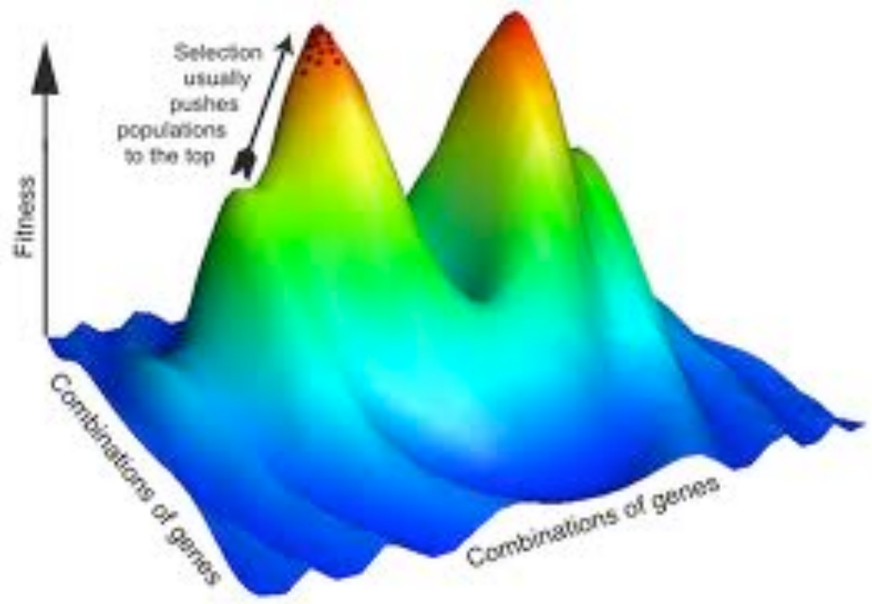
Alternative splicing, recombination, miRNA, piRNA, ...

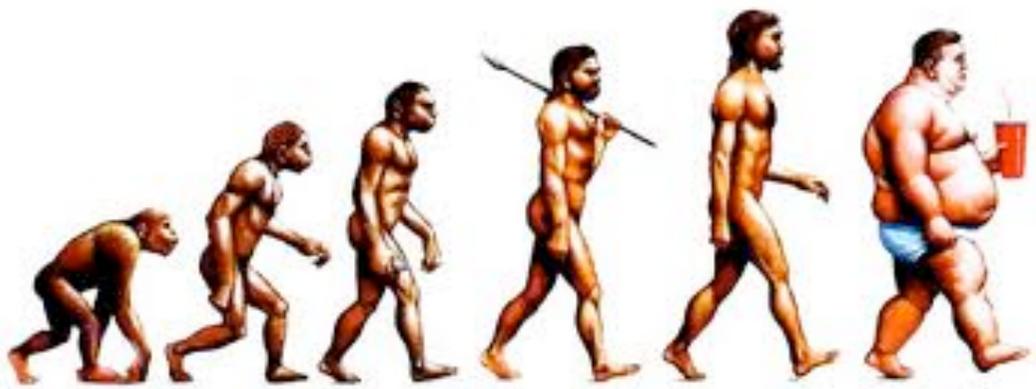
It’s all about REGULATION!!!

S. Huang / Progress in Biophysics and Molecular Biology 110 (2012) 69–86

Evolution

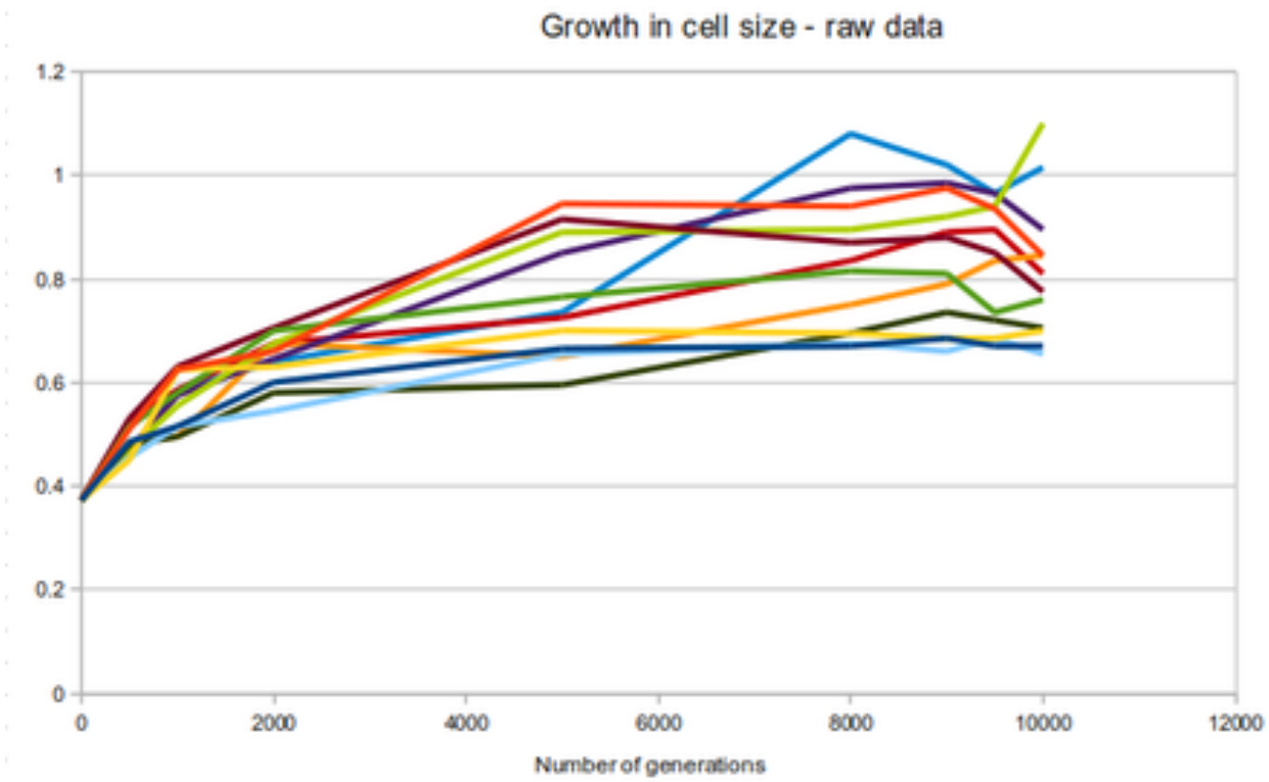
- Natural selection - Survival of the fittest
- Mutation - Malicious / Beneficial mutation





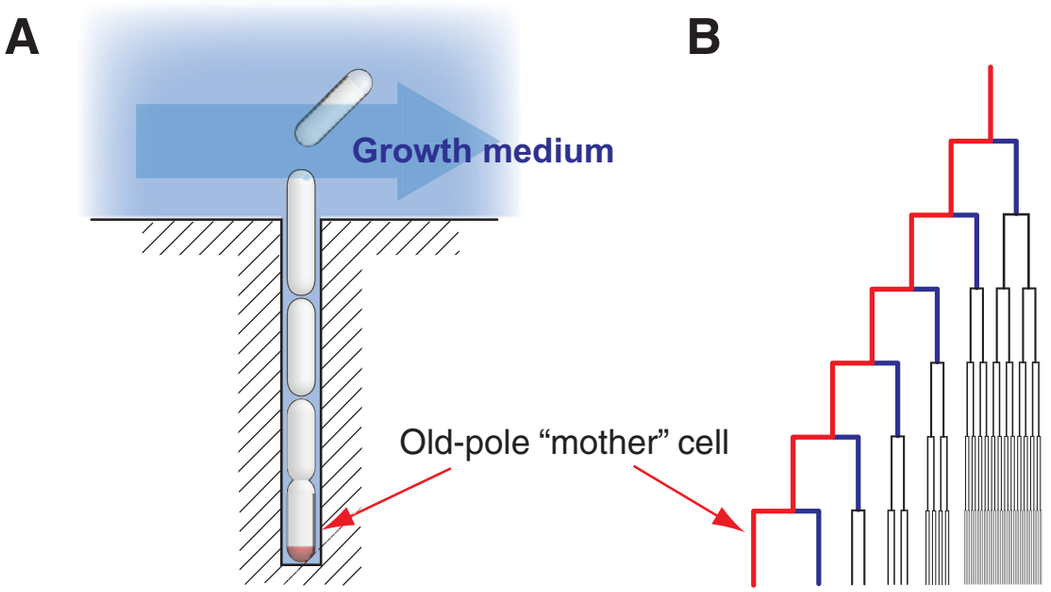
The ***E. coli* long-term evolution experiment** is an ongoing study in experimental evolution led by Richard Lenski that has been tracking genetic changes in 12 initially identical populations of asexual Esherichia coli bacteria since 24 February 1988. The populations reached the milestone of 50,000 generations in February 2010.

Since the experiment's inception, Lenski and his colleagues have reported a wide array of genetic changes; some evolutionary adaptation have occurred in all 12 populations, while others have only appeared in one or a few populations. One particularly striking adaption was the evolution of a strain of *E. coli* that was able to use citric acid as a carbon source in an aerobic environment.



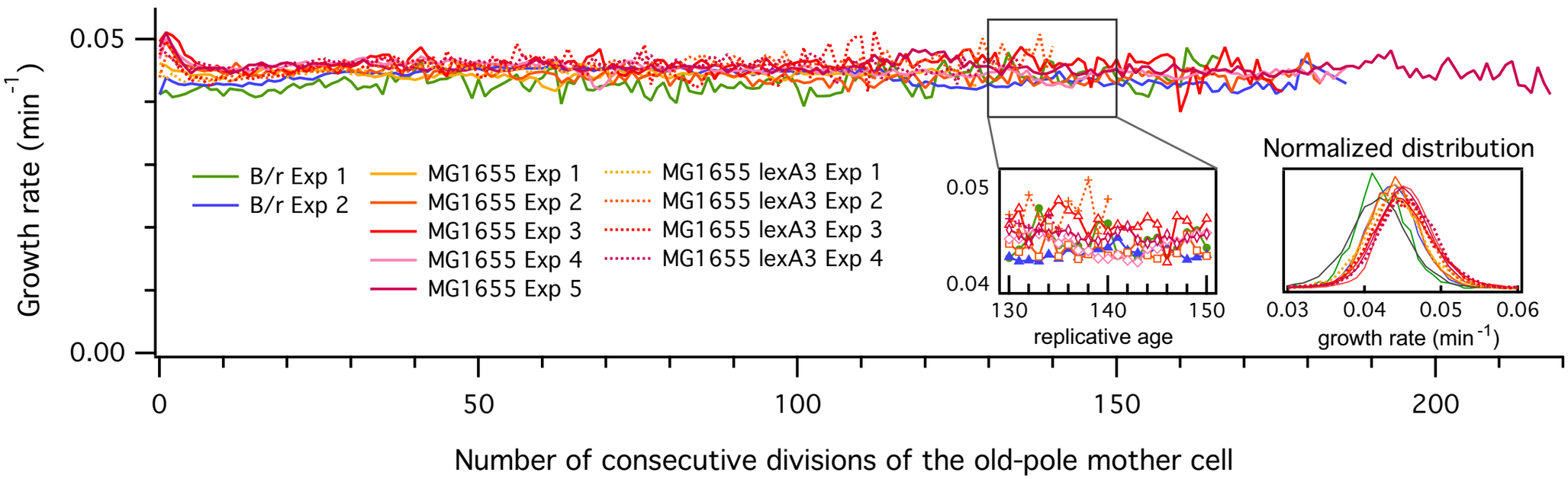
Growth in cell size of bacteria in the Lenski experiment

Growth rate (=Fitness)



[redacted]

mother machine



Current Biology *20*, 1099–1103,

3. Foldings and Functions of RNA

non-coding RNAs

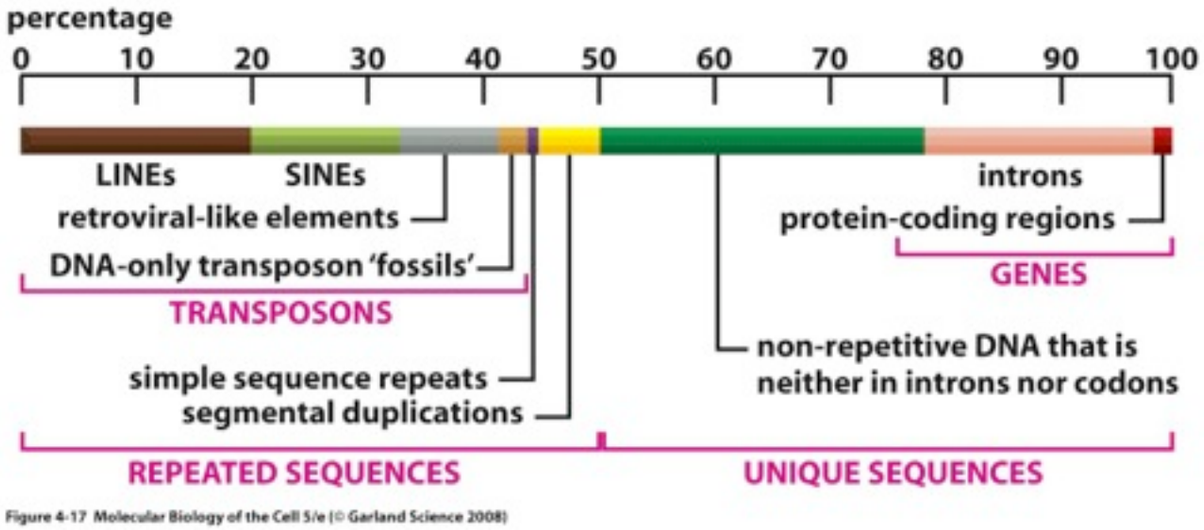


Figure 4-17 Molecular Biology of the Cell 5/e (© Garland Science 2008)

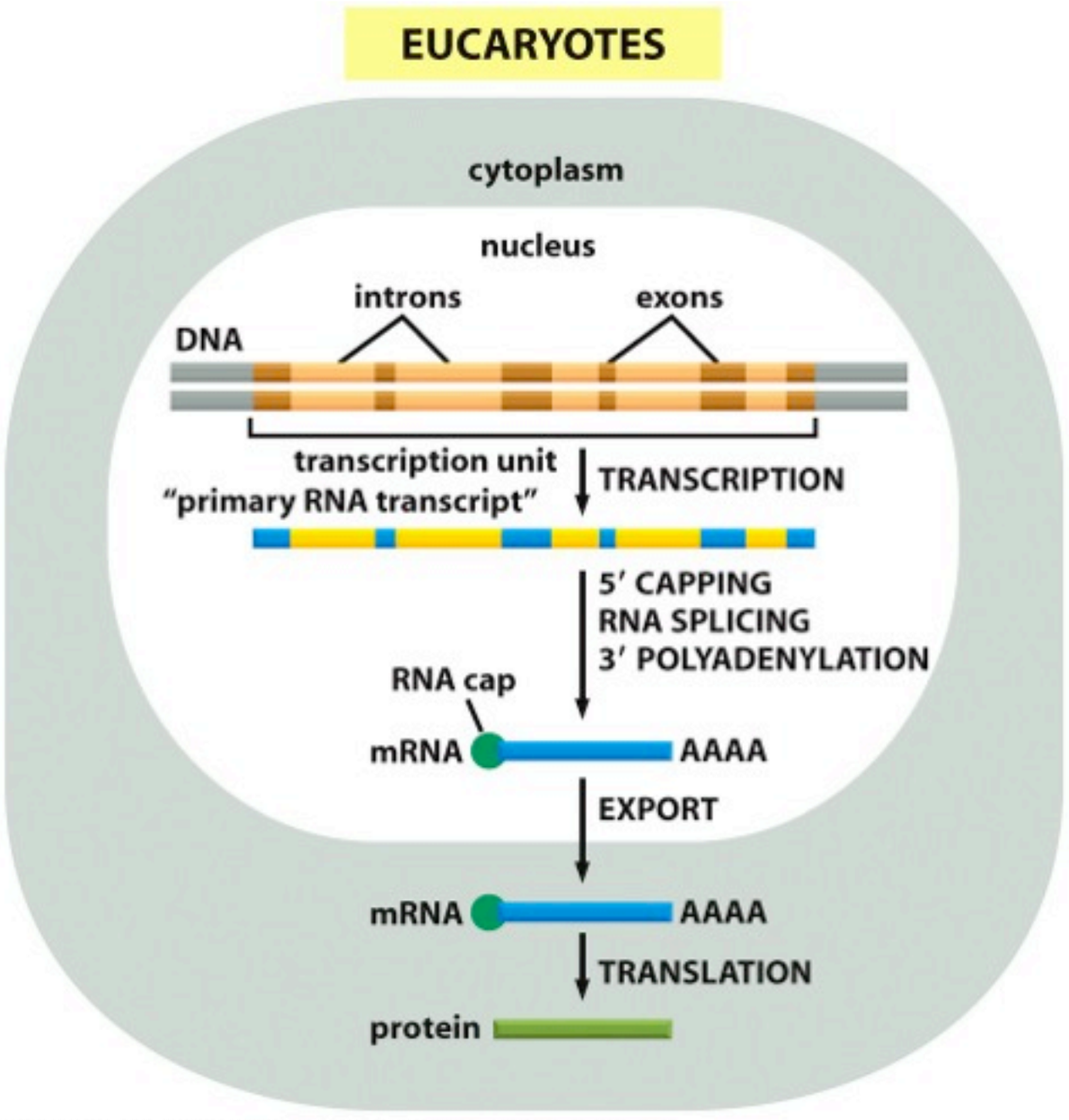
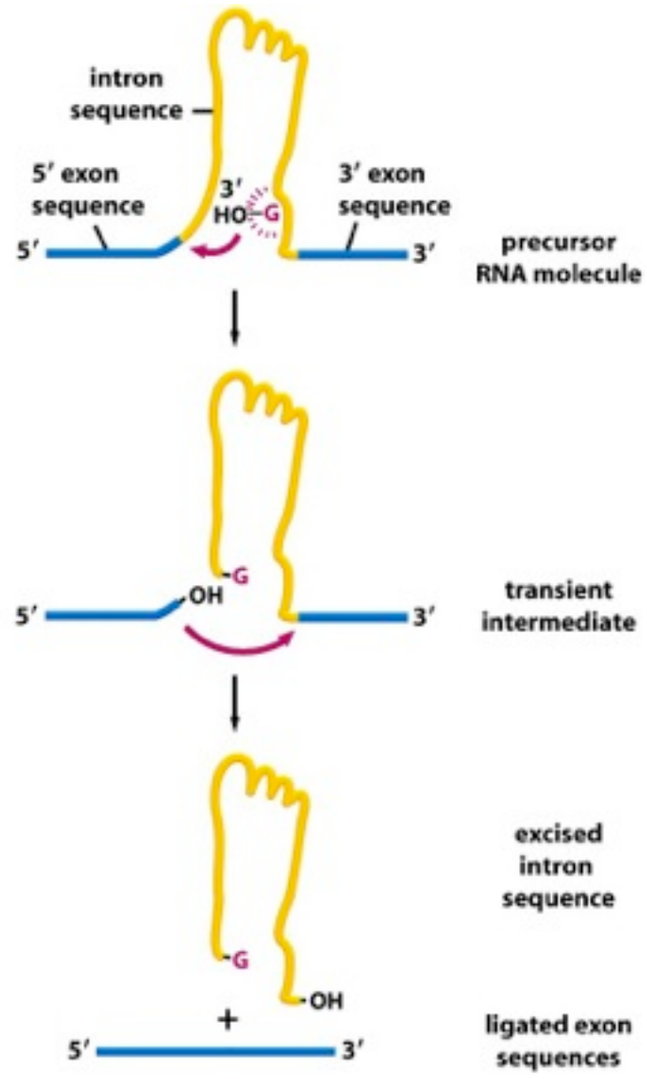
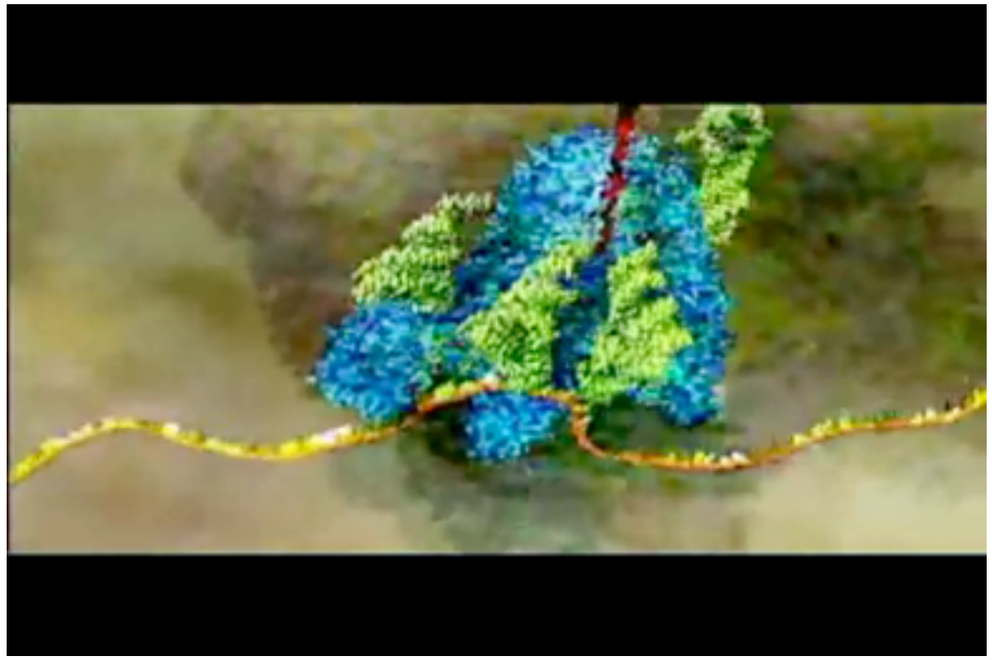


