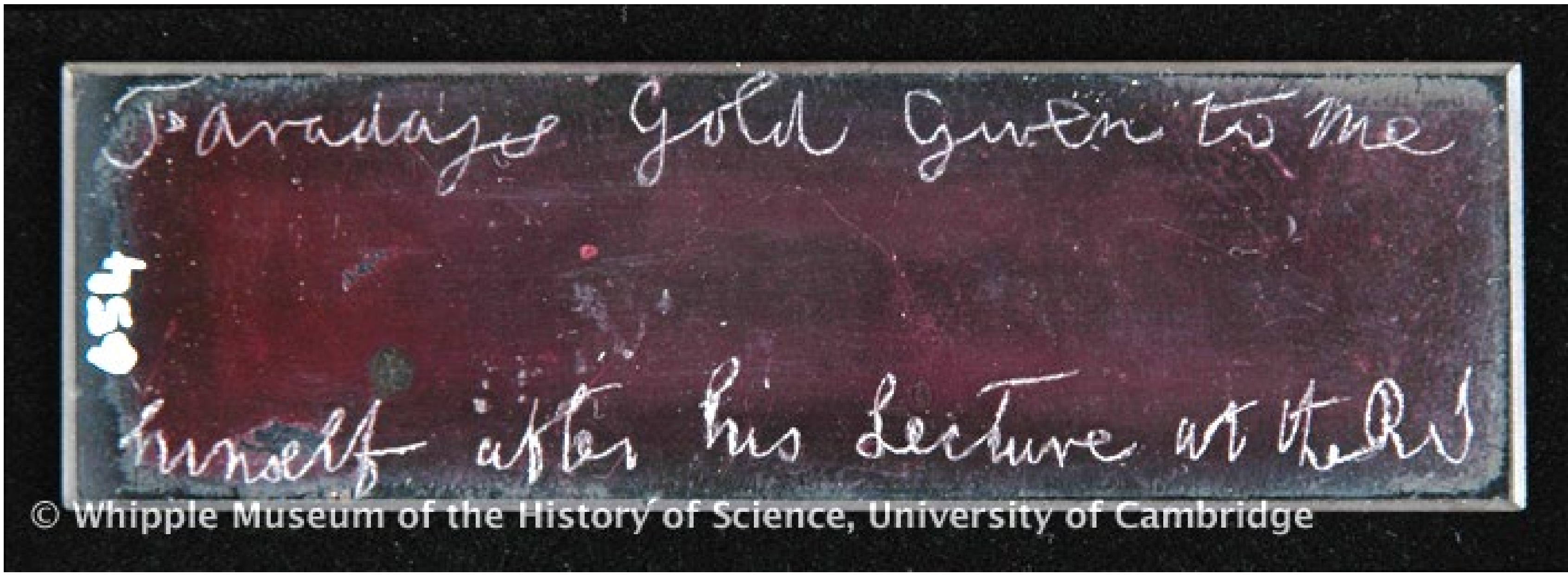
Advanced Nanotechnology (B)

- Class hour: every Friday, 09:30-12:10
- Classroom: P101, 1F, IoP

What is Nanotechnology?

- Nanotechnology is the science, engineering, and technology conducted at the nanoscale, which is about 1 to 100 nanometers.
- Significance: Manipulating matter on an atomic, molecular, and supramolecular scale for the fabrication of macro-scale products.
- Applications: medicine, electronics, biomaterials, energy production, and environmental protection

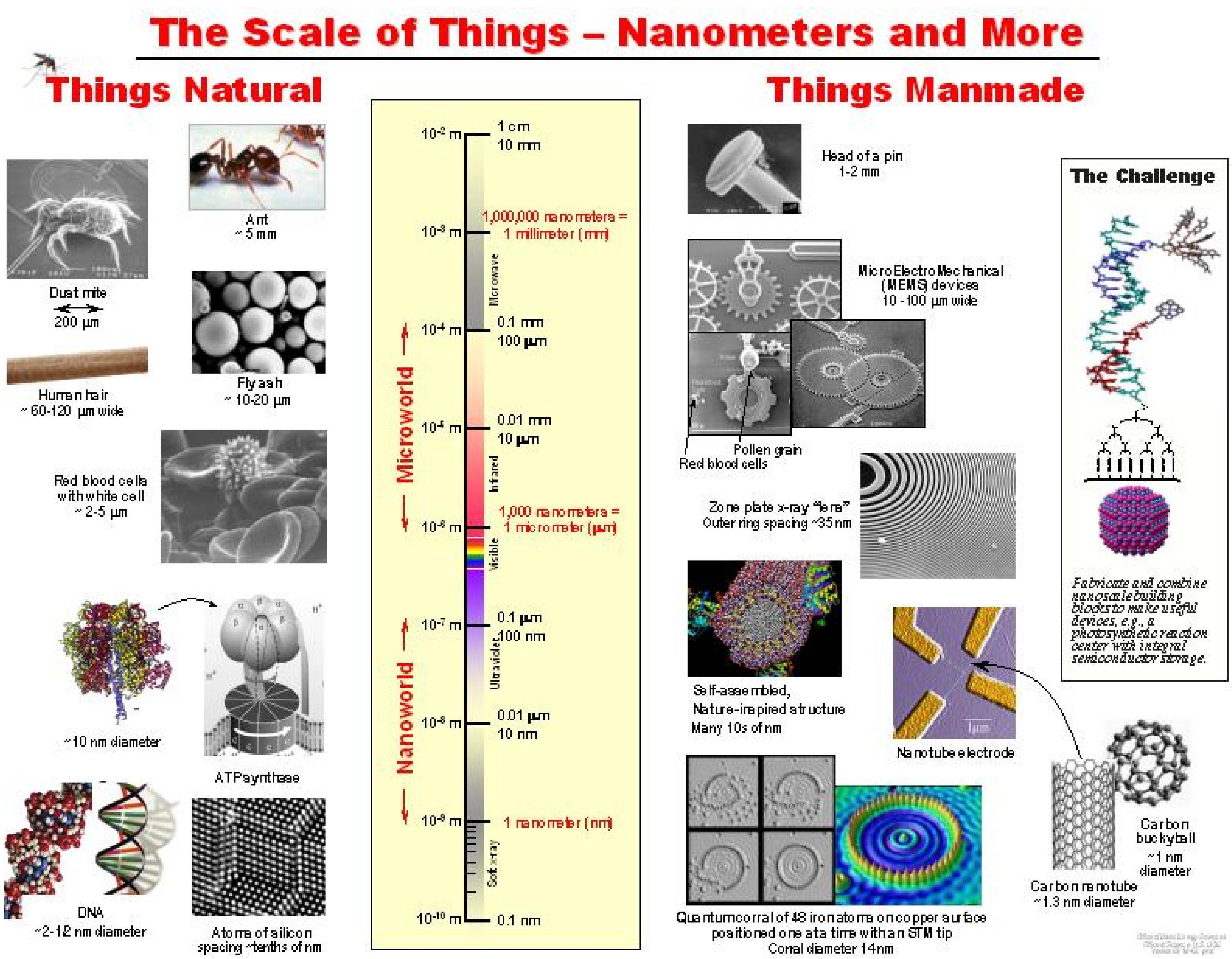
Faraday’s Gold Sol



1856
20-40 nm gold

http://cohesion.rice.edu/CentersAndInst/CNST/emplibrary/Scale%20of%20Nanotechnology.jpg

What is nano?



Richard Feynman,1959

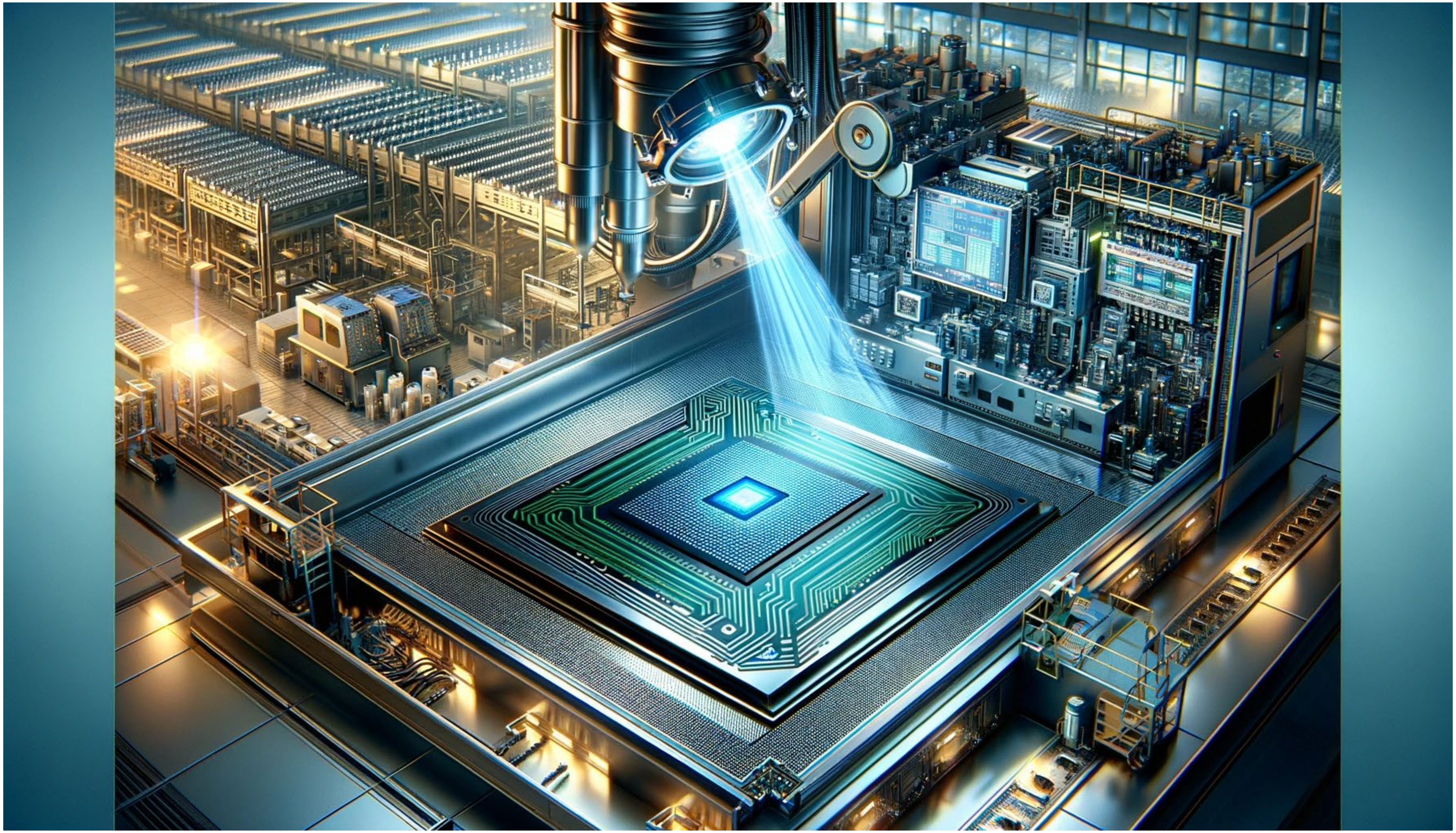
There's plenty of room at the bottom

President Clinton, 2000

The National Nanotechnology Initiative

"Imagine the possibilities: materials with ten times the strength of steel and only a small fraction of the weight -- shrinking all the information housed at the Library of Congress into a device the size of a sugar cube -- detecting cancerous tumors when they are only a few cells in size."

Chips fabricated by the Optical Lithography



Lithography Roadmap

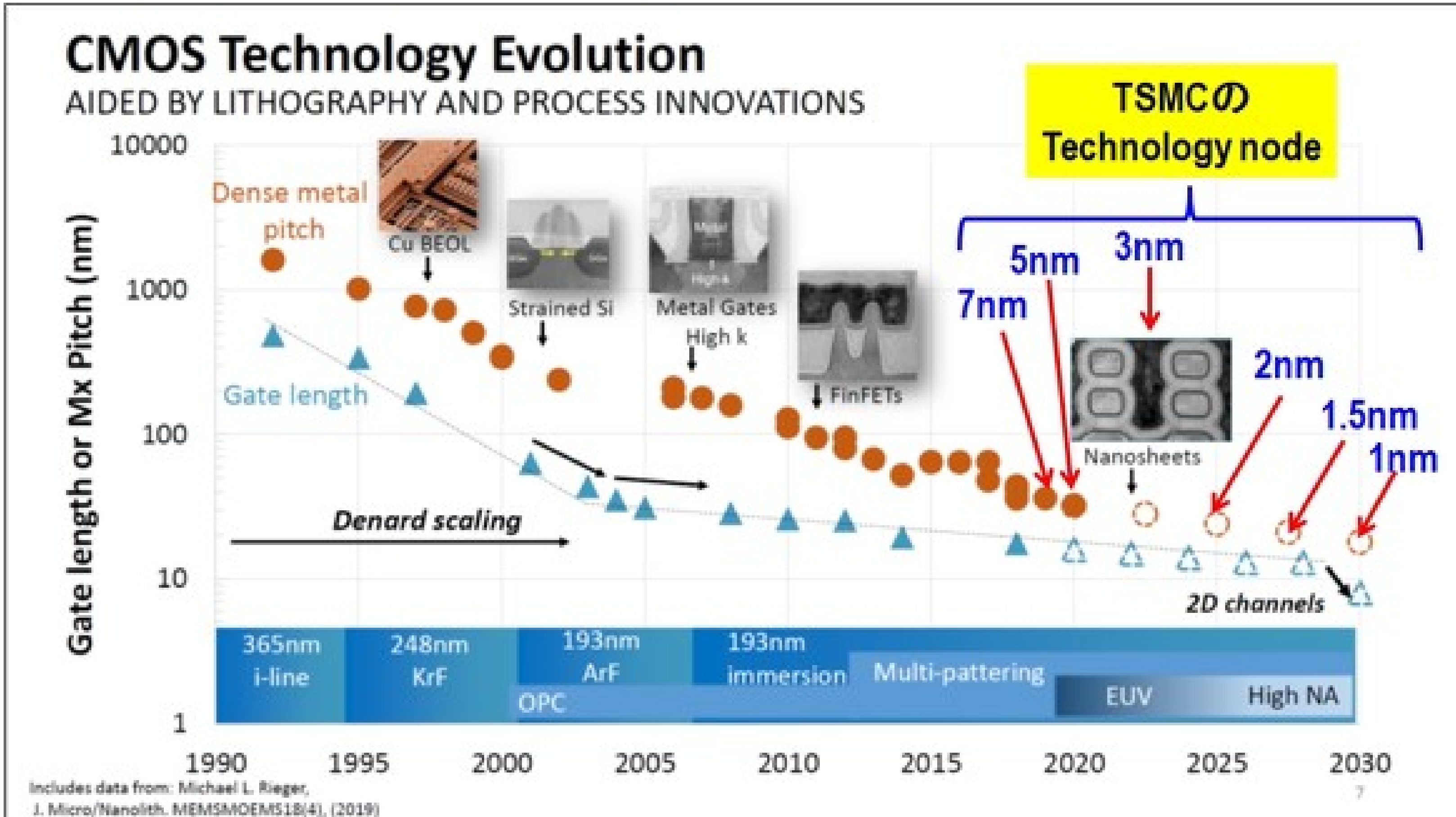
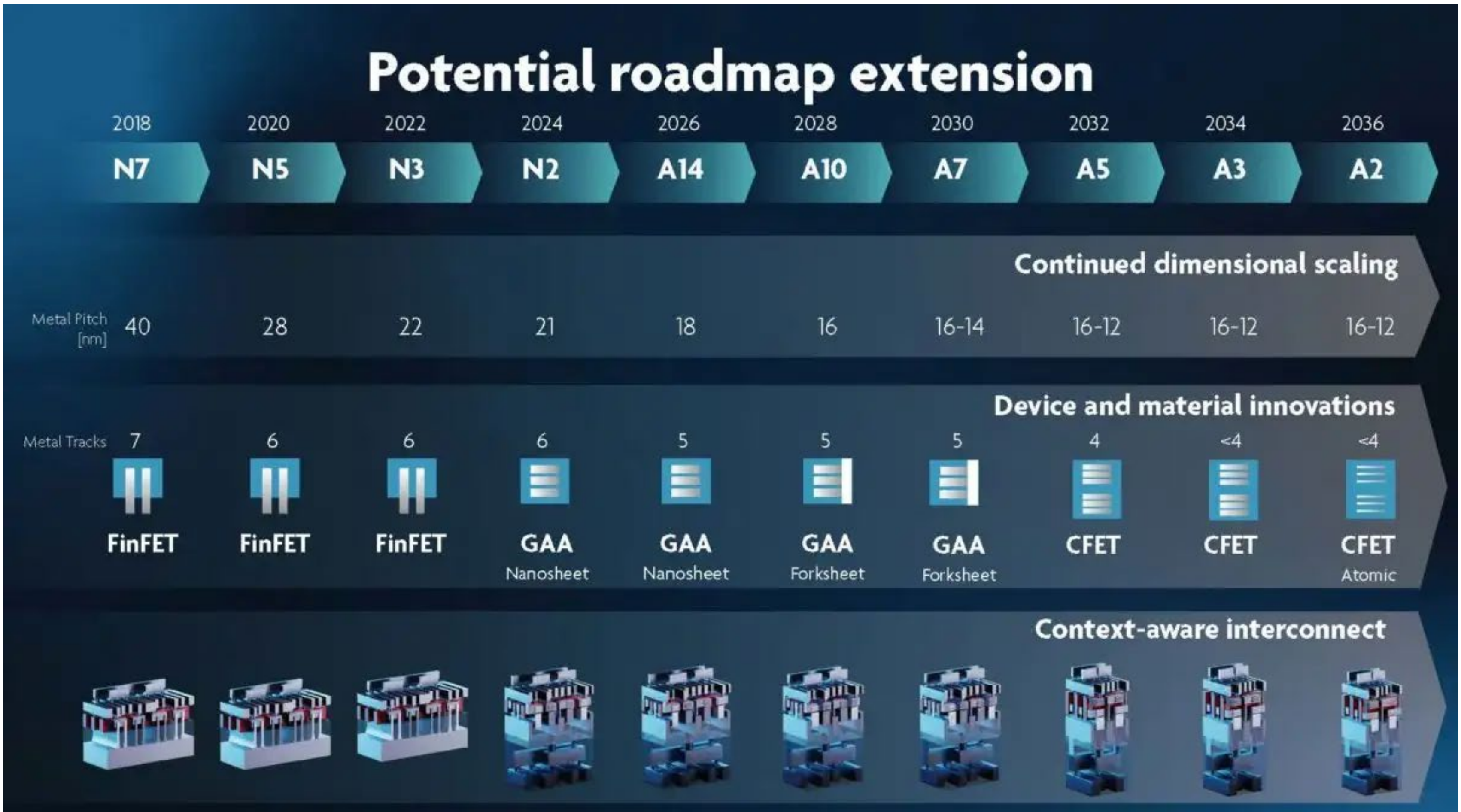


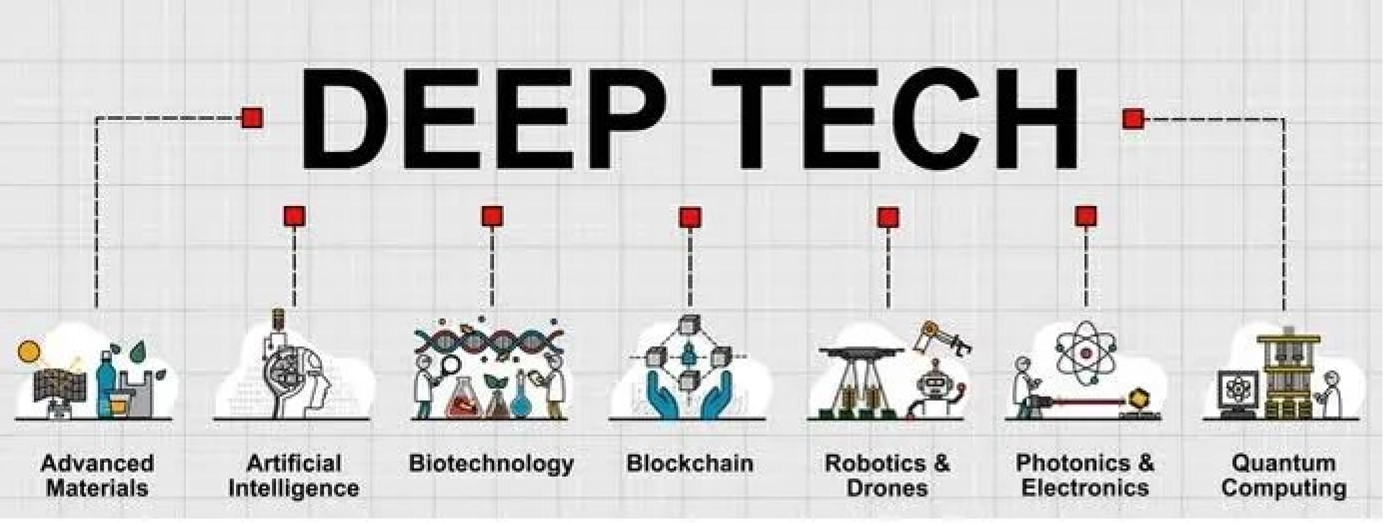
図5 ロジックのCMOS技術の進化

出所：Sri Samavedam（imec）, "Future Logic Scaling: Towards Atomic Channels and Deconstructed Chips", IEDM2020, Plenaryに筆者が加筆.

Lithography Roadmap



Deep Tech



Nano + Info

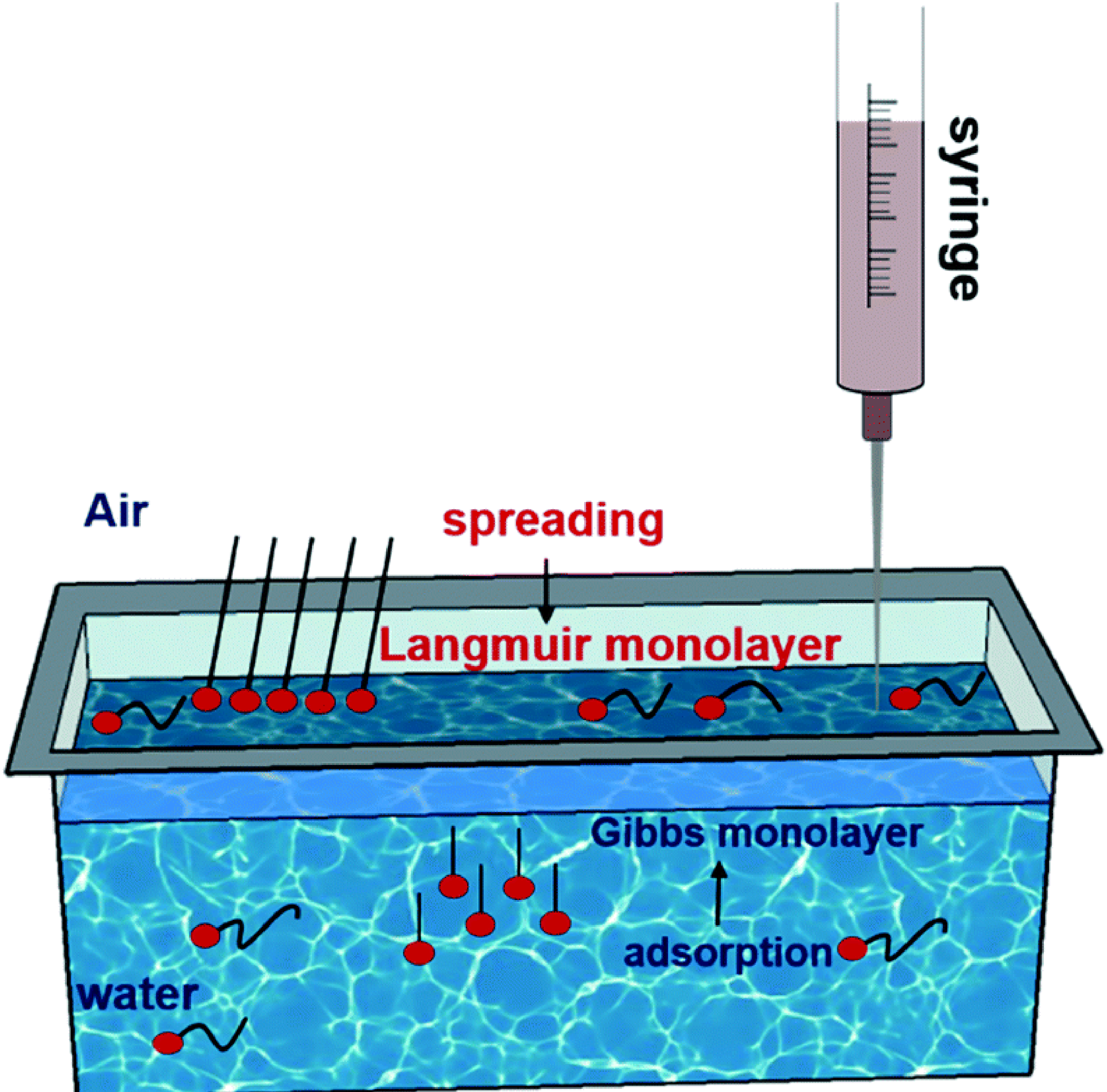
AI


THE NOBEL PRIZE
IN PHYSICS 2024
Illustrations: Niklas Elmehed
John J. Hopfield
Geoffrey E. Hinton
"for foundational discoveries and inventions that enable machine learning with artificial neural networks"
THE ROYAL SWEDISH ACADEMY OF SCIENCES
THE NOBEL PRIZE
IN CHEMISTRY 2024
Illustrations: Niklas Elmehed
David Baker
Demis Hassabis
John M. Jumper
"for computational protein design"
"for protein structure prediction"
THE ROYAL SWEDISH ACADEMY OF SCIENCES

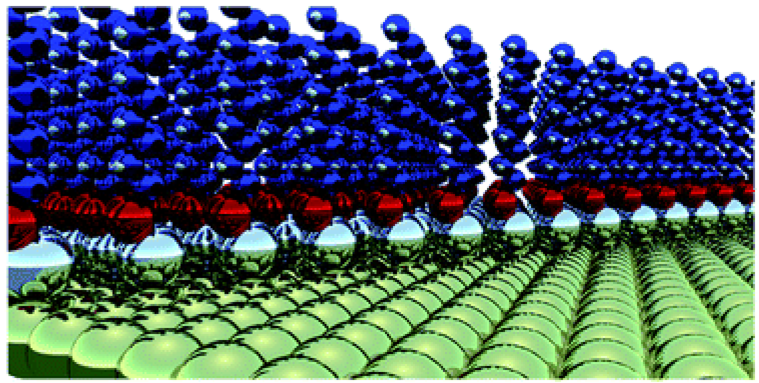
Tentative Schedule

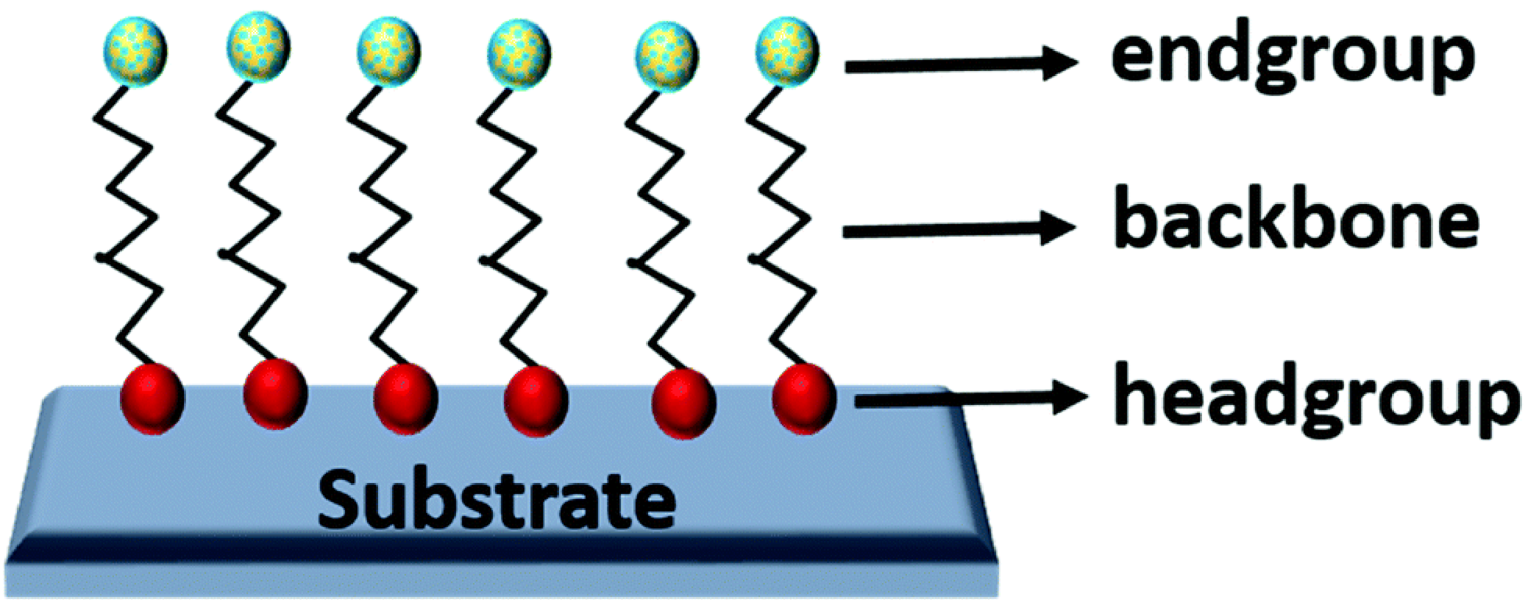
- 3/7 Introduction: Nuclear Acids and Amino Acids, Proteins
- 3/14 Glycobiology and lipid
- 3/21 Dr. Tung
- 3/28 Dr. Tung
- 4/11 NPs Synthesis and Characterization
- 4/18 NP for Diagnostic and Therapeutic
- 4/25 Imaging or other updated topic
- 5/2 AI
- 5/16 AI