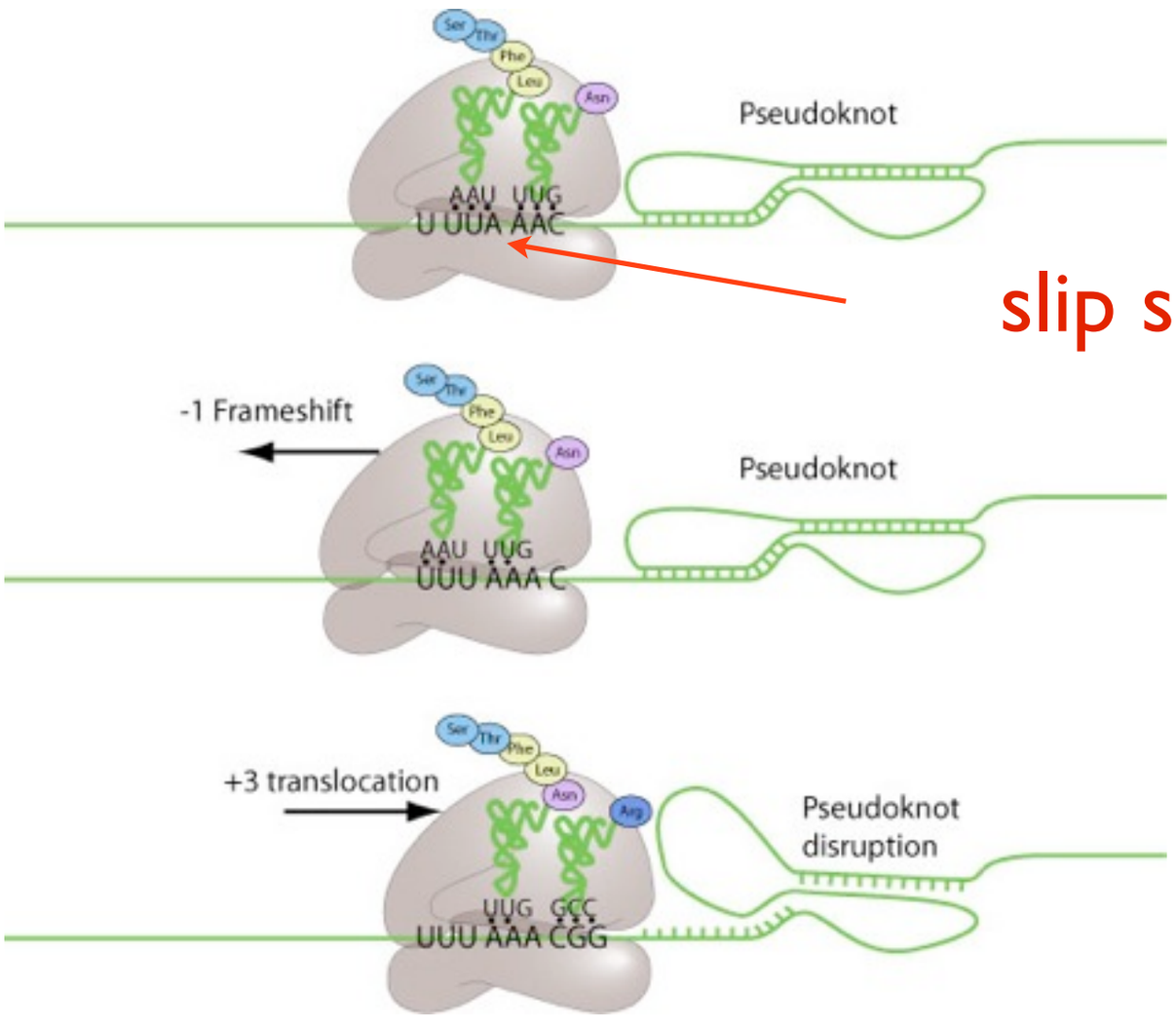
Figure 6-21a Molecular Biology of the Cell 5/e (© Garland Science 2008)



DNA → RNA → protein

mRNA (coding region)

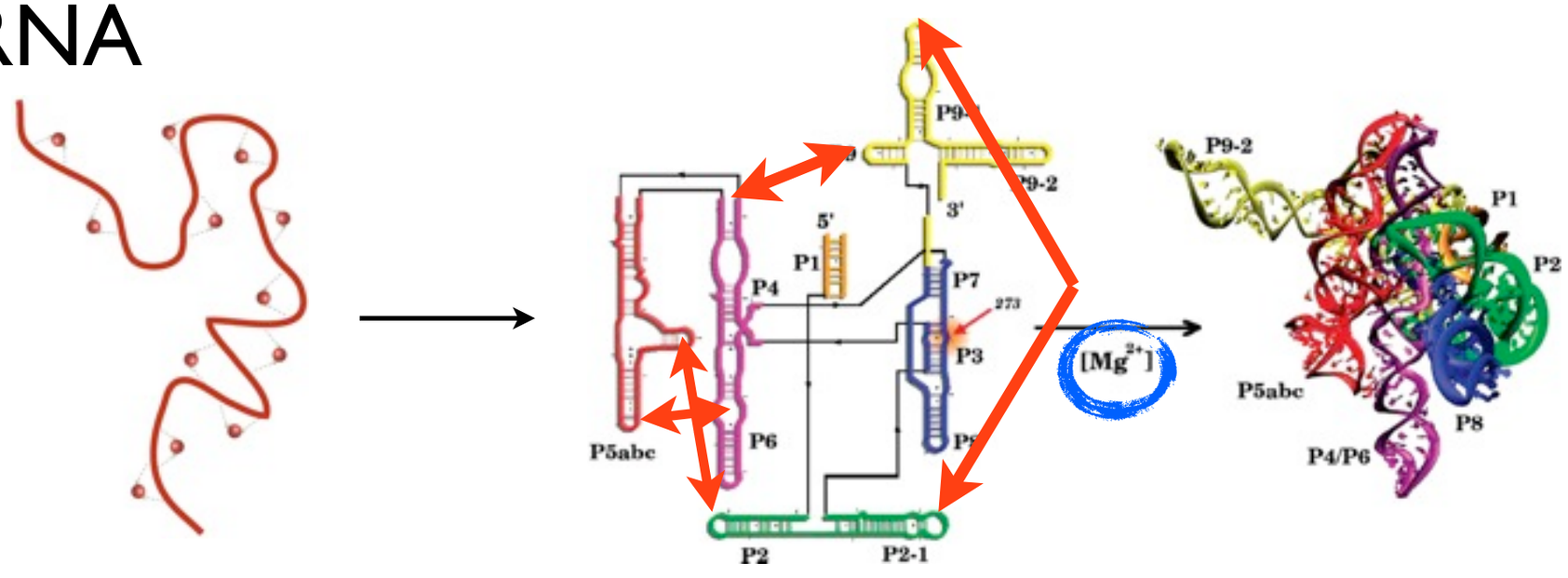




slip sequence

RNA folding

ncRNA

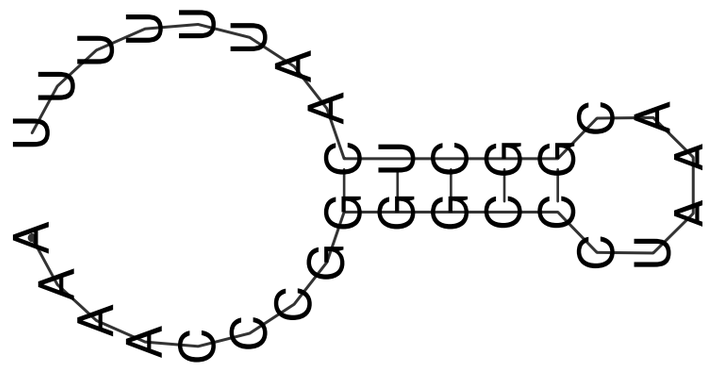


Primary to Secondary to 3D structure formation

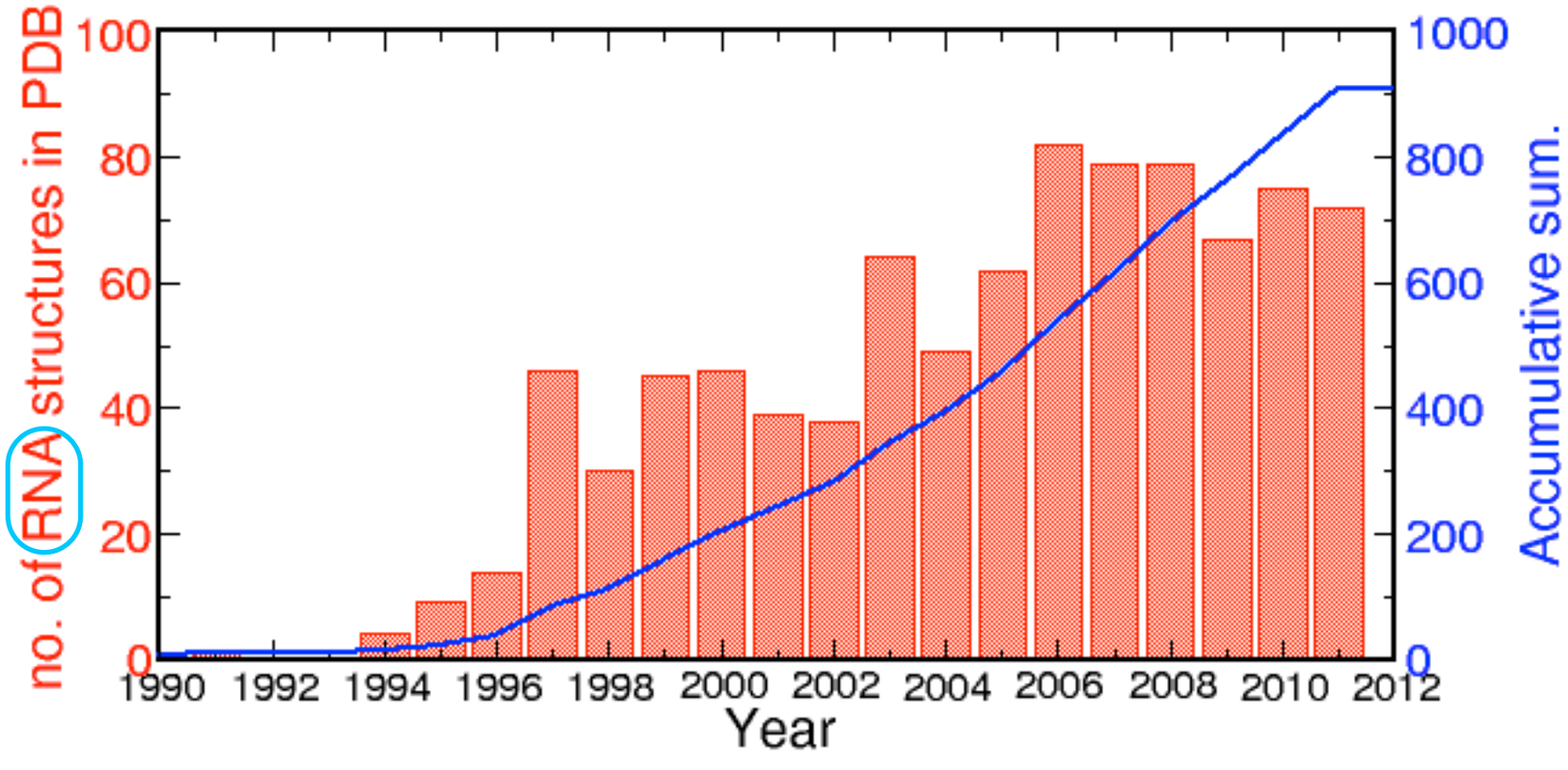
- Characteristic base-pairing rule for RNA
  (stacking, loop, bulge, internal multiloop free energy, ...)
  →1970s (Tinoco, Uhlenbeck, Crothers, Porschke...)

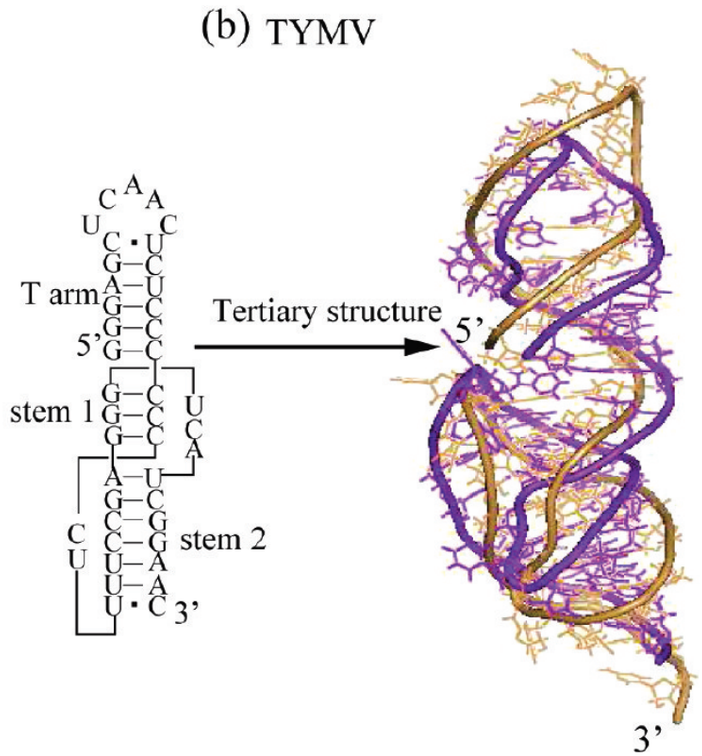
>>AAAACCCGGGGCCCUA
AACGGCUCAAUUUUUU

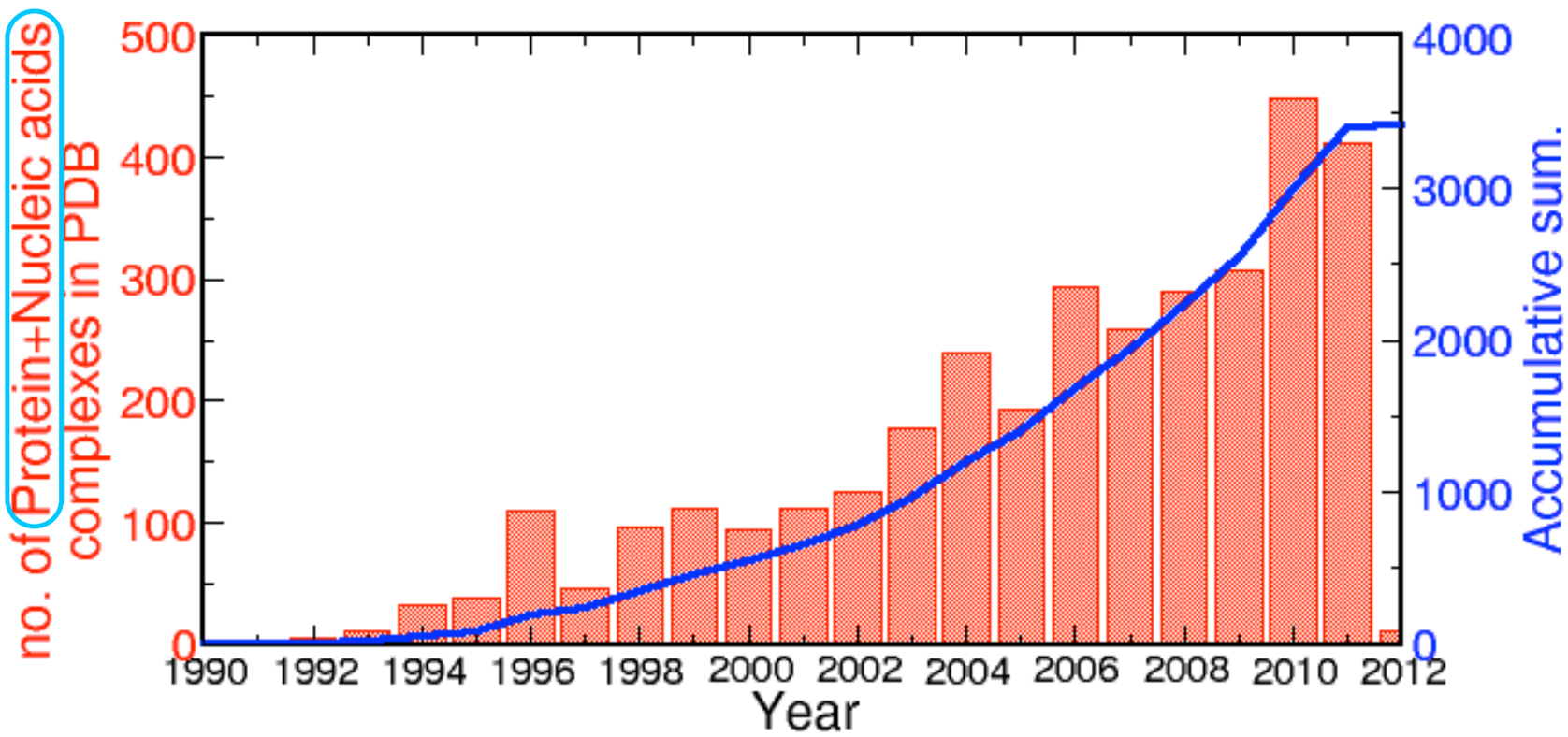
mfold,
Vienna RNA package



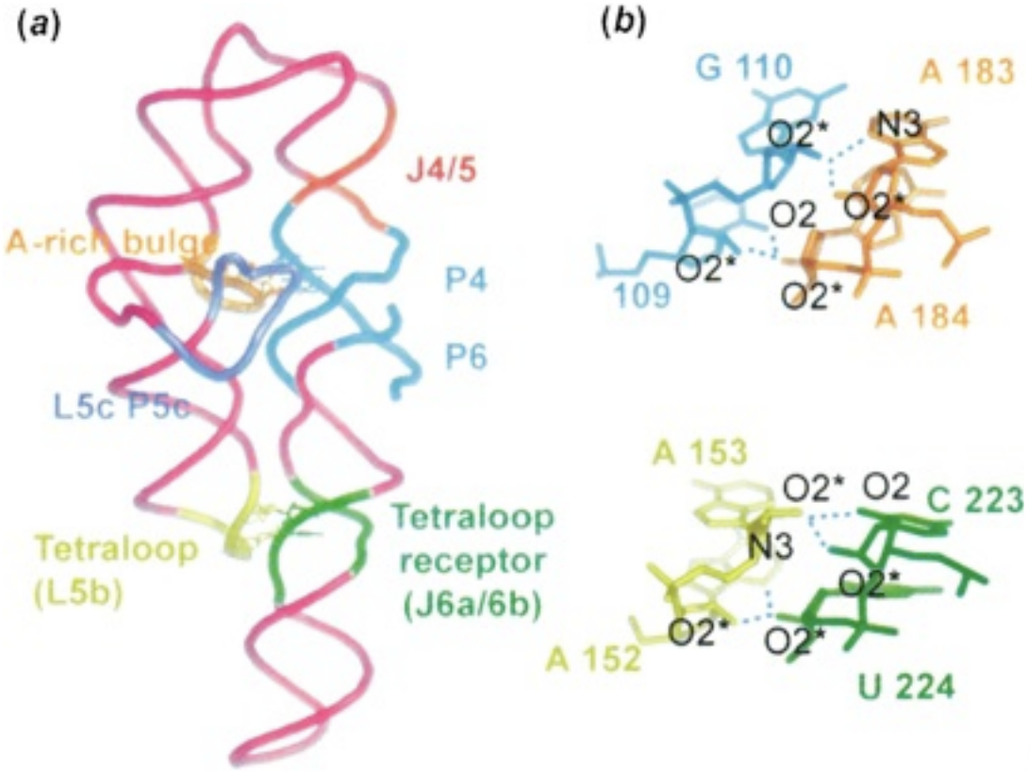
Statistics of RNA structures in PDB .... It is about time...