Self-Assembly Monolayers

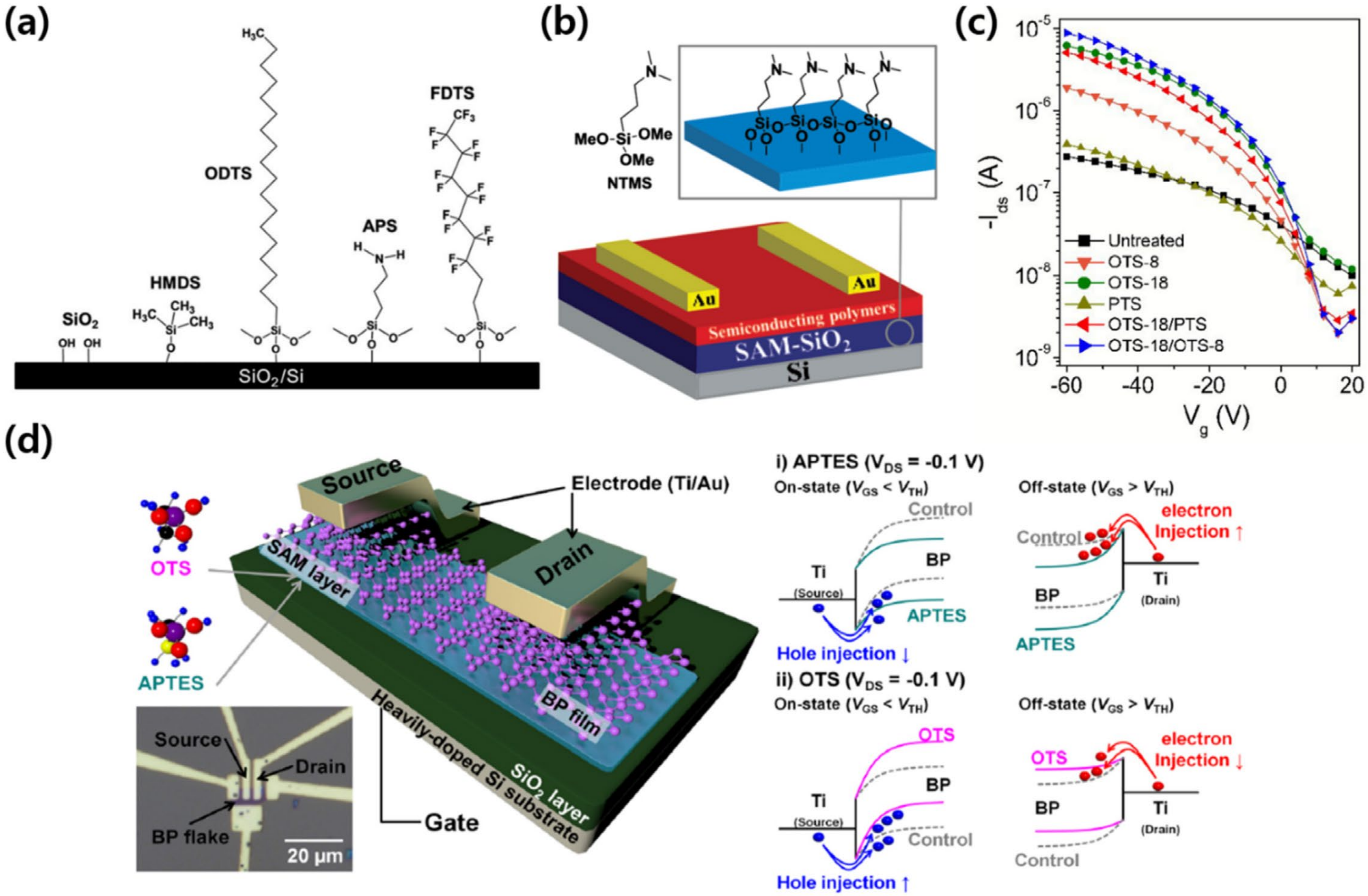


Self-Assembly Monolayers



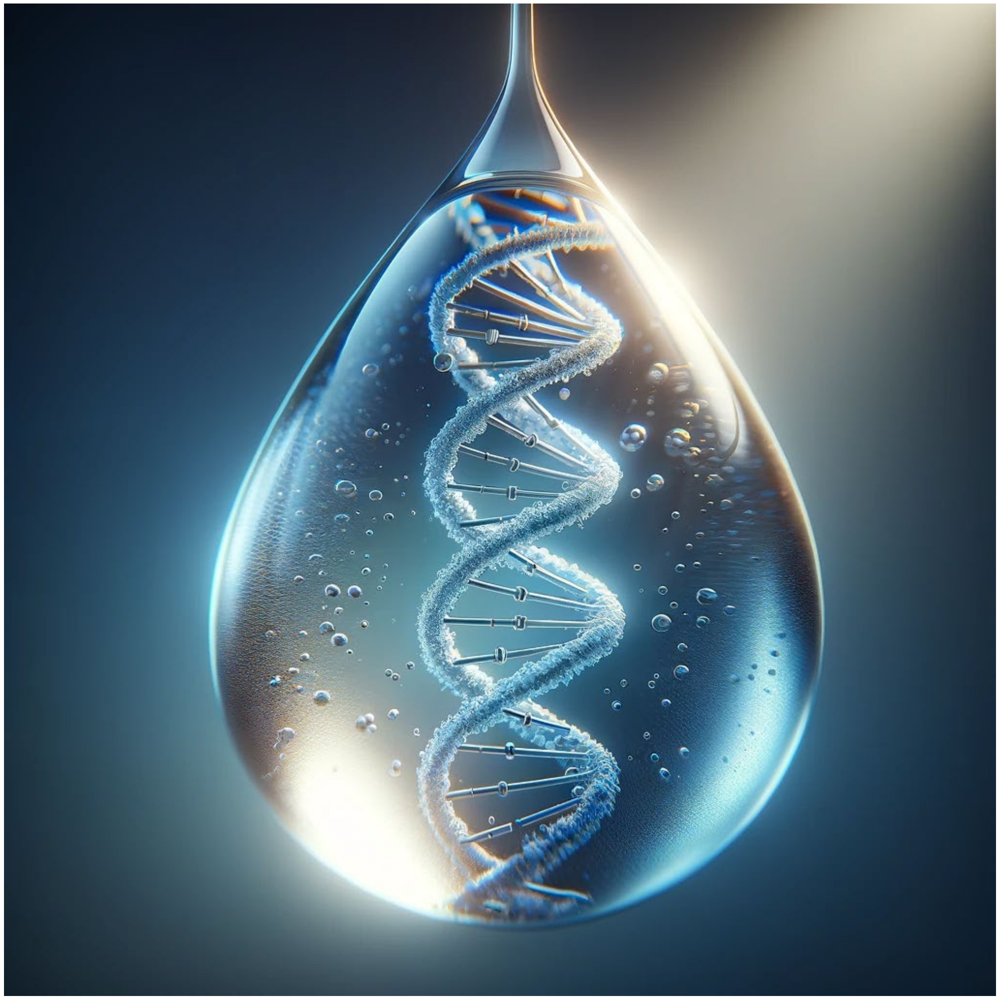


Molecular Electronic



A busy cellular factory

One molecule in one drop of water

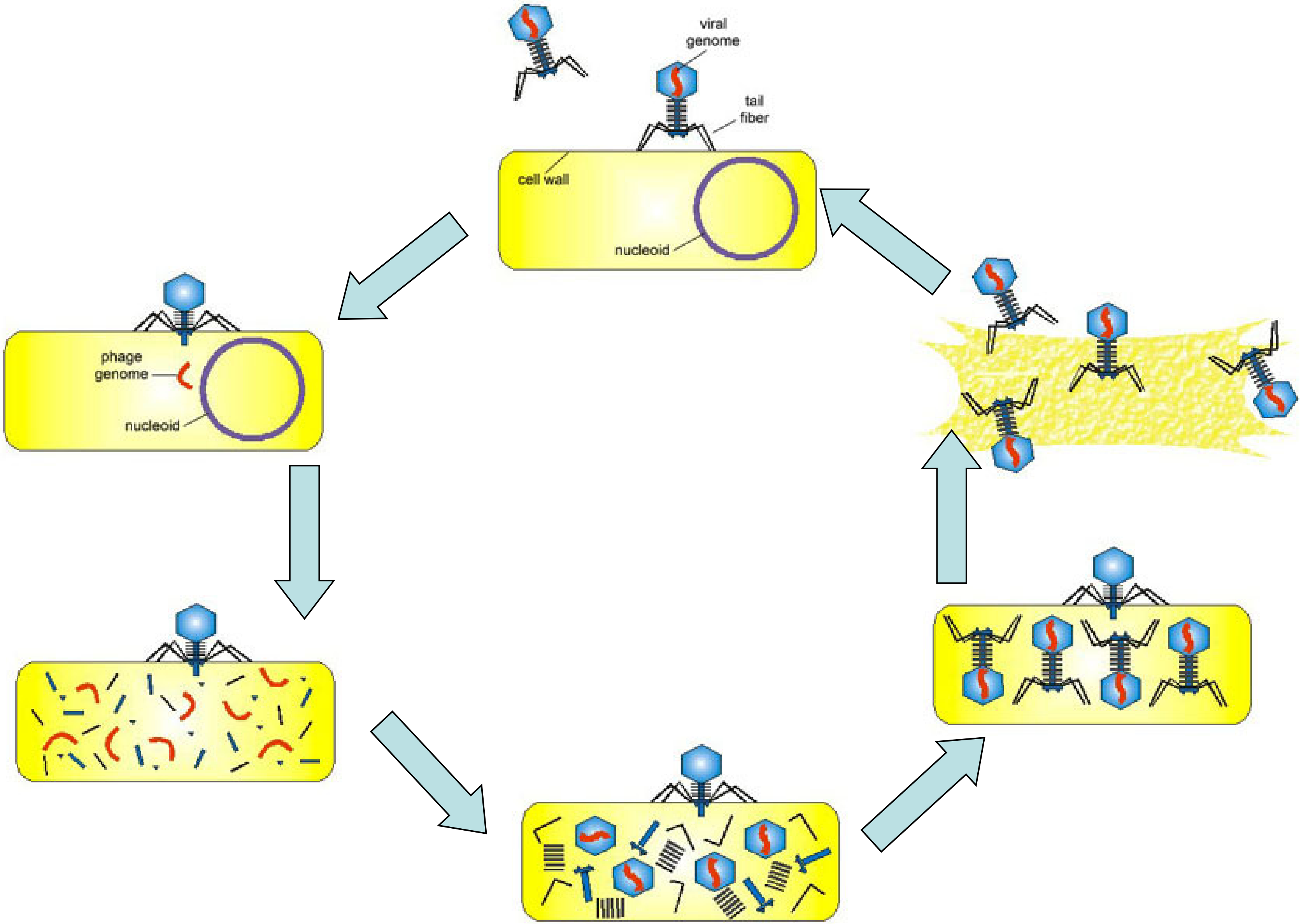


1 drop of water = 1/20 cc

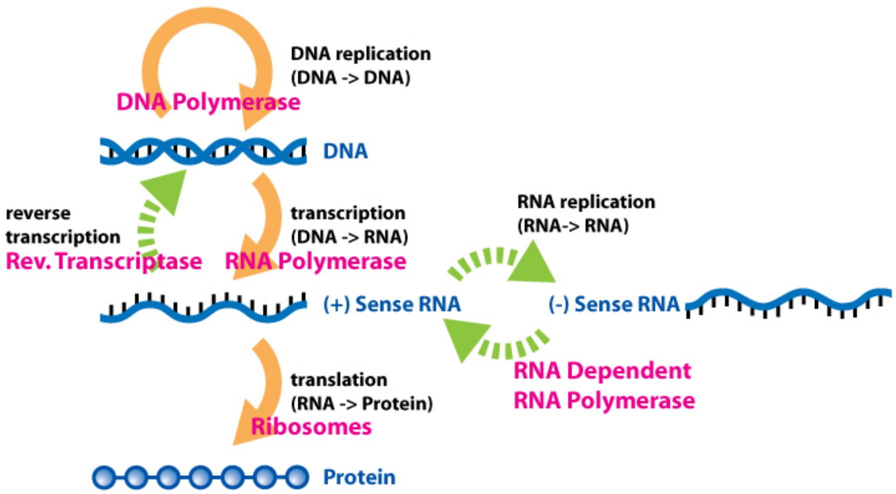
1/(6 x 10^ 23)/(1/20) = 3 x 10^-20 M

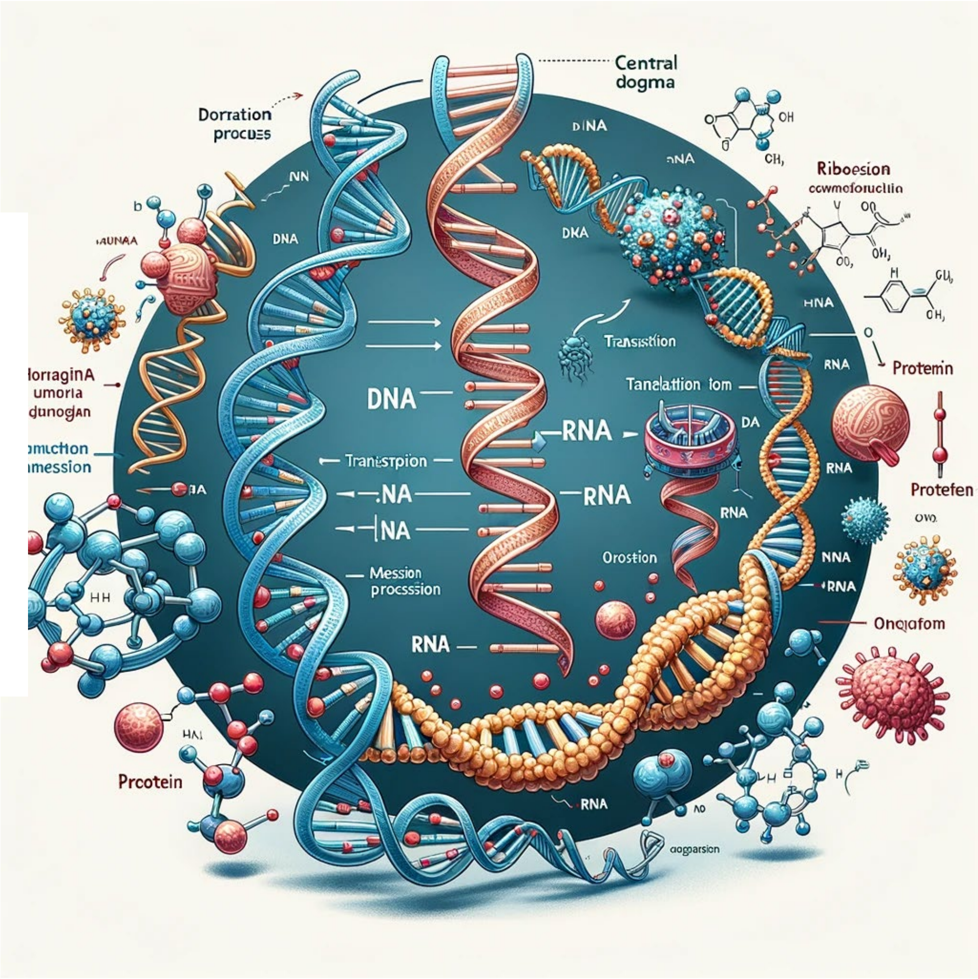
(1 um)^3 = 10^-12cc=10^-15l
1/(6 x 10^ 23)/10^-15l = **1.6 nM**

Virus Reproduction

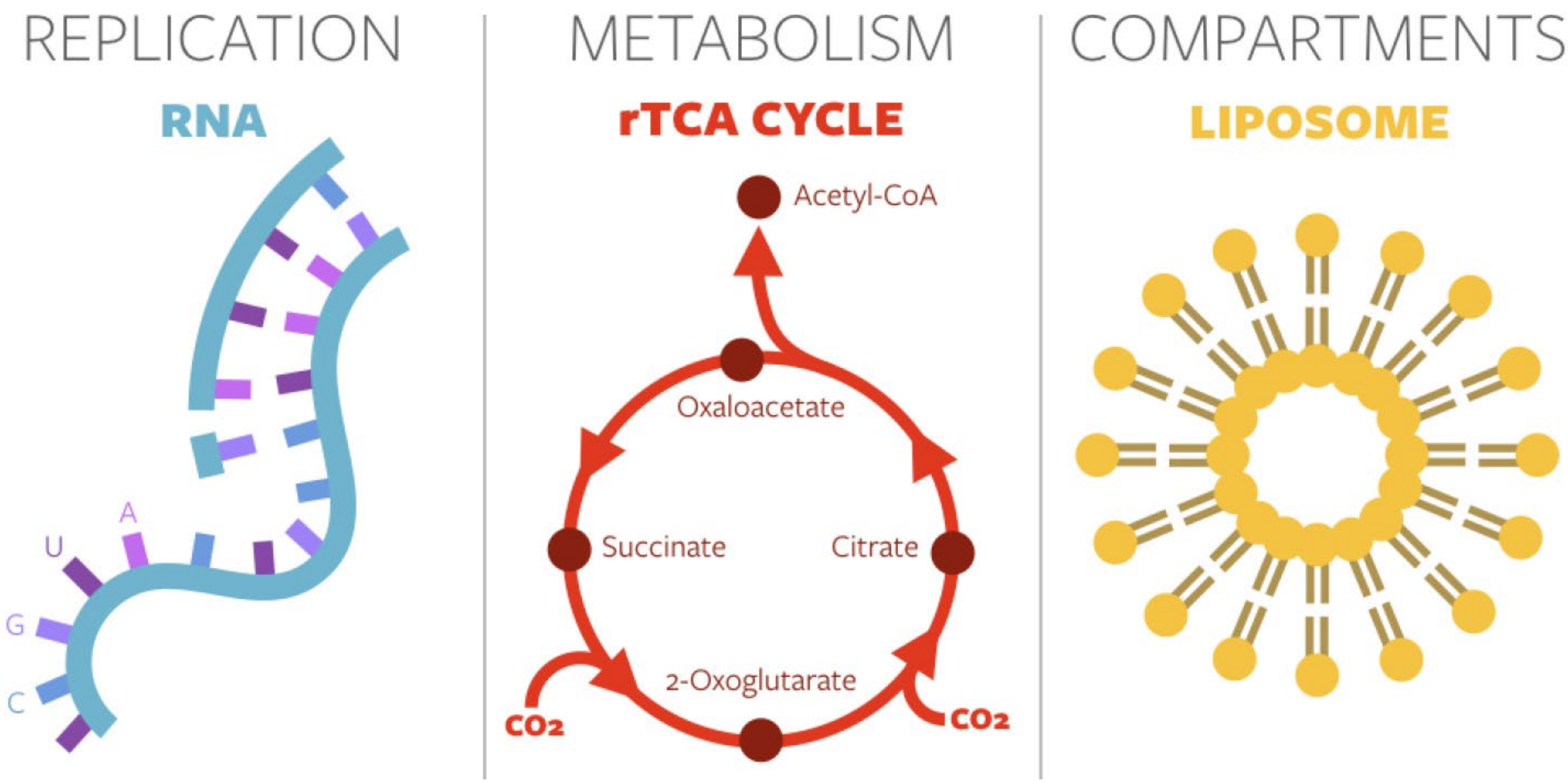


Central Dogma



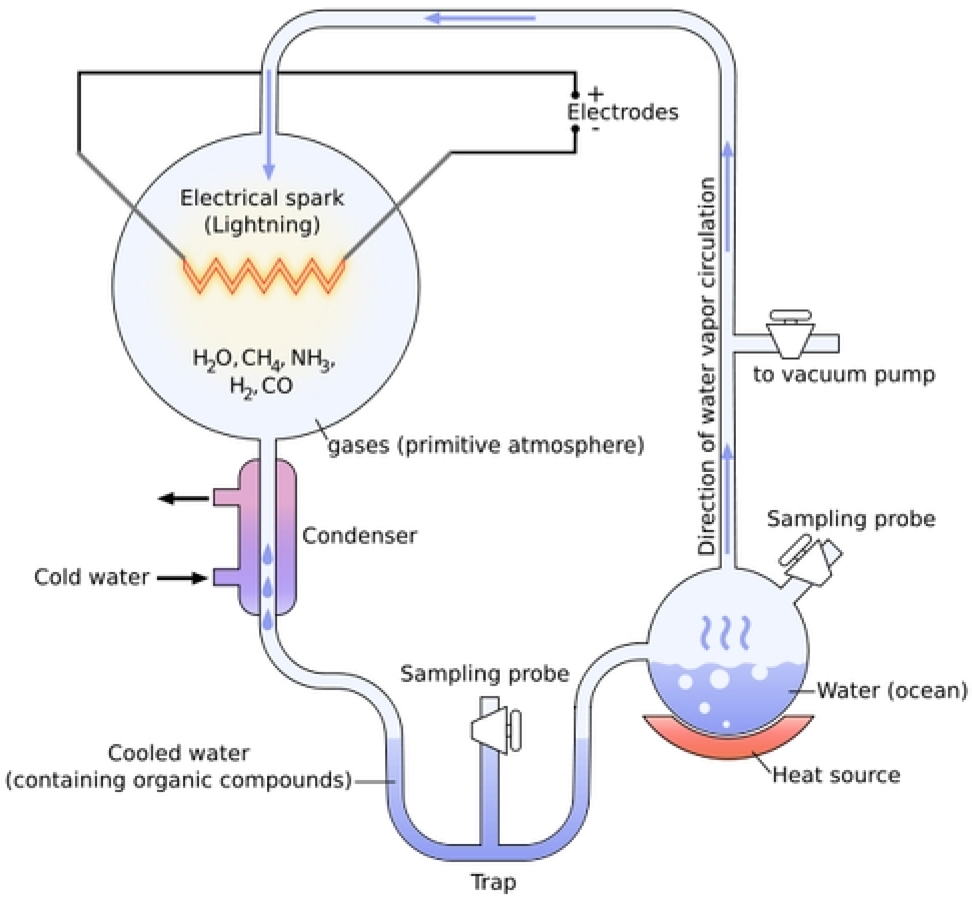


Definition of Life



Early Live on Earth

Miller-Urey experiment (1953)



Amino acids

Where Do We Come From? What Are We? Where Are We Going?

French: D'où venons-nous ? Que sommes-nous ? Où allons-nous ?

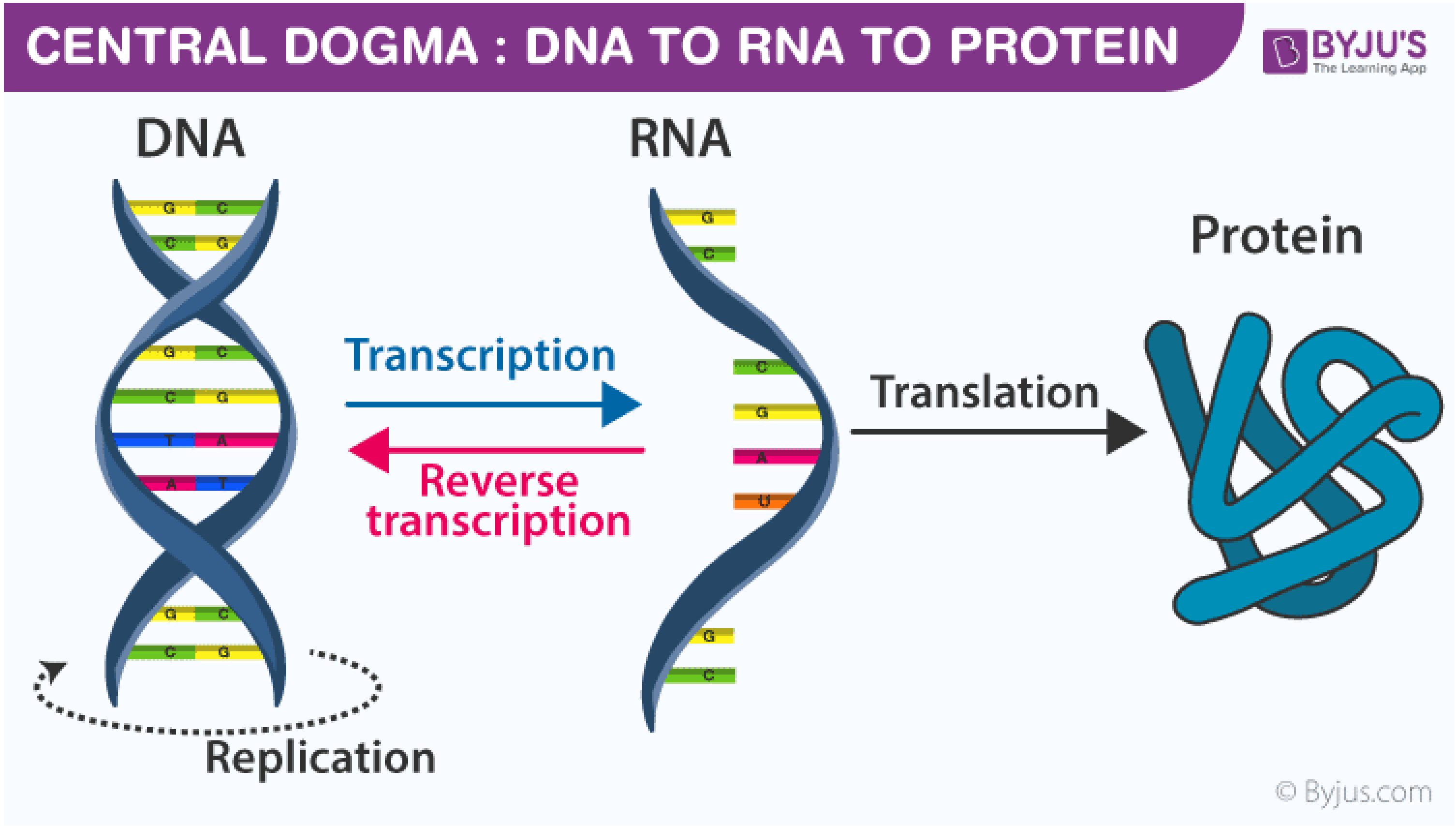
Artist	Paul Gauguin
Year	1897–1898

A Common Language through Lives

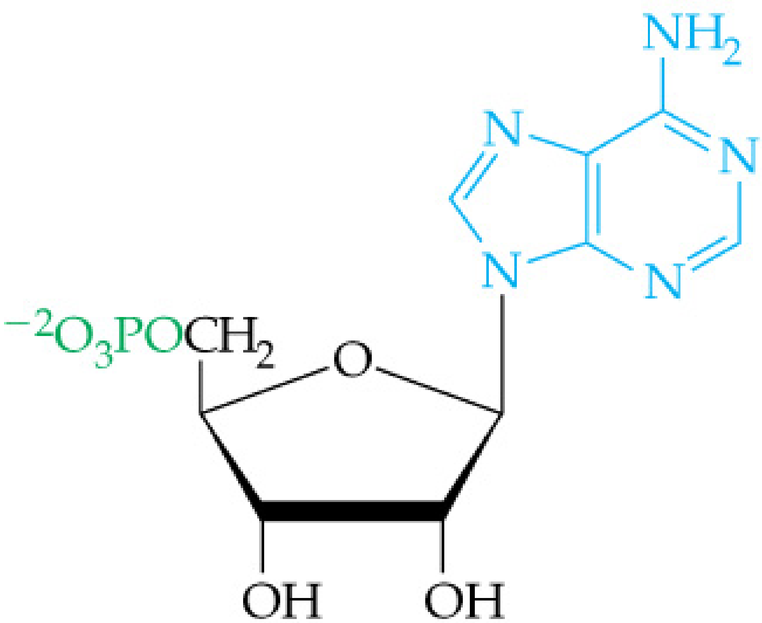
Second letter

First letter	U		C		A		G		Third letter
U	UUU	Phe	UCU		UAU	Tyr	UGU	Cys	U
	UUC		UCC	Ser	UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU		CCU		CAU	His	CGU		U
	CUC	Leu	CCC	Pro	CAC		CGC	Arg	C
	CUA		CCA		CAA	Gin	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU		ACU		AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU		GCU		GAU	Asp	GGU		U
	GUC	Val	GCC	Ala	GAC		GGC	Gly	C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

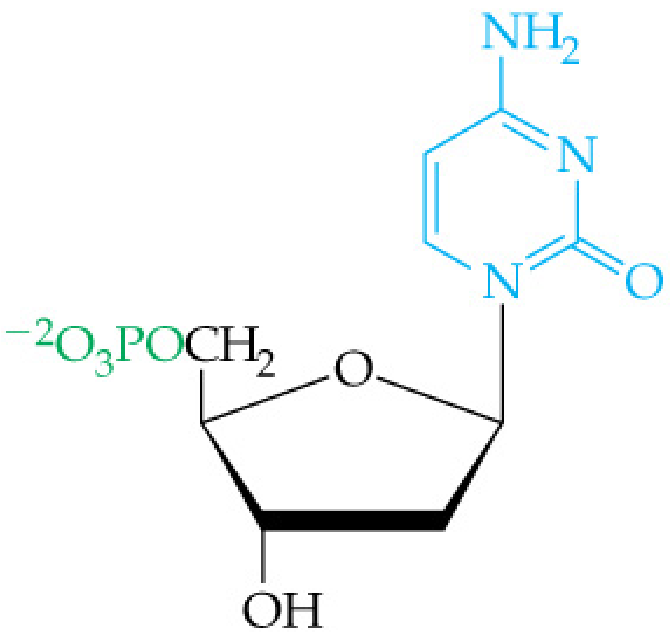
Central Dogma of Molecular Biology



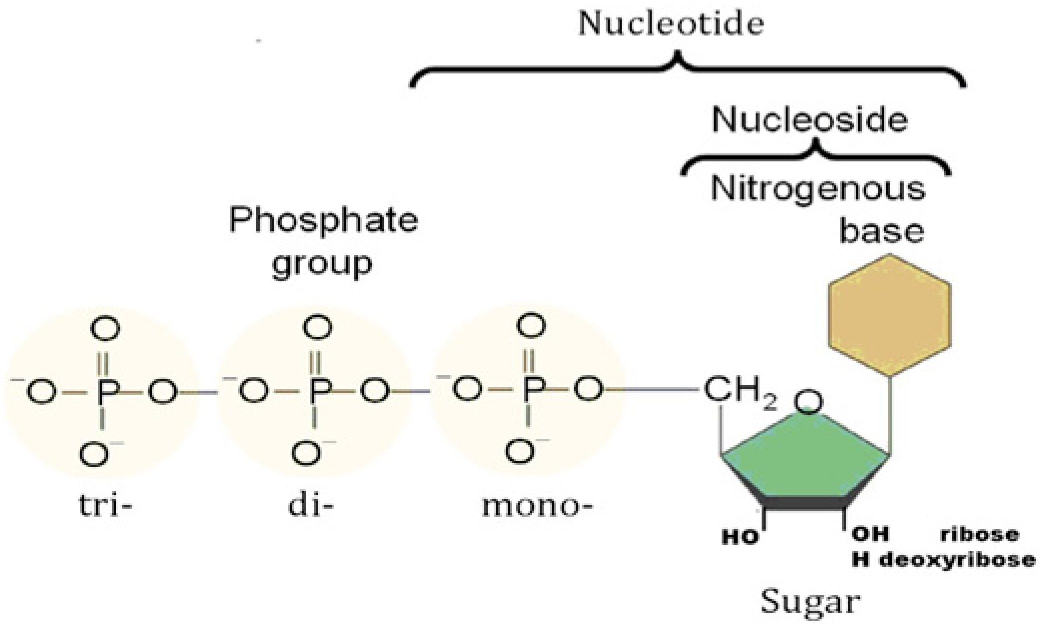
- In RNA, the sugar is ribose.
- In DNA, the sugar is deoxyribose.

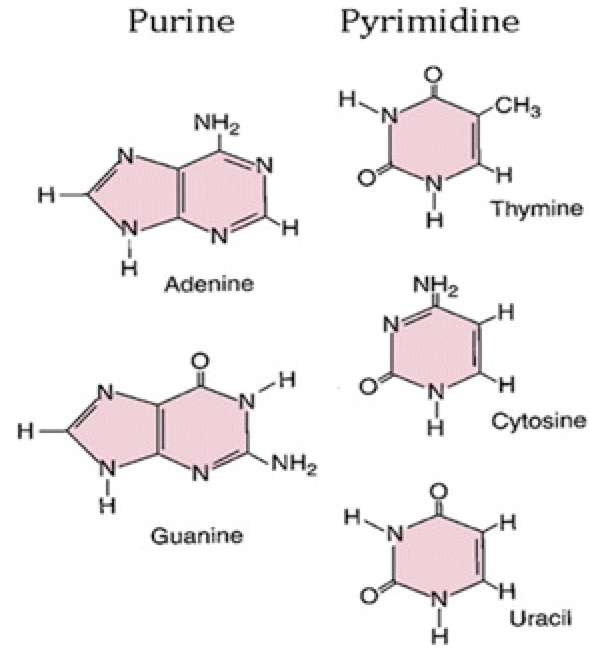


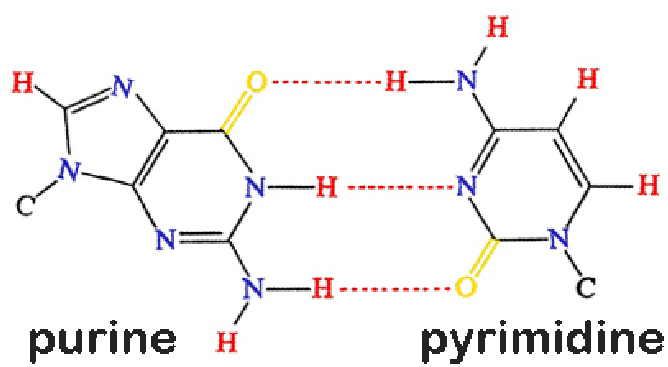
Adenosine 5′-monophosphate (AMP)
(a ribonucleotide)



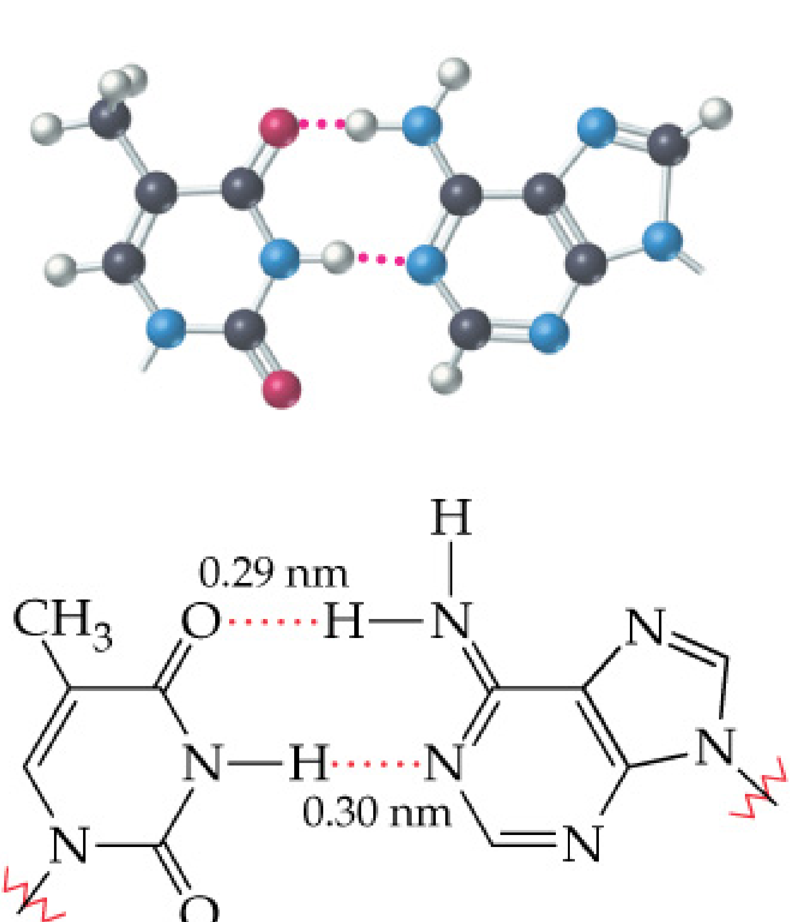
Deoxycytidine 5′-monophosphate (dCMP)
(a deoxyribonucleotide)


Nucleotide
Nucleoside
Nitrogenous base
Phosphate group
tri-
di-
mono-
HO
OH ribose
H deoxyribose
Sugar