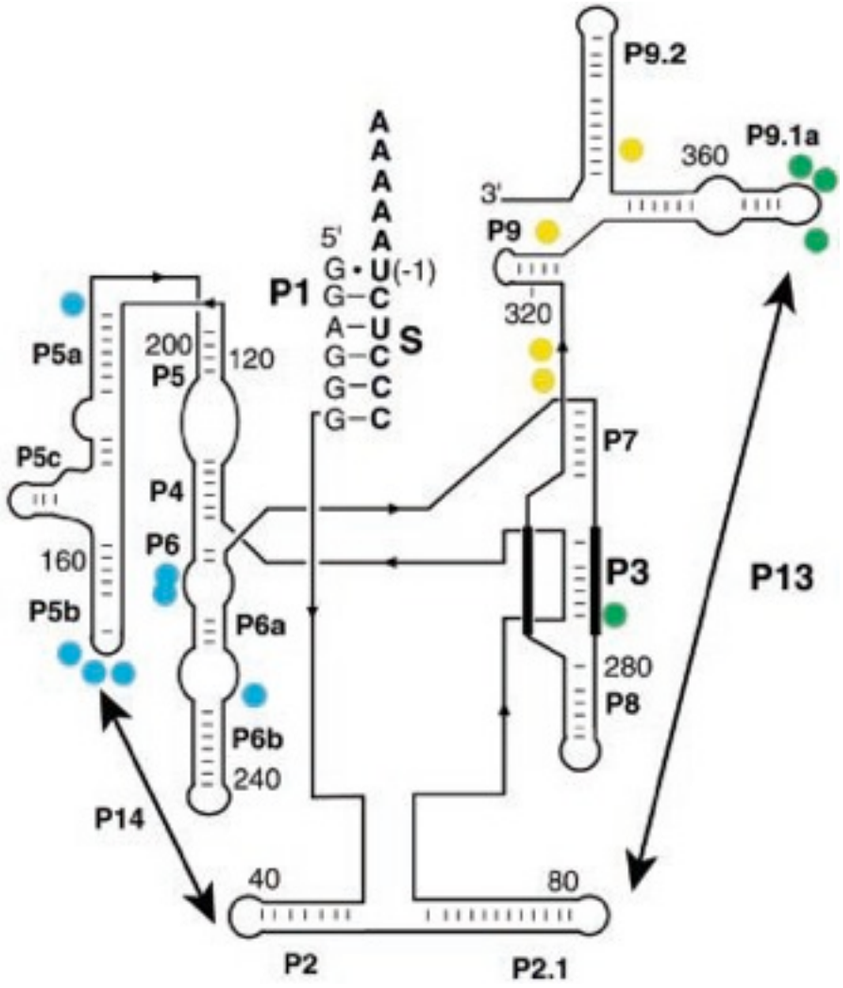


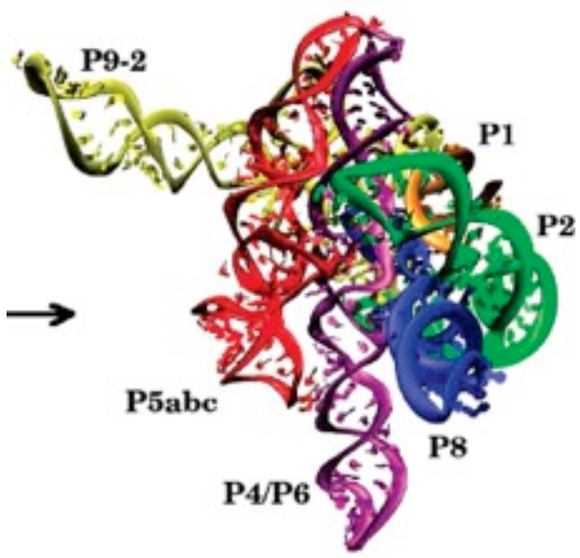


RNA motifs for tertiary structure formation formation



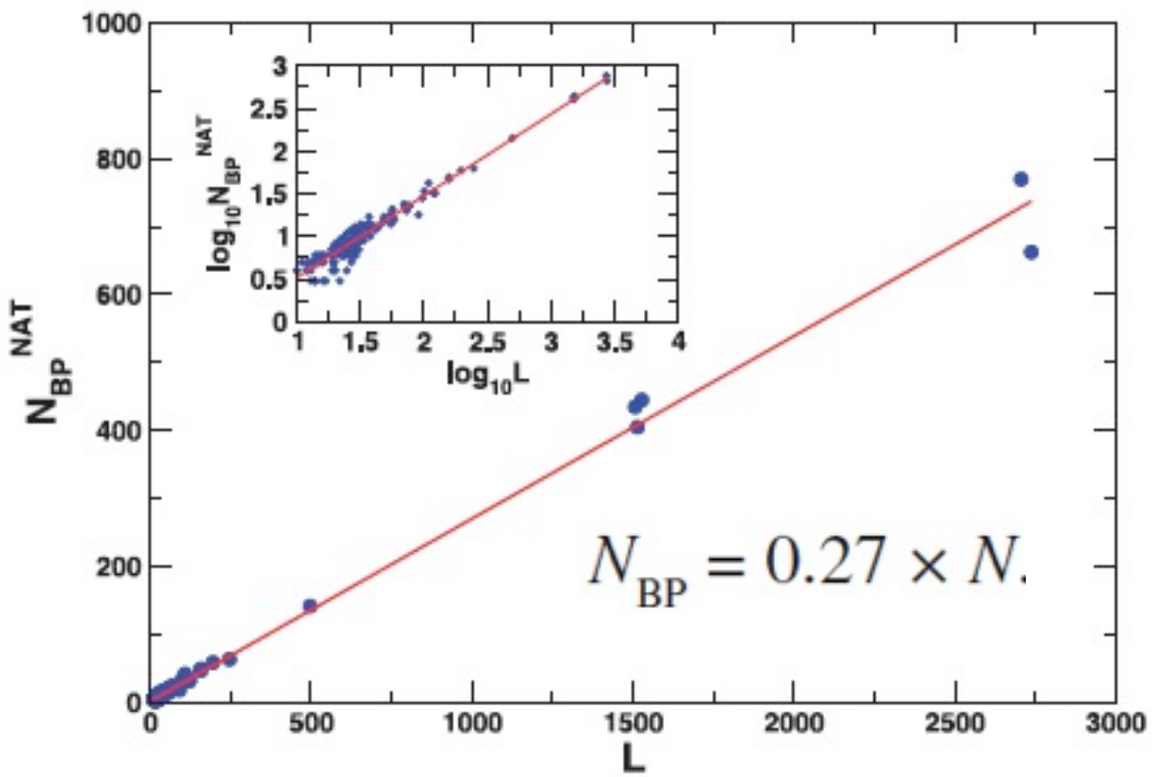
Ribose zipper,
kissing hairpin,
Coaxial stack, ...



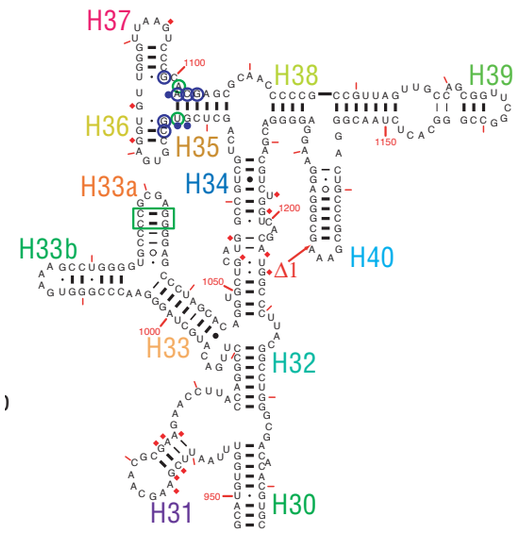


Prediction of RNA native state (RNA folding problem) is difficult

1. Building block of RNA : Chemically similar 4-nt (instead of chemically dissimilar 20 aa)
2. Polyelectrolyte nature (charged phosphate group)
3. Uniformity of hydrophilic backbone along with the lack of diversity in the bases makes RNA closer to “homopolymer” than polypeptide chains -> Make the stability gap smaller and alternate structure possible.



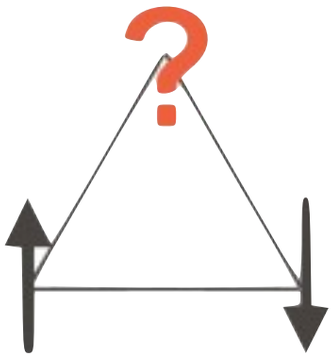
4. The plot on the left shows that 46 % the sequence consists of non-pairing regions such as bulges, loops, dangling ends and other motifs

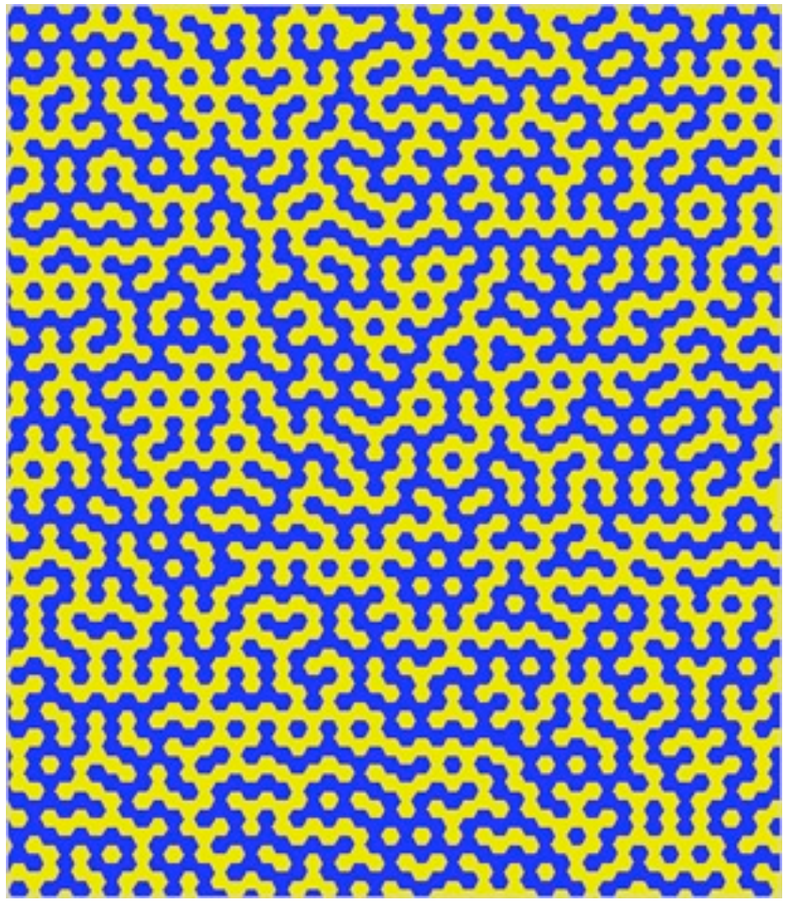


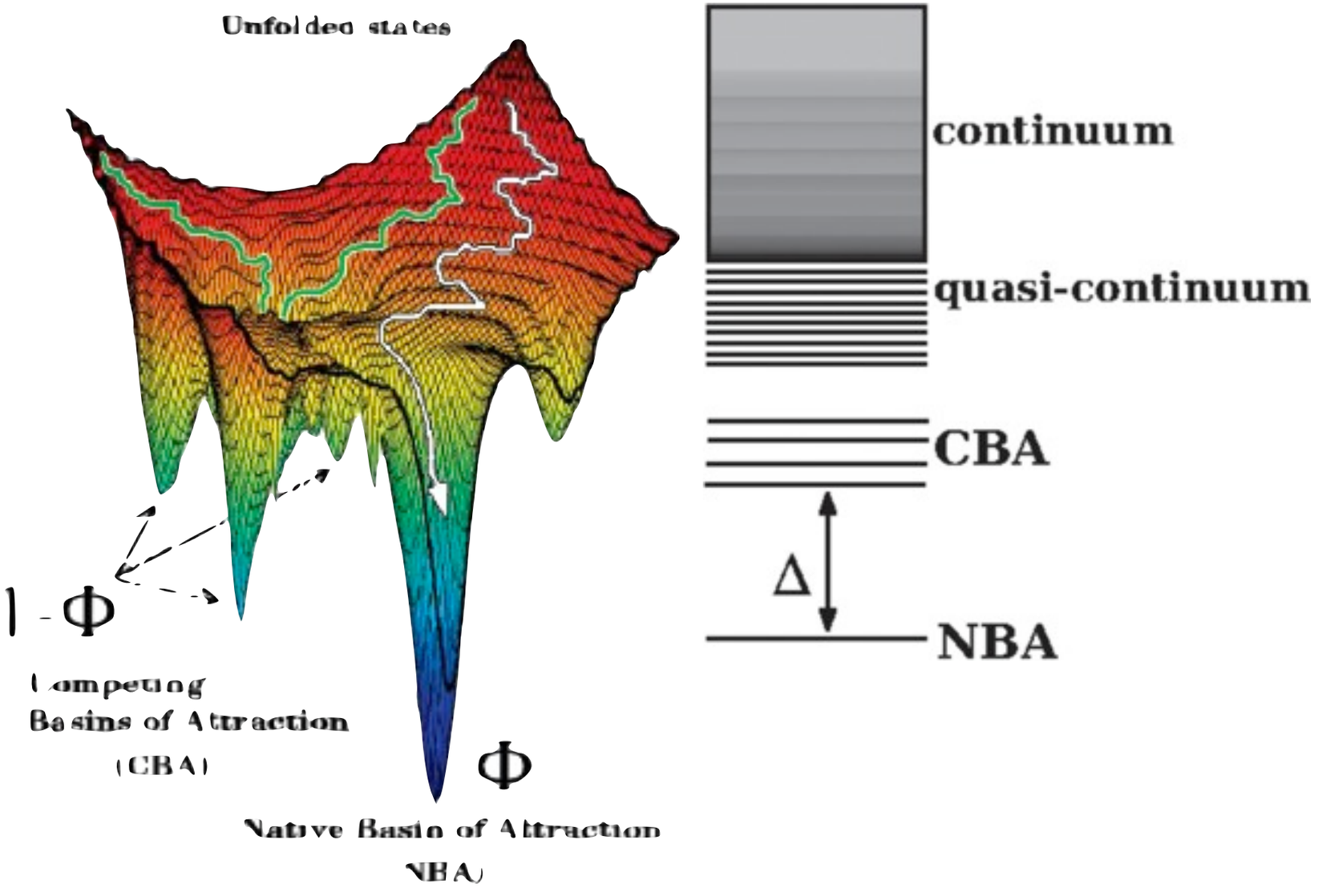
See http://www.major.iric.ca/MC-Pipeline/ (NCM, nucleotide cyclic motifs).

RNA energy landscape is rugged

One dimensional chain + Heterogeneous interactions → Topological / Energetic frustration

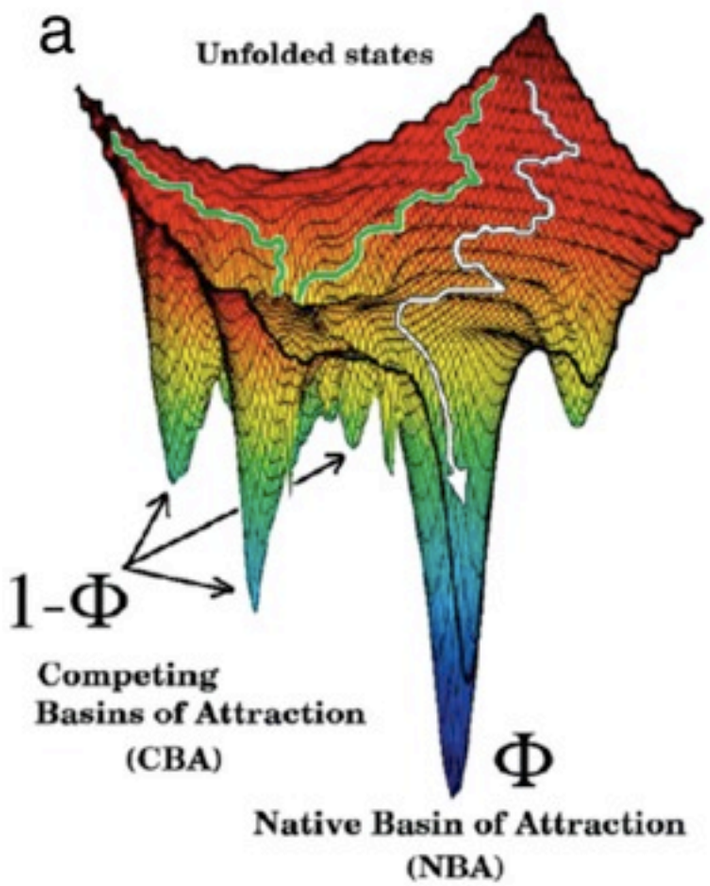


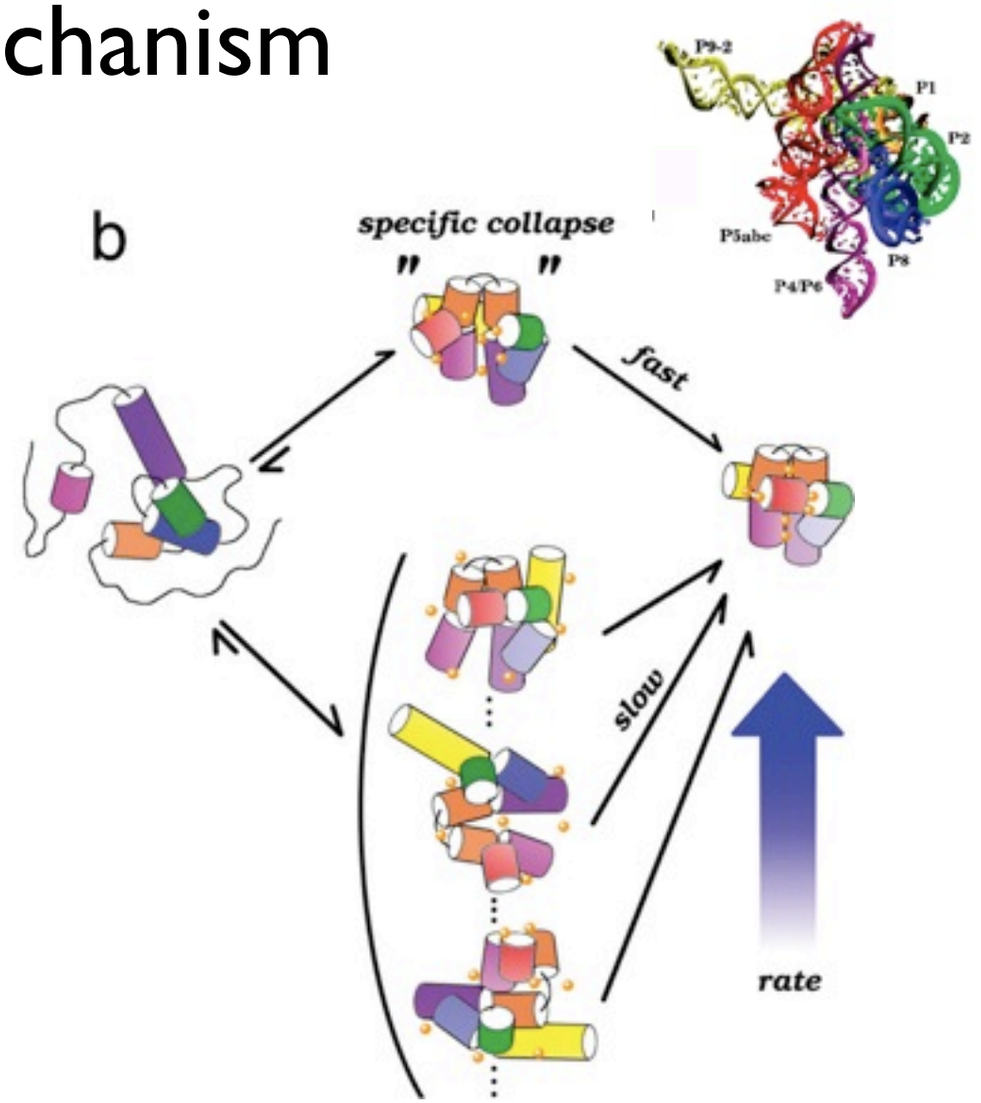




Thirumalai and Hyeon, Biochemistry (2005)

RNA energy landscape
and
kinetic partitioning mechanism



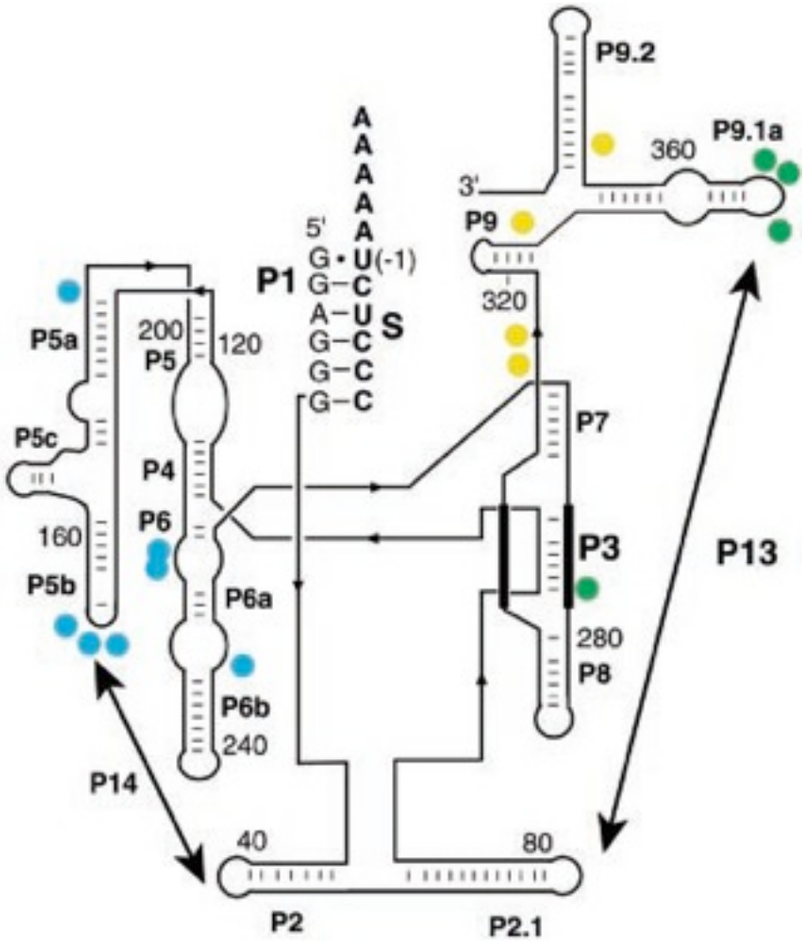


Thirumalai and Hyeon, Biochemistry (2005)

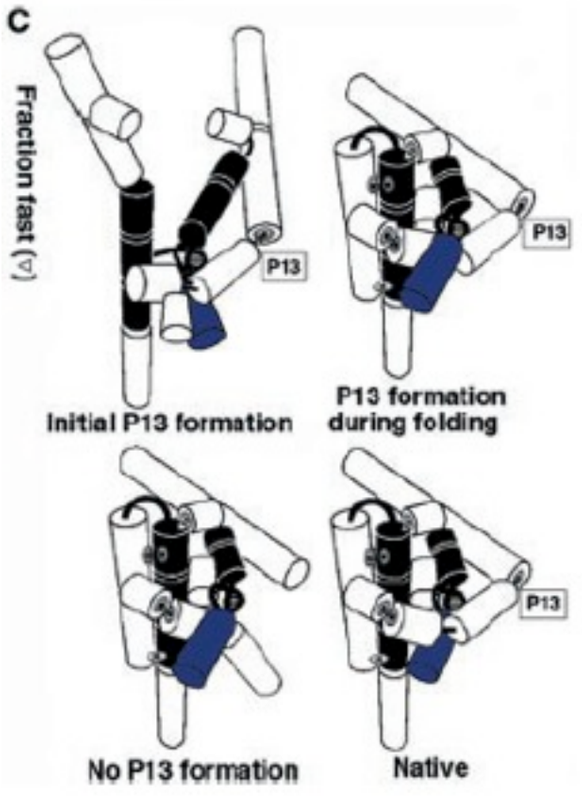
$$f_N(t) = \Phi_{max} - \Phi e^{-k_{fast}t} - \sum_i A_i e^{-k_i t}$$

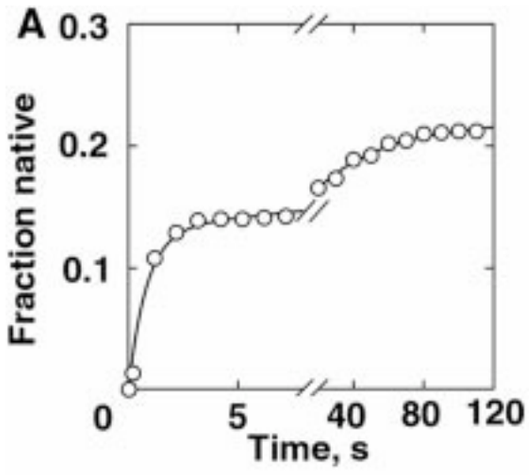
The value of Φ depends on point mutation, ionic environment, and initial condition ...

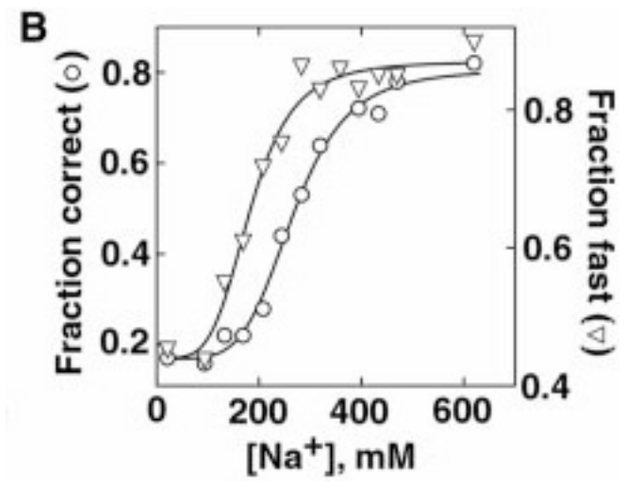
Initial condition dependent folding routes of T. Ribozyme



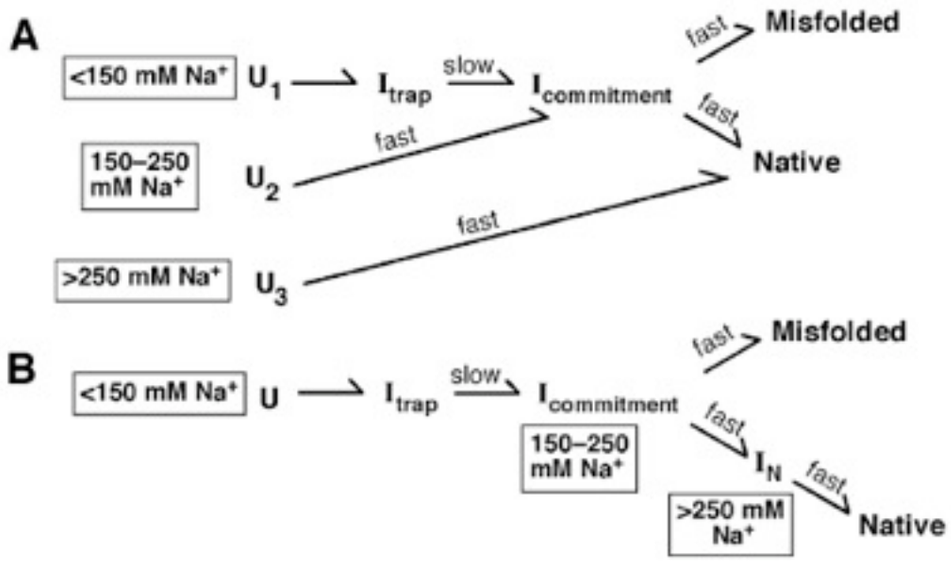
Russell et al. PNAS (2001)

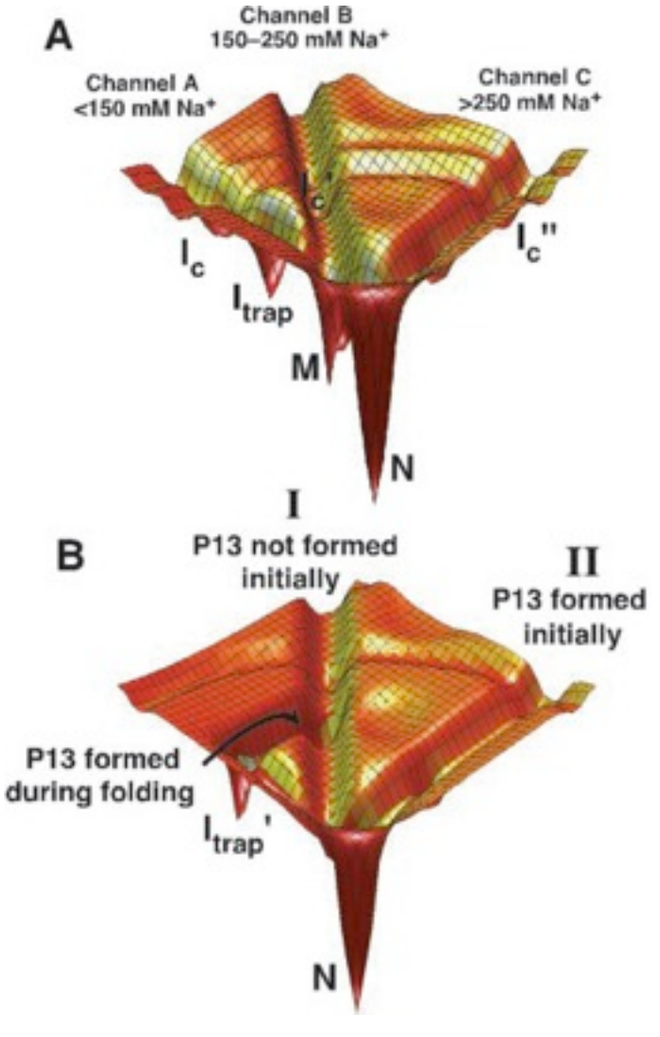




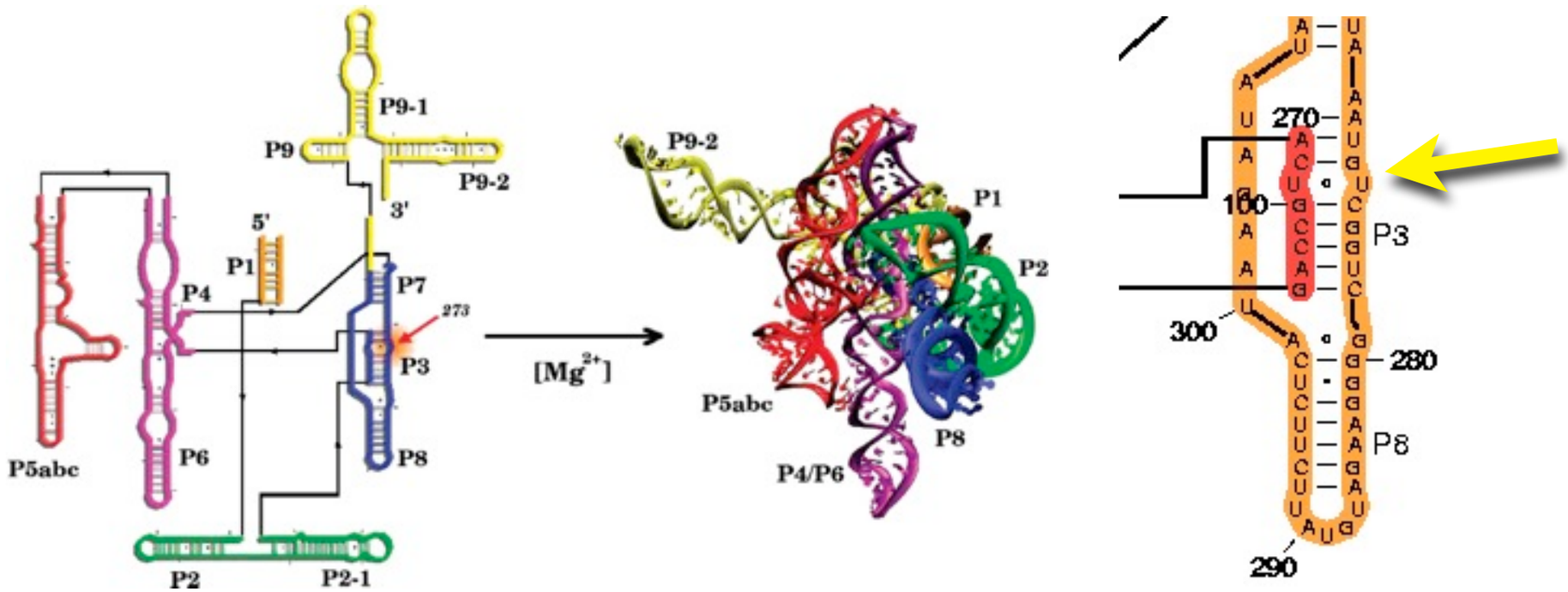


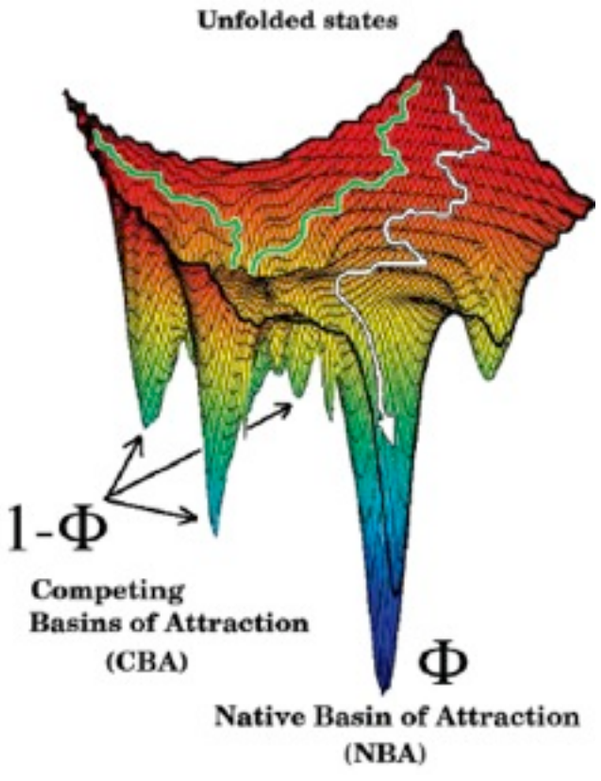
Varying the Na^+ concentration in the preincubation from 20 to 620 mM increased the fraction that folded correctly from 0.2 to ~0.8 (Fig. 1*B*) even though folding was always under the same conditions (5 mM Mg^{2+}, 20 mM Na^+). Also, the fraction of the correct-folders that folded fast (1 s^{-1}) increased from ~0.45 to ~0.9.

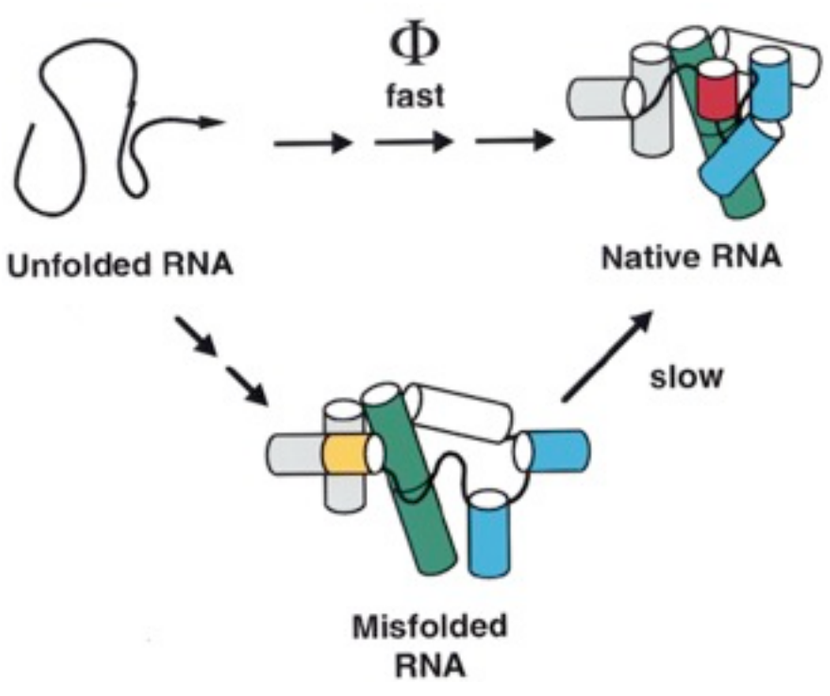




Point mutation





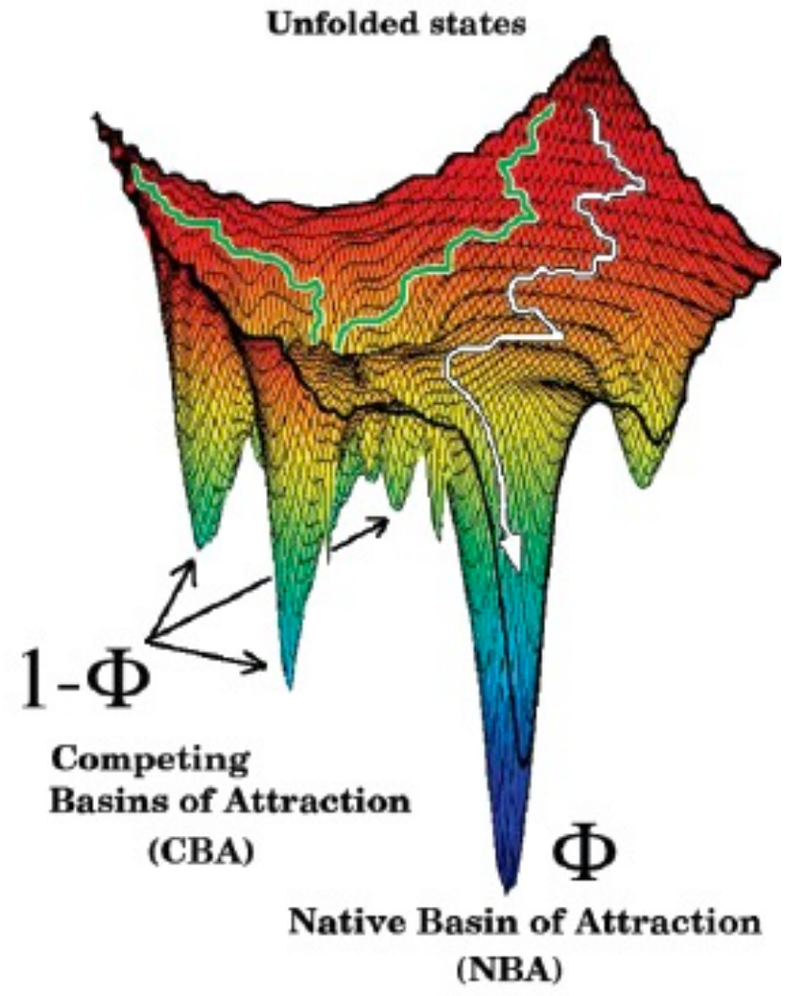


Pan et al. (1997) J. Mol. Biol.