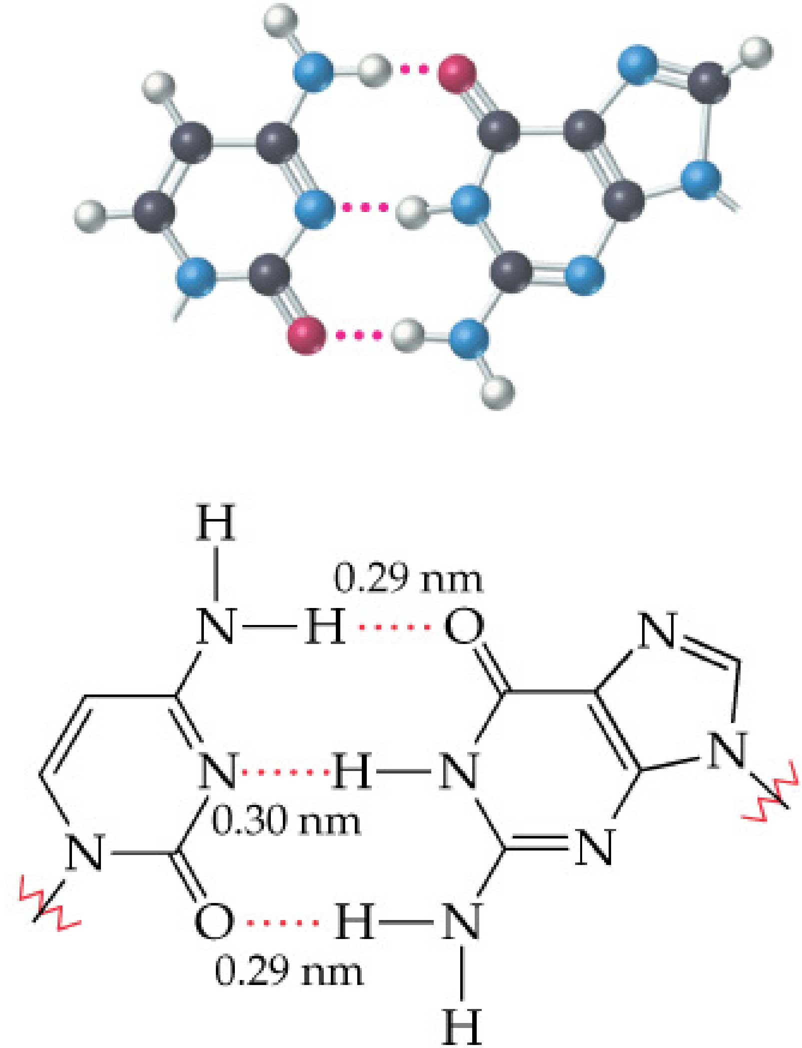

Purine
Pyrimidine
Adenine
Guanine
Thymine
Cytosine
Uracil

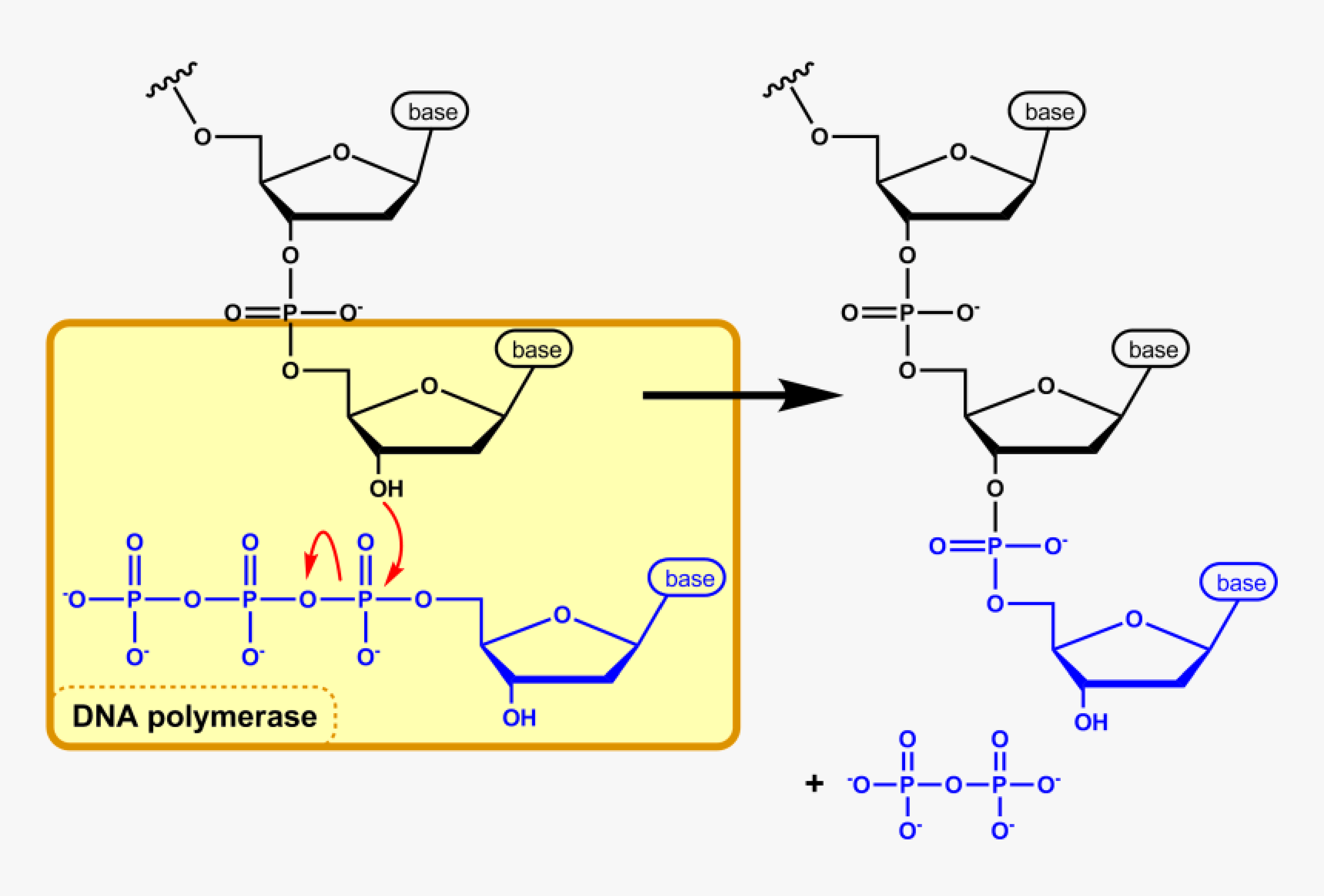

purine
pyrimidine

Thymine–Adenine

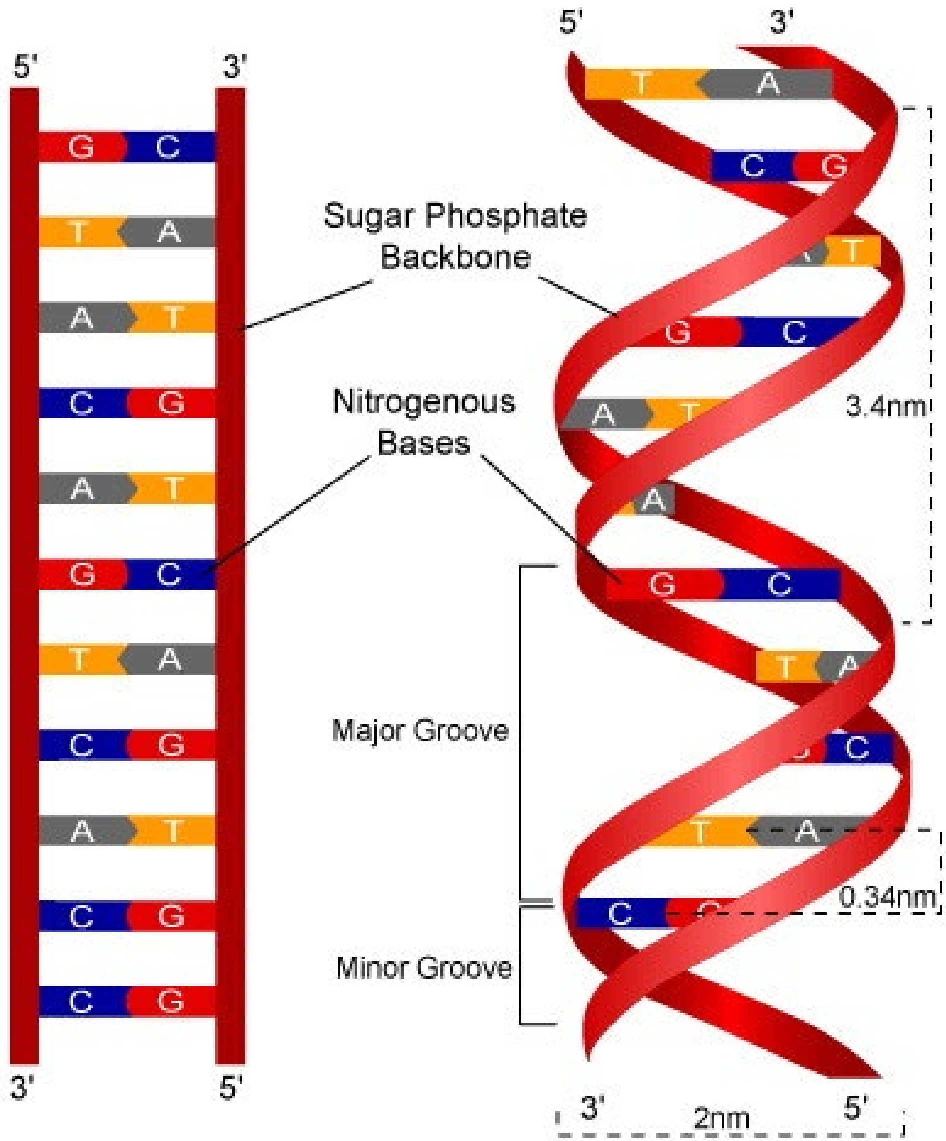


Cytosine–Guanine




base
DNA polymerase
OH

DNA Double Helix Structure



Chemical Bond Energy

Table 7.1 Average Bond Dissociation Energies

Bond	Bond Dissociation Energy kcal/mol (kJ/mol)	Bond	Bond Dissociation Energy kcal/mol (kJ/mol)	Bond	Bond Dissociation Energy kcal/mol (kJ/mol)
C—H	99 (413)	N—H	93 (391)	C=C	147 (614)
C—C	83 (347)	N—N	38 (160)	C≡C	201 (839)
C—N	73 (305)	N—Cl	48 (200)	C=O*	178 (745)
C—O	86 (358)	N—O	48 (201)	O=O	119 (498)
C—Cl	81 (339)	H—H	103 (432)	N=O	145 (607)
Cl—Cl	58 (243)	O—H	112 (467)	O≡N	213 (891)
H—Cl	102 (427)	O—Cl	49 (203)	N≡N	226 (946)

*The C=O bond dissociation energies in CO_2 are 191 kcal/mol (799 kJ/mol).

Hydrogen Bond Energy

Table 3.1 H-bond and its bond strength.

H-bond	Bond Strength (kcal/mol)
F–H·····F	7
O–H·····O	4.5–7.6
O–H·····N	4–7
C–H·····pi electrons	2–4
C–H·····O	2–3
N–H·····O	2–3
N–H·····N	1.3

Strong hydrogen bonds of 20-40 kcal/mole

Weak hydrogen bonds of 1-5 kcal/mole

Normal hydrogen bond 3 - 12 kcal/mole

Disulfide Bond Energy

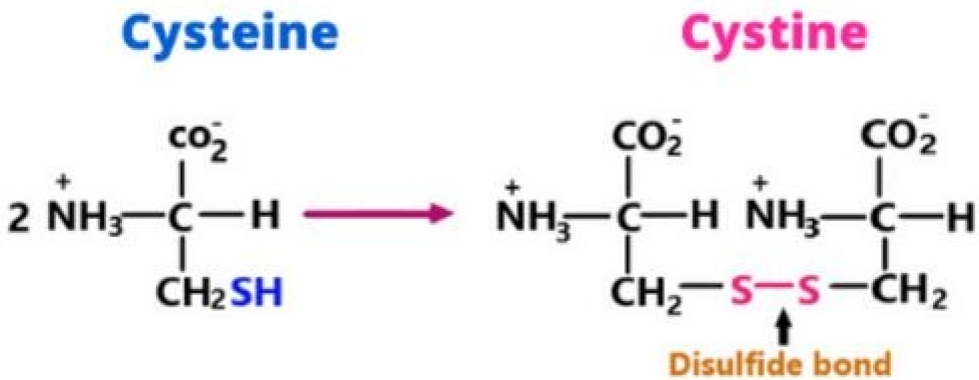
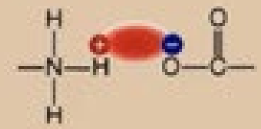
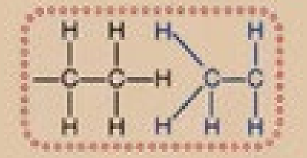
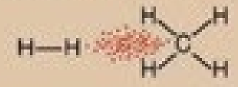


Fig: Disulfide bond in protein

This bond length is 2.2 Å and bond energy is 60 kcal/mol.

Chemical Bonds & Interactions

NAME	BASIS OF INTERACTION	STRUCTURE	BOND ENERGY[a] (KCAL/MOL)
Covalent bond	Sharing of electron pairs		50–110
Ionic bond	Attraction of opposite charges		3–7
Hydrogen bond	Sharing of H atom		3–7
Hydrophobic interaction	Interaction of nonpolar substances in the presence of polar substances (especially water)		1–2
van der Waals interaction	Interaction of electrons of nonpolar substances		1

[a]Bond energy is the amount of energy needed to separate two bonded or interacting atoms under physiological conditions.

Thermal Energy

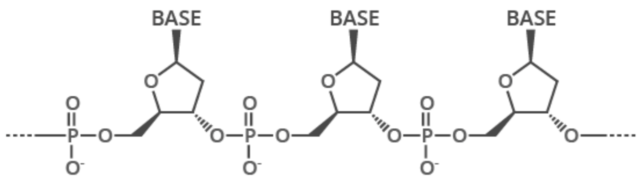
$E=3/2\ RT$

$R= 1.987$ cal/mol/K

$E=3/2\ 1.987 \times 300 \sim 0.9$ Kcal

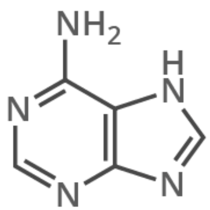
THE CHEMICAL STRUCTURE OF DNA

THE SUGAR PHOSPHATE 'BACKBONE'

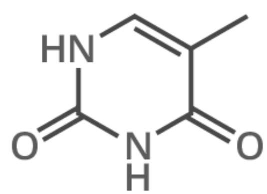


DNA is a polymer made up of units called nucleotides. The nucleotides are made of three different components: a sugar group, a phosphate group, and a base. There are four different bases: adenine, thymine, guanine and cytosine.

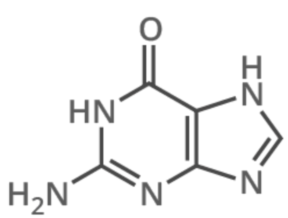
A ADENINE



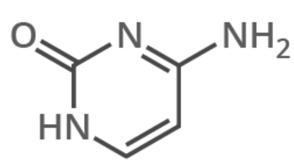
T THYMINE



G GUANINE

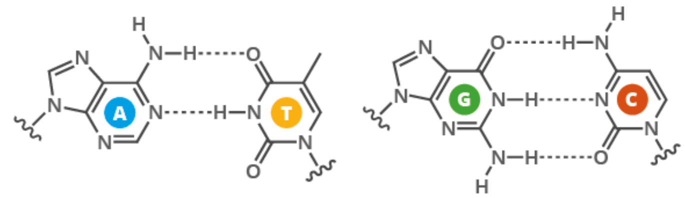


C CYTOSINE



WHAT HOLDS DNA STRANDS TOGETHER?

DNA strands are held together by hydrogen bonds between bases on adjacent strands. Adenine (A) always pairs with thymine (T), while guanine (G) always pairs with cytosine (C). Adenine pairs with uracil (U) in RNA.



FROM DNA TO PROTEINS

The bases on a single strand of DNA act as a code. The letters form three letter codons, which code for amino acids - the building blocks of proteins.

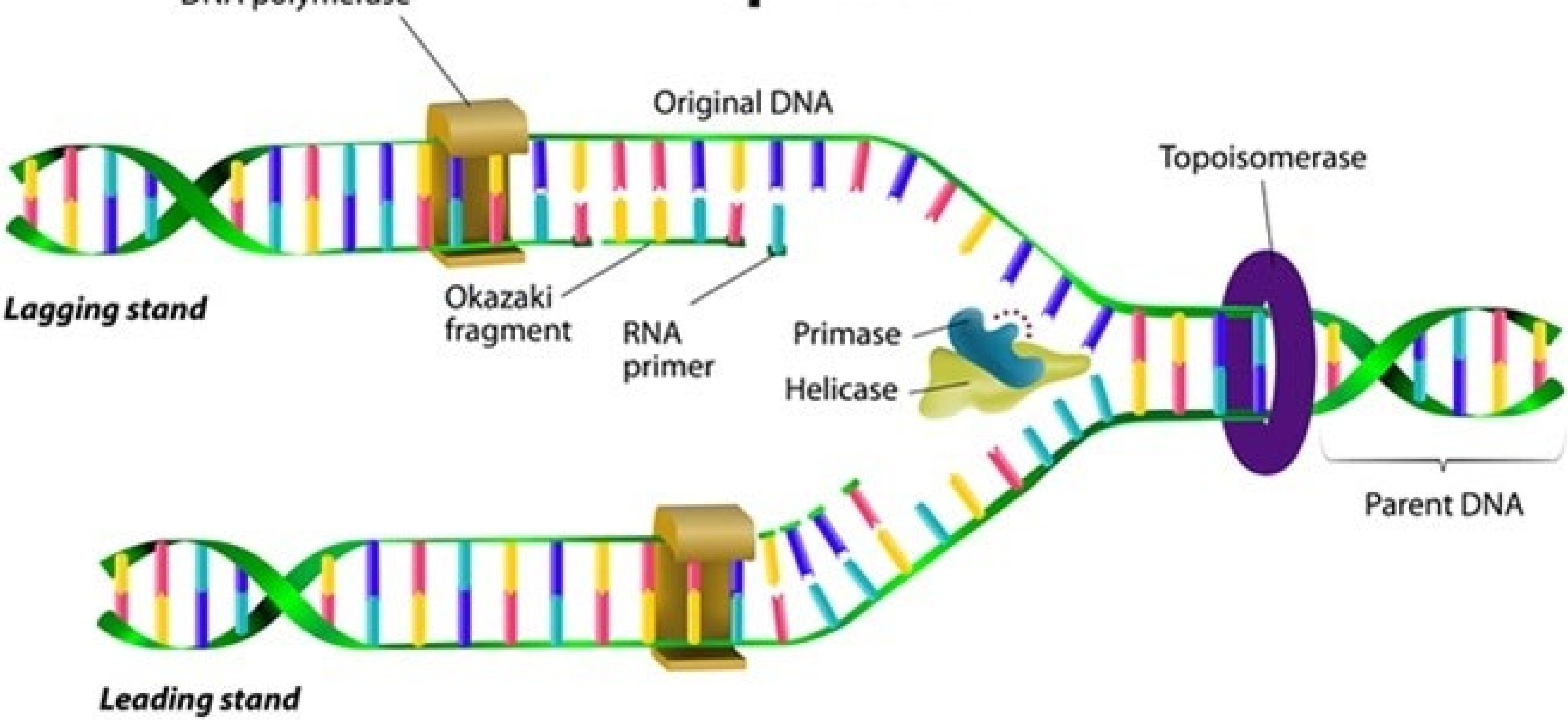
DNA $\xrightarrow{\text{TRANSCRIPTION}}$ RNA $\xrightarrow{\text{TRANSLATION}}$ PROTEIN

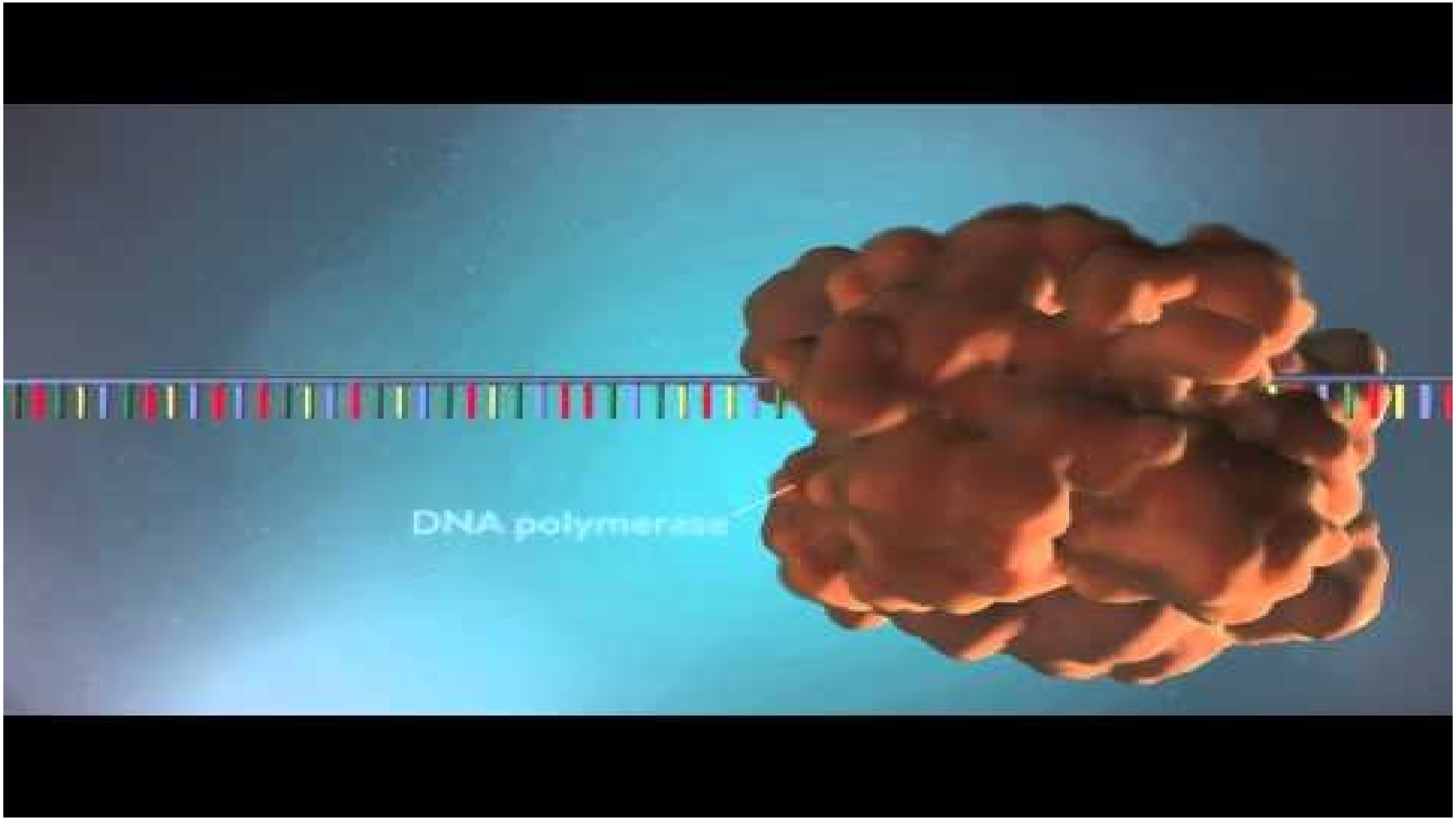
An enzyme, RNA polymerase, transcribes DNA into mRNA (messenger ribonucleic acid). It splits apart the two strands that form the double helix, then reads a strand and copies the sequence of nucleotides. The only difference between the RNA and the original DNA is that in the place of thymine (T), another base with a similar structure is used: uracil (U).

DNA SEQUENCE	T	T	C	C	T	G	A	A	C	C	C	G	T	T	A
mRNA SEQUENCE	U	U	C	C	U	G	A	A	C	C	C	G	U	U	A
AMINO ACID	Phenylalanine			Leucine			Asparagine			Proline			Leucine		

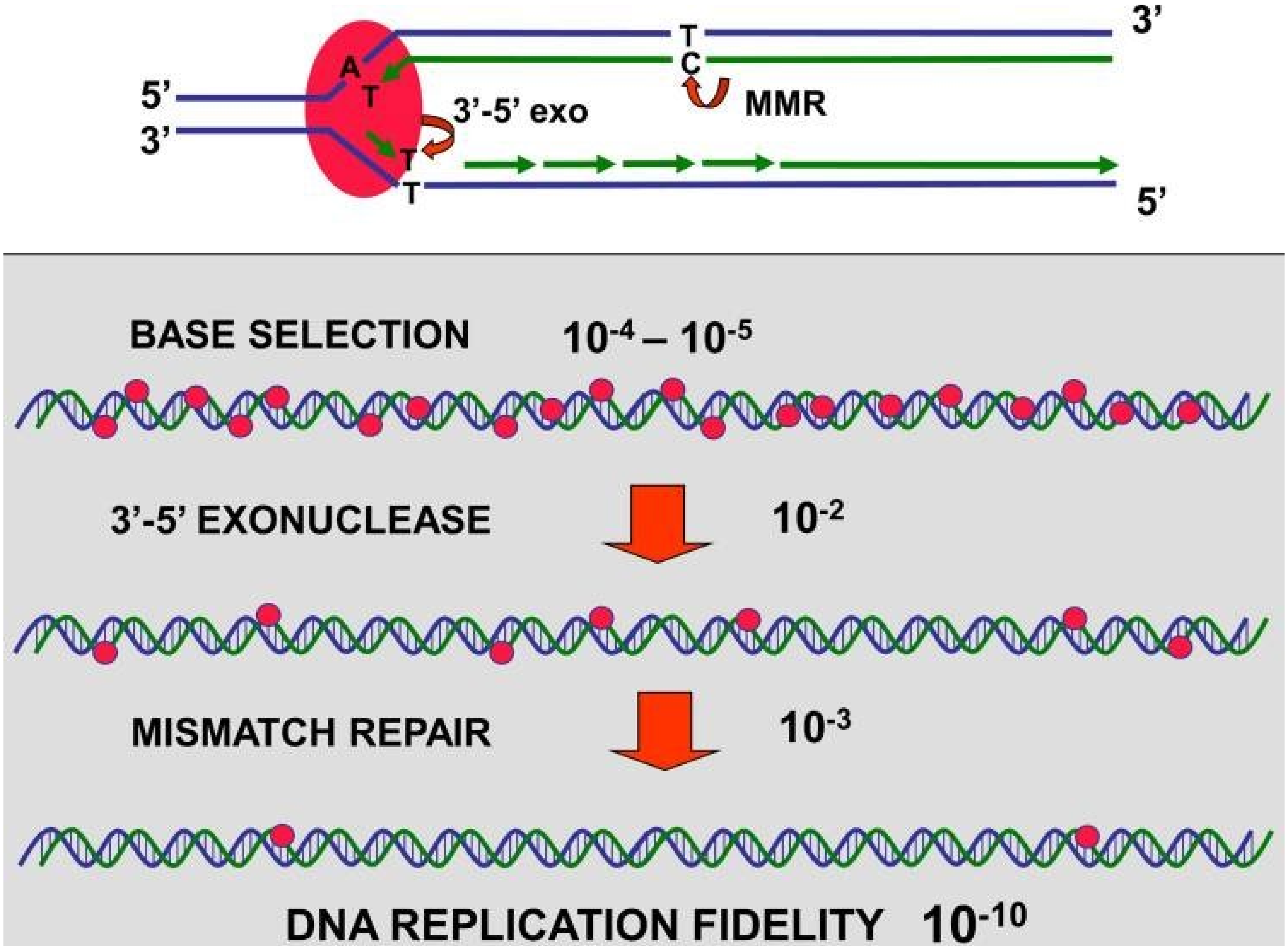
In multicellular organisms, the mRNA carries genetic code out of the cell nucleus, to the cytoplasm. Here, protein synthesis takes place. 'Translation' is the process of turning the mRNA's 'code' into proteins. Molecules called ribosomes carry out this process, building up proteins from the amino acids coded for.

DNA replication



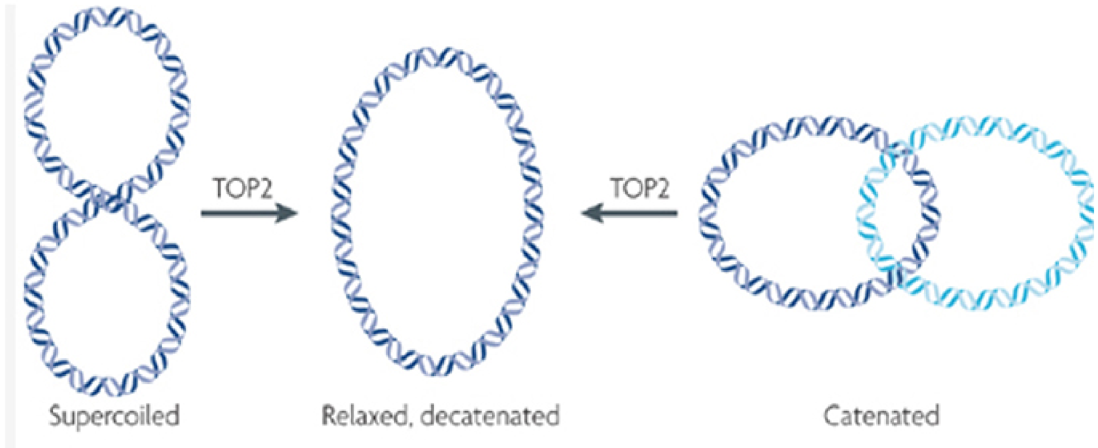


https://youtu.be/TNKWgcFPHqw

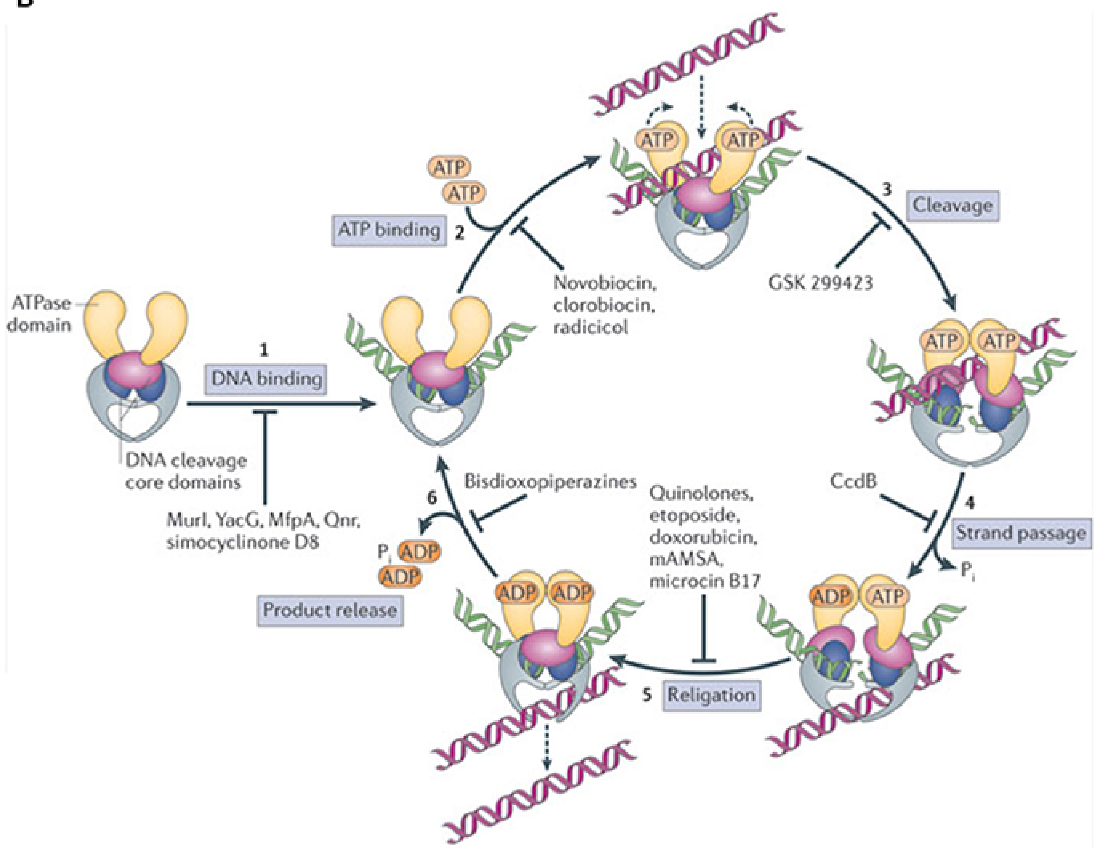


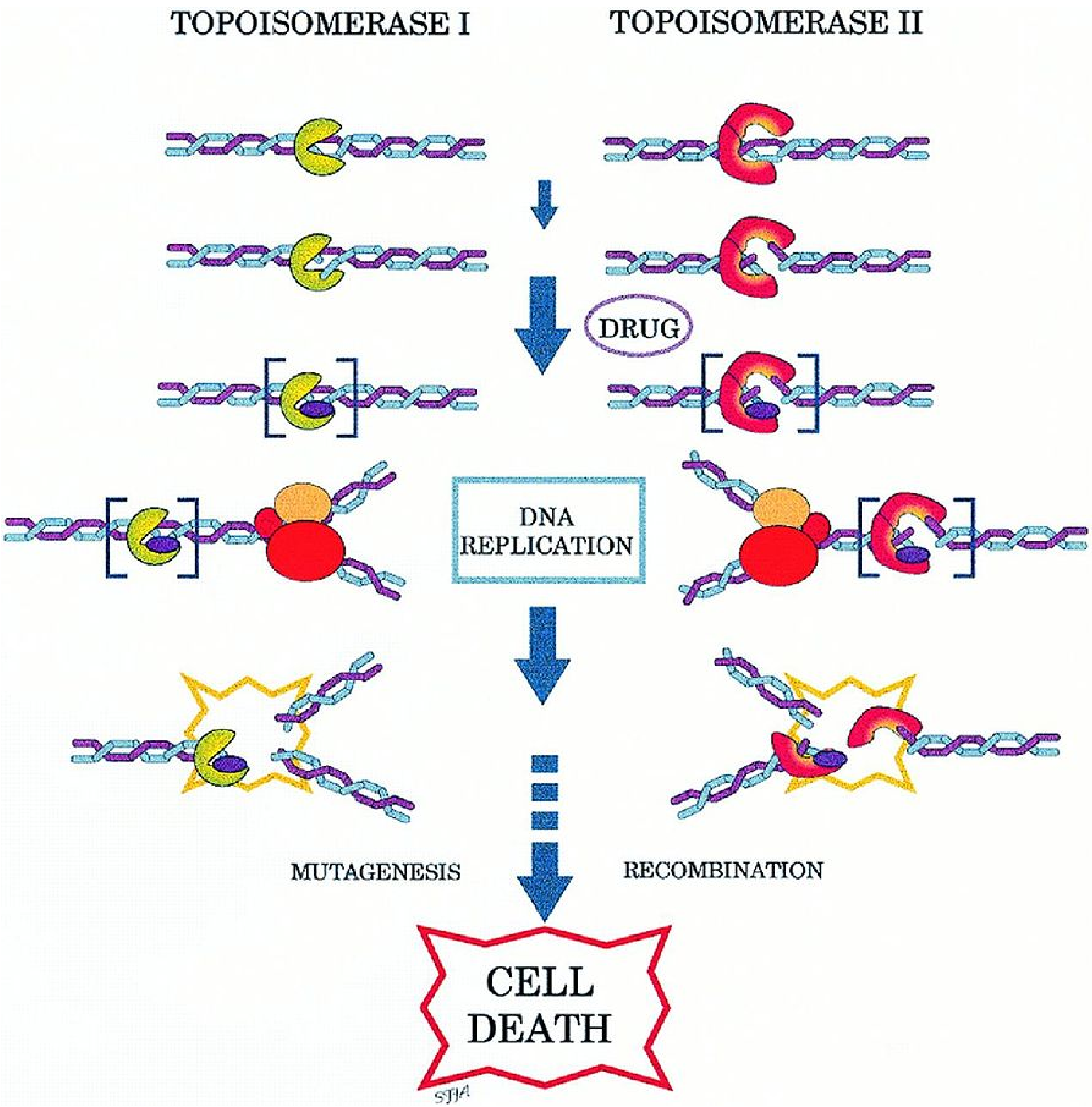
3'
T
C
5'
A
T
MMR
3'-5' exo
3'
T
T
5'
BASE SELECTION $10^{-4} – 10^{-5}$
3'-5' EXONUCLEASE 10^{-2}
MISMATCH REPAIR 10^{-3}
DNA REPLICATION FIDELITY 10^{-10}