Φ=0.06-0.1 for
Tetrahymena ribozyme.
But, a point mutation
U273A in P3 increases Φ
to 0.8 !

Rugged folding landscape of RNA and functions

- RNA consisting of four chemically similar building blocks (A,C, G, U)
  - → RNA are closer to homopolymers
  - → rugged landscape
  - → adopt alternative structures
  - → kinetic partitioning mechanism

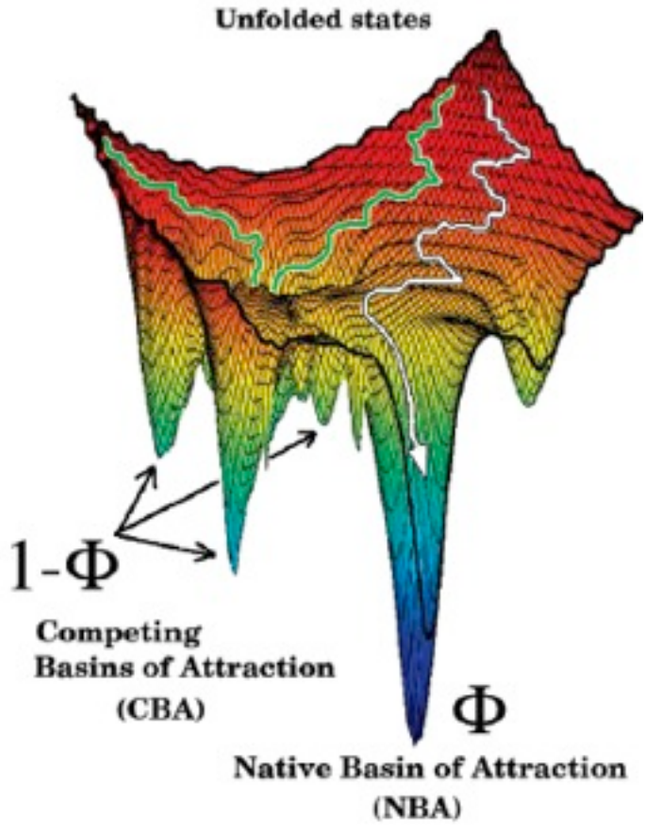


Guo & Thirumalai (1995) Biopolymers

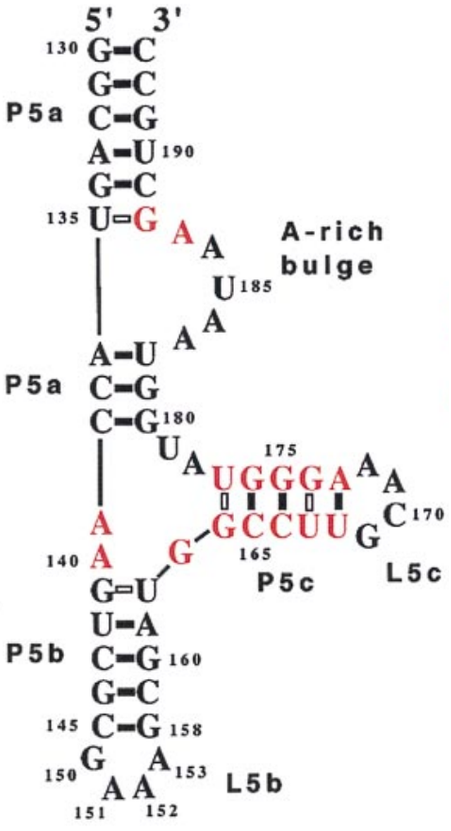
Pan, Thirumalai, Woodson (1997) JMB

Thirumalai & Hyeon (2005) Biochemistry

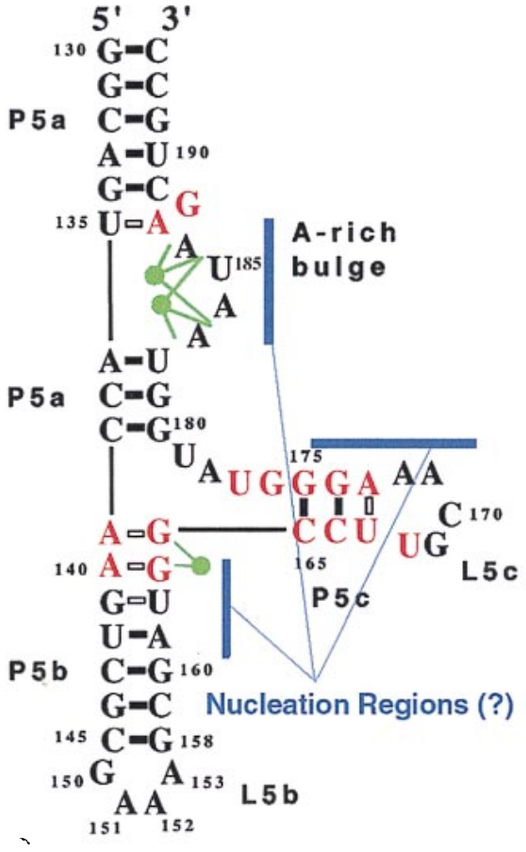
- Foldings and functions of RNA are associated with conformational adaptation (CBA↔NBA) in response to cellular signals (metal ions, metabolites, proteins, nucleic acids,...)

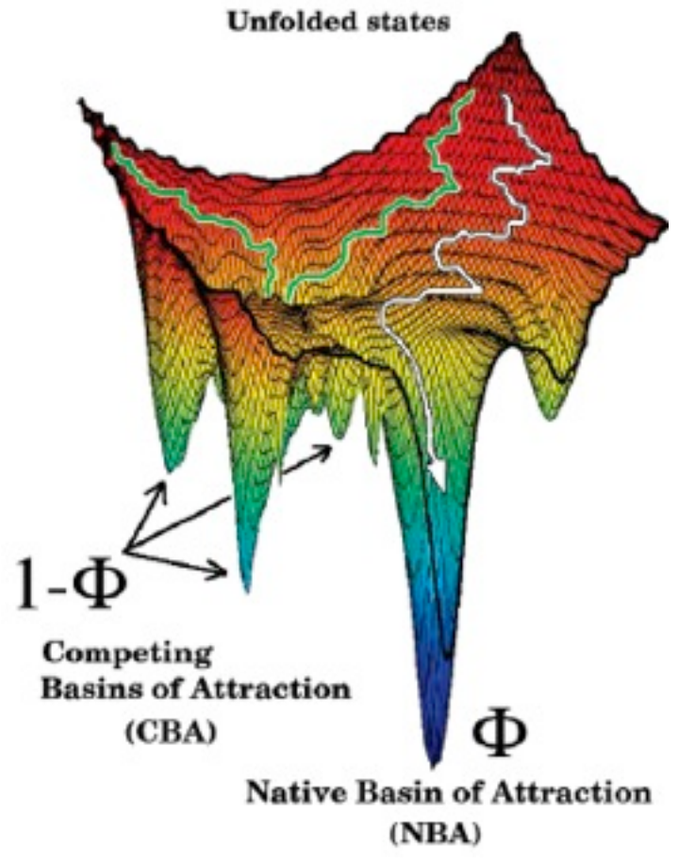


divalent metal-ion induced secondary structure rearrangement

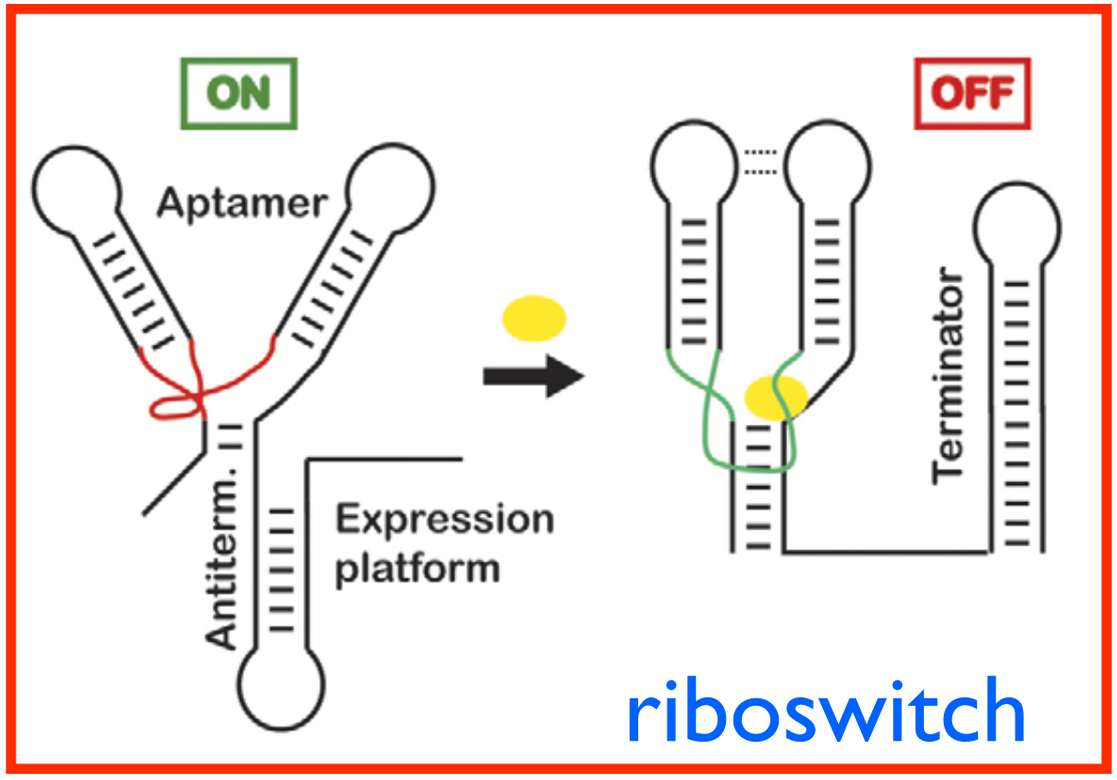


Mg^{2+}



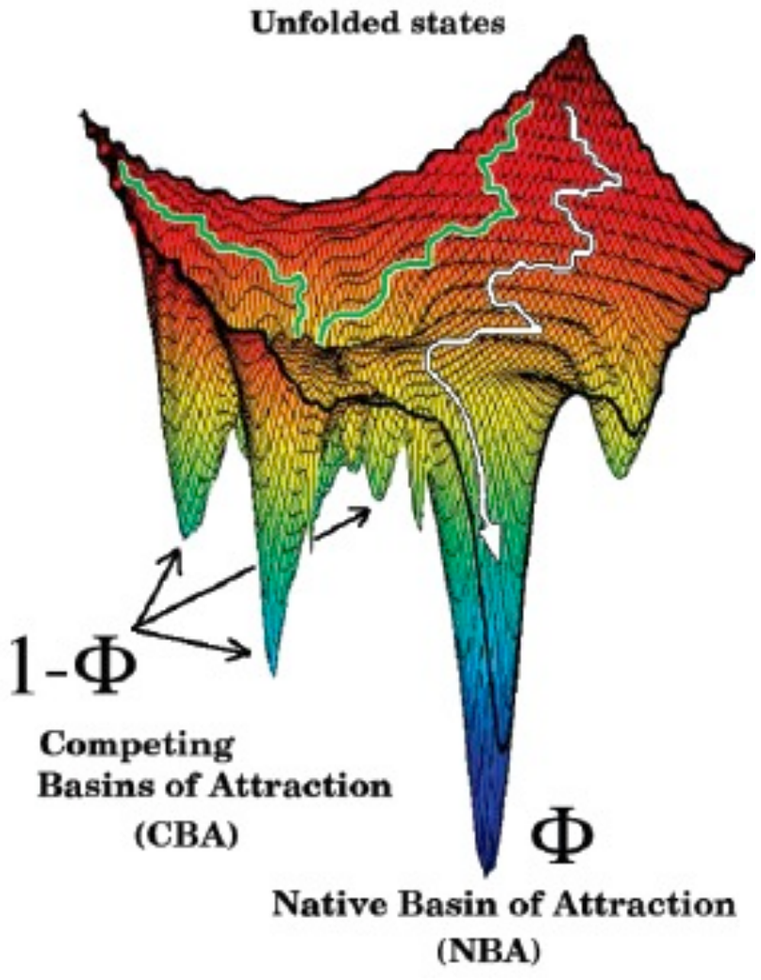


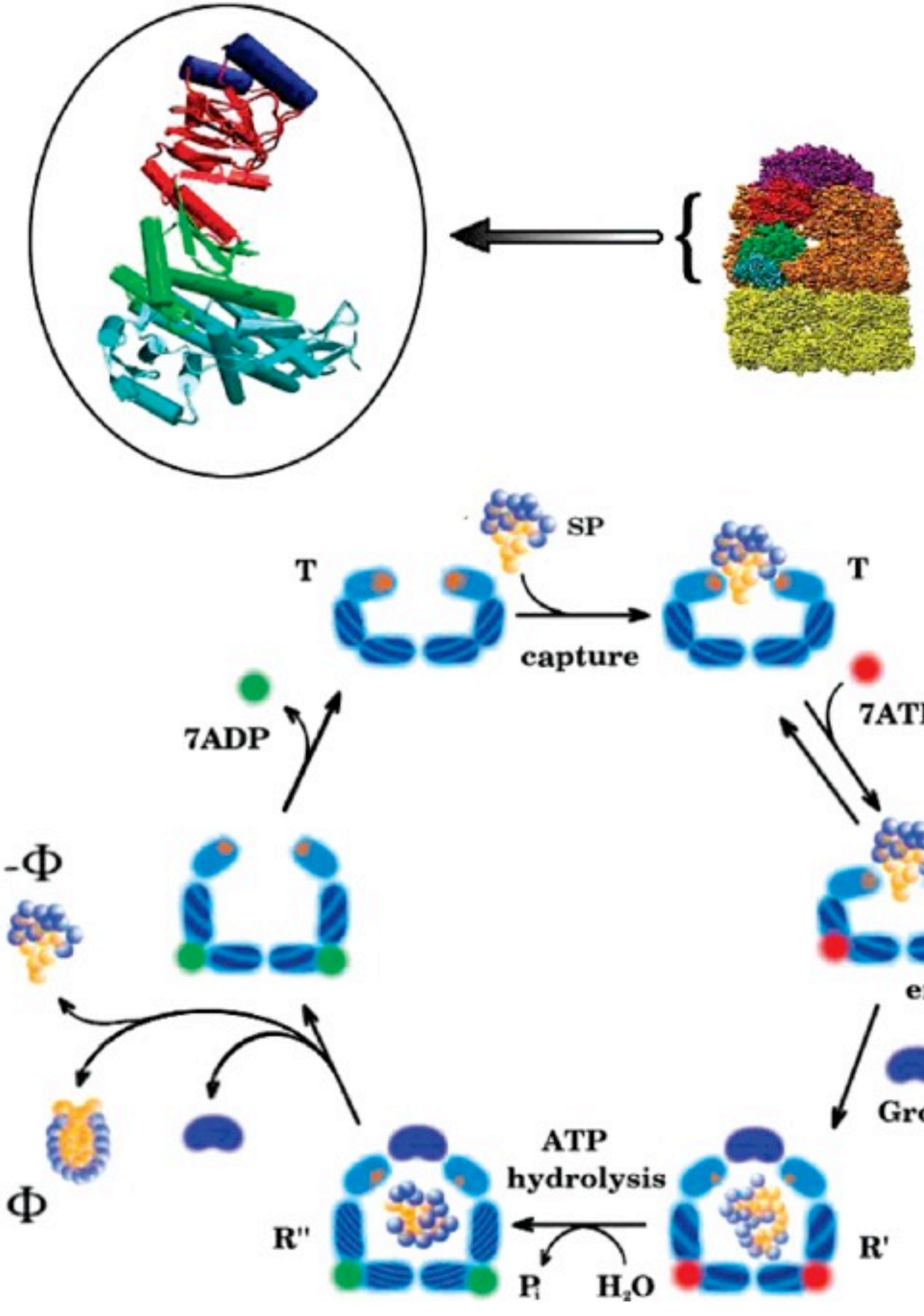
Functions of RNA are often dictated by the conformational switches between the alternative states



Al-Hashimi & Walter (2008) COSB

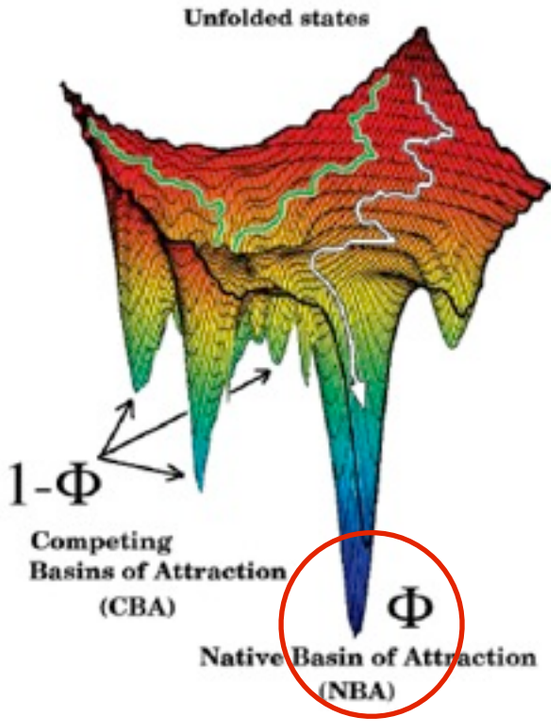
Protein chaperone: GroEL-GroES chaperonin system