A

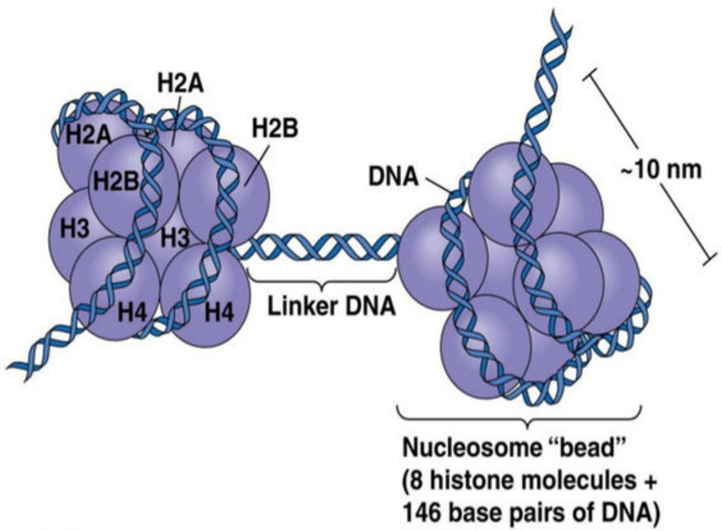

TOP2
TOP2
Supercoiled
Relaxed, decatenated
Catenated

B

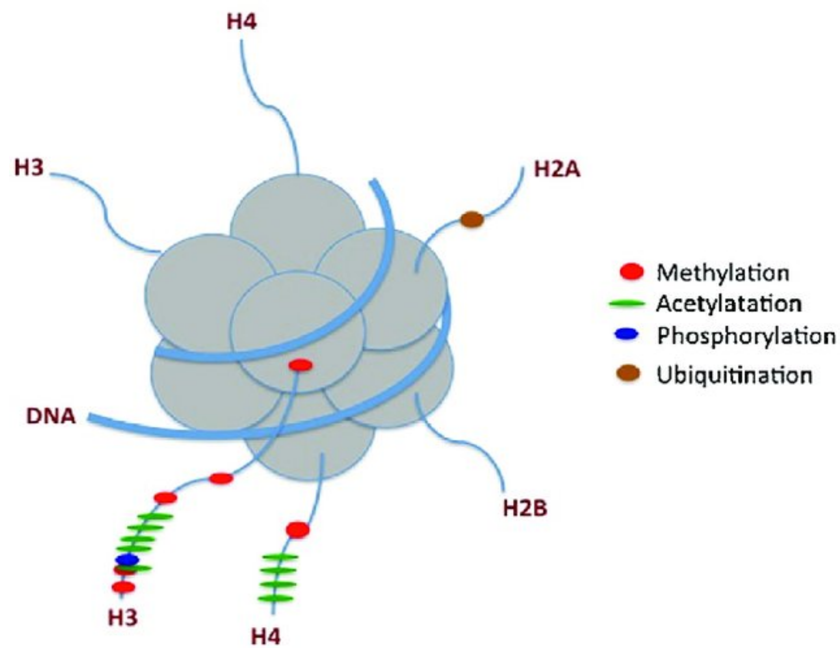

ATPase domain
DNA cleavage core domains
1
DNA binding
MurI, YacG, MfpA, Qnr, simocyclinone D8
ATP
ATP
ATP binding
2
Novobiocin, clorobiocin, radicicol
ATP
ATP
3
Cleavage
GSK 299423
ATP
ATP
CcdB
4
Strand passage
P_i
ADP
ATP
Quinolones, etoposide, doxorubicin, mAMSA, microcin B17
5
Religation
ADP
ADP
Bisdioxopiperazines
6
P_i ADP
ADP
Product release

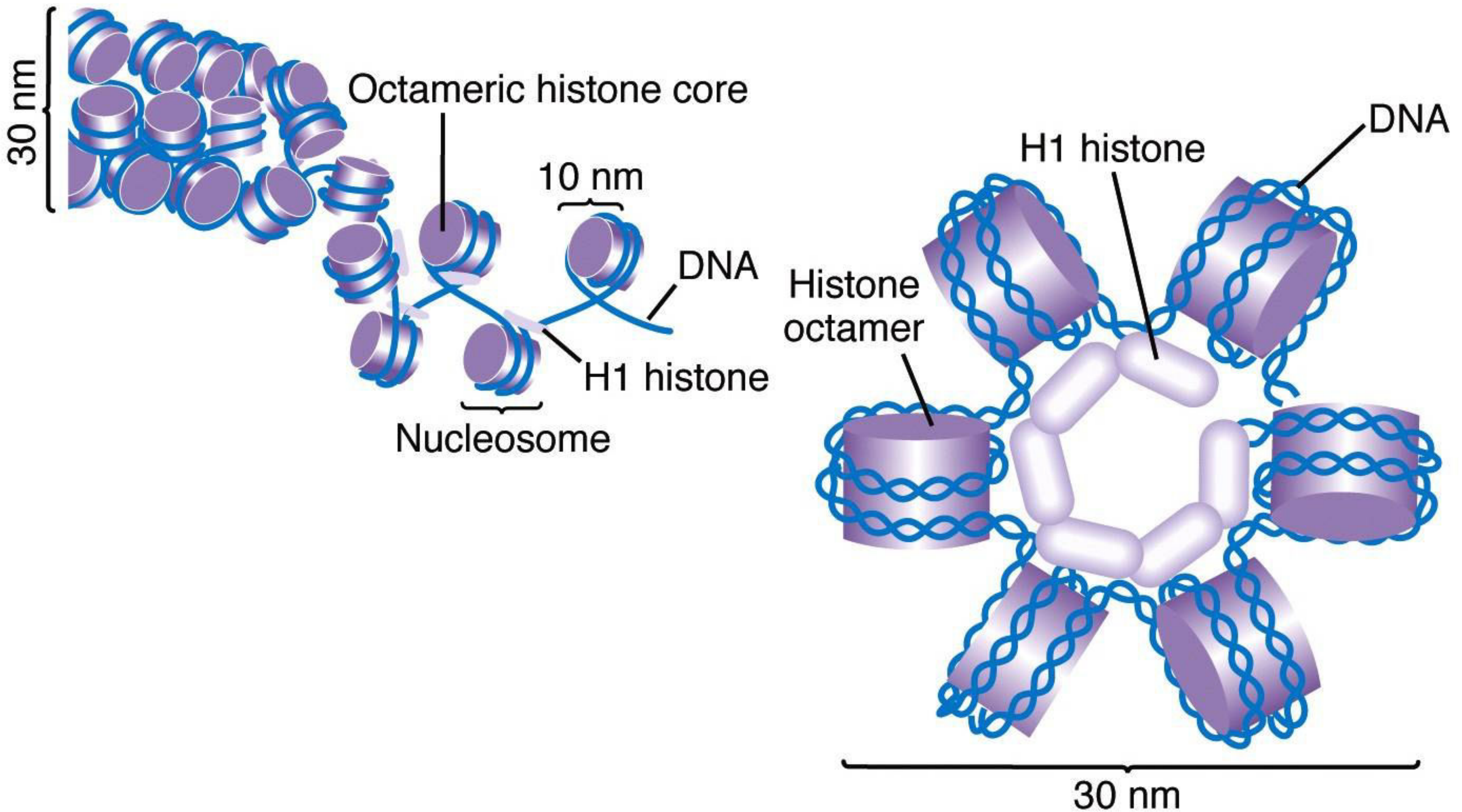


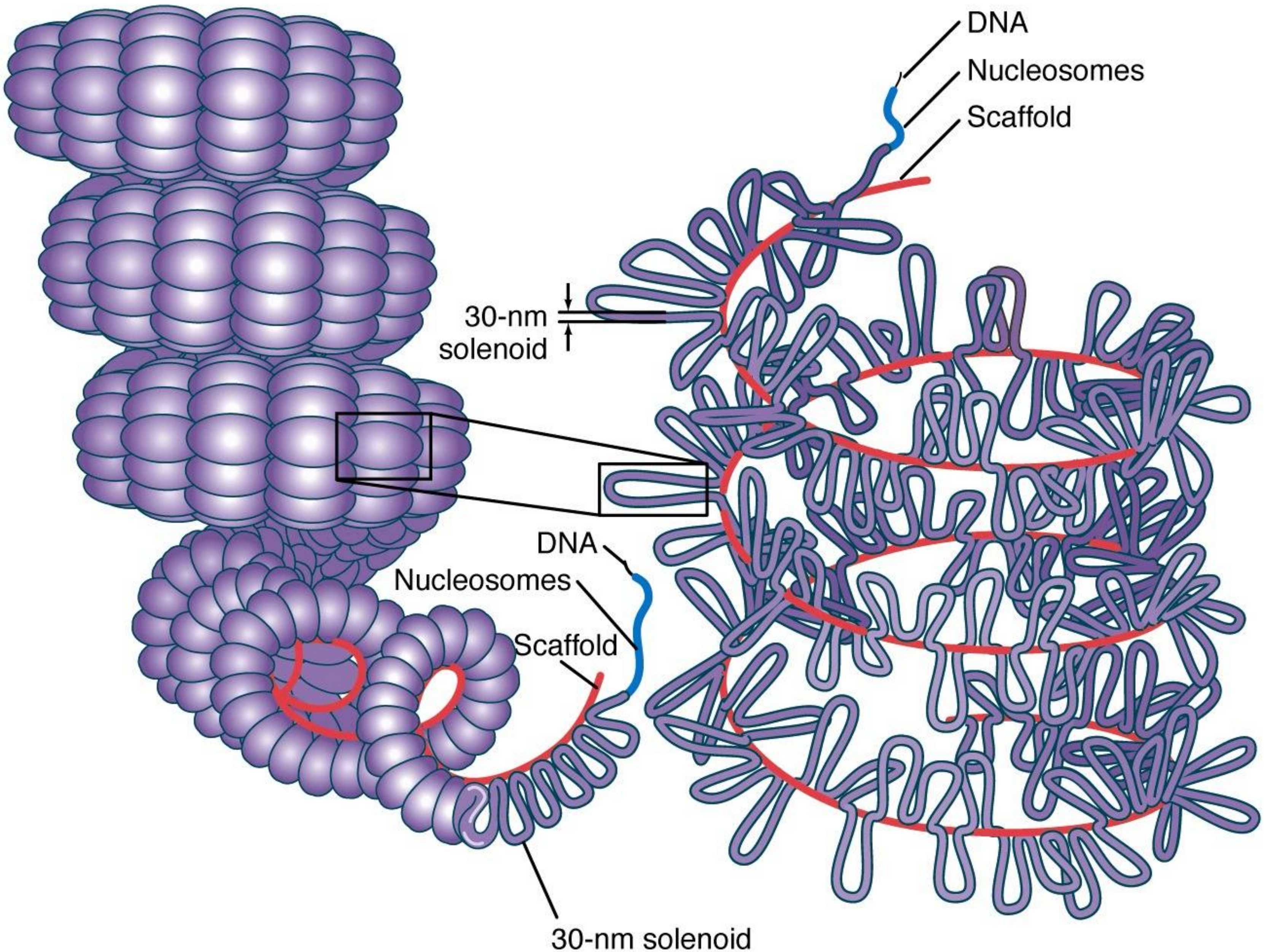
TOPOISOMERASE I
TOPOISOMERASE II
DRUG
DNA REPLICATION
MUTAGENESIS
RECOMBINATION
CELL DEATH
STJA

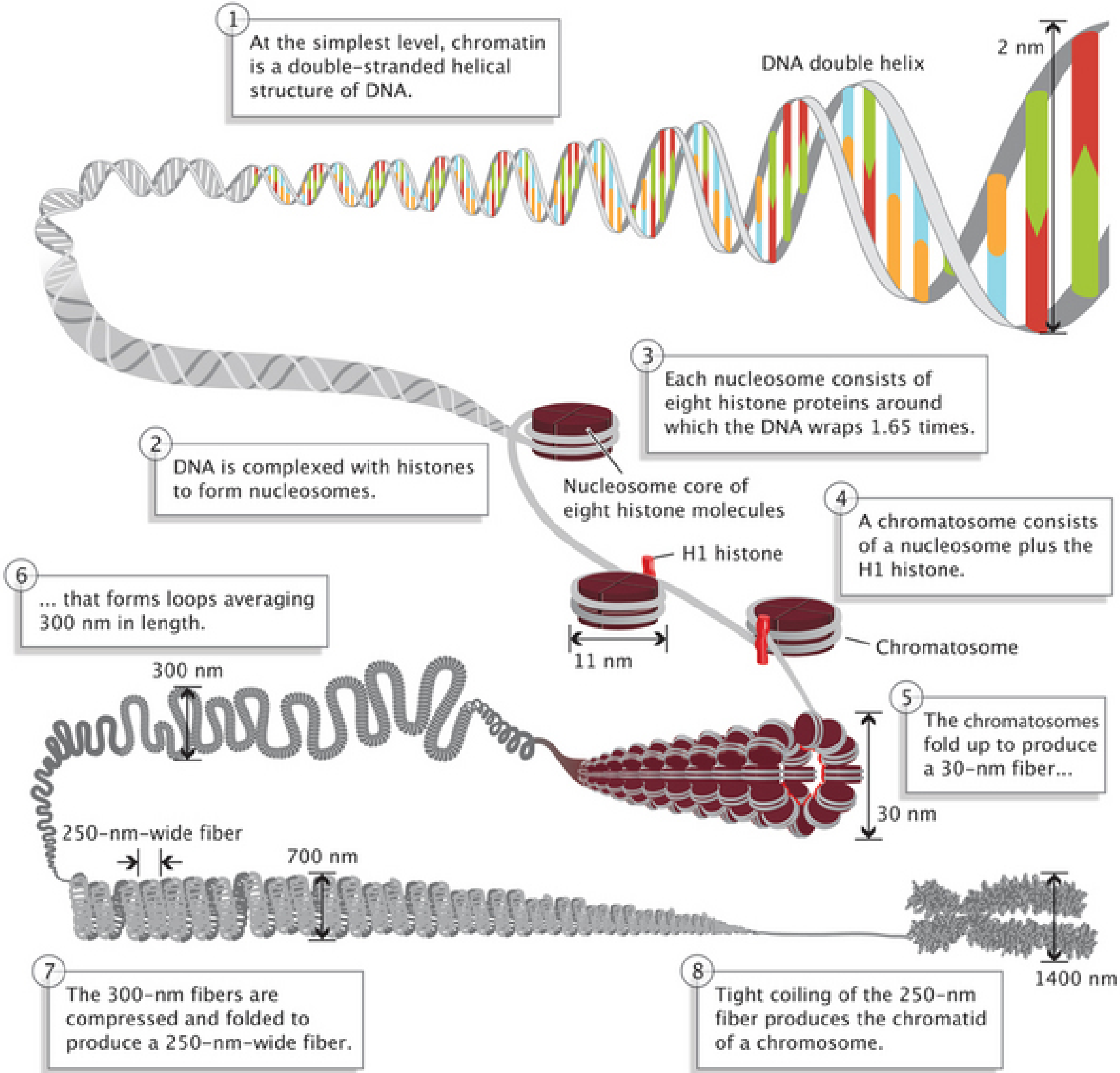


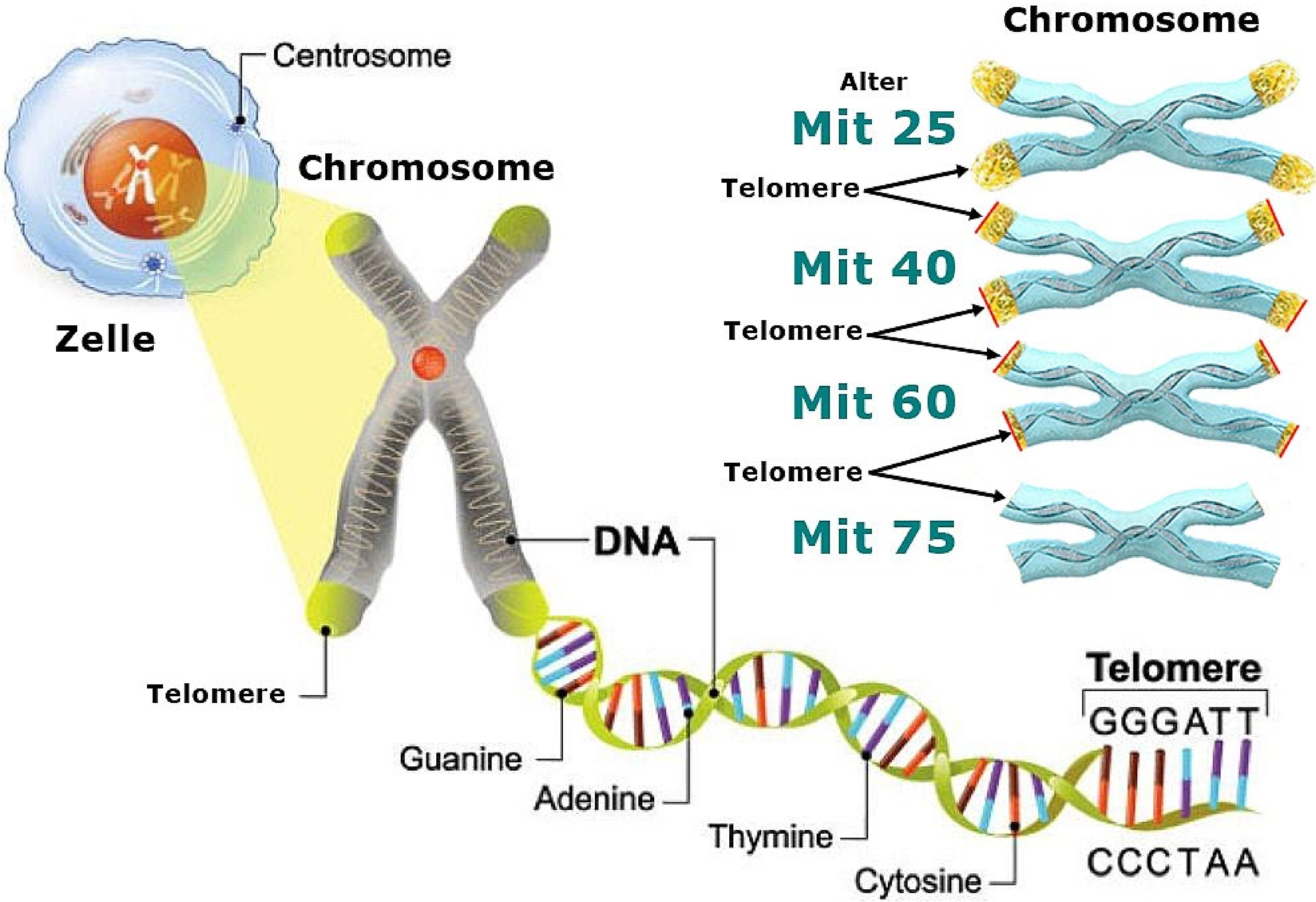
https://www.slideshare.net/jannatiftikhar/role-of-histone-in-dna-packaging



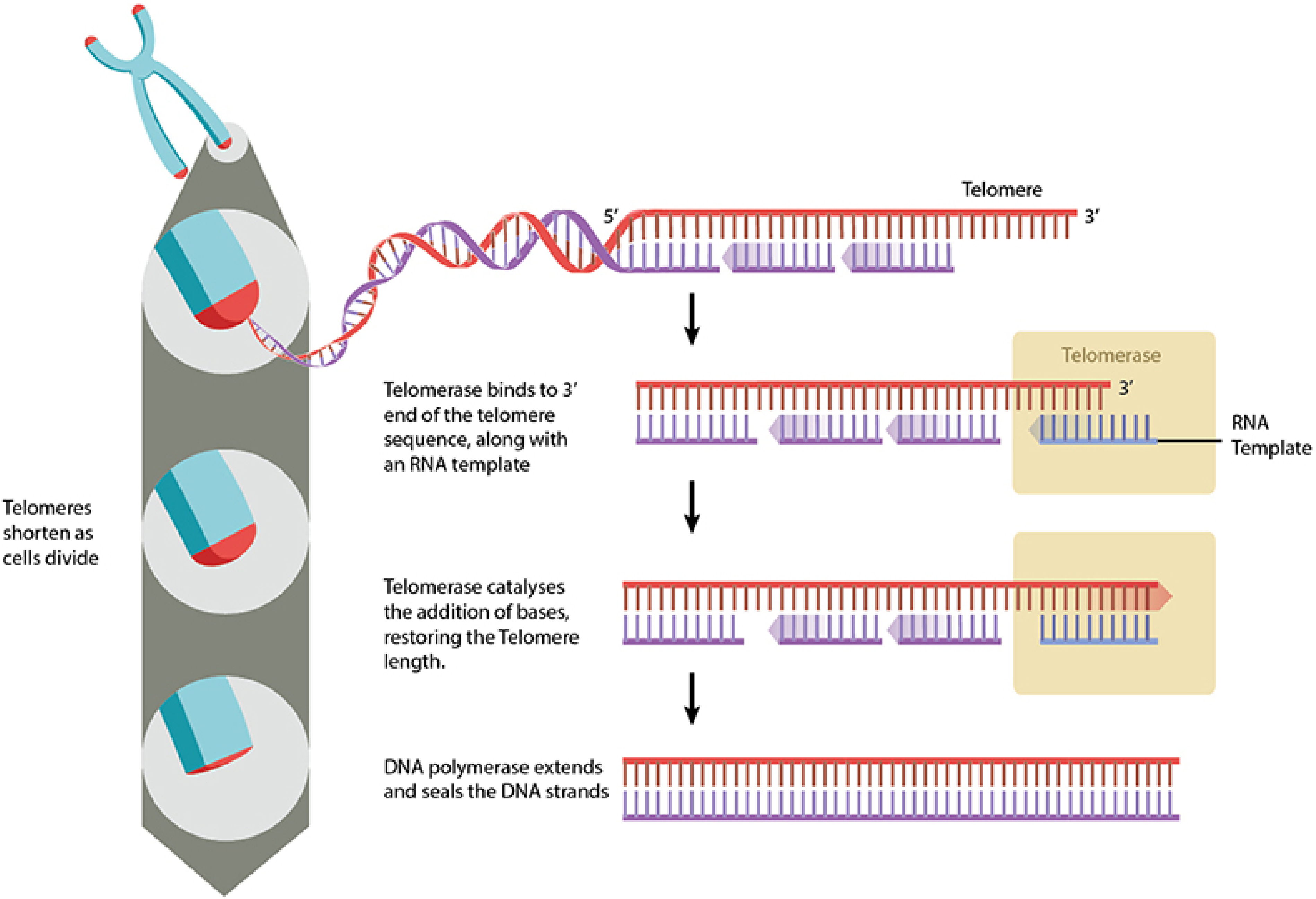

30 nm
Octameric histone core
10 nm
DNA
H1 histone
Nucleosome
H1 histone
DNA
Histone
octamer
30 nm


DNA
Nucleosomes
Scaffold
30-nm
solenoid
DNA
Nucleosomes
Scaffold
30-nm solenoid

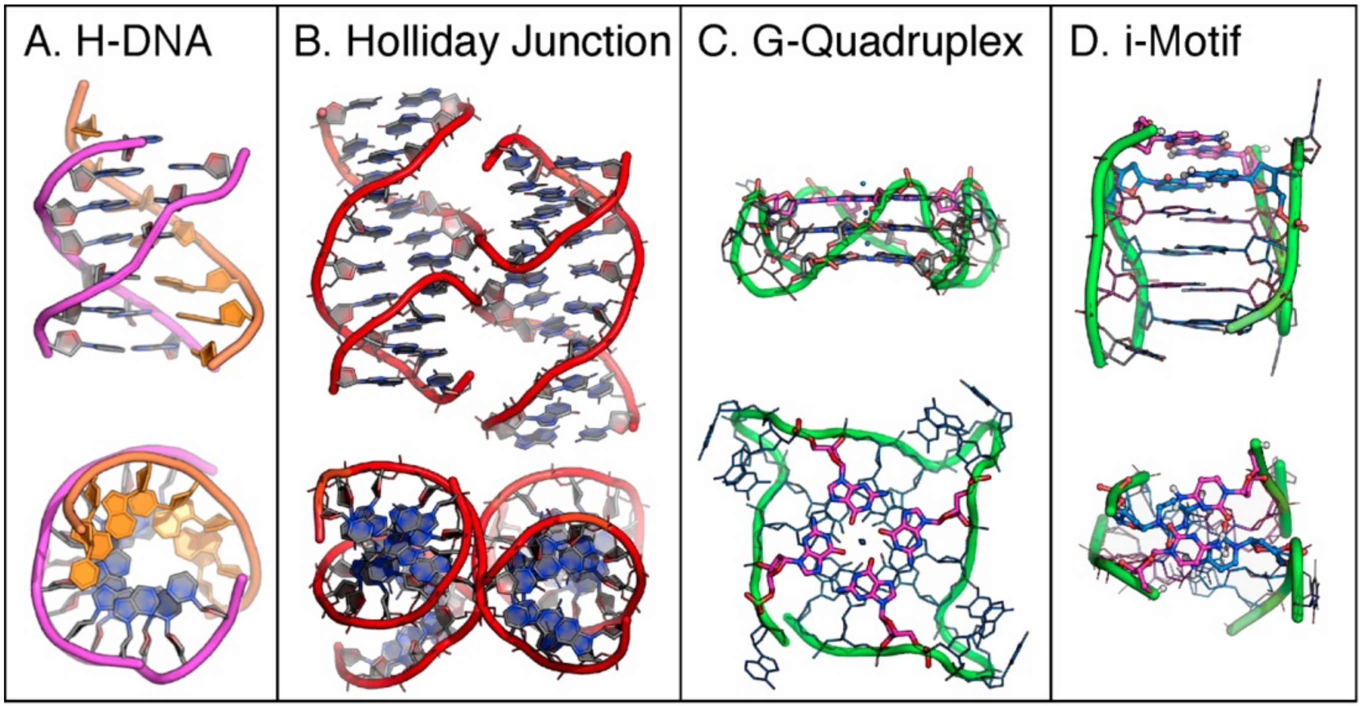

1 At the simplest level, chromatin is a double-stranded helical structure of DNA.
DNA double helix
2 nm
2 DNA is complexed with histones to form nucleosomes.
3 Each nucleosome consists of eight histone proteins around which the DNA wraps 1.65 times.
Nucleosome core of eight histone molecules
H1 histone
4 A chromatosome consists of a nucleosome plus the H1 histone.
11 nm
Chromatosome
6 ... that forms loops averaging 300 nm in length.
300 nm
5 The chromatosomes fold up to produce a 30-nm fiber...
30 nm
250-nm-wide fiber
700 nm
1400 nm
7 The 300-nm fibers are compressed and folded to produce a 250-nm-wide fiber.
8 Tight coiling of the 250-nm fiber produces the chromatid of a chromosome.



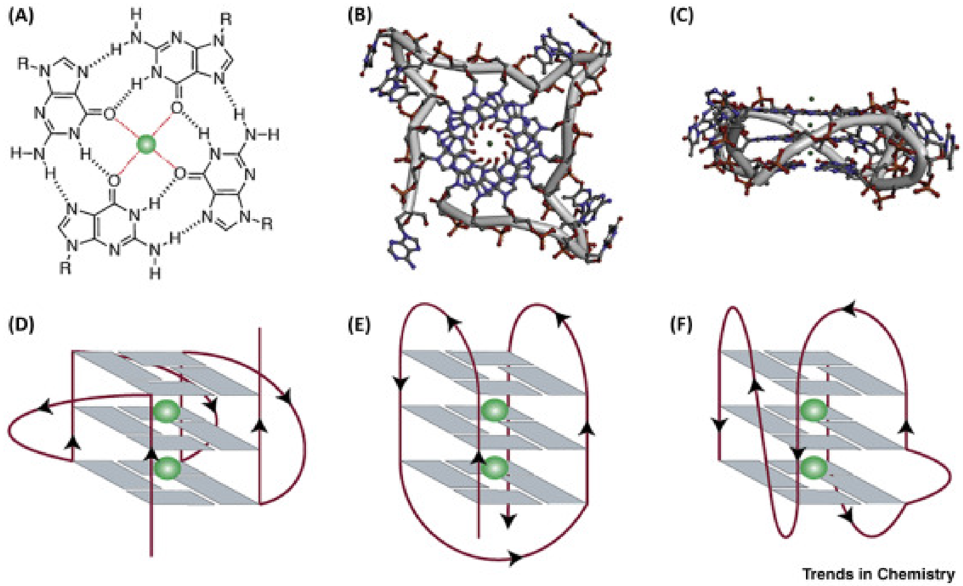
Centrosome
Chromosome
Zelle
Telomere
DNA
Guanine
Adenine
Thymine
Cytosine
Chromosome
Alter
Mit 25
Telomere
Mit 40
Telomere
Mit 60
Telomere
Mit 75
Telomere
GGGATT
CCCTAA

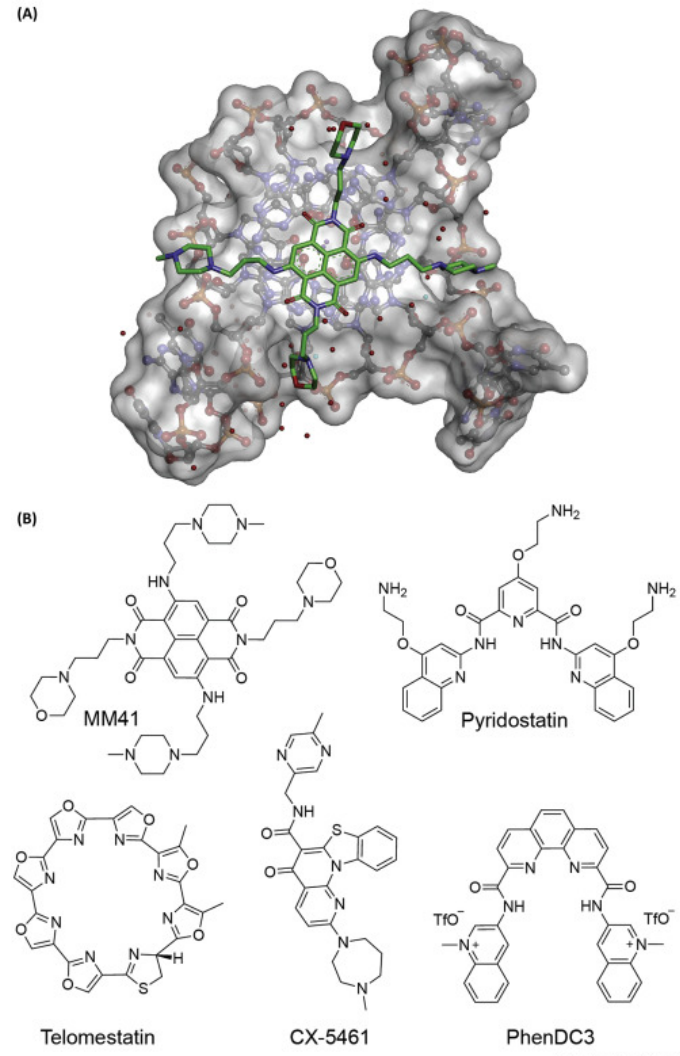

Telomere
5′
3′
Telomerase
Telomerase binds to 3′ end of the telomere sequence, along with an RNA template
3′
RNA Template
Telomeres shorten as cells divide
Telomerase catalyses the addition of bases, restoring the Telomere length.
DNA polymerase extends and seals the DNA strands

Triple and Quadruple Strained DNA

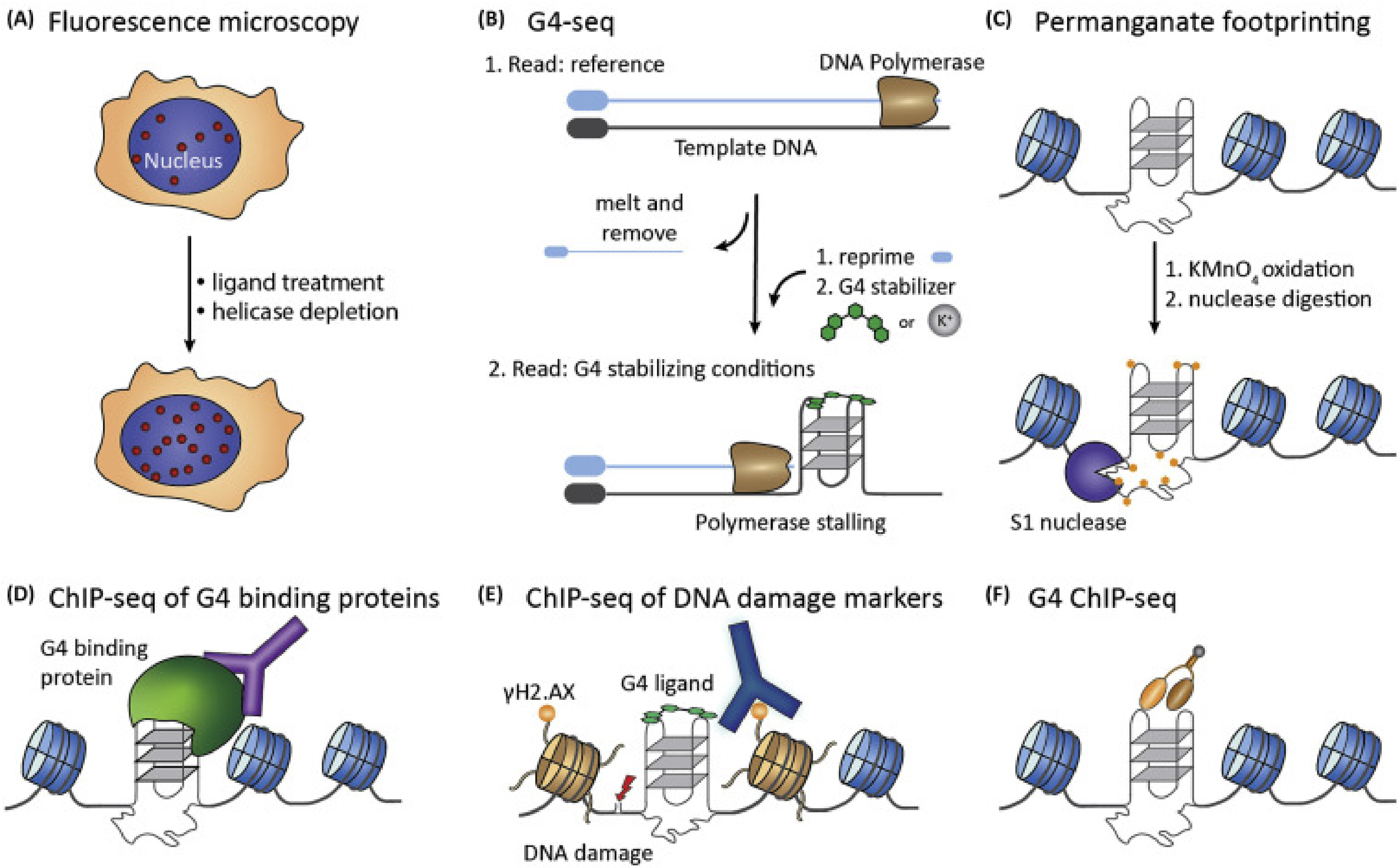


DNA G-quadruplex (G4)

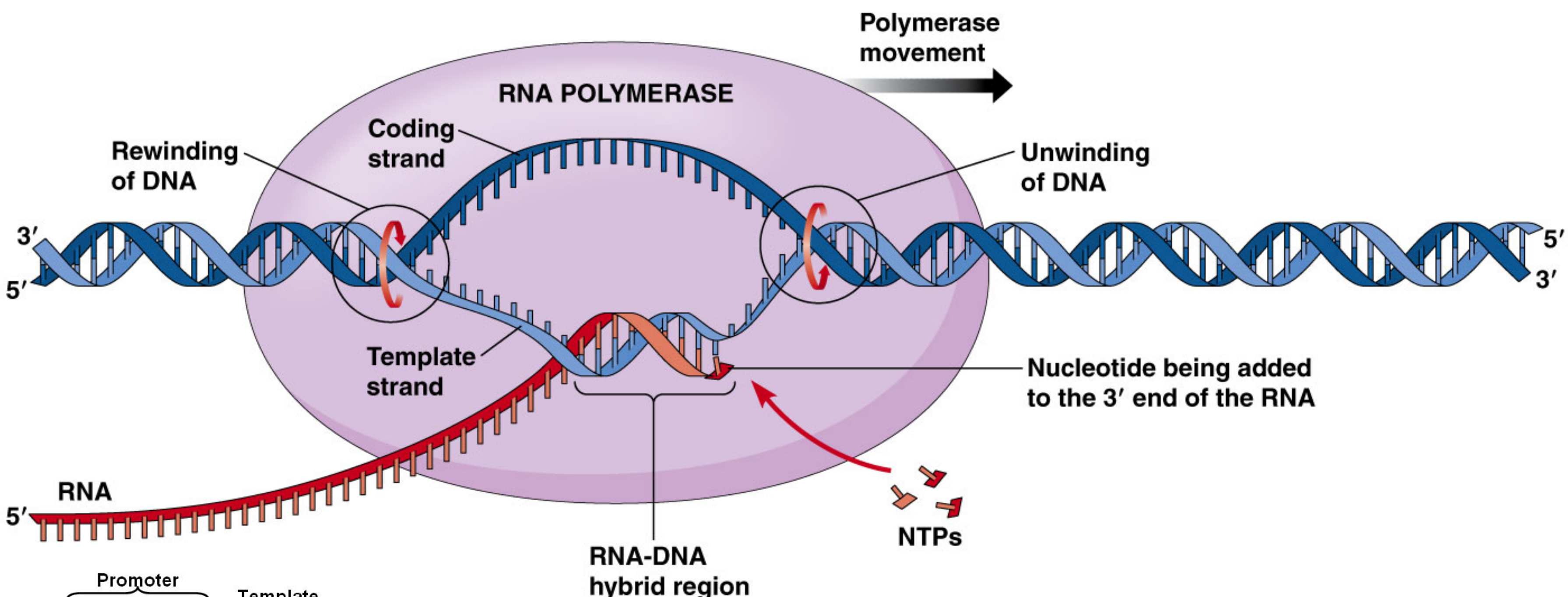


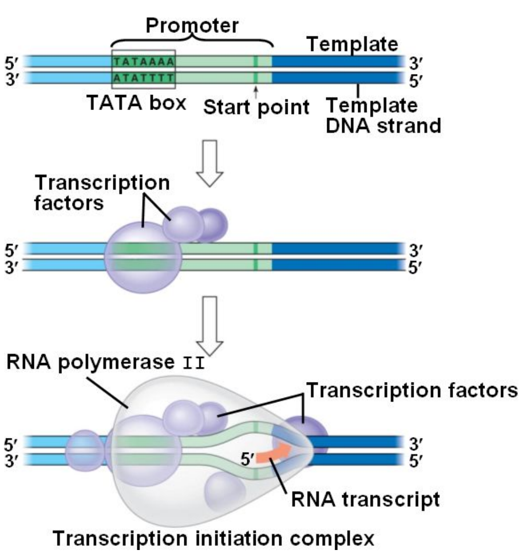


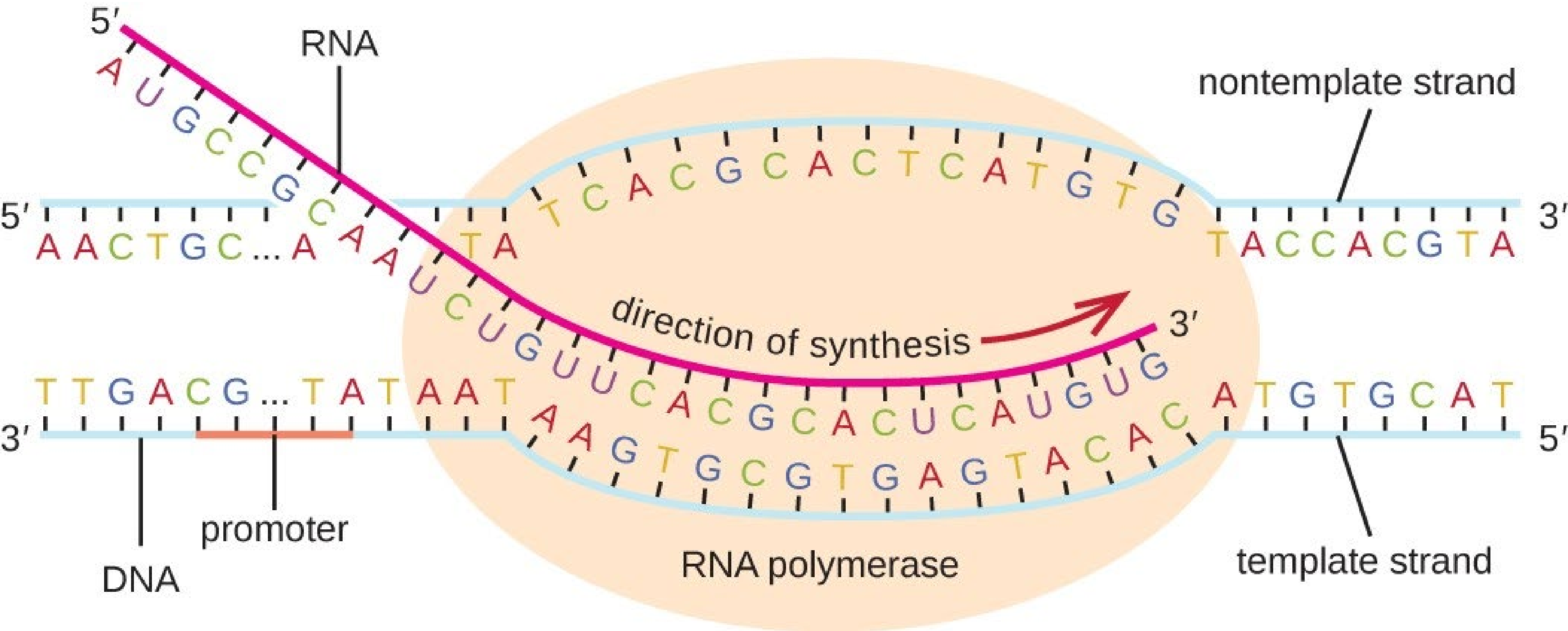
Trends in Chemistry, February 2020, Vol. 2, No. 2 P123

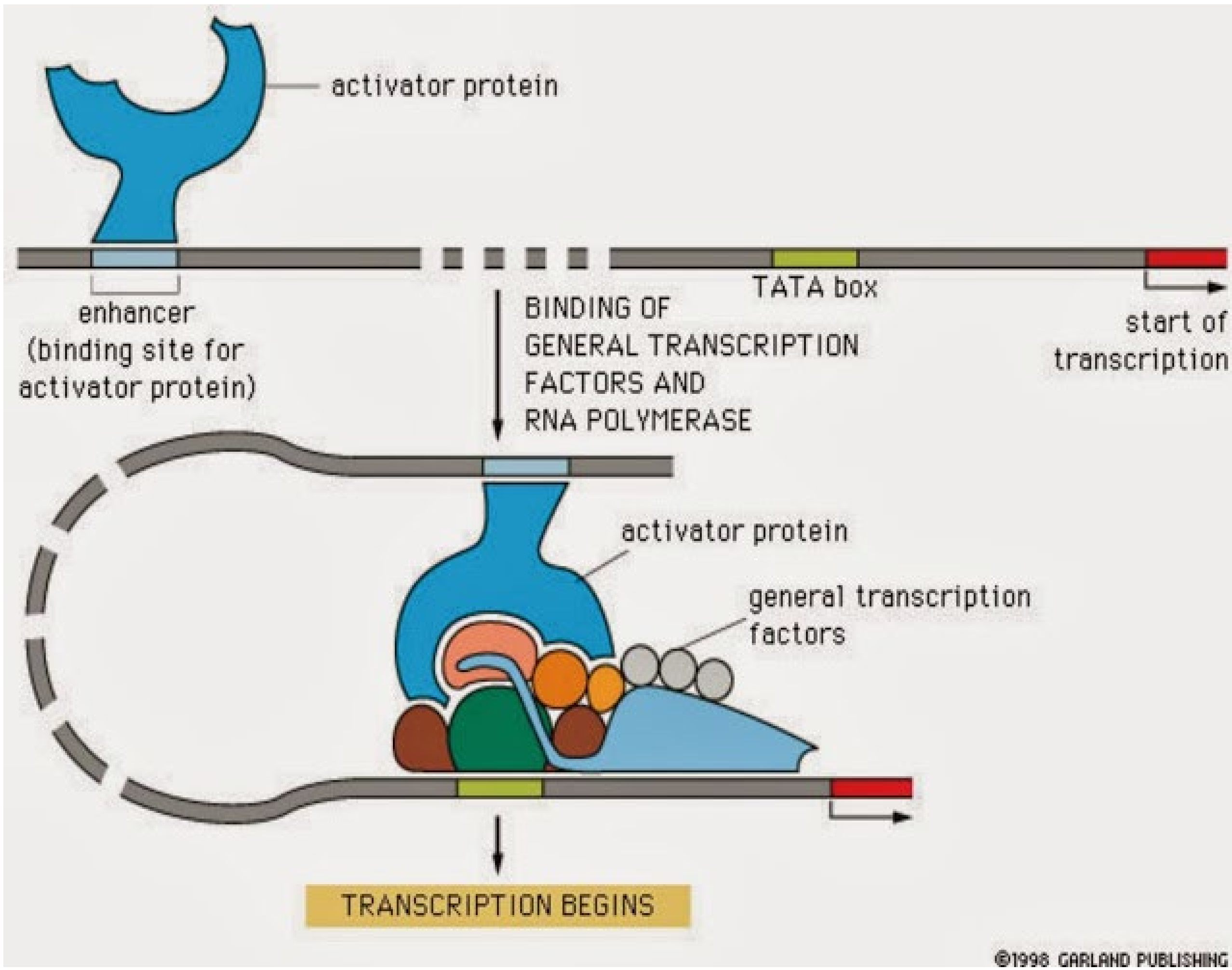


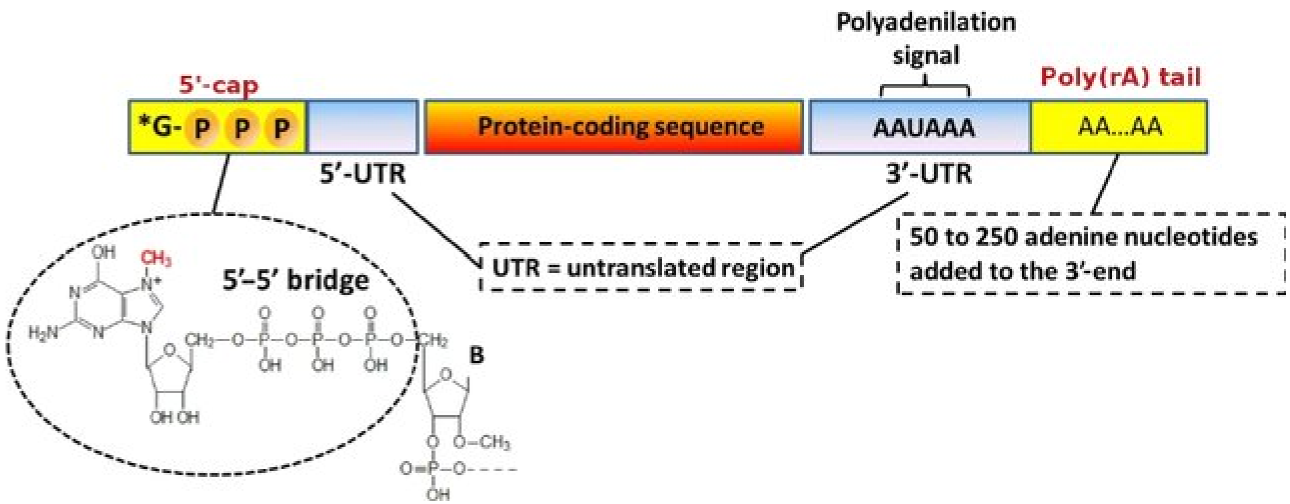
Trends in Chemistry


Polymerase movement
RNA POLYMERASE
Coding strand
Rewinding of DNA
Unwinding of DNA
3′
5′
5′
3′
Template strand
Nucleotide being added to the 3′ end of the RNA
RNA
5′
NTPs
RNA-DNA hybrid region


Promoter
Template
5′
3′
TATAAAA
ATATTTT
3′
5′
TATA box
Start point
Template DNA strand
Transcription factors
5′
3′
3′
5′
RNA polymerase II
Transcription factors
5′
3′
5′
3′
5′
RNA transcript
Transcription initiation complex

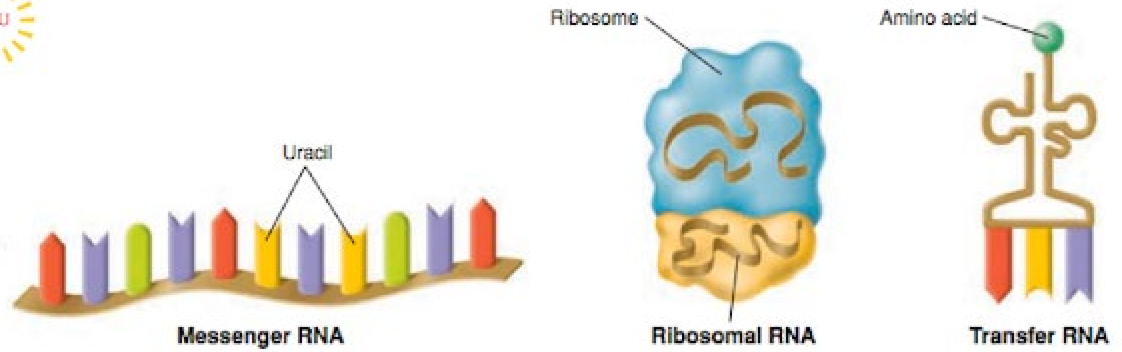

5′
RNA
nontemplate strand
5′
3′
direction of synthesis
3′
3′
5′
DNA
promoter
RNA polymerase
template strand


activator protein
TATA box
enhancer
(binding site for
activator protein)
BINDING OF
GENERAL TRANSCRIPTION
FACTORS AND
RNA POLYMERASE
start of
transcription
activator protein
general transcription
factors
TRANSCRIPTION BEGINS
©1998 GARLAND PUBLISHING



Polyadenilation signal
5'-cap
Poly(rA) tail
*G- P P P
Protein-coding sequence
AAUAAA
AA...AA
5'-UTR
3'-UTR
UTR = untranslated region
50 to 250 adenine nucleotides added to the 3'-end
5'–5' bridge
B

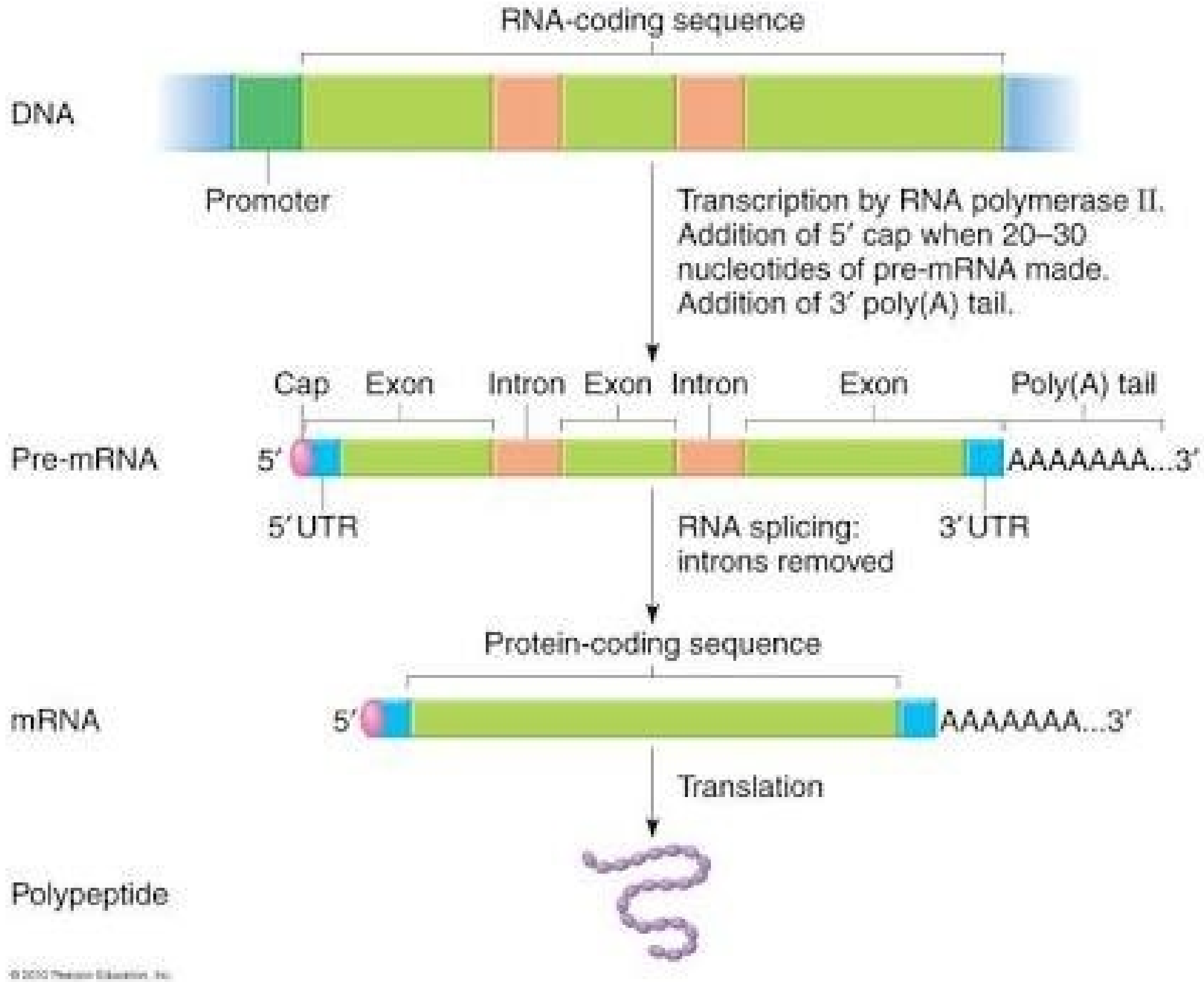
11 DIFFERENT TYPES OF RNA IN A CELL

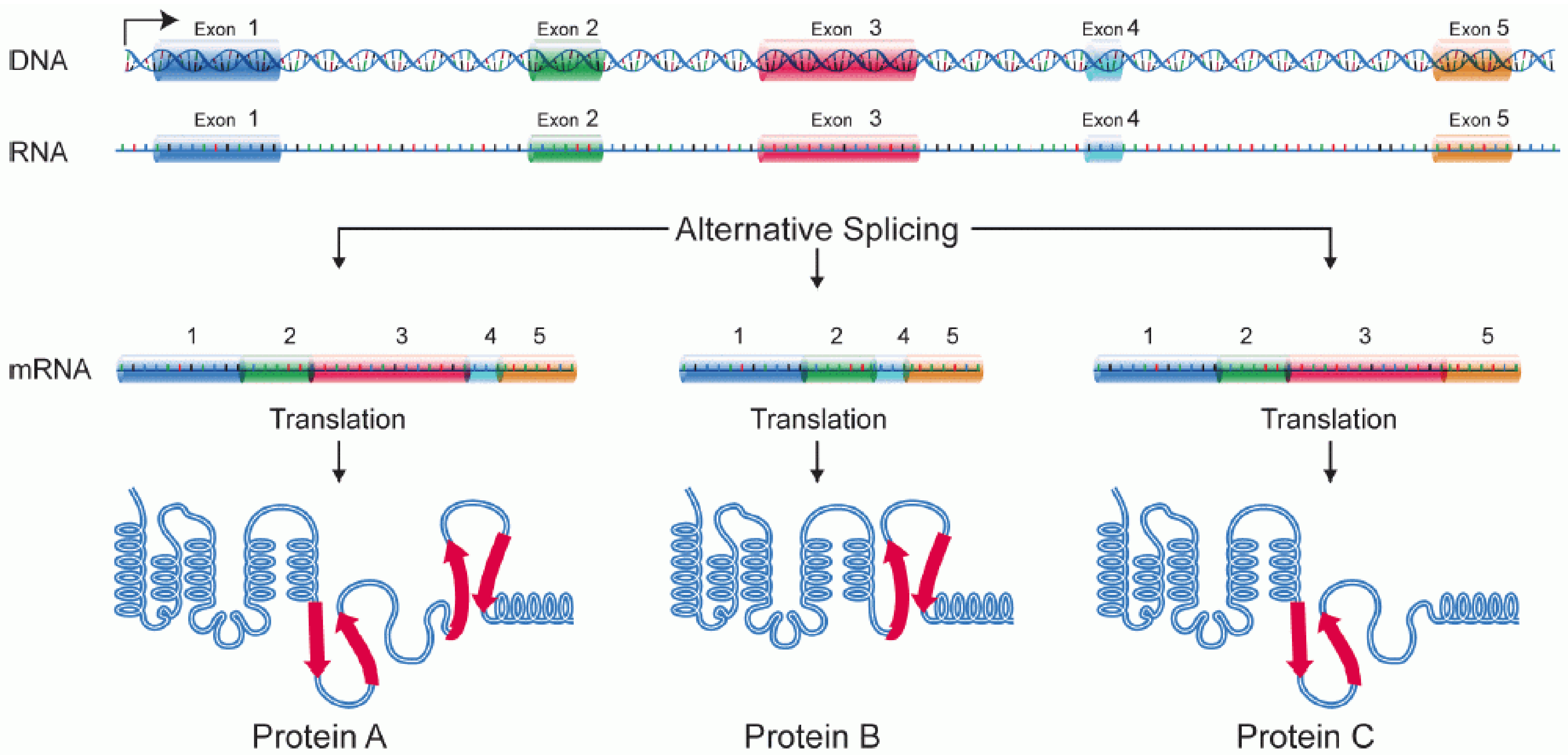


WWW.BIOLOGYEXAMS4U.COM

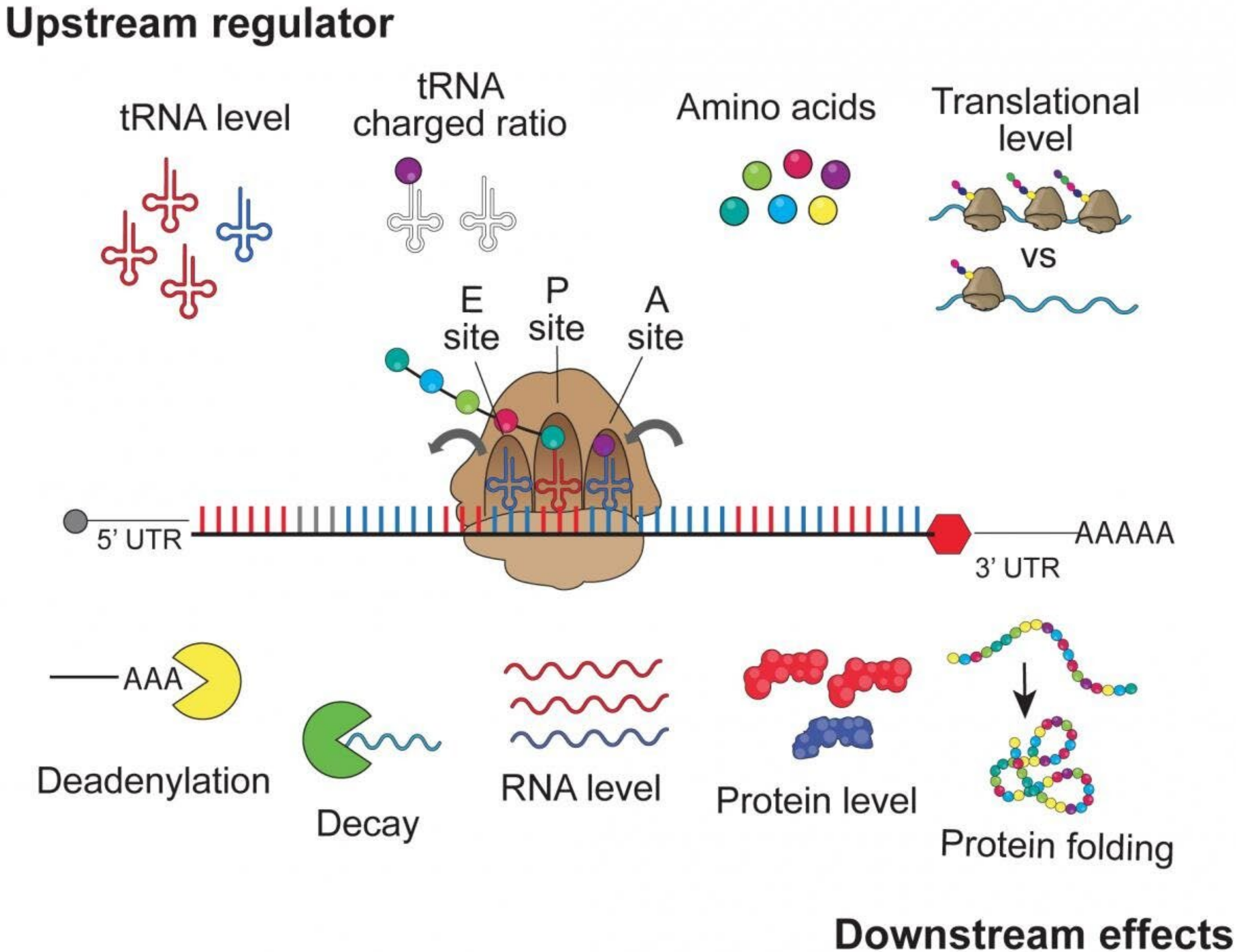
Post Transcription Modification of RNA

1. RNA capping
2. PolyA tail
3. Splicing



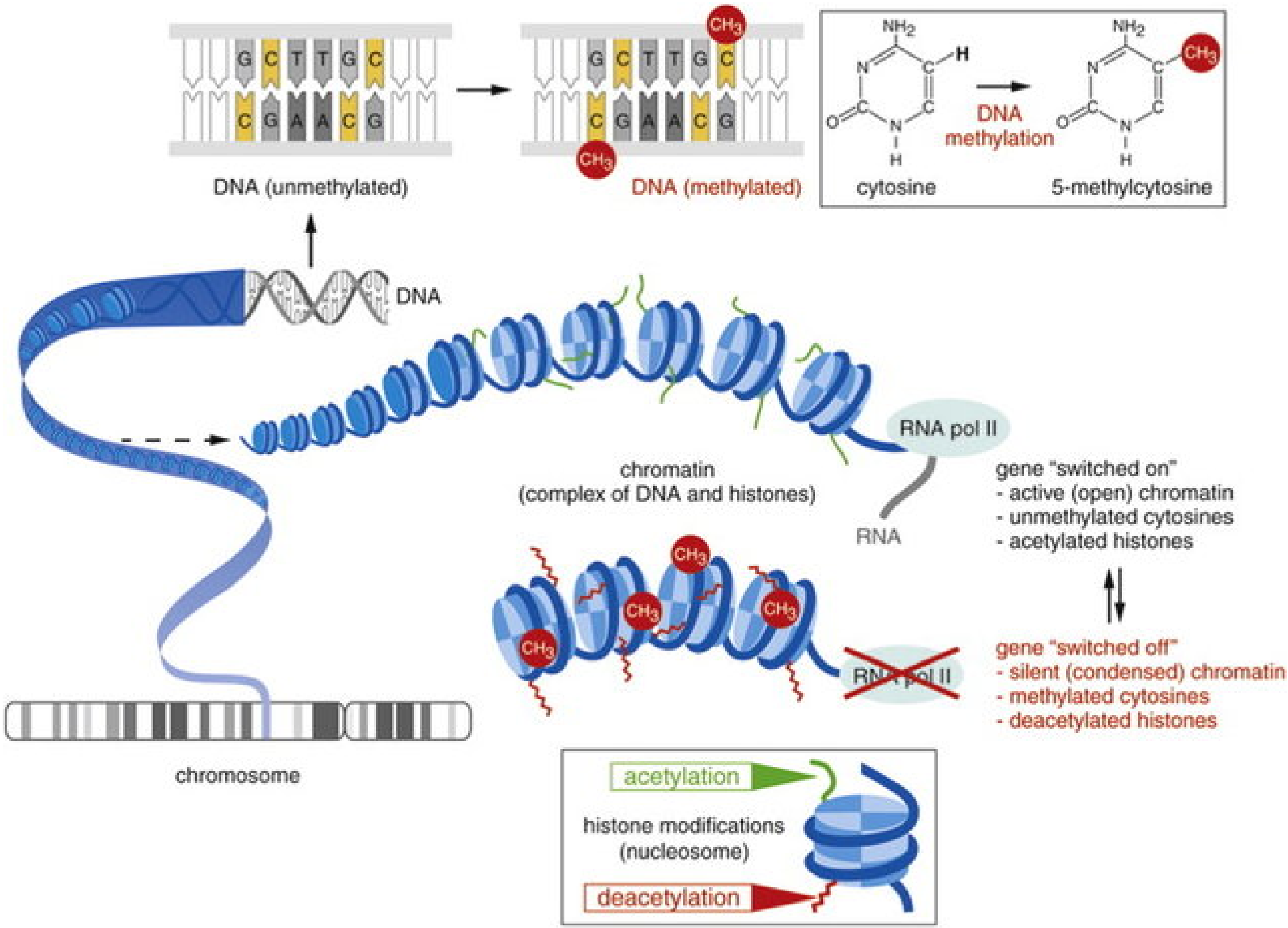


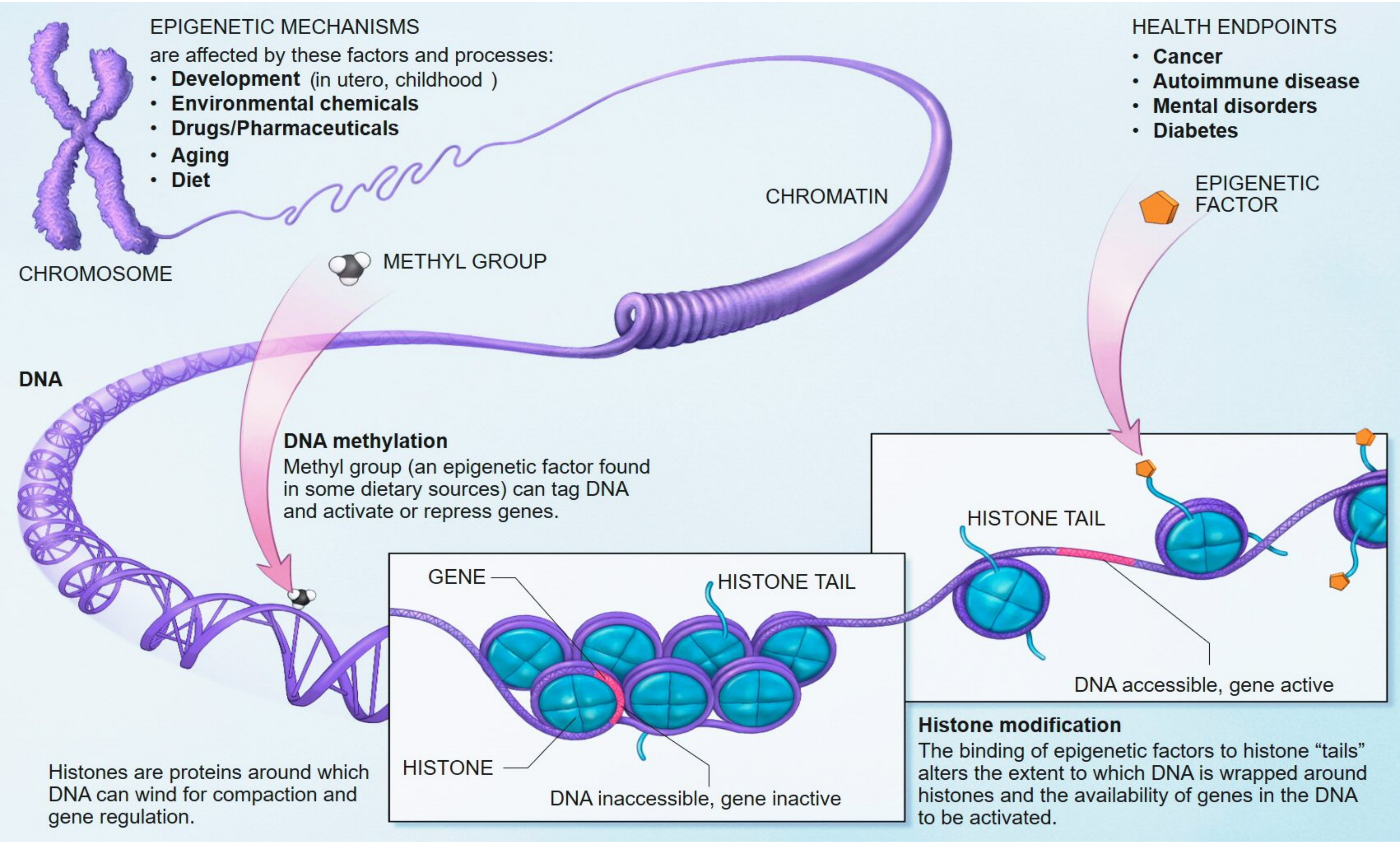
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
DNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
RNA
Alternative Splicing
1 2 3 4 5
mRNA
1 2 4 5
1 2 3 5
Translation
Translation
Translation
Protein A
Protein B
Protein C



Upstream regulator
tRNA level
tRNA
charged ratio
Amino acids
Translational
level
vs
E
site
P
site
A
site
5' UTR
AAAAA
3' UTR
AAA
Deadenylation
Decay
RNA level
Protein level
Protein folding
Downstream effects

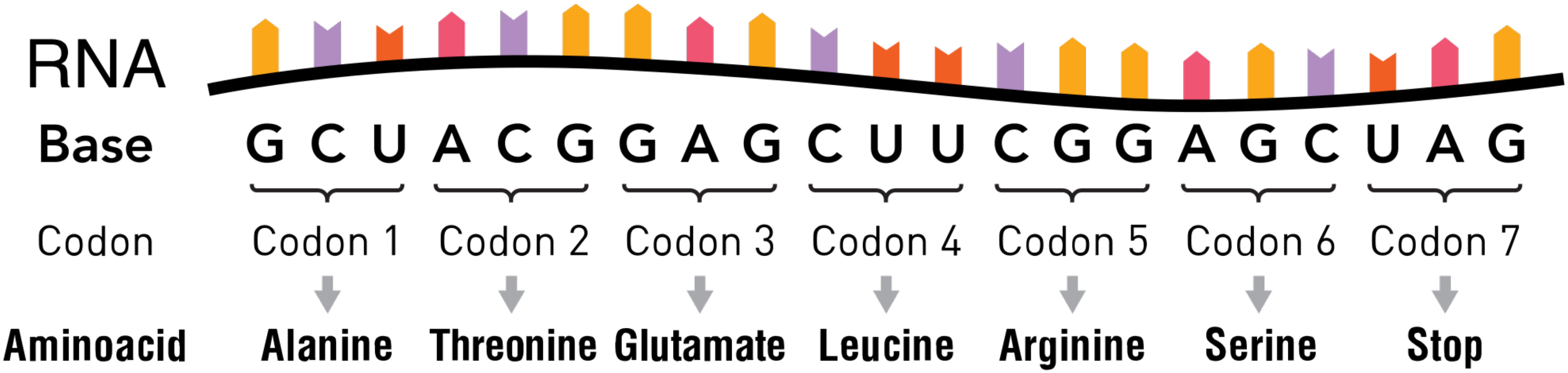
DNA Methylation and Histone Acetylation




EPIGENETIC MECHANISMS
are affected by these factors and processes:
• Development (in utero, childhood)
• Environmental chemicals
• Drugs/Pharmaceuticals
• Aging
• Diet
HEALTH ENDPOINTS
• Cancer
• Autoimmune disease
• Mental disorders
• Diabetes
CHROMOSOME
CHROMATIN
EPIGENETIC FACTOR
METHYL GROUP
DNA
DNA methylation
Methyl group (an epigenetic factor found in some dietary sources) can tag DNA and activate or repress genes.
HISTONE TAIL
GENE
HISTONE TAIL
HISTONE
DNA inaccessible, gene inactive
DNA accessible, gene active
Histone modification
The binding of epigenetic factors to histone “tails” alters the extent to which DNA is wrapped around histones and the availability of genes in the DNA to be activated.
Histones are proteins around which DNA can wind for compaction and gene regulation.