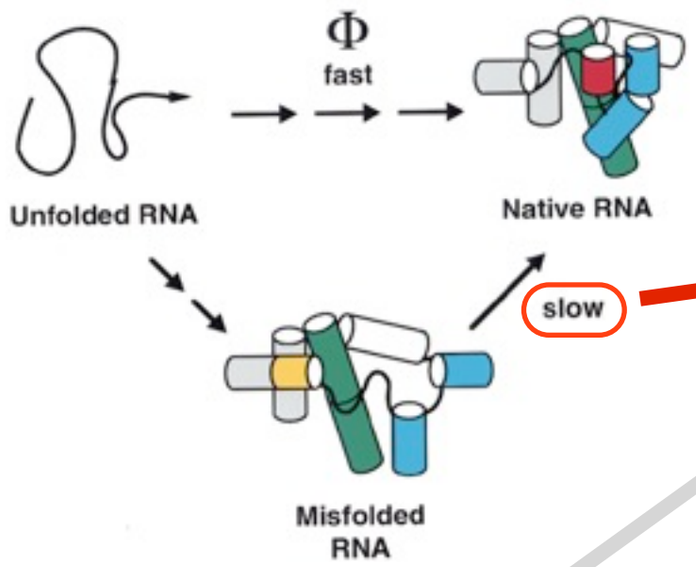




Molecular chaperonin system increases the yield of folds via Iterative annealing mechanism (IAM)

$$Yield = 1 - (1 - \Phi)^N$$



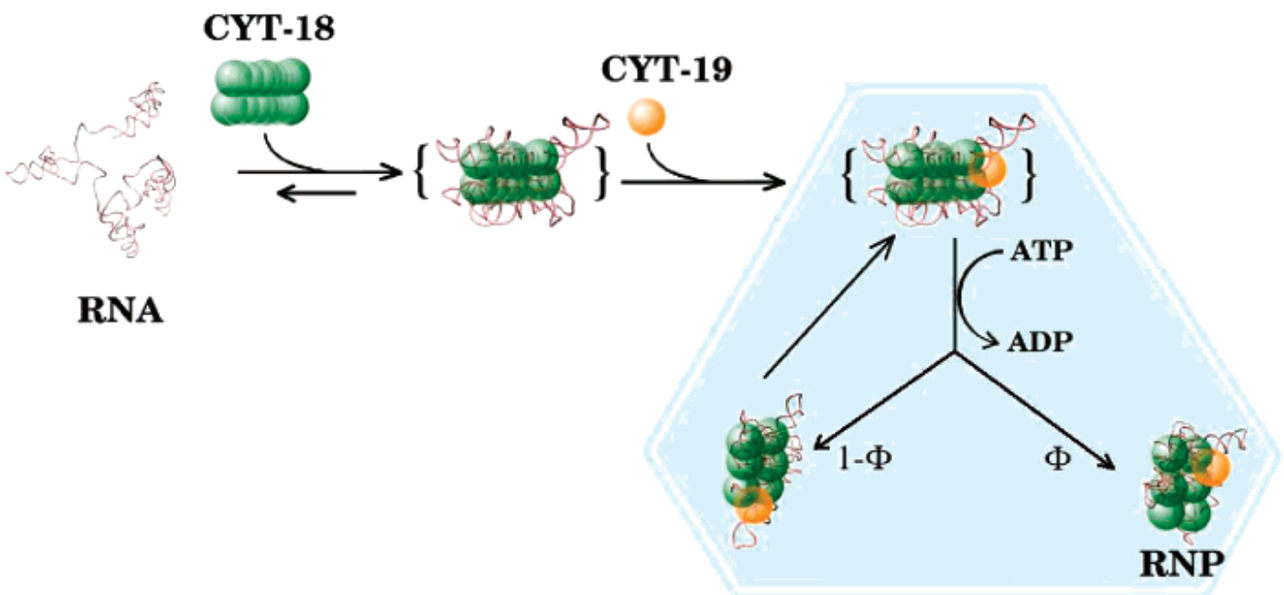


biologically inaccessible

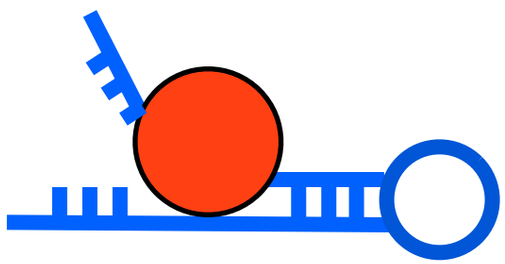
$$\text{Yield Folded} = 1-(1-\Phi)^N$$

after N iterations

$$1 \rightarrow \left\{ \begin{array}{l} \Phi \\ \boxed{1-\Phi} \end{array} \right. \xrightarrow[\text{ATP ADP}]{} \left\{ \begin{array}{l} \Phi(1-\Phi) \\ \boxed{(1-\Phi)^2} \end{array} \right. \xrightarrow[\text{ATP ADP}]{} \left\{ \begin{array}{l} \Phi(1-\Phi)^2 \\ \boxed{(1-\Phi)^3} \end{array} \right. \xrightarrow[\text{ATP ADP}]{} \left\{ \begin{array}{l} \Phi(1-\Phi)^3 \\ \cdots \end{array} \right.$$



DEAD-box proteins assist secondary structure rearrangement by acting as RNA strand separators (or helicases).

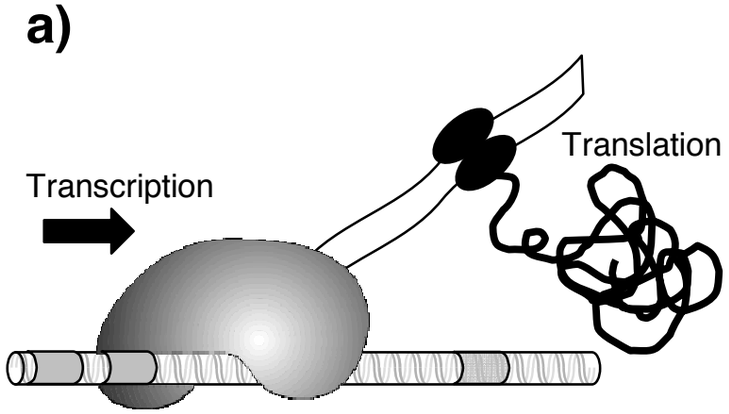


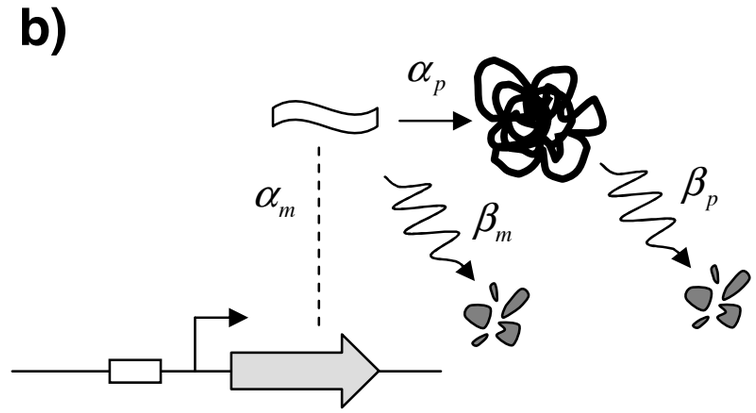
"DEAD-box 'Helicase' Proteins as Chaperones of RNA Folding"

Conclusions

- RNAs are more primitive and homopolymer-like than proteins.
- RNA folding landscapes are rugged.
- Kinetic partitioning mechanism
- Ionic environment is crucial for RNA folding and function
- RNA chaperone is essential to facilitate RNA folding

4. Gene expression, Genetic switch



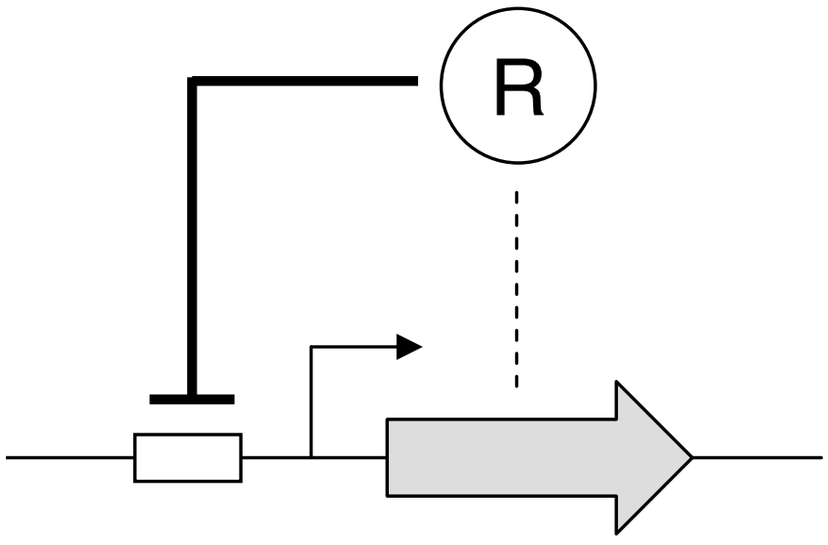


$$\text{DNA} \xrightarrow{\alpha_m} \text{RNA (m)} \xrightarrow{\alpha_p} \text{protein (p)}$$

$$\beta_m \downarrow \varnothing \qquad \beta_p \downarrow \varnothing$$

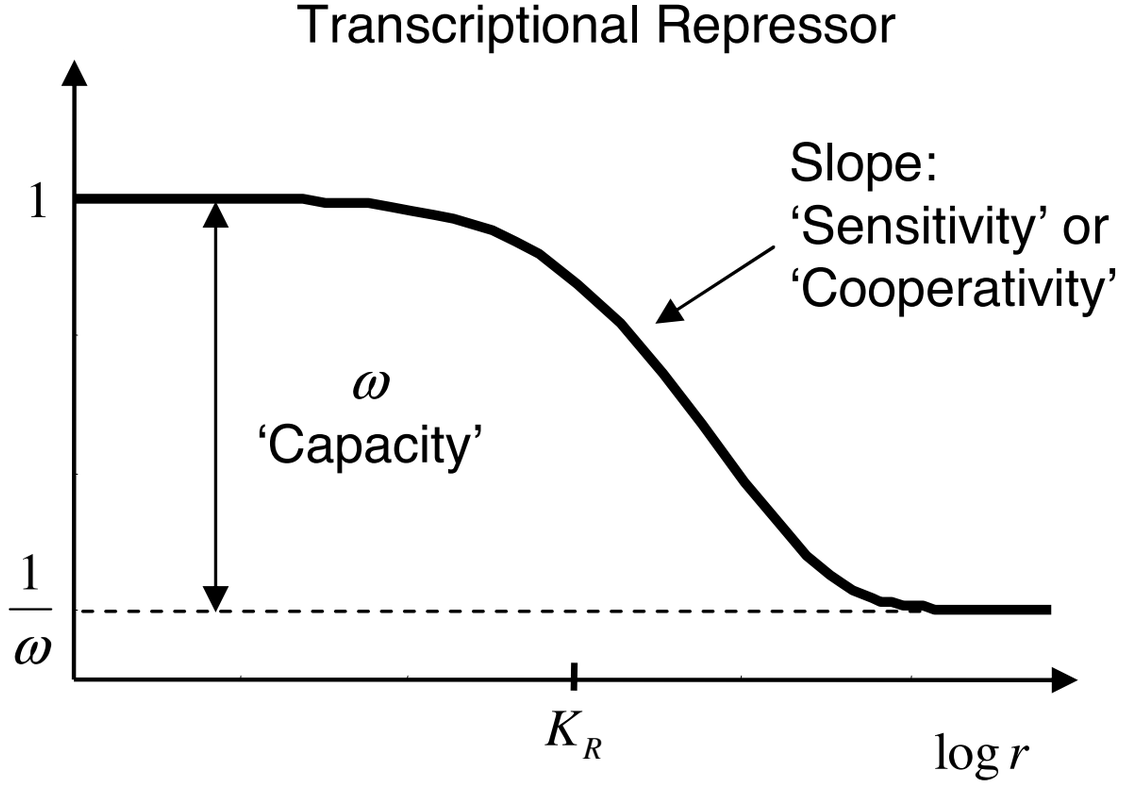
$$\frac{dm}{dt} = \alpha_m - \beta_m m \qquad \frac{dp}{dt} = \alpha_p m - \beta_p p$$

$$m^* = \frac{\alpha_m}{\beta_m} \approx (1-10)nM \qquad p^* = \frac{\alpha_m \alpha_p}{\beta_m \beta_p} = (10^2 - 10^3)nM$$



$$\frac{dm}{dt} = \alpha_m g_R(R) - \beta_m m$$

$$\frac{dR}{dt} = \alpha_p m - \beta_p R$$



$$g_R(r) \sim \frac{1}{1+(r/K_R)^n}$$

$$DR_n \rightleftharpoons D + nR$$

$$(K_R)^n = \frac{[D][R]^n}{[DR_n]}$$

For *Ecoli*, mRNA degrades much faster than proteins, and reaches the SS values. $\beta_m \gg \beta_p$

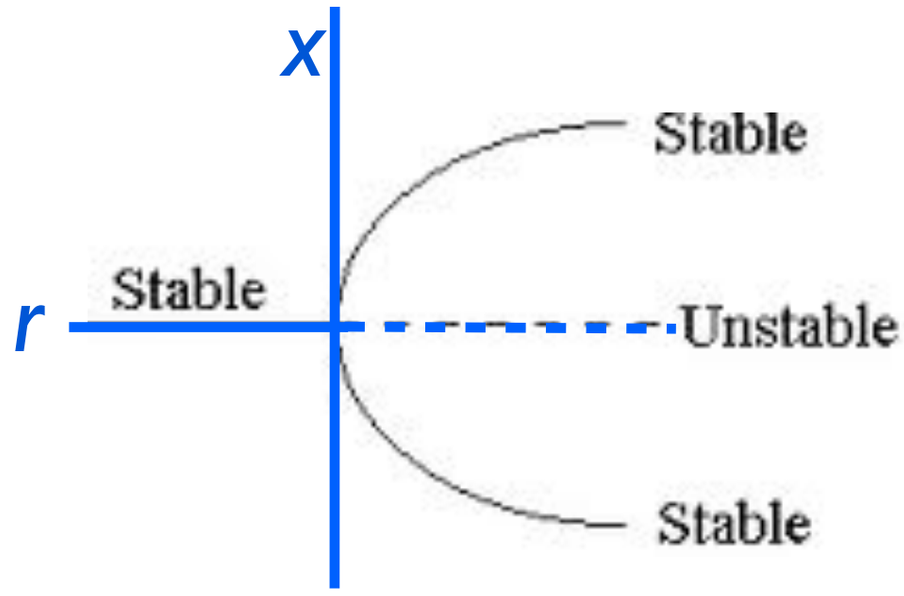
$$\frac{dm}{dt} \approx 0; \quad m^* = \frac{\alpha_m}{\beta_m} g_R(R)$$

$$\frac{dR}{dt} = \alpha_p m^* - \beta_p R$$

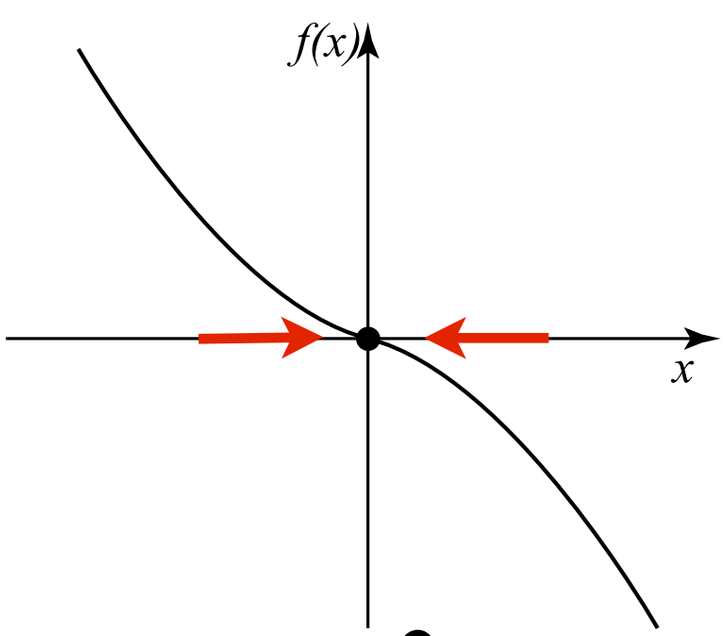
Since $g_R(R) \sim 1$ for $R \ll K_R$ and $g_R(R) \sim 0$ for $R \gg K_R$ it is expected that R(t) increase slowly when $R > K_R$ and the steady state value of R is smaller than unrepressed case.

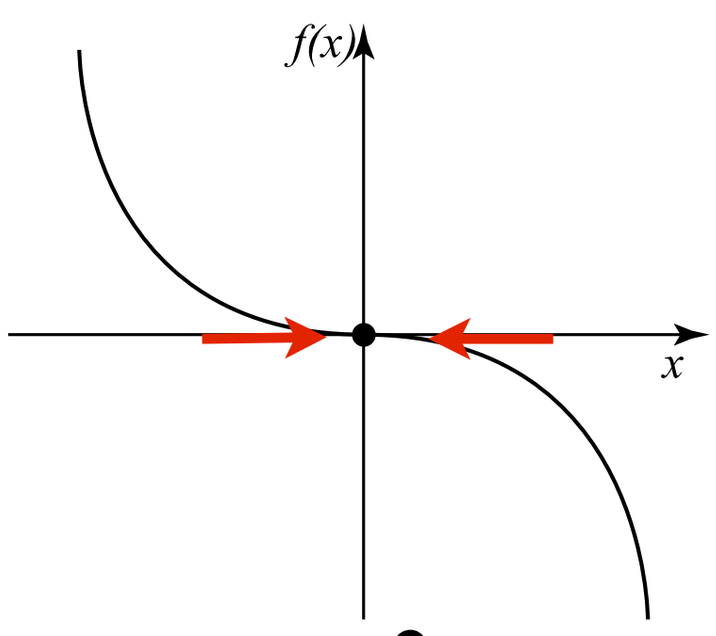
In 1D

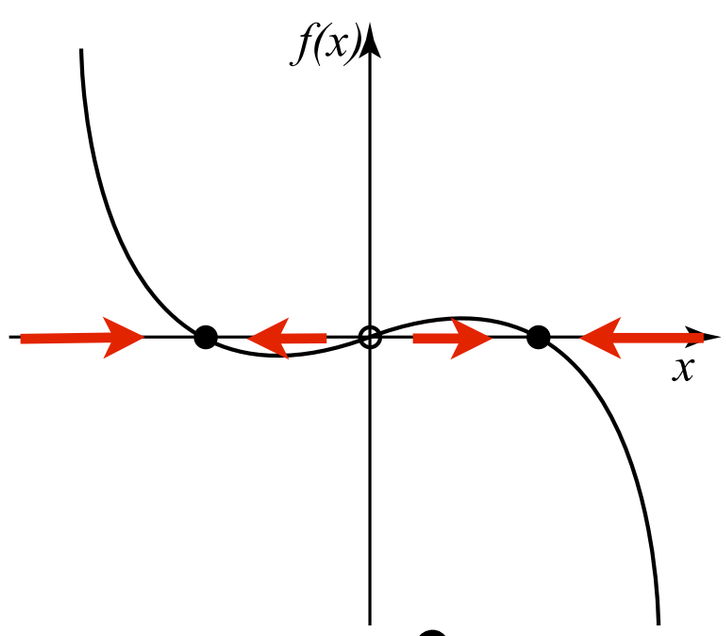
$$\dot{x} = f(x)$$



$f(x) = rx - x^3$ (pitchfork bifurcation)