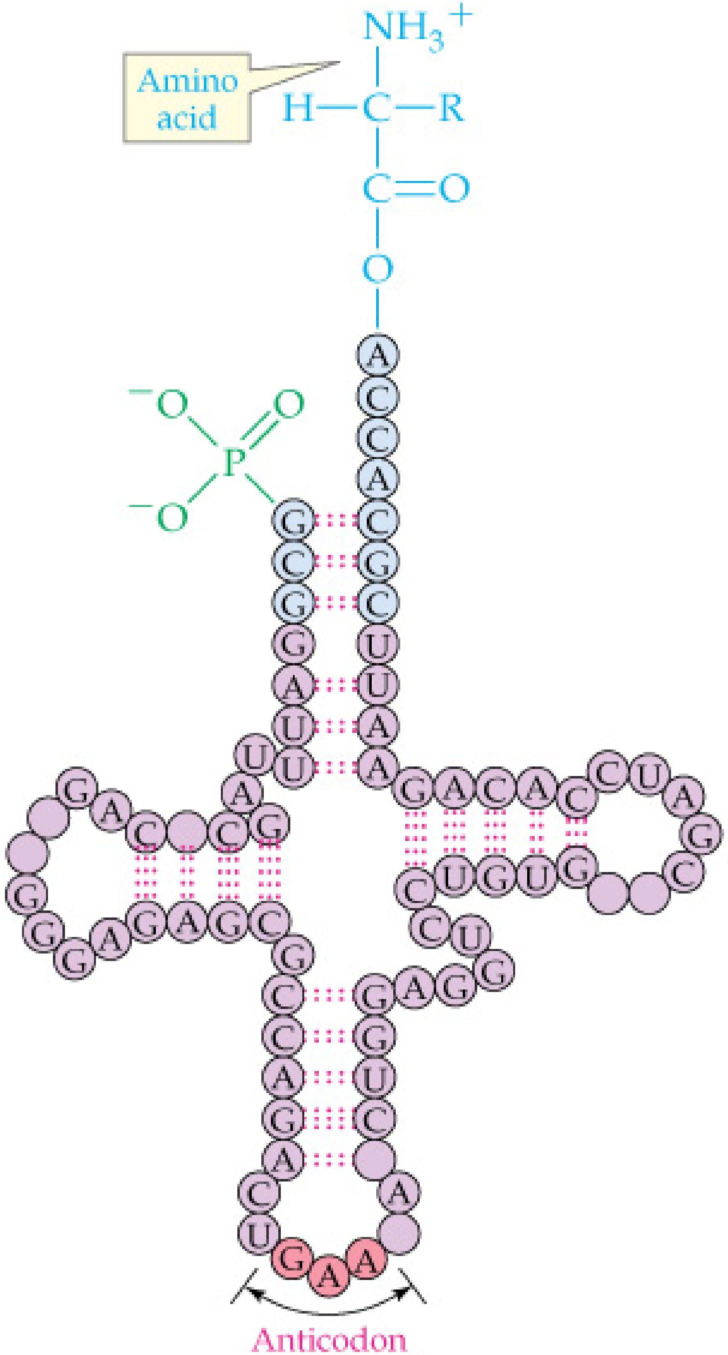
RNA Sequence

First letter	Second letter: U	C	A	G	Third letter
U	UUU, UUC: Phe UUA, UUG: Leu	UCU, UCC, UCA, UCG: Ser	UAU, UAC: Tyr UAA: Stop UAG: Stop	UGU, UGC: Cys UGA: Stop UGG: Trp	U C A G
C	CUU, CUC, CUA, CUG: Leu	CCU, CCC, CCA, CCG: Pro	CAU, CAC: His CAA, CAG: Gln	CGU, CGC, CGA, CGG: Arg	U C A G
A	AUU, AUC, AUA: Ile AUG: Met	ACU, ACC, ACA, ACG: Thr	AAU, AAC: Asn AAA, AAG: Lys	AGU, AGC: Ser AGA, AGG: Arg	U C A G
G	GUU, GUC, GUA, GUG: Val	GCU, GCC, GCA, GCG: Ala	GAU, GAC: Asp GAA, GAG: Glu	GGU, GGC, GGA, GGG: Gly	U C A G



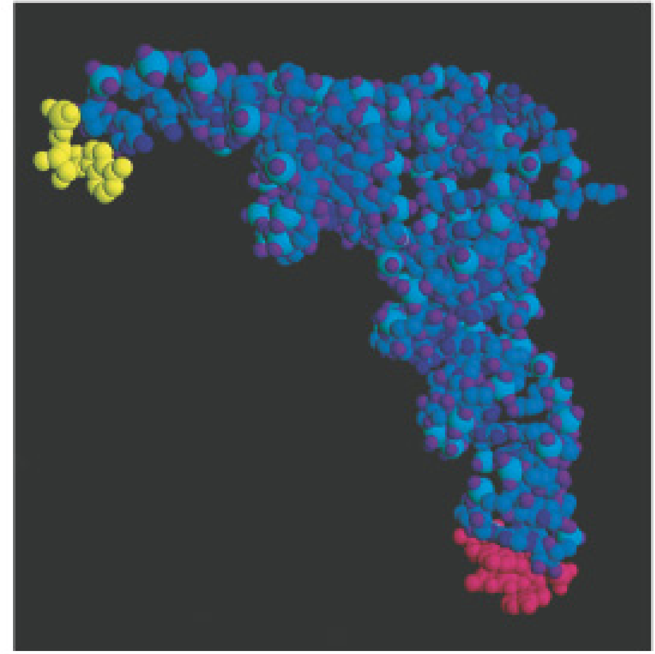
Second letter

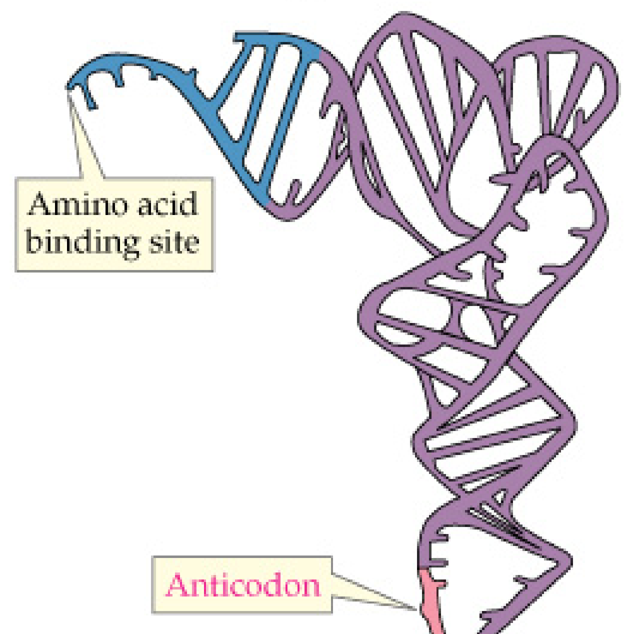
First letter	U	C	A	G	Third letter
U	UUU } Phe UUC UUA } Leu UUG	UCU UCC } Ser UCA UCG	UAU } Tyr UAC **UAA Stop** **UAG Stop**	UGU } Cys UGC **UGA Stop** UGG Trp	U C A G
C	CUU CUC } Leu CUA CUG	CCU CCC } Pro CCA CCG	CAU } His CAC CAA } Gln CAG	CGU CGC } Arg CGA CGG	U C A G
A	AUU AUC } Ile AUA **AUG Met**	ACU ACC } Thr ACA ACG	AAU } Asn AAC AAA } Lys AAG	AGU } Ser AGC AGA } Arg AGG	U C A G
G	GUU GUC } Val GUA GUG	GCU GCC } Ala GCA GCG	GAU } Asp GAC GAA } Glu GAG	GGU GGC } Gly GGA GGG	U C A G



NH3+
Amino acid
H—C—R
C=O
O
Anticodon

(a)

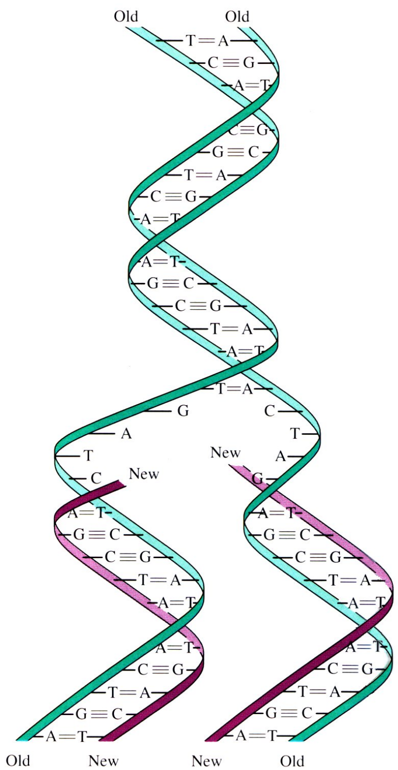

(b)

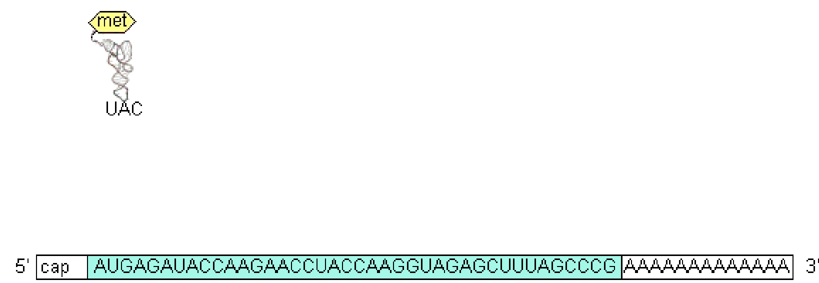


Amino acid binding site
Anticodon

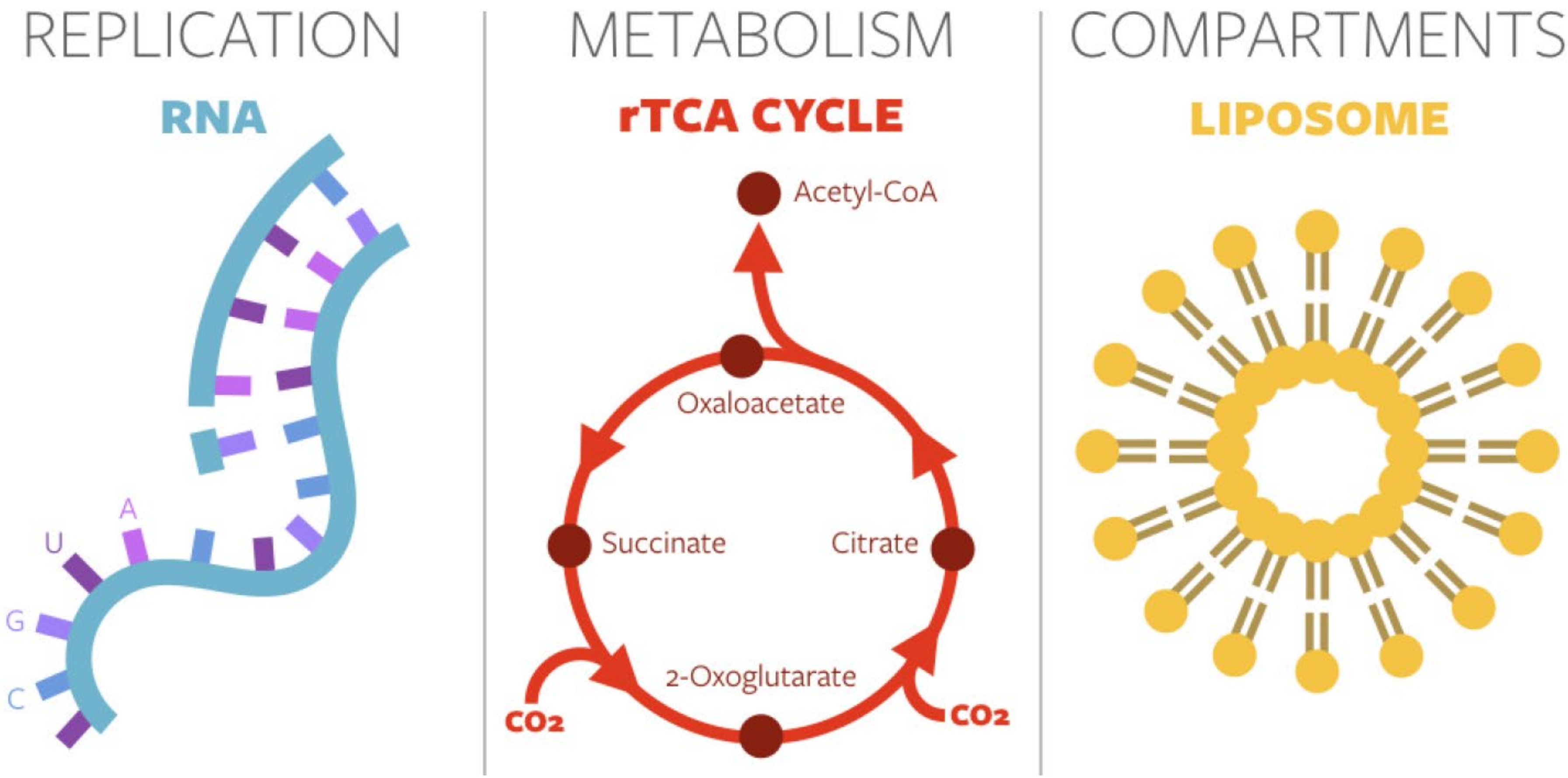
(c)

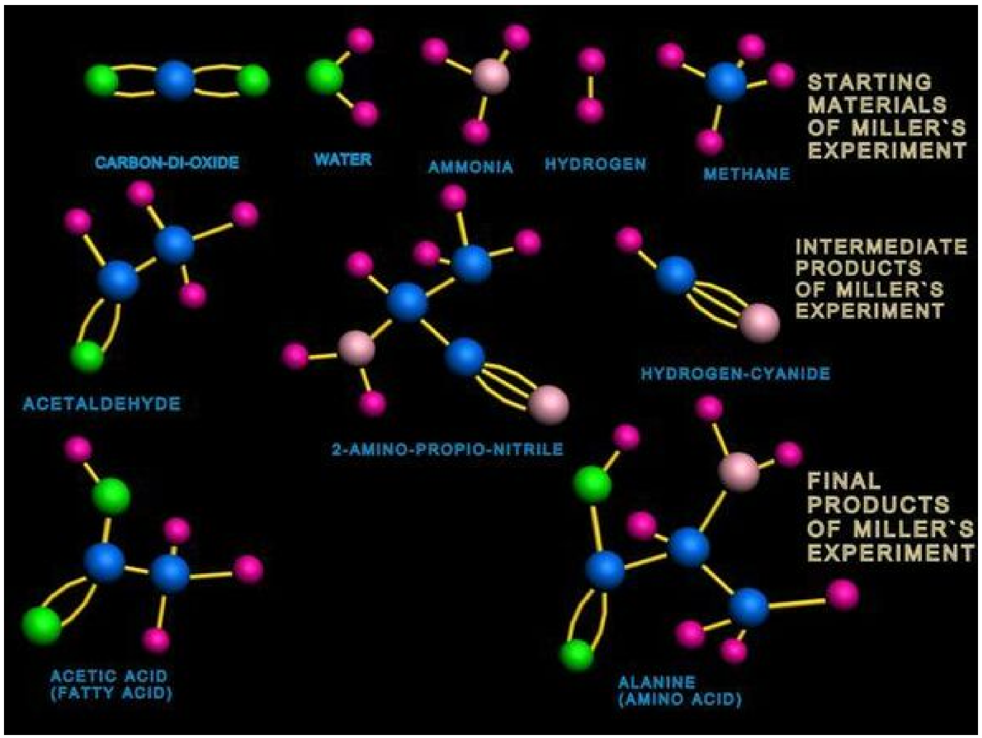
Self-Assembly Process in Nature





Definition of Life





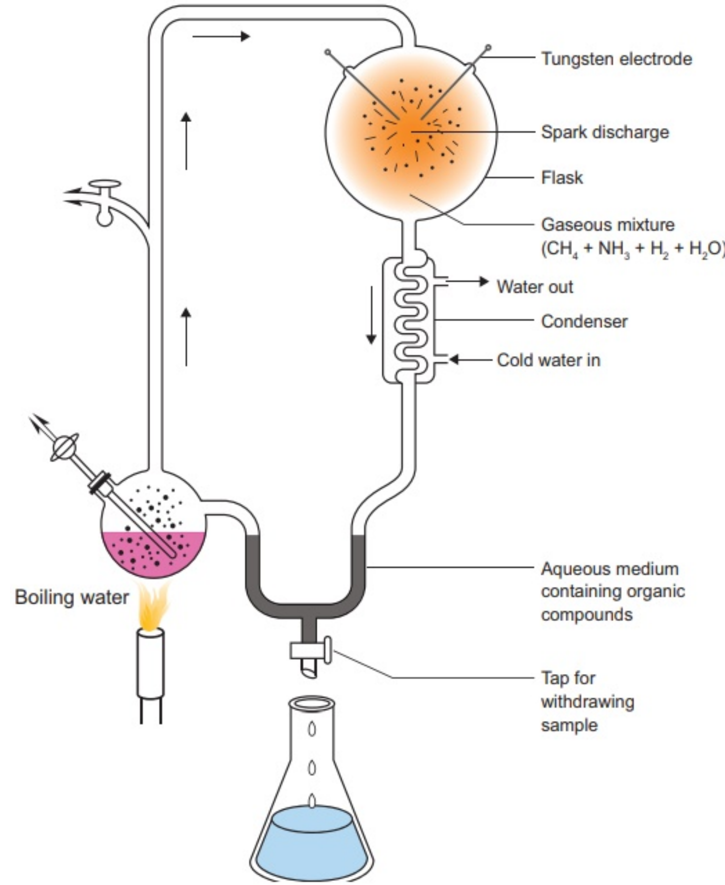
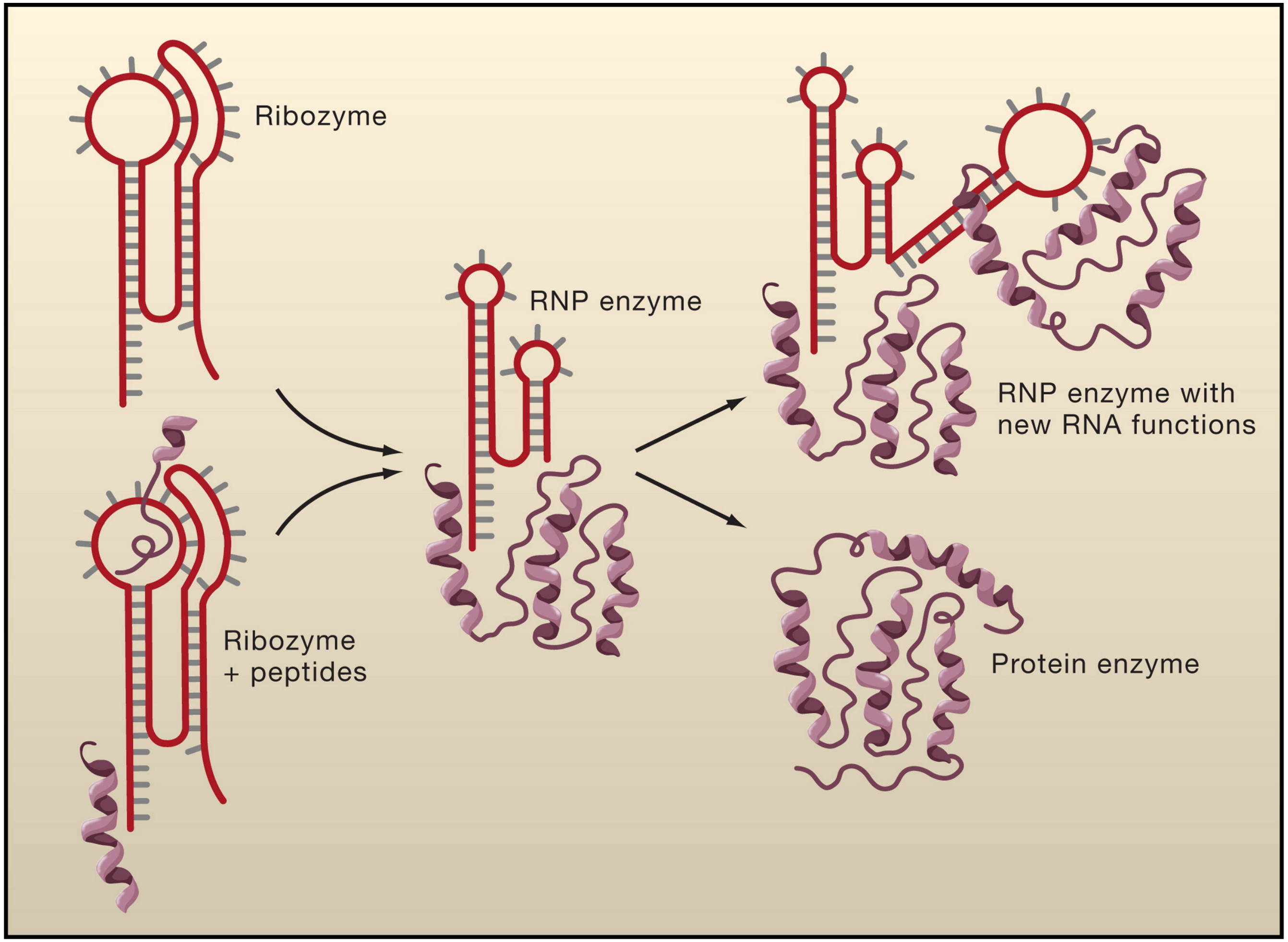
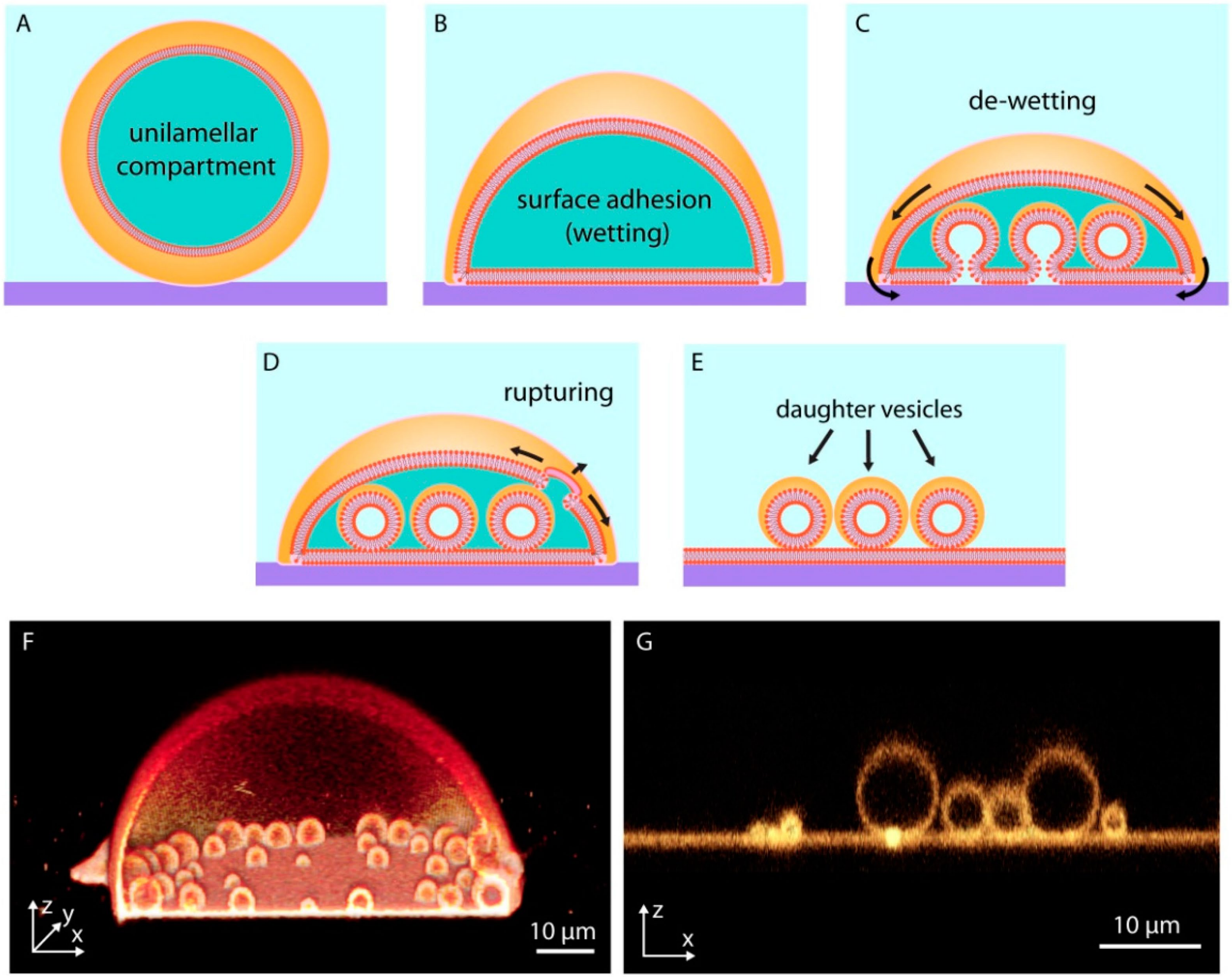


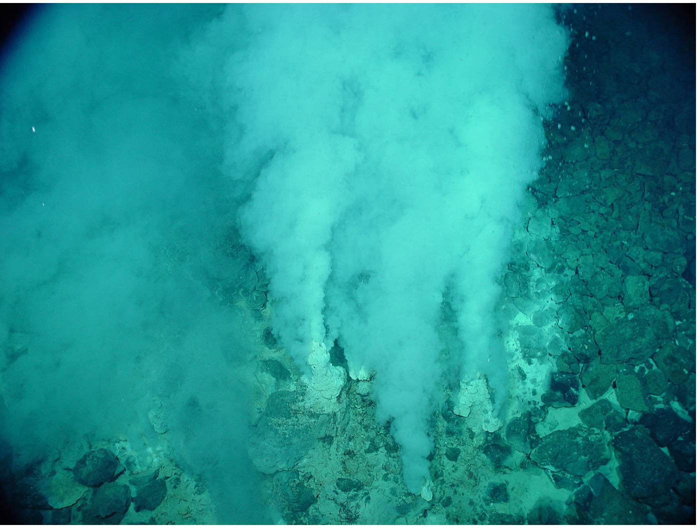
Fig. 6.1 Diagrammatic representation of Urey-Miller's experiment

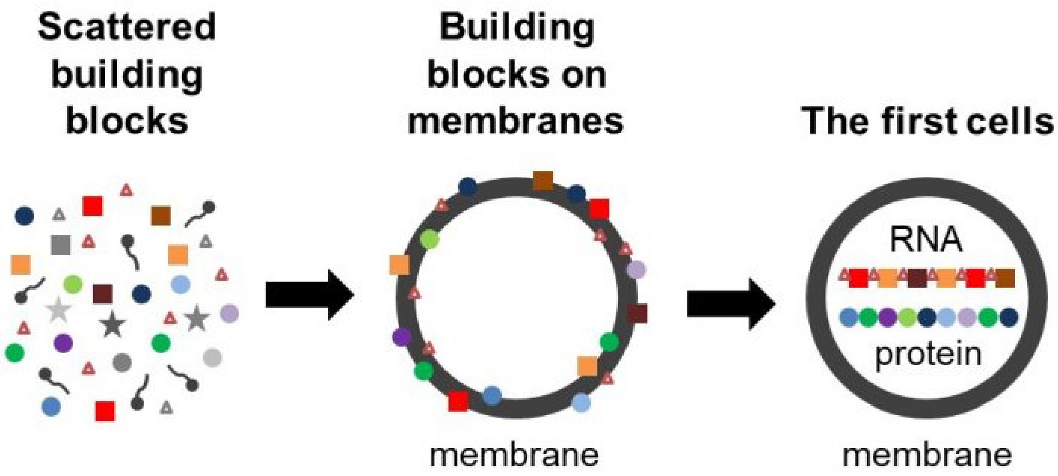
RNA World



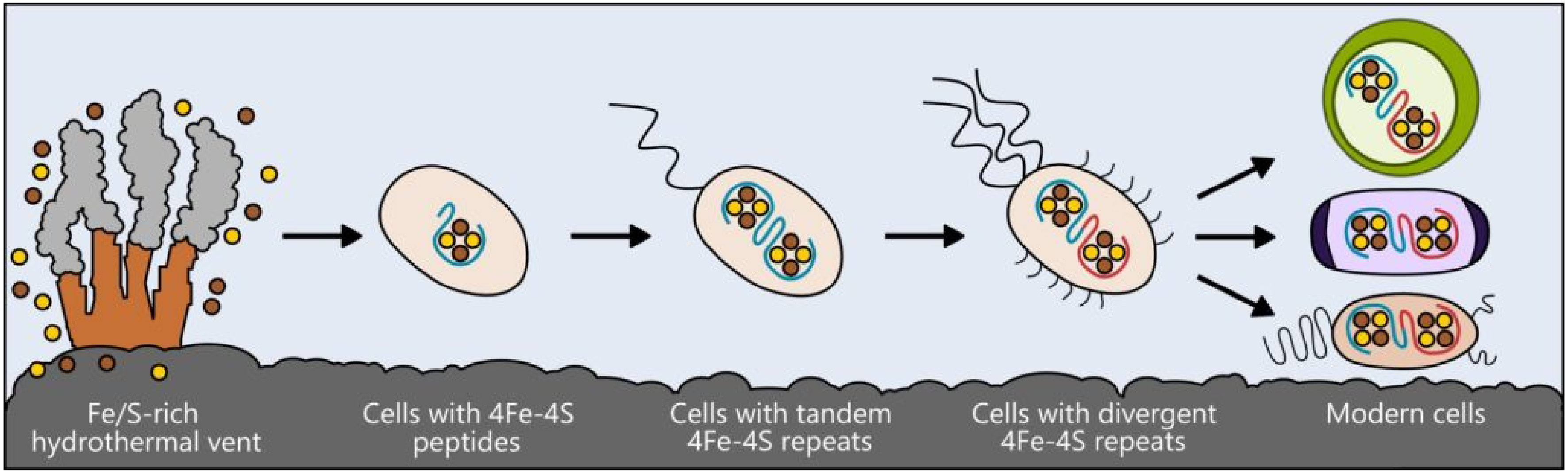
Cell 2009 136599-602DOI: (10.1016/j.cell.2009.02.002)


A
unilamellar
compartment
B
surface adhesion
(wetting)
C
de-wetting
D
rupturing
E
daughter vesicles
F
z
y
x
10 µm
G
z
x
10 µm

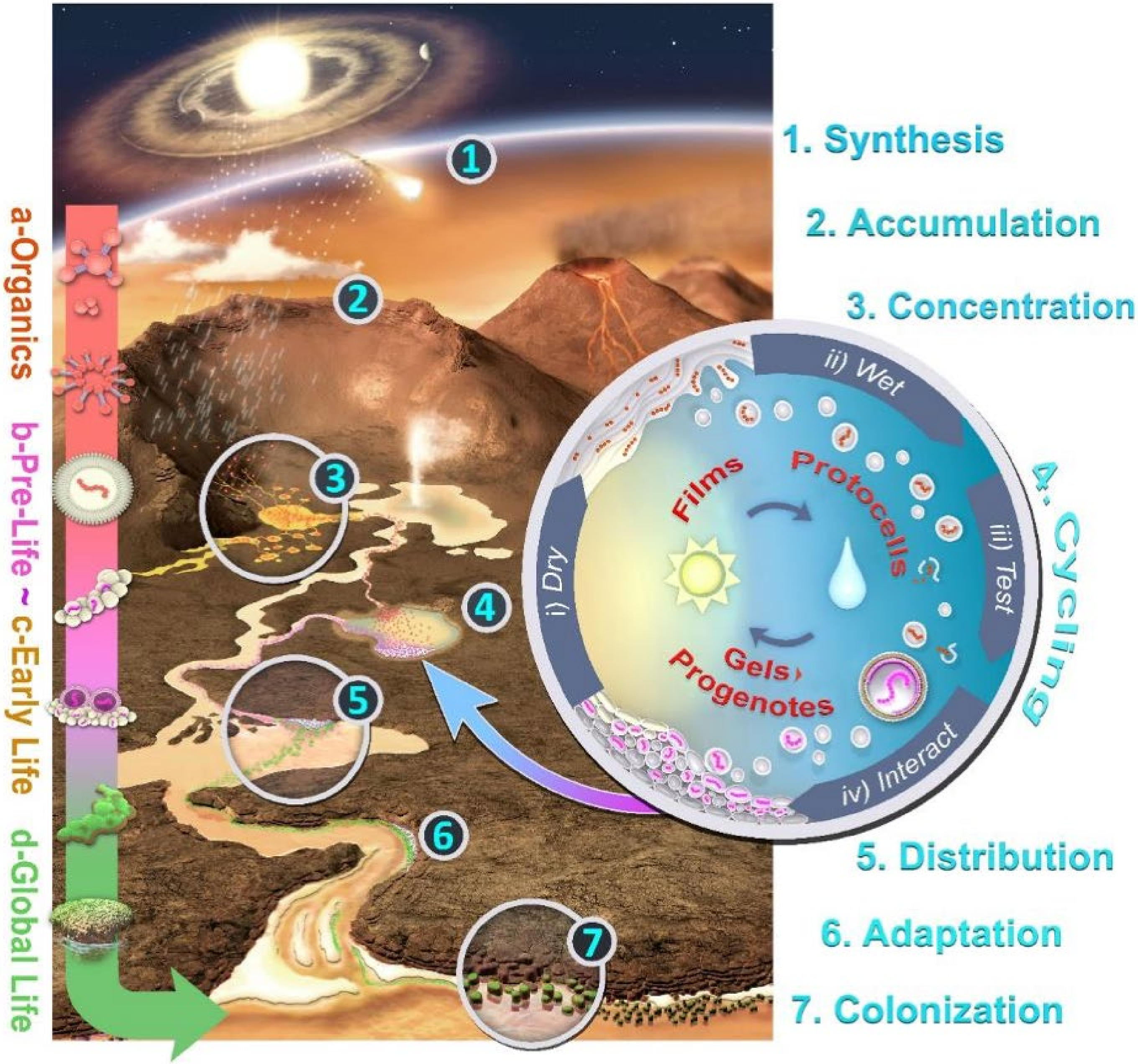




Scattered building blocks
Building blocks on membranes
The first cells
RNA
protein
membrane
membrane

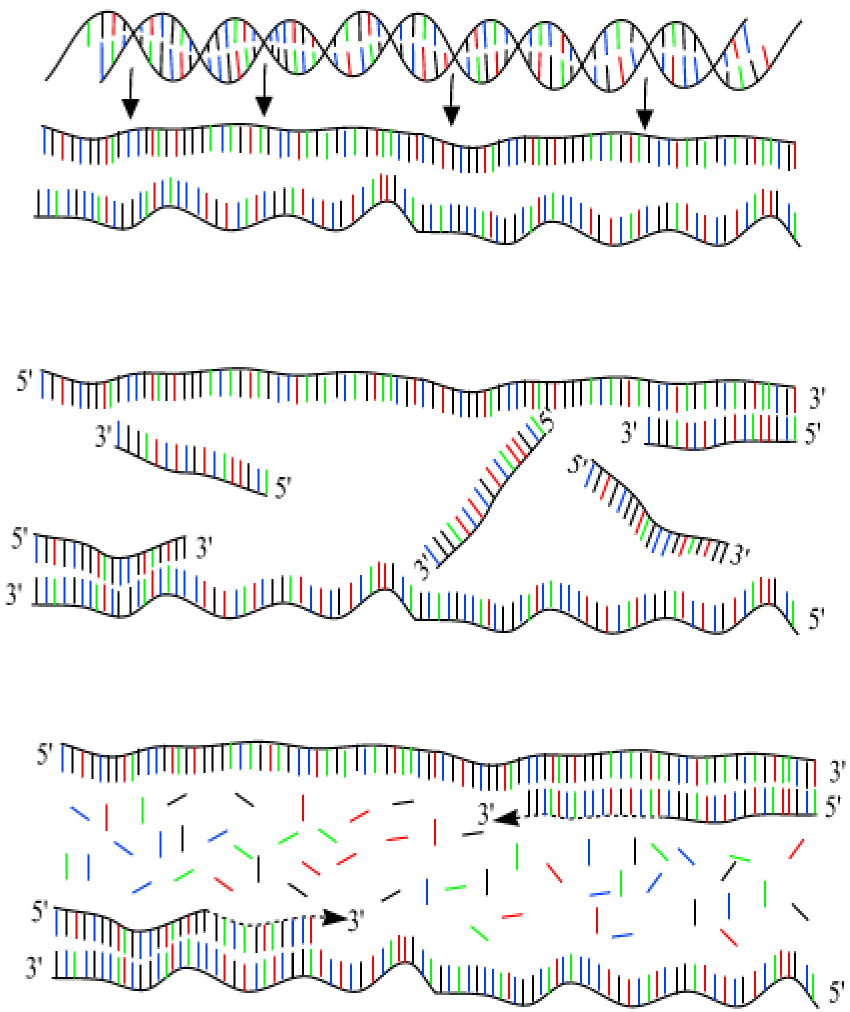


Fe/S-rich hydrothermal vent
Cells with 4Fe-4S peptides
Cells with tandem 4Fe-4S repeats
Cells with divergent 4Fe-4S repeats
Modern cells


1. Synthesis
2. Accumulation
3. Concentration
4. Cycling
5. Distribution
6. Adaptation
7. Colonization
a-Organics
b-Pre-Life ~ c-Early Life
d-Global Life
i) Dry
ii) Wet
iii) Test
iv) Interact
Films
Protocells
Gels
Progenotes

PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :



Step 1 : denaturation

1 minut 94 °C

Step 2 : annealing

45 seconds 54 °C

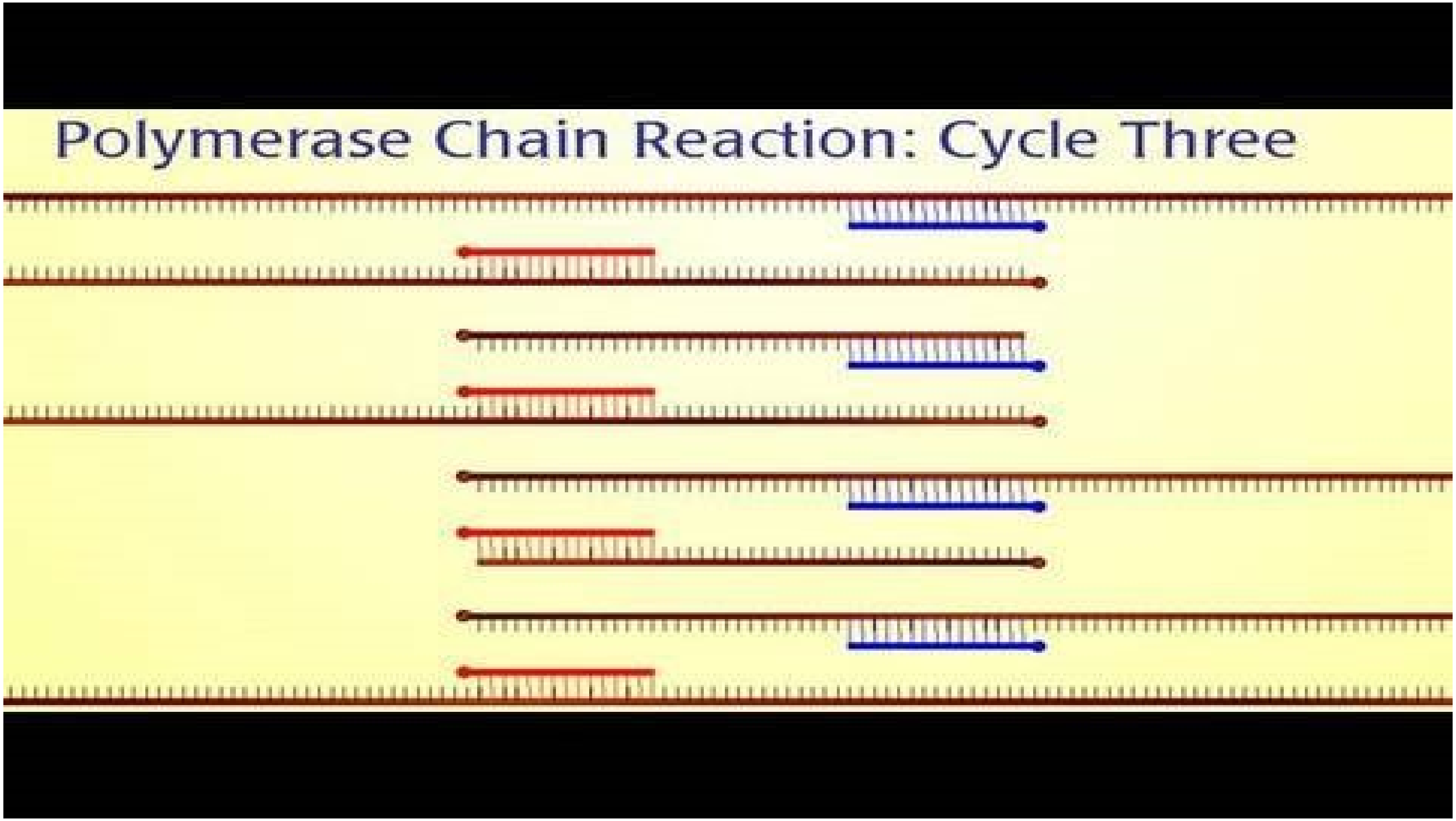
forward and reverse primers !!!

Step 3 : extension

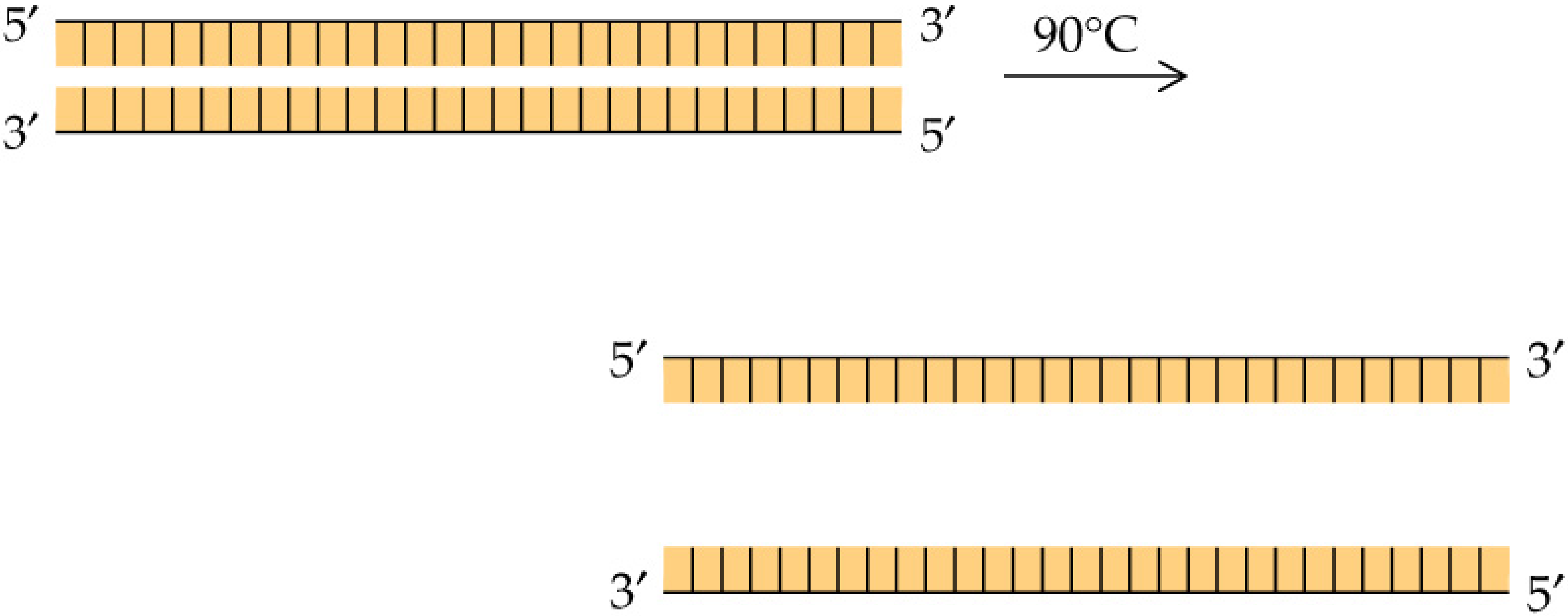
2 minutes 72 °C
only dNTP's

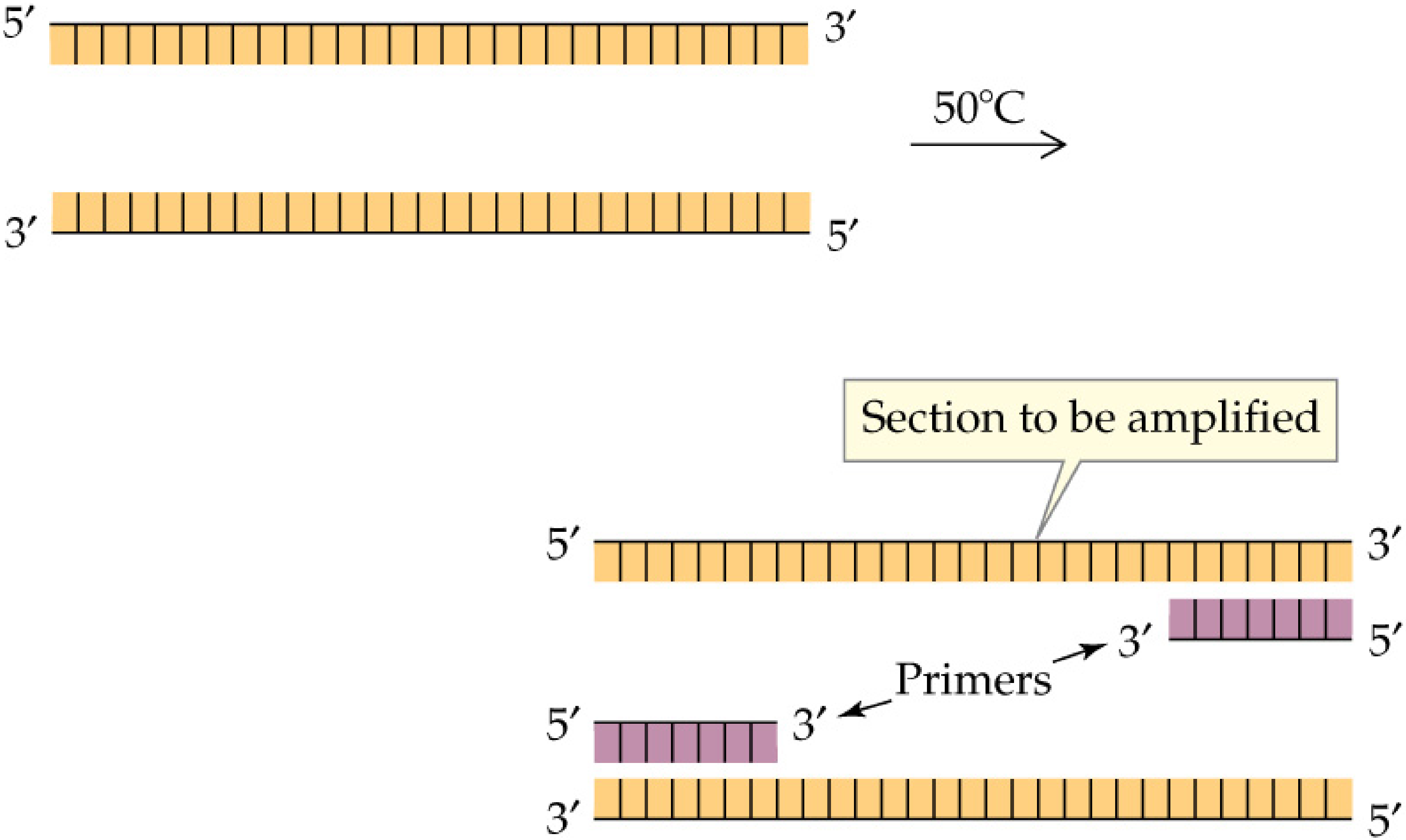
(Andy Vierstraete 1999)

PCR

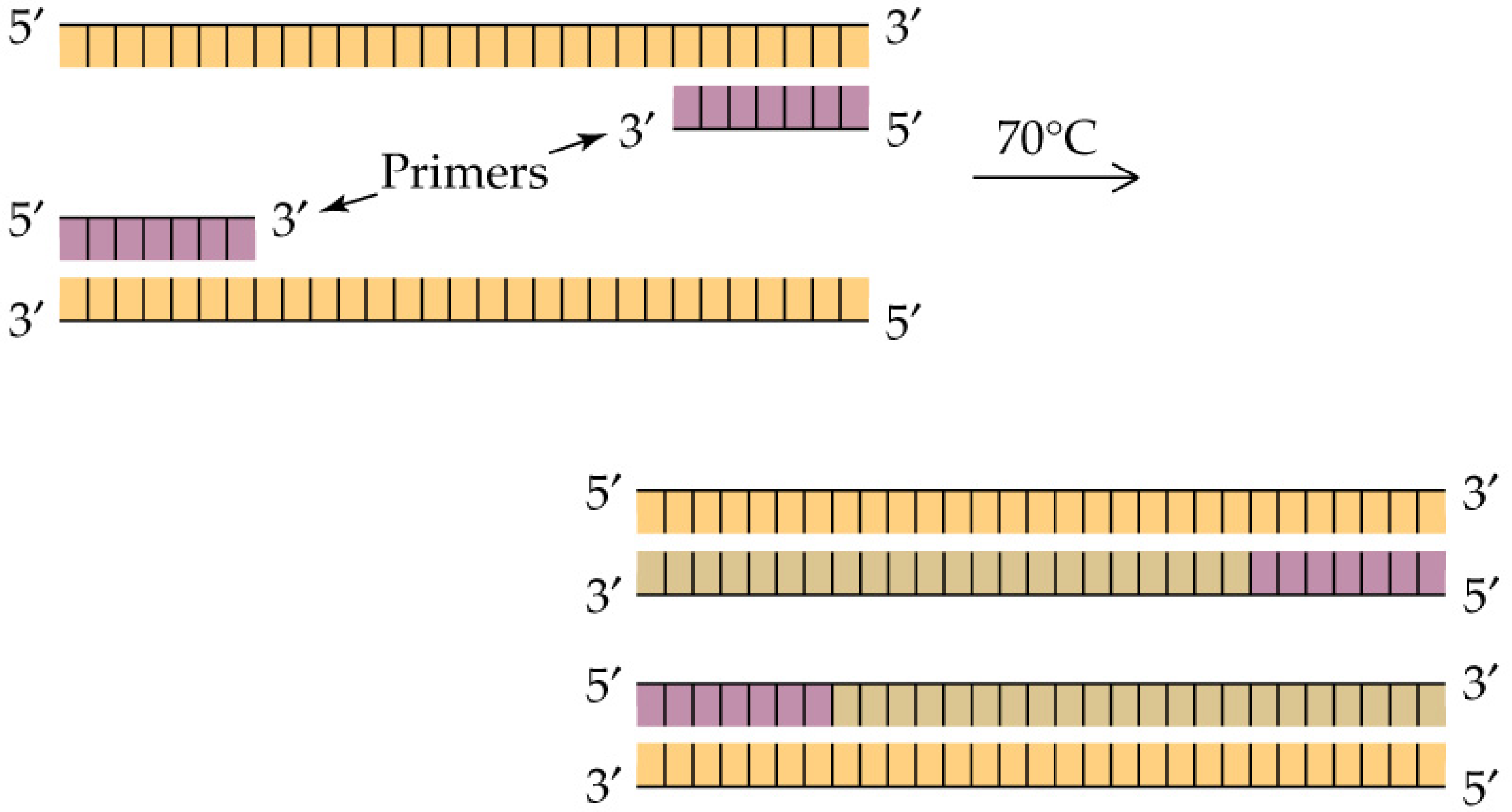


https://www.youtube.com/watch?v=JRAA4C2OPwg

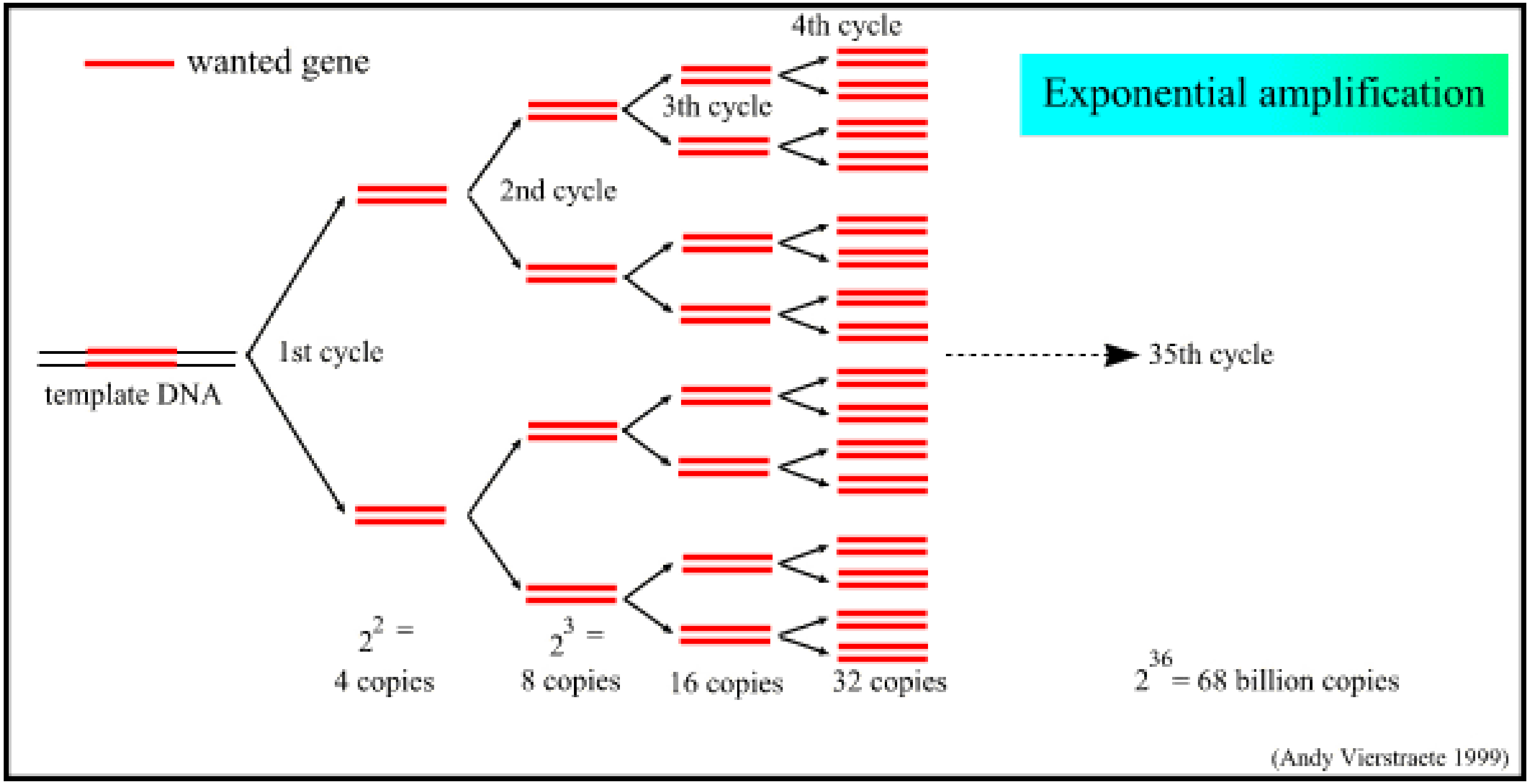

5′
3′
90°C
3′
5′
5′
3′
3′
5′



5′
3′
50°C
3′
5′
Section to be amplified
5′
3′
3′
5′
Primers
5′
3′
3′
5′

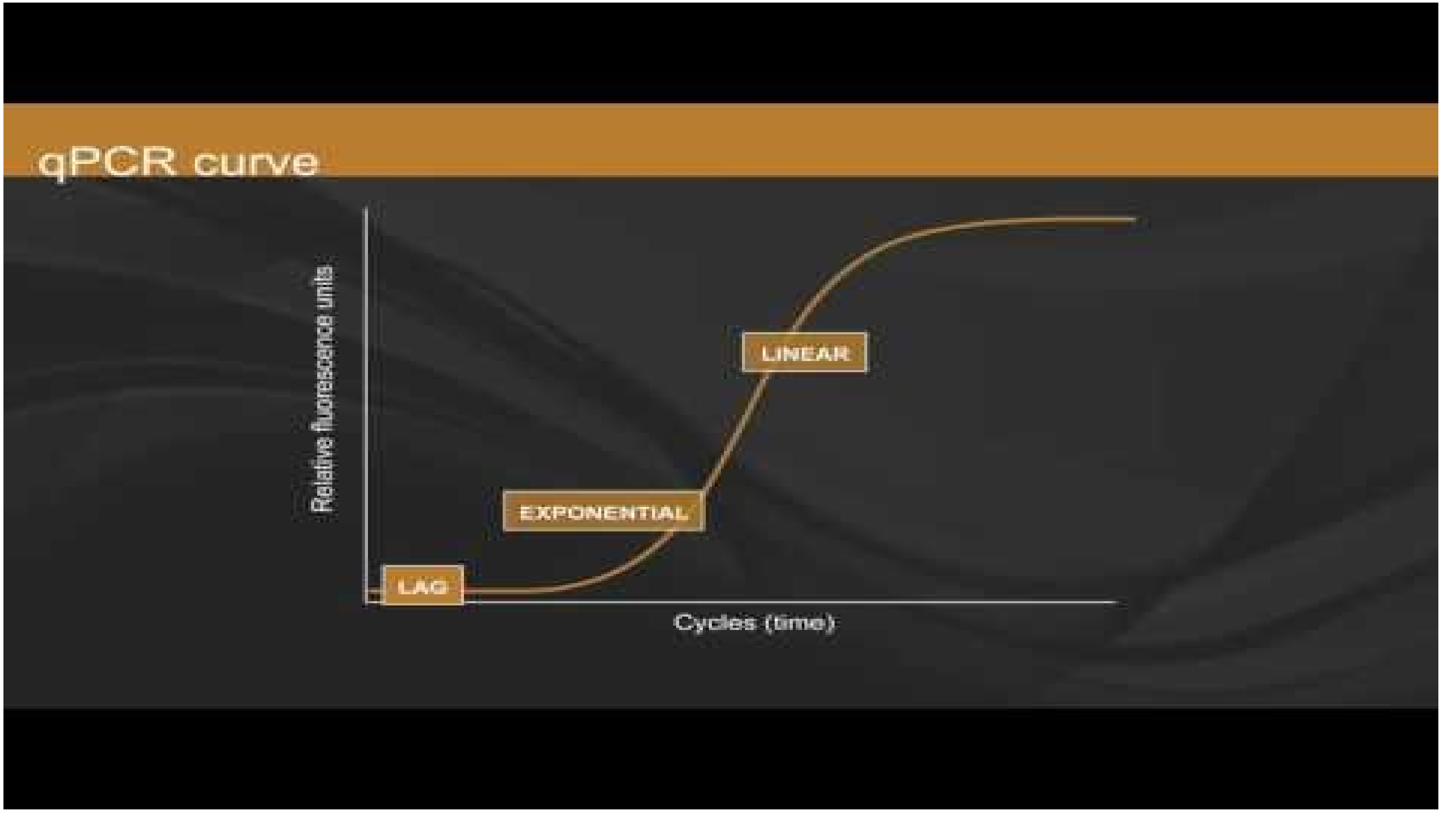


5′
3′
3′
5′
Primers
5′
3′
3′
5′
70°C
5′
3′
3′
5′
5′
3′
3′
5′

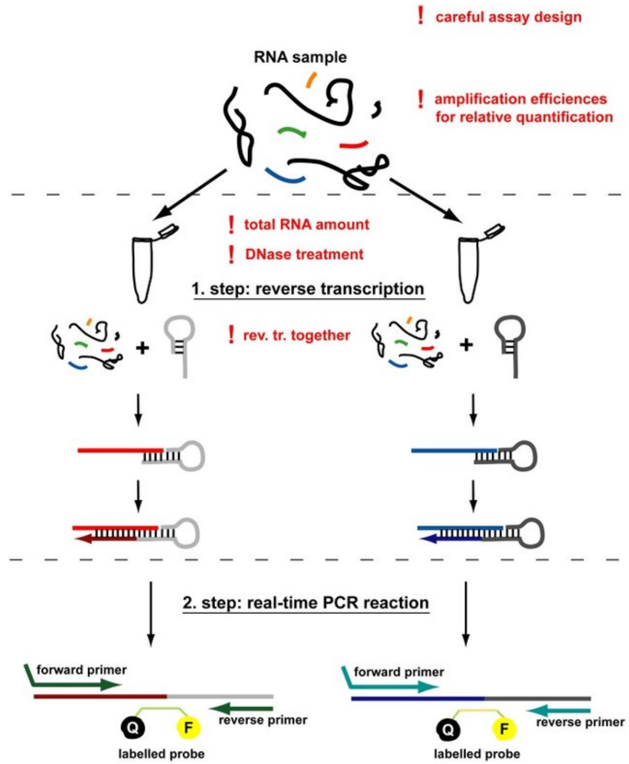


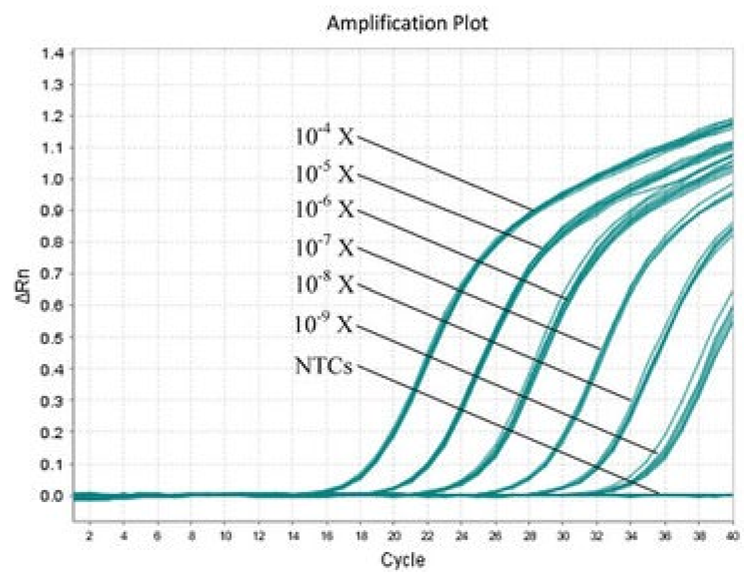
wanted gene
4th cycle
Exponential amplification
3th cycle
2nd cycle
1st cycle
template DNA
35th cycle
2^2 =
4 copies
2^3 =
8 copies
16 copies
32 copies
2^{36} = 68 billion copies
(Andy Vierstraete 1999)

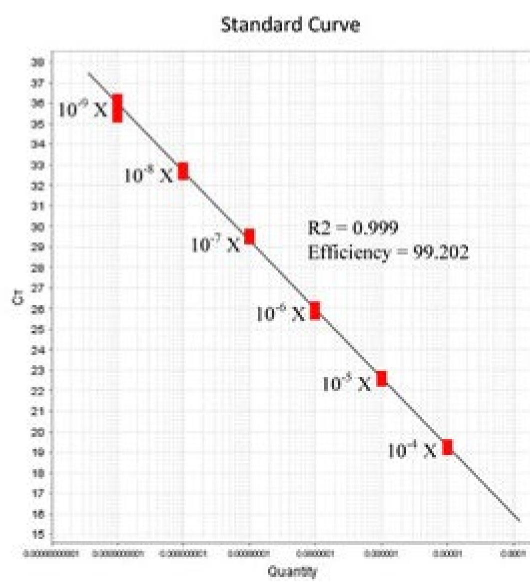
Real-time PCR



https://www.youtube.com/watch?v=1kvy17ugI4w


! careful assay design
RNA sample
! amplification efficiences for relative quantification
! total RNA amount
! DNase treatment
1. step: reverse transcription
! rev. tr. together
2. step: real-time PCR reaction
forward primer
reverse primer
labelled probe
Q
F
forward primer
reverse primer
labelled probe
Q
F


Amplification Plot
10-4 X
10-5 X
10-6 X
10-7 X
10-8 X
10-9 X
NTCs
ΔRn
Cycle


Standard Curve
10-9 X
10-8 X
10-7 X
10-6 X
10-5 X
10-4 X
R2 = 0.999
Efficiency = 99.202
CT
Quantity

PROTOCOL OF SARS-COV-2 DETECTION USING REAL-TIME RT-PCR

Target gene → RdRp gene (Corman *et al.* 2020)

PCR amplification regions → nCoV_IP2 / 12621-12727 and nCoV_IP4 / 14010-14116 (Institut Pasteur, Paris)

Primer sets and probes → designed based on the first sequences of SARS-CoV-2 available on the GISAID database

RNA extraction → NucleoSpin® RNA Virus or viral RNA mini kit (QIAGEN)

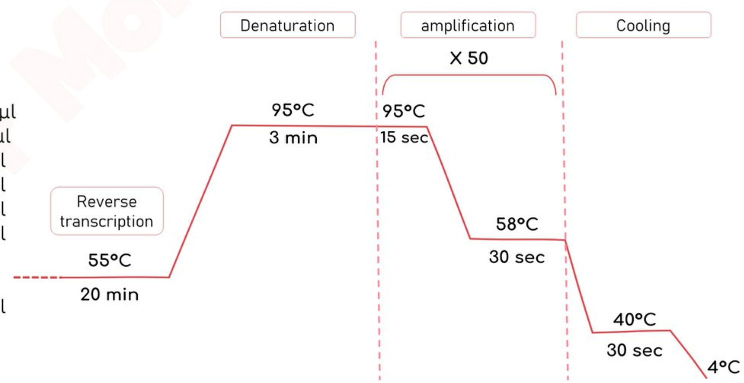
Sample lysis	Binding of viral RNA	Washing	Elution of viral RNA
5 min incubation of sample in Lysis Buffer containing Proteinase K	Ethanol addition and transfer of lysate to Column	1st Wash Buffer (high salt concentration) 2nd Wash Buffer (low salt concentration)	Elution in 20-50 µl RNase-free water or Elution Buffer

Real-time Multiplex RT-PCR (Institut Pasteur, Paris)

Multiplex Mix (nCoV_IP2&IP4)

Sample RNA	5 µl
H2O	1.3 µl
Reaction mix 2X	12.50 µl
MgSO4 (50mM)	0.40 µl
Forward Primer1 (10µM)	1.00 µl
Reverse Primer1 (10µM)	1.00 µl
Forward Primer2 (10µM)	1.00 µl
Reverse Primer2 (10µM)	1.00 µl
Probe 1 (10µM)	0.4 µl
Probe 2 (10µM)	0.4 µl
SuperscriptIII RT/Platinum Taq Mix	1.00 µl

Amplification Cycles (Lightcycler System)