In 2D

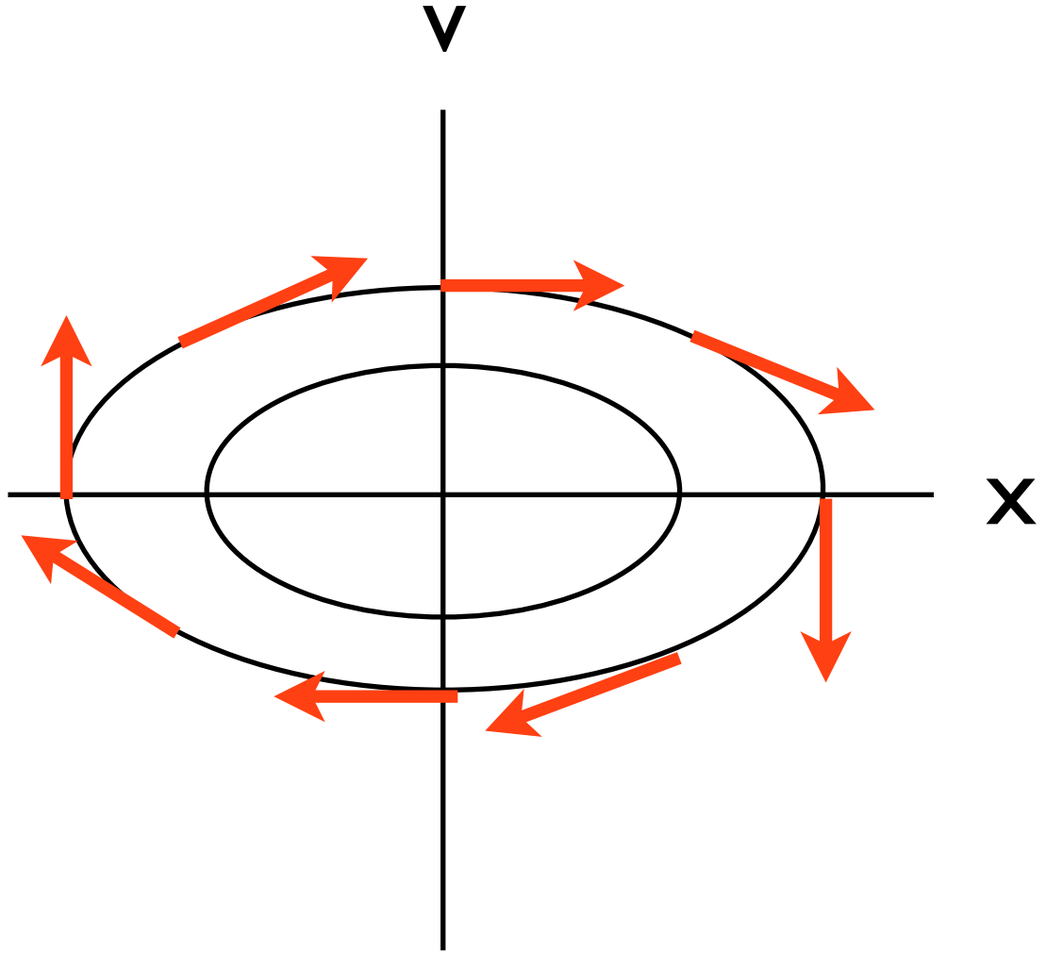
$$\dot{\vec{x}} = f(\vec{x})$$

$$\dot{q}_i = \frac{\partial H}{\partial p_i}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}$$

$$\dot{x} = v$$

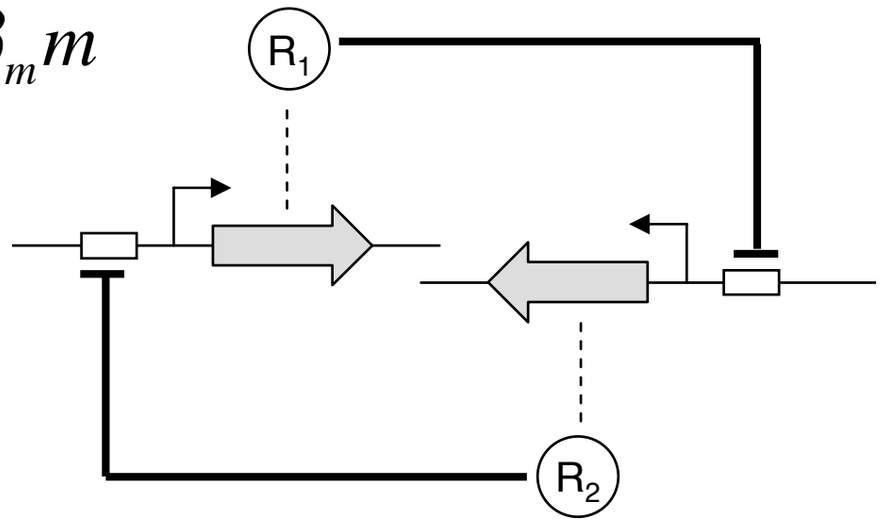
$$\dot{v} = -m\omega^2 x$$



Toggle switch

$$\frac{dm_1}{dt} = \alpha_m g_R(r_2) - \beta_m m$$

$$\frac{dR_1}{dt} = \alpha_p m_1 - \beta_p R_1$$



$$\frac{dm_2}{dt} = \alpha_m g_R(R_2) - \beta_m m_2$$

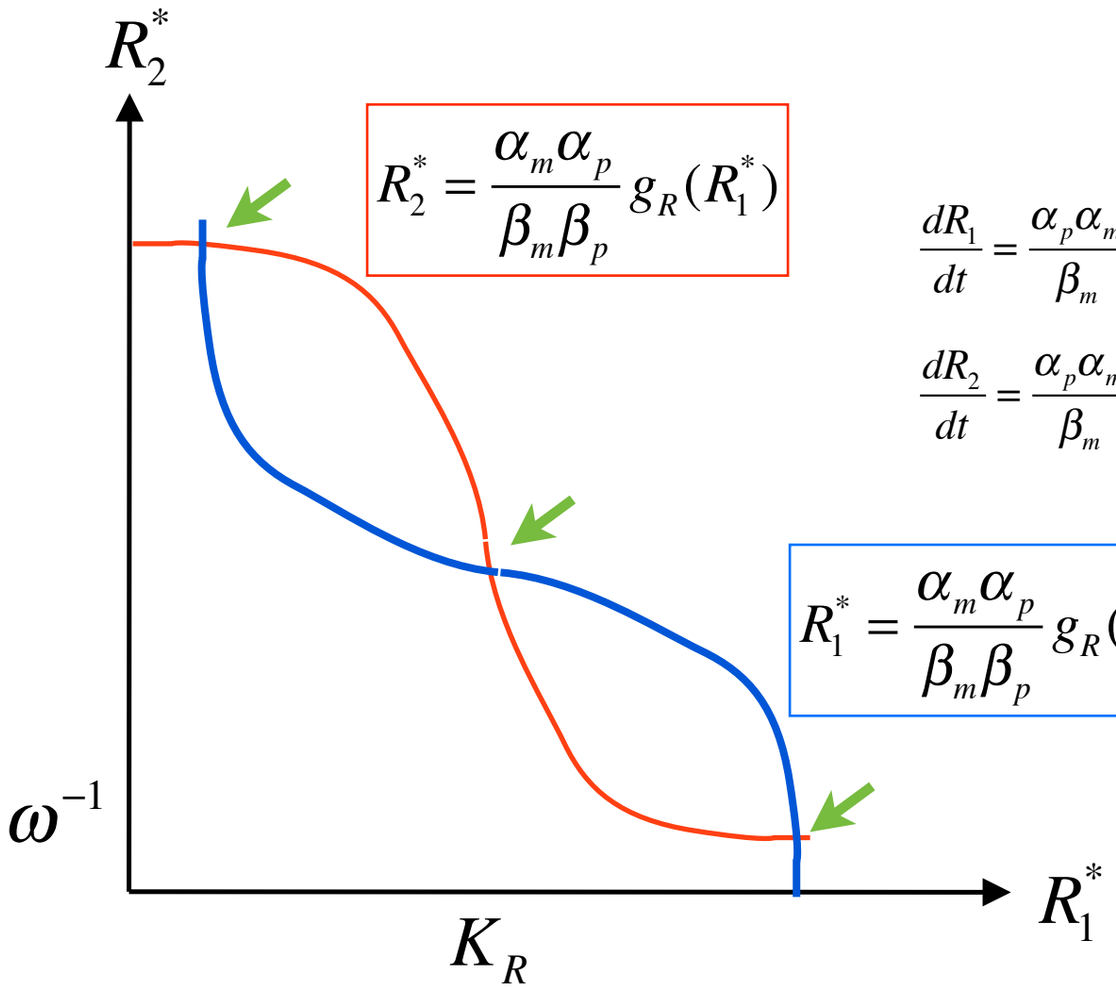
$$\frac{dR_2}{dt} = \alpha_p m_2 - \beta_p R_2$$

For *Ecoli*, mRNA degrades much faster than proteins, and reaches the SS values. $\beta_m \gg \beta_p$

$$m_1^*(R_2) = \frac{\alpha_m}{\beta_m} g_R(R_2) \qquad m_2^*(R_1) = \frac{\alpha_m}{\beta_m} g_R(R_1)$$

$$\frac{dR_1}{dt} = \frac{\alpha_p \alpha_m}{\beta_m} g_R(R_2) - \beta_p R_1 \xrightarrow{ss} R_1^* = \frac{\alpha_m \alpha_p}{\beta_m \beta_p} g_R(R_2^*)$$

$$\frac{dR_2}{dt} = \frac{\alpha_p \alpha_m}{\beta_m} g_R(R_1) - \beta_p R_2 \xrightarrow{ss} R_2^* = \frac{\alpha_m \alpha_p}{\beta_m \beta_p} g_R(R_1^*)$$



$$R_2^* = \frac{\alpha_m \alpha_p}{\beta_m \beta_p} g_R(R_1^*)$$

$$R_1^* = \frac{\alpha_m \alpha_p}{\beta_m \beta_p} g_R(R_2^*)$$

$$\frac{dR_1}{dt} = \frac{\alpha_p \alpha_m}{\beta_m} g_R(R_2) - \beta_p R_1$$

$$\frac{dR_2}{dt} = \frac{\alpha_p \alpha_m}{\beta_m} g_R(R_1) - \beta_p R_2$$

At crossing points

$$g_R(R_1^*) = g_R(R_2^*)$$

$$\frac{dR_1}{dt} = \frac{dR_2}{dt} = 0$$

... to assess the stability of these fixed points, consider $\delta R_i = R_i - R_i^*$

$$\frac{dR_1}{dt} = \frac{\alpha_p \alpha_m}{\beta_m} g_R(R_2) - \beta_p R_1 = F(R_1, R_2)$$

$$\frac{dR_2}{dt} = \frac{\alpha_p \alpha_m}{\beta_m} g_R(R_1) - \beta_p R_2 = G(R_1, R_2)$$

$$\rightarrow \quad \frac{d}{dt}\begin{pmatrix} \delta R_1 \\ \delta R_2 \end{pmatrix} = \begin{pmatrix} \frac{dF}{dR_1} & \frac{dF}{dR_2} \\ \frac{dG}{dR_1} & \frac{dG}{dR_2} \end{pmatrix}_{R_1^*, R_2^*} \begin{pmatrix} \delta R_1 \\ \delta R_2 \end{pmatrix}$$

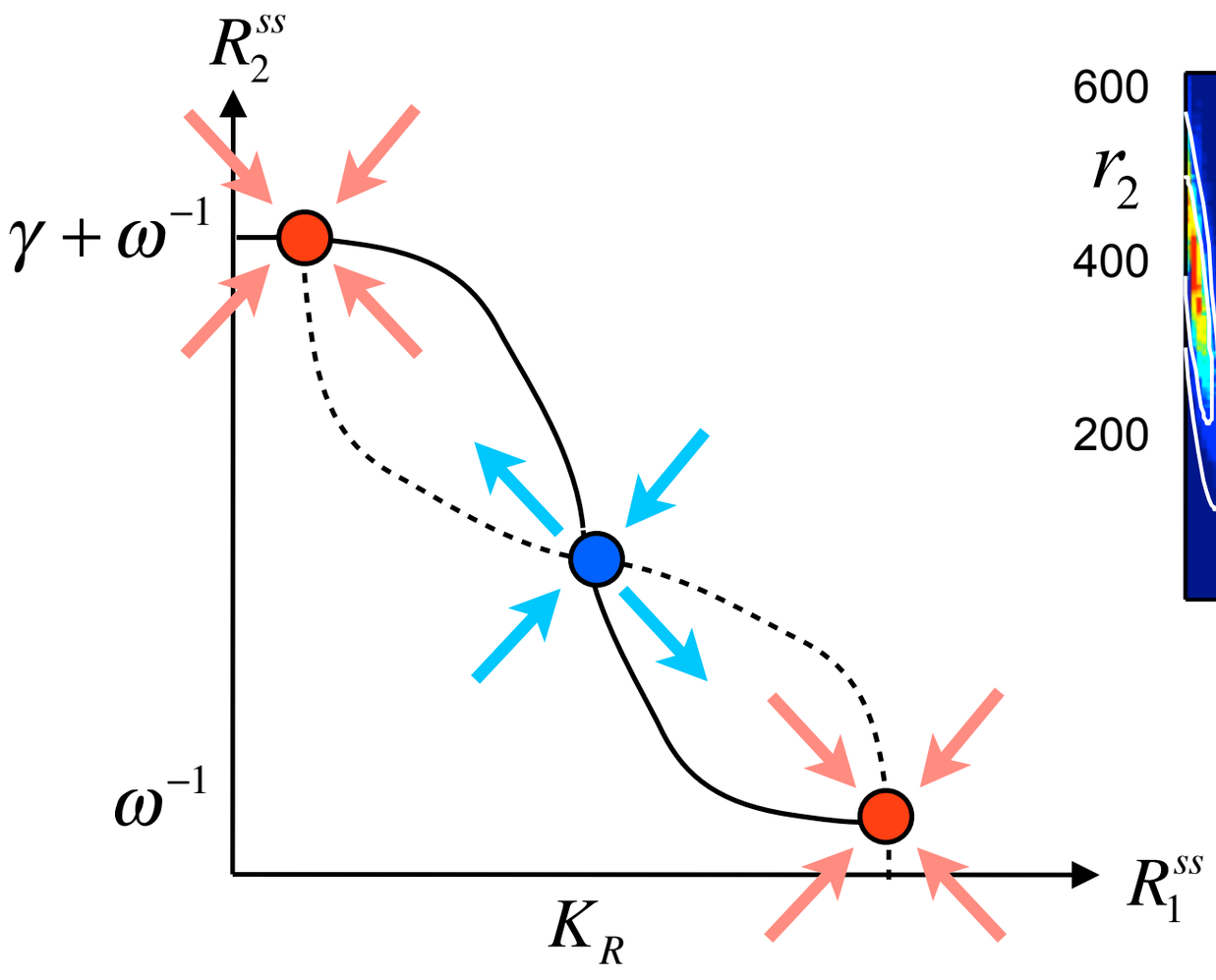
$$\rightarrow \begin{pmatrix} \delta R_1 \\ \delta R_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} e^{\lambda_1 t} + \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} e^{\lambda_2 t}$$

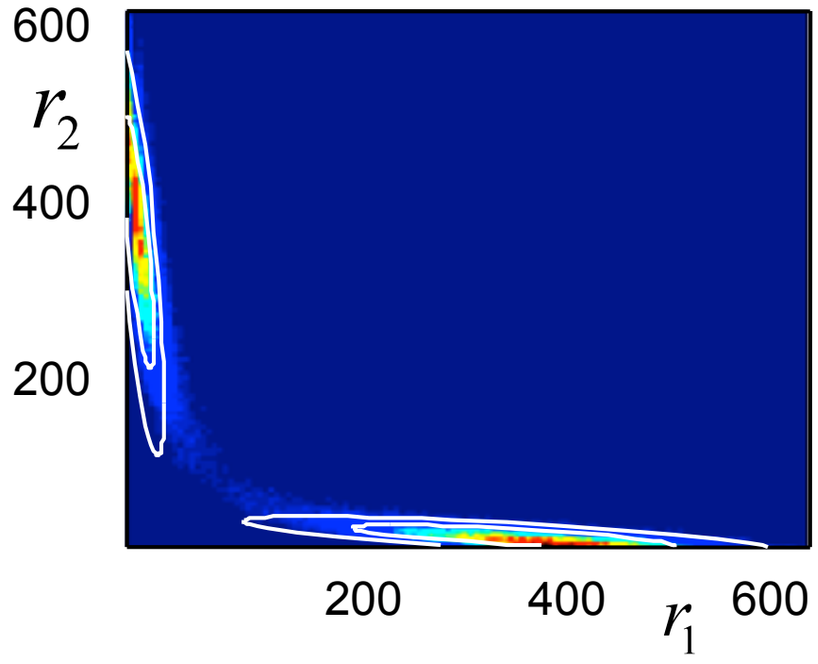
$$\begin{pmatrix} -\beta_p - \lambda & \frac{\alpha_p \alpha_m}{\beta_m} g'_R(R_2^*) \\ \frac{\alpha_p \alpha_m}{\beta_m} g'_R(R_1^*) & -\beta_p - \lambda \end{pmatrix}_{R_1^*, R_2^*} \begin{pmatrix} X \\ Y \end{pmatrix} = 0$$

$$\lambda_{1,2} = -\beta_p \pm \frac{\alpha_m \alpha_p}{\beta_m} \sqrt{g_R'(R_1^*) g_R'(R_2^*)}$$

Then, at each fixed point

- $g_R'(R_1^*) = g_R'(R_2^*) \approx 0 \rightarrow \lambda_{1,2} < 0$ stable fixed point
- $g_R'(R_1^*) = g_R'(R_2^*) \approx 1 \rightarrow \lambda_1 > 0, \lambda_2 < 0$ unstable



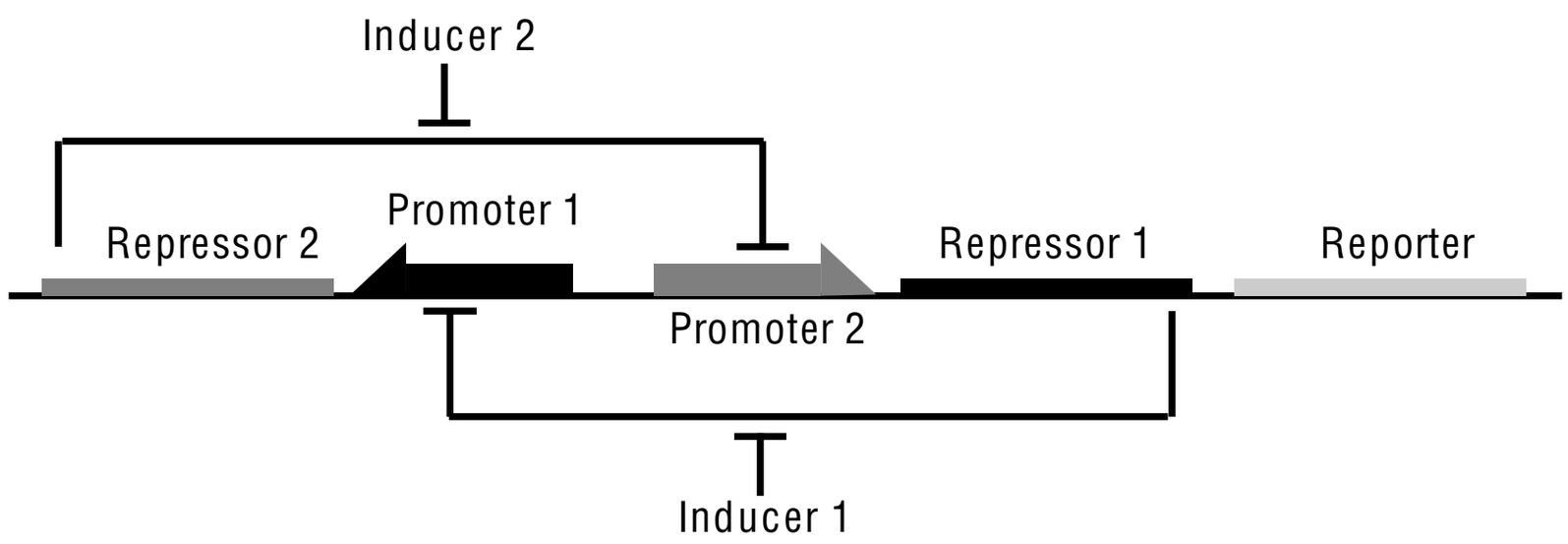


Construction of a genetic toggle switch in *Escherichia coli*

Timothy S. Gardner*†, Charles R. Cantor* & James J. Collins*†

** Department of Biomedical Engineering, † Center for BioDynamics and ‡ Center for Advanced Biotechnology, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA*

Nature (2000) 403:339

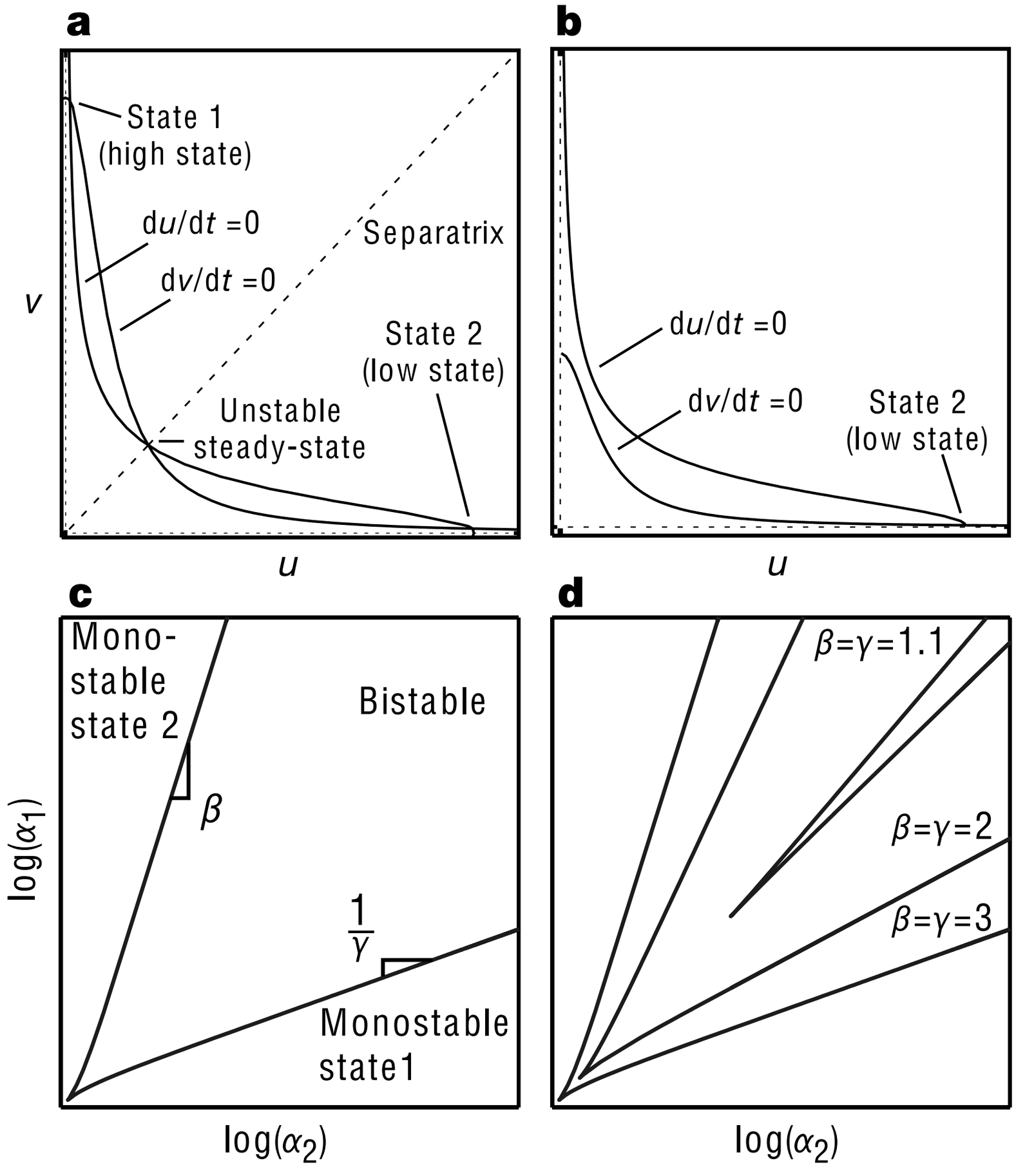


$$\frac{du}{dt} = \frac{\alpha_1}{1 + v^{\beta}} - u$$

$$\frac{dv}{dt} = \frac{\alpha_2}{1 + u^{\gamma}} - v$$

$$\frac{du}{dt} = \frac{\alpha_1}{1 + v^{\beta}} - u$$

$$\frac{dv}{dt} = \frac{\alpha_2}{1 + u^{\gamma}} - v$$

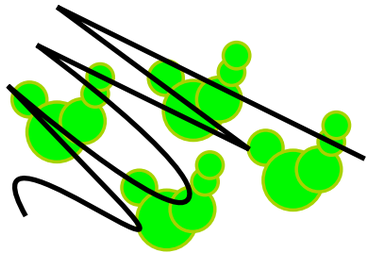


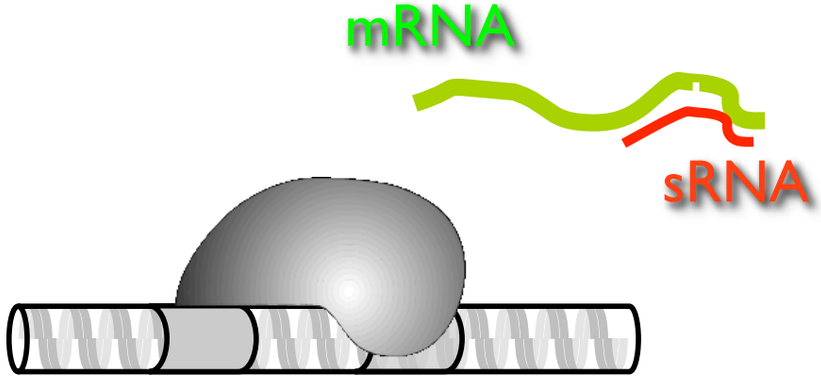
OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

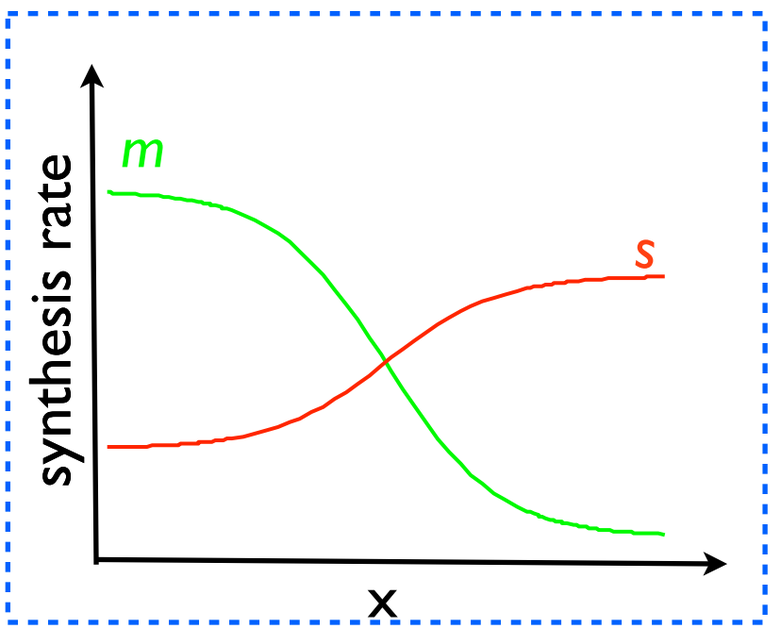
Small Regulatory RNAs May Sharpen Spatial Expression Patterns

Erel Levine, **Peter McHale**, **Herbert Levine***

Center for Theoretical Biological Physics, University of California San Diego, La Jolla, California, United States of America



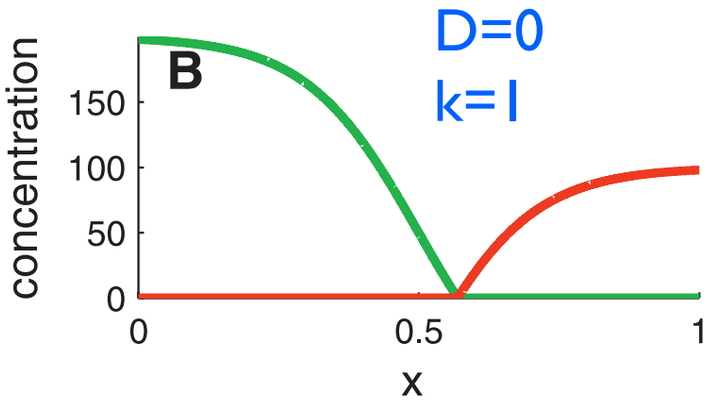


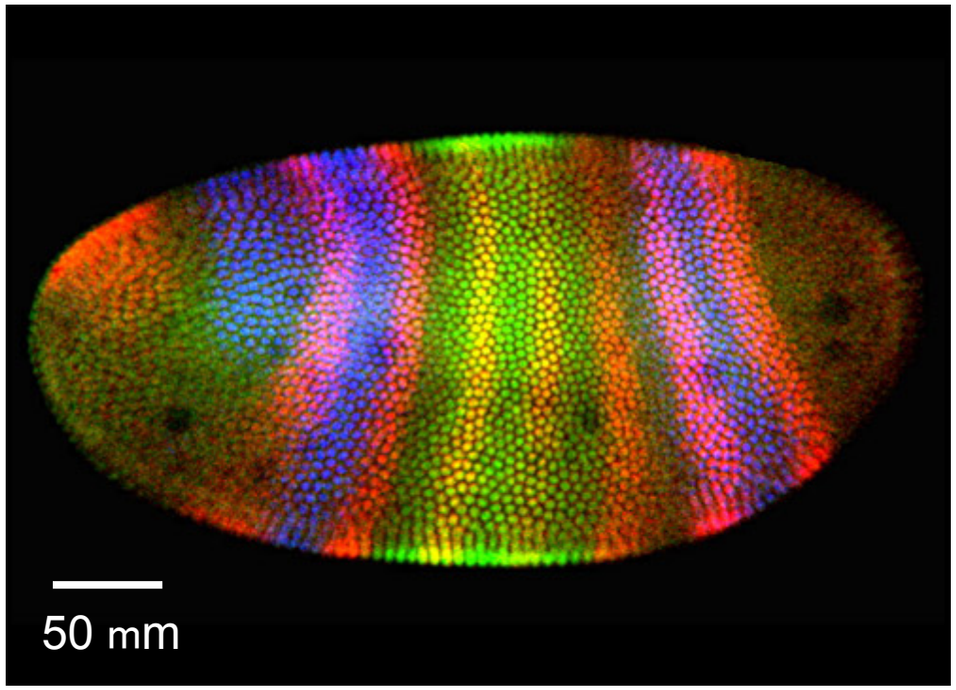


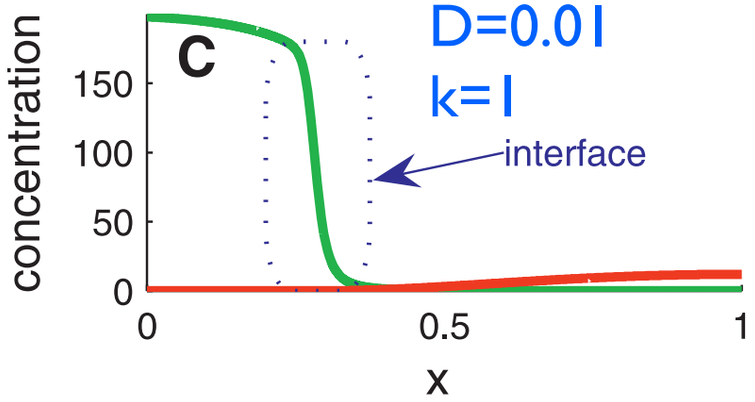
x-dependent synthesis rate

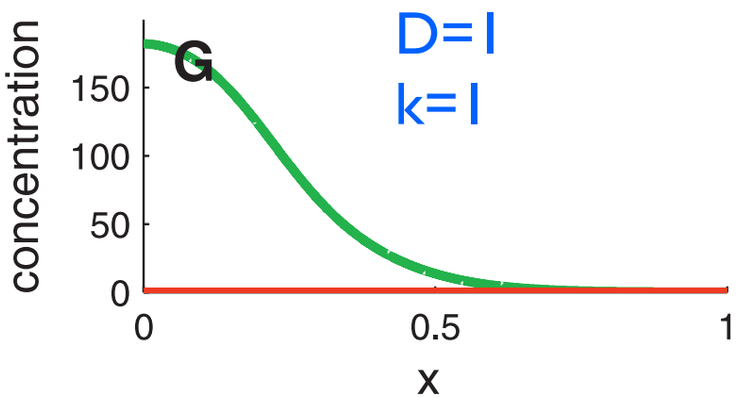
$$\frac{\partial m}{\partial t} = \alpha_m(x) - \beta_m \cdot m - km \cdot s$$

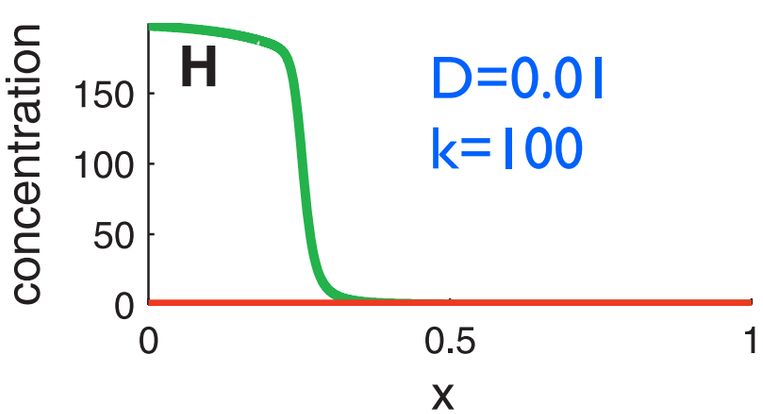
$$\frac{\partial s}{\partial t} = \alpha_s(x) - \beta_s \cdot s - km \cdot s + D\nabla_x^2 s$$

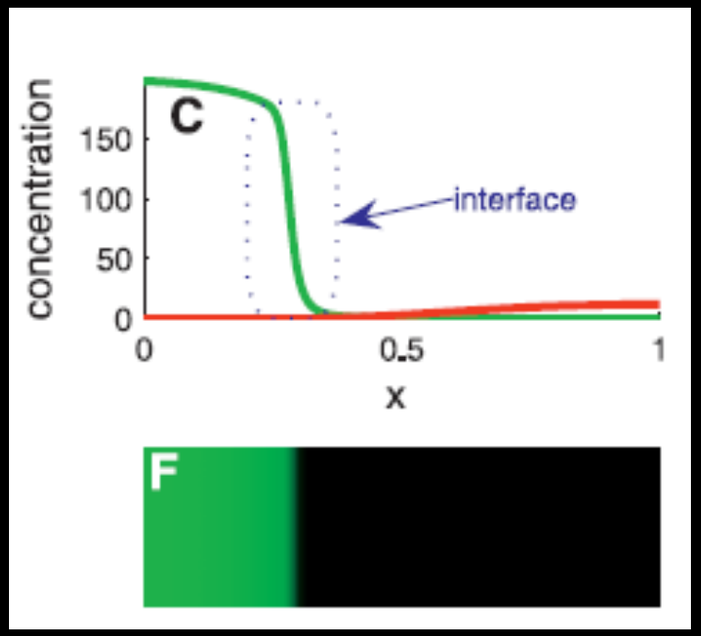

B
D=0
k=1
concentration
150
100
50
0
0
0.5
1
x


50 mm

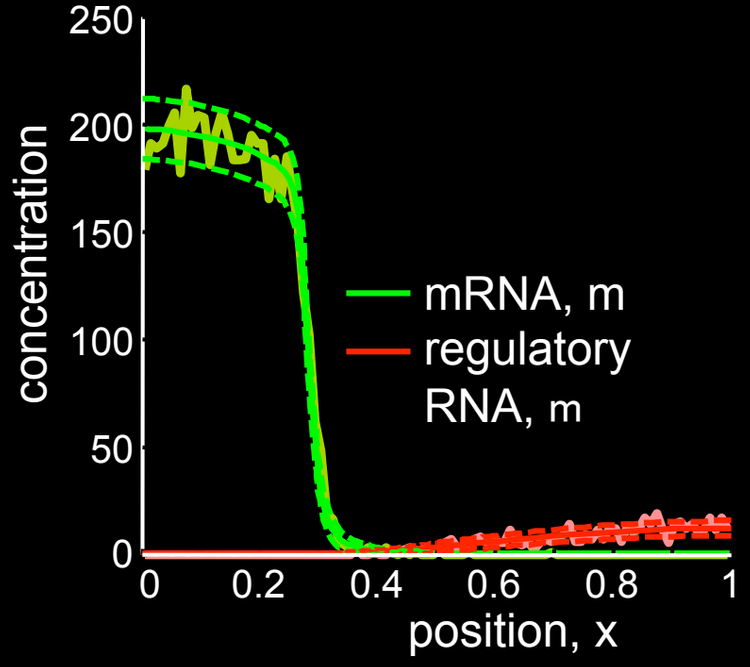

C
D=0.01
k=1
interface
concentration
150
100
50
0
0
0.5
1
x

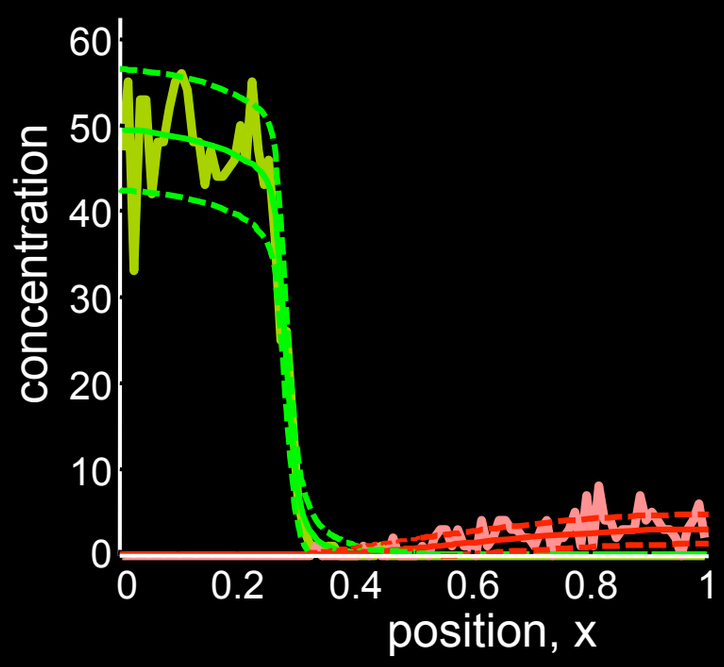

G
D=1
k=1
concentration
150
100
50
0
0
0.5
1
x

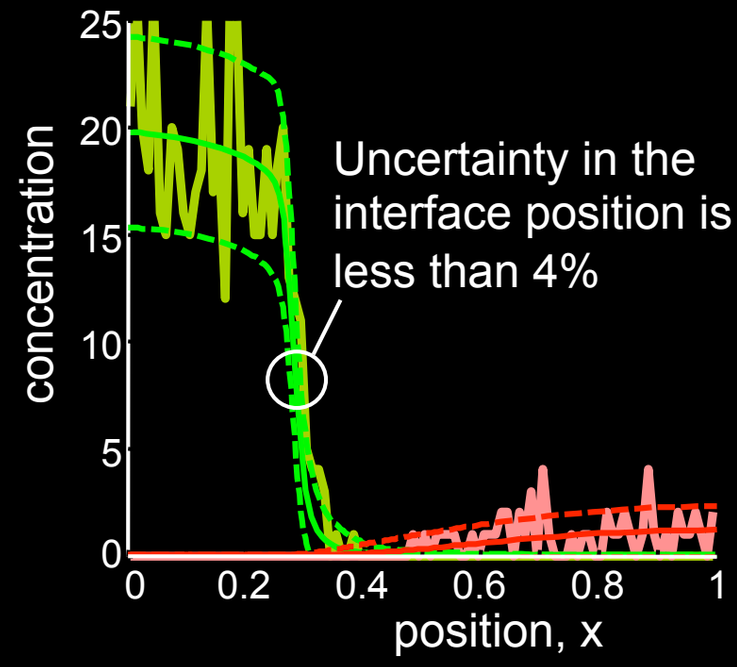

H
D=0.01
k=100
concentration
150
100
50
0
0
0.5
1
x



Look at the mean and variance about a spatially *inhomogeneous* steady-state.







Patterning mechanism remains robust in *very small* populations