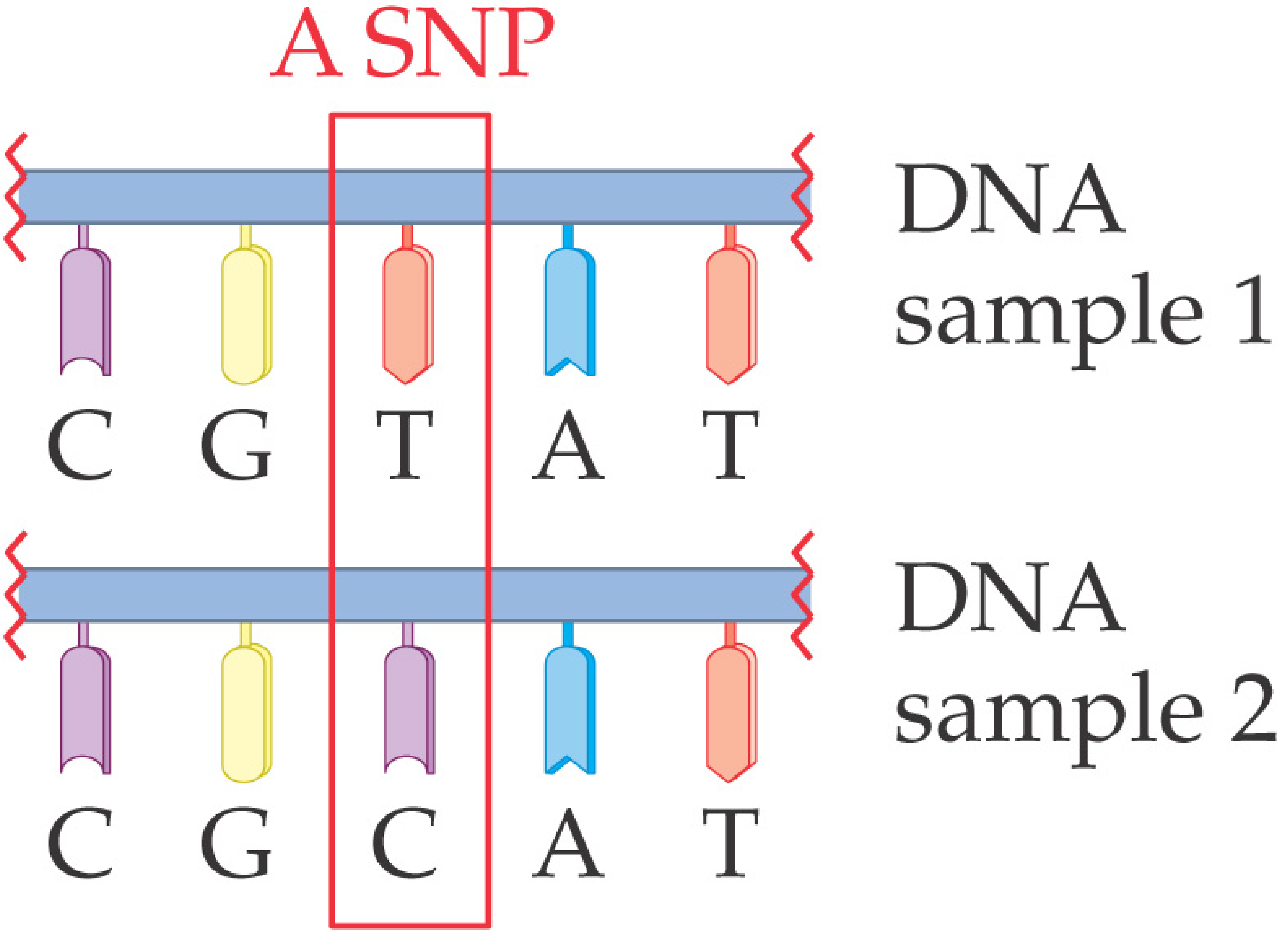


POSITIVE CONTROL

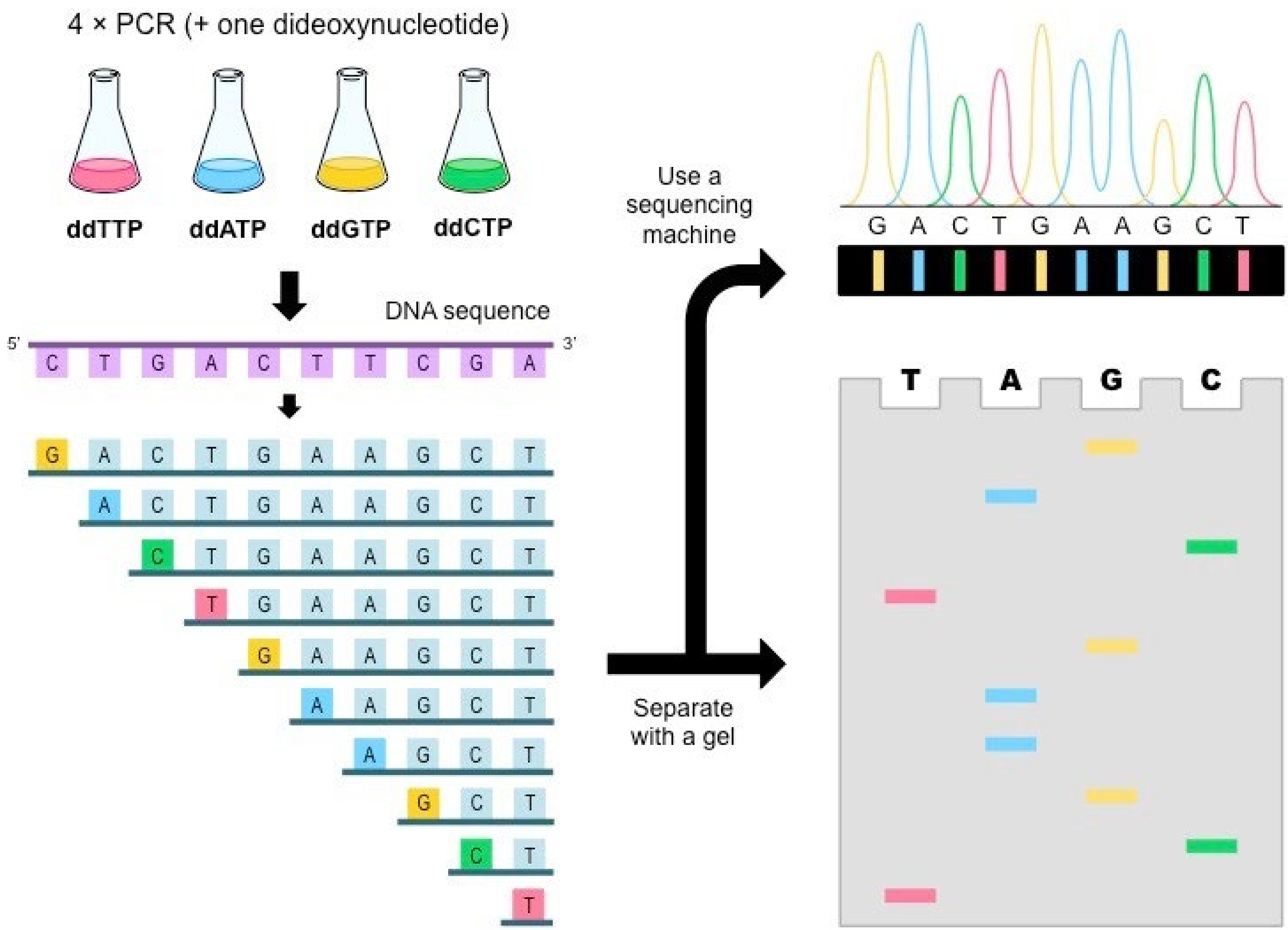
Positive control for real-time RT-PCR is the in vitro transcribed RNA derived from strain BetaCoV_Wuhan_WIV04_2019. The transcript contains the **amplification regions** of the RdRp and E gene as positive strand.

References

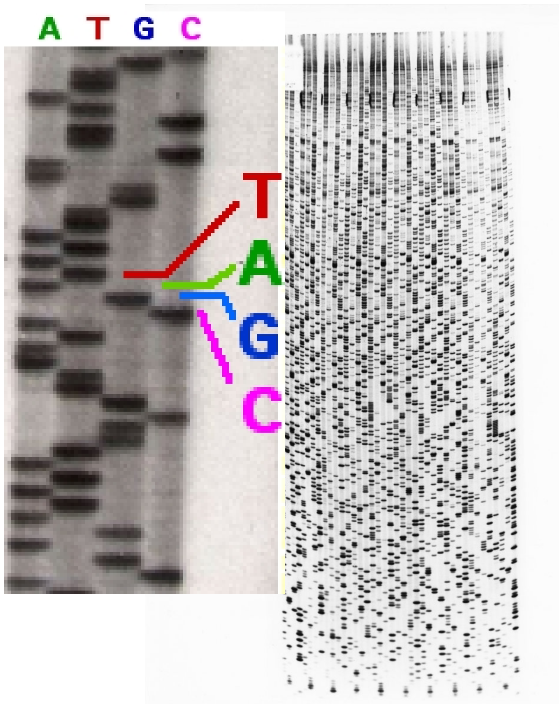
1. Institut Pasteur, Paris, « Protocol: Real-time RT-PCR assays for the detection of SARS-CoV-2 », OMS, 2 mars 2020.
2. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill 2020;25.

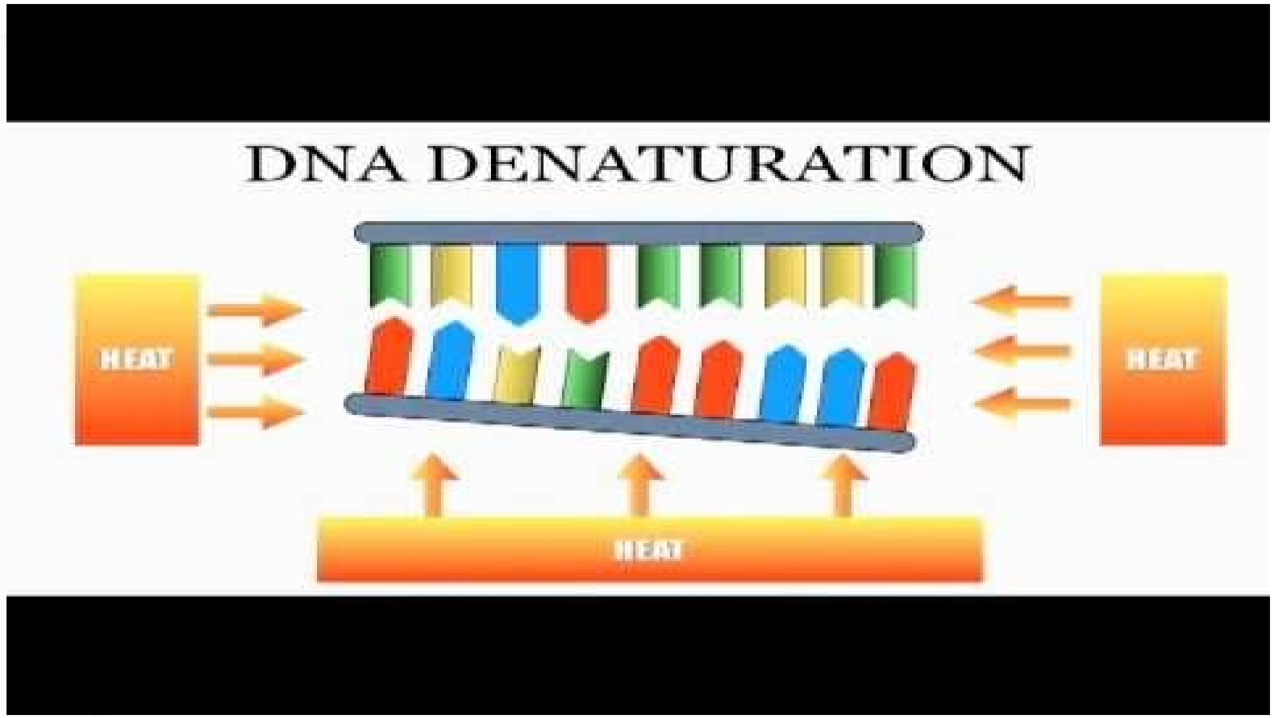

A SNP
DNA
sample 1
C G T A T
DNA
sample 2
C G C A T

DNA Sequencing



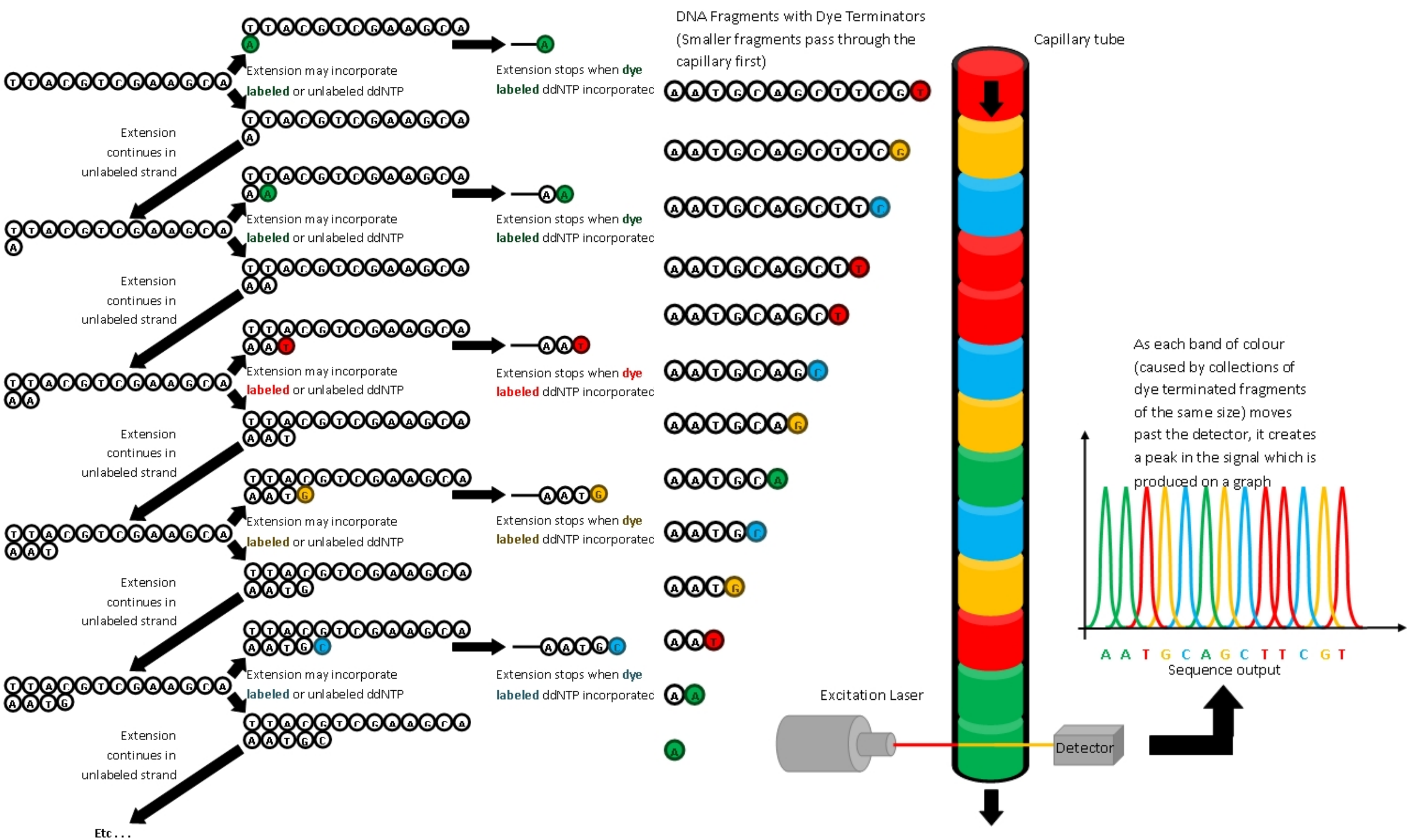
DNA Sequencing

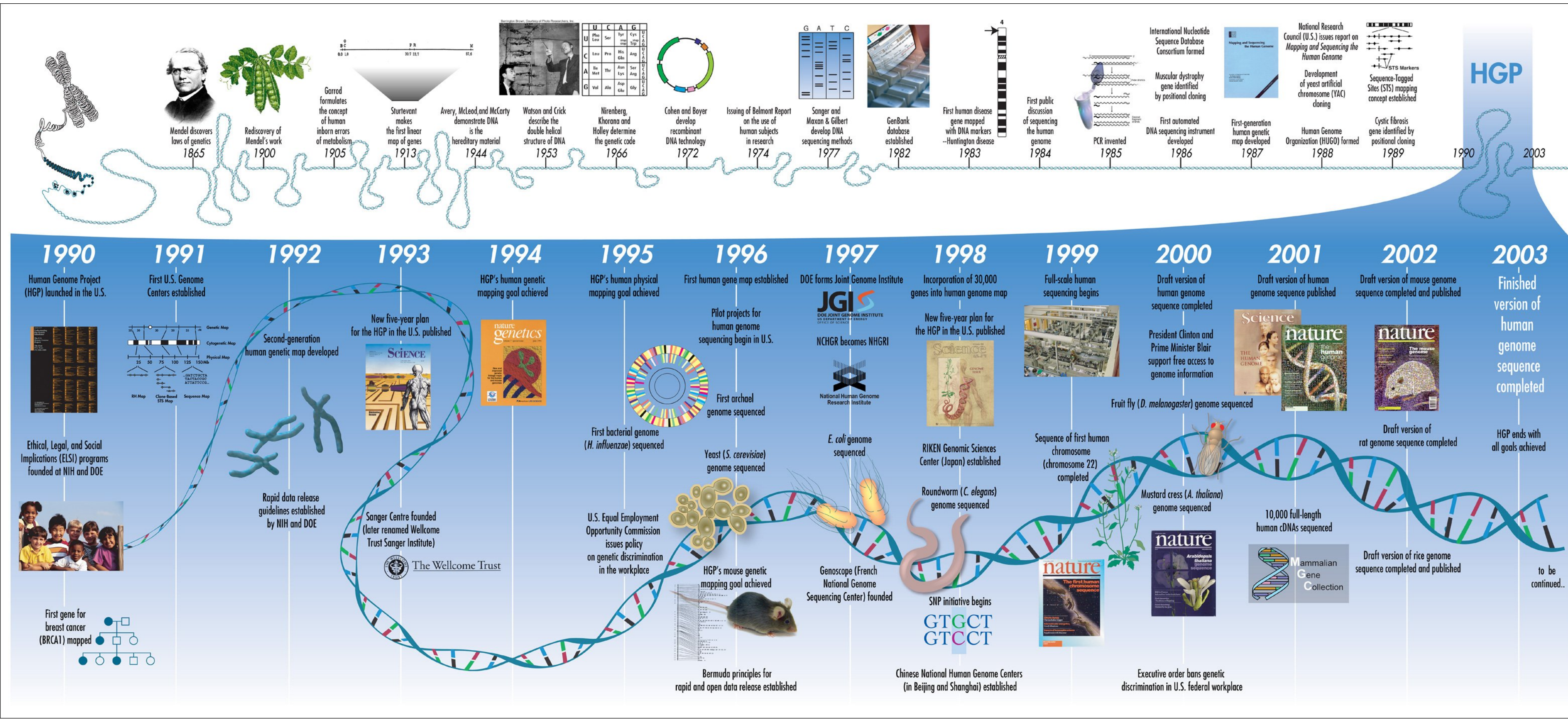




https://www.youtube.com/watch?v=vK-HIMaitnE

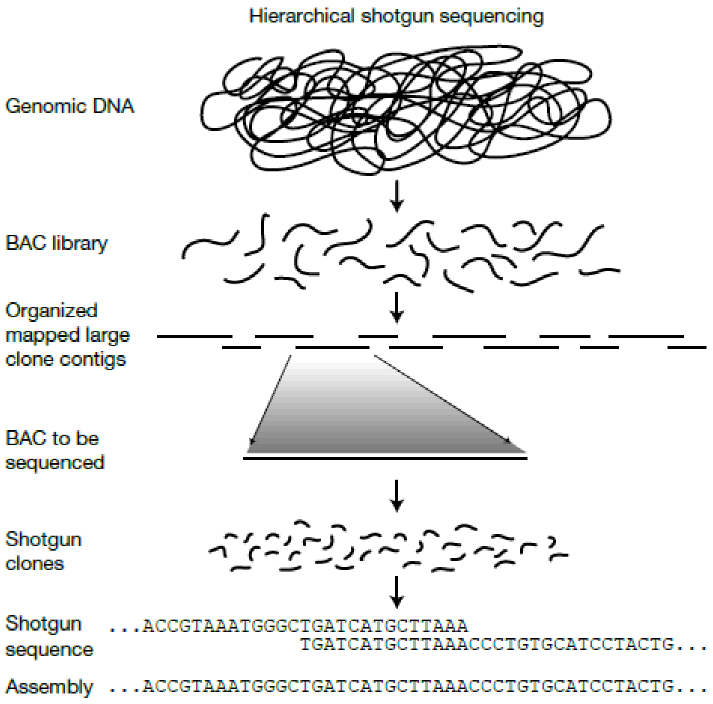
Dye Terminations





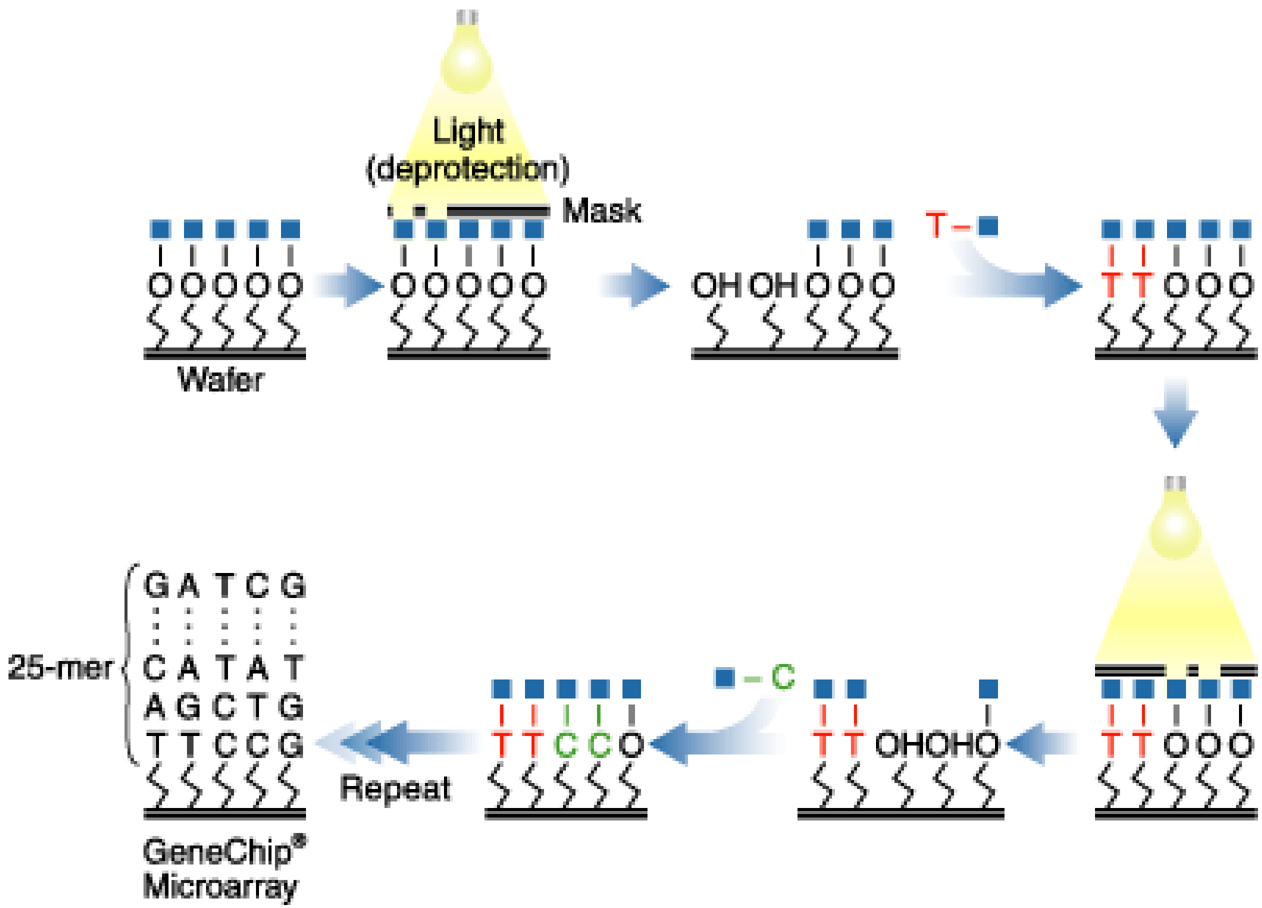
Mendel discovers laws of genetics
1865
Rediscovery of Mendel's work
1900
Garrod formulates the concept of human inborn errors of metabolism
1905
Sturtevant makes the first linear map of genes
1913
Avery, McLeod, and McCarty demonstrate DNA is the hereditary material
1944
Watson and Crick describe the double helical structure of DNA
1953
Nirenberg, Khorana and Holley determine the genetic code
1966
Cohen and Boyer develop recombinant DNA technology
1972
Issuing of Belmont Report on the use of human subjects in research
1974
Sanger and Maxam & Gilbert develop DNA sequencing methods
1977
GenBank database established
1982
First human disease gene mapped with DNA markers –Huntington disease
1983
First public discussion of sequencing the human genome
1984
PCR invented
1985
International Nucleotide Sequence Database Consortium formed
Muscular dystrophy gene identified by positional cloning
First automated DNA sequencing instrument developed
1986
First-generation human genetic map developed
1987
National Research Council (U.S.) issues report on *Mapping and Sequencing the Human Genome*
Development of yeast artificial chromosome (YAC) cloning
Human Genome Organization (HUGO) formed
1988
Sequence-Tagged Sites (STS) mapping concept established
Cystic fibrosis gene identified by positional cloning
1989
HGP
1990
2003
1990
Human Genome Project (HGP) launched in the U.S.
Ethical, Legal, and Social Implications (ELSI) programs founded at NIH and DOE
First gene for breast cancer (BRCA1) mapped
1991
First U.S. Genome Centers established
1992
Second-generation human genetic map developed
Rapid data release guidelines established by NIH and DOE
1993
New five-year plan for the HGP in the U.S. published
Sanger Centre founded (later renamed Wellcome Trust Sanger Institute)
The Wellcome Trust
1994
HGP's human genetic mapping goal achieved
1995
HGP's human physical mapping goal achieved
First bacterial genome (*H. influenzae*) sequenced
U.S. Equal Employment Opportunity Commission issues policy on genetic discrimination in the workplace
1996
First human gene map established
Pilot projects for human genome sequencing begin in U.S.
First archael genome sequenced
Yeast (*S. cerevisiae*) genome sequenced
HGP's mouse genetic mapping goal achieved
Bermuda principles for rapid and open data release established
1997
DOE forms Joint Genome Institute
JGI
NCHGR becomes NHGRI
National Human Genome Research Institute
E. coli genome sequenced
Genoscope (French National Genome Sequencing Center) founded
1998
Incorporation of 30,000 genes into human genome map
New five-year plan for the HGP in the U.S. published
RIKEN Genomic Sciences Center (Japan) established
Roundworm (*C. elegans*) genome sequenced
SNP initiative begins
GTGCT
GTCCT
Chinese National Human Genome Centers (in Beijing and Shanghai) established
1999
Full-scale human sequencing begins
Sequence of first human chromosome (chromosome 22) completed
2000
Draft version of human genome sequence completed
President Clinton and Prime Minister Blair support free access to genome information
Fruit fly (*D. melanogaster*) genome sequenced
Mustard cress (*A. thaliana*) genome sequenced
Executive order bans genetic discrimination in U.S. federal workplace
2001
Draft version of human genome sequence published
10,000 full-length human cDNAs sequenced
Mammalian Gene Collection
2002
Draft version of mouse genome sequence completed and published
Draft version of rat genome sequence completed
Draft version of rice genome sequence completed and published
2003
Finished version of human genome sequence completed
HGP ends with all goals achieved
to be continued..

Human Genome Project



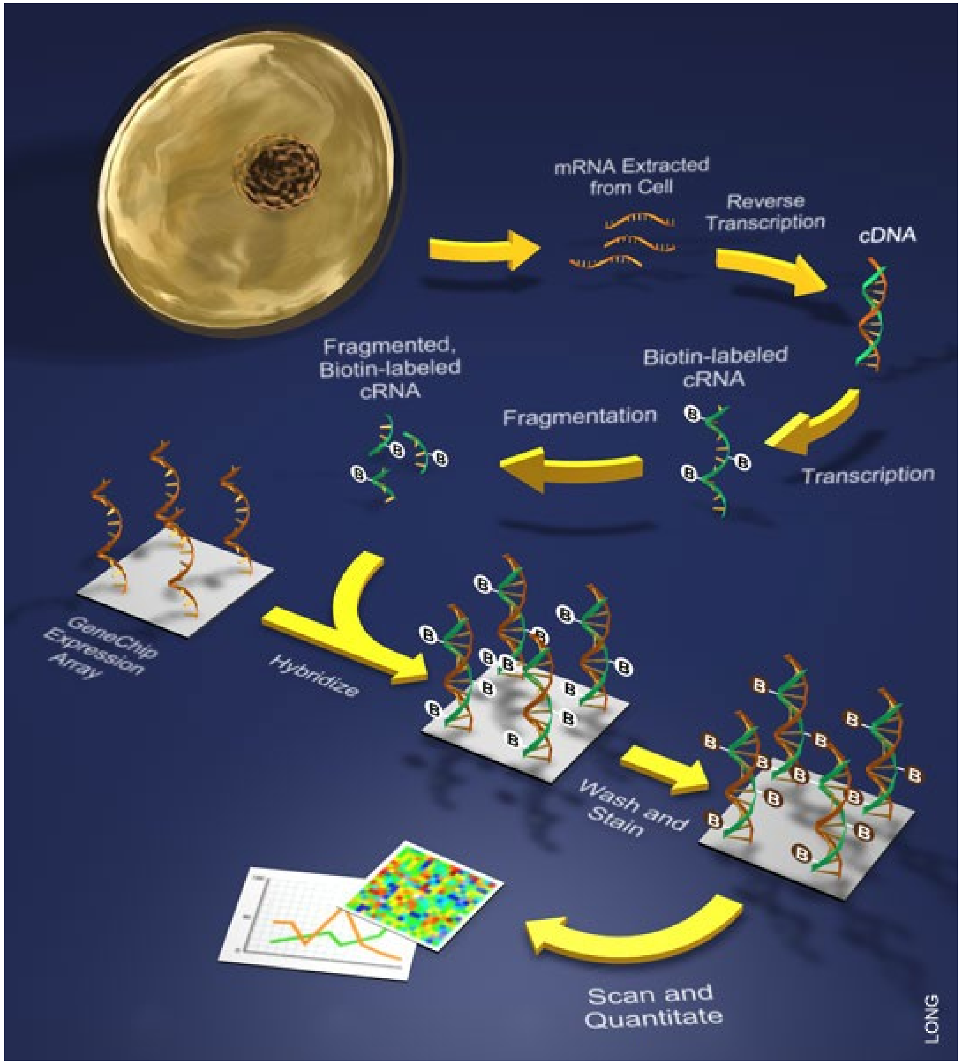
1990 15-year project 3B USD
20 groups

GeneChip

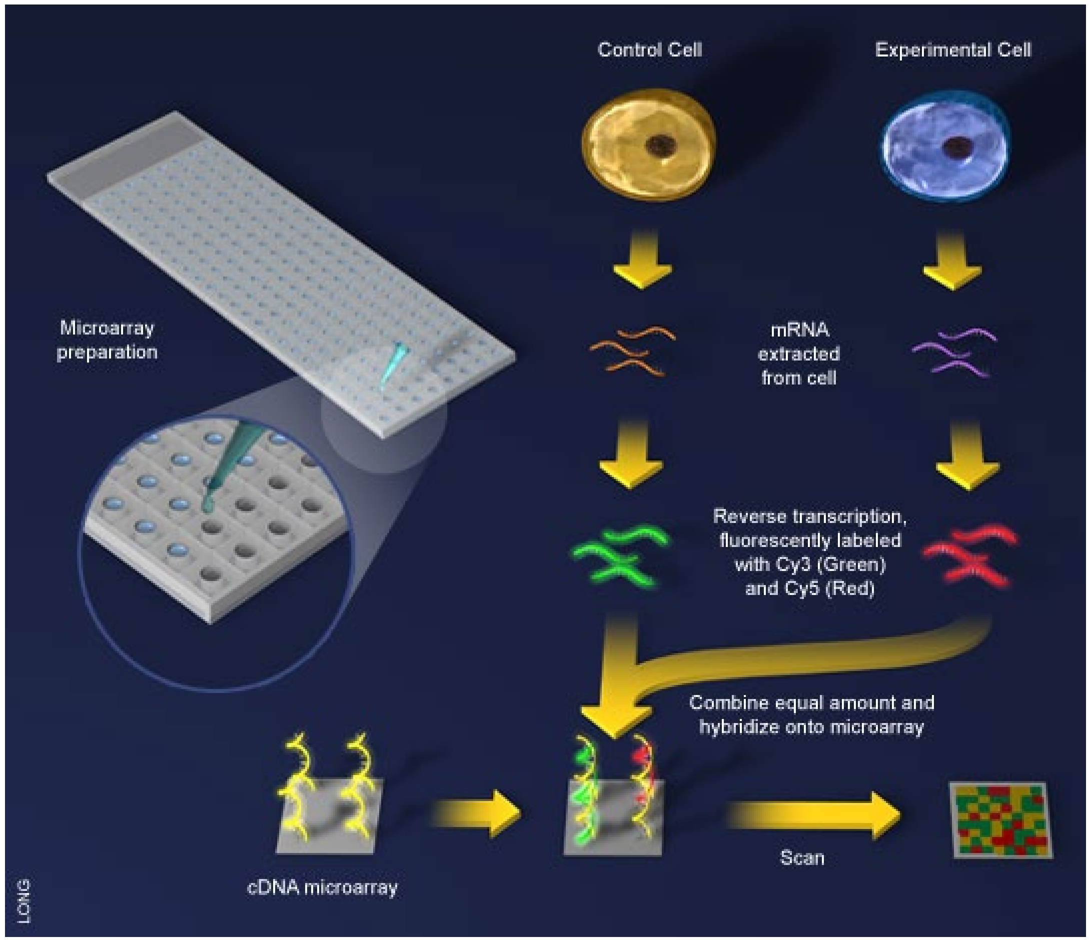


Light
(deprotection)
Mask
Wafer
T –
– C
25-mer
G A T C G
C A T A T
A G C T G
T T C C G
Repeat
GeneChip®
Microarray
OH OH O O O
T T OHOHO
T T C C O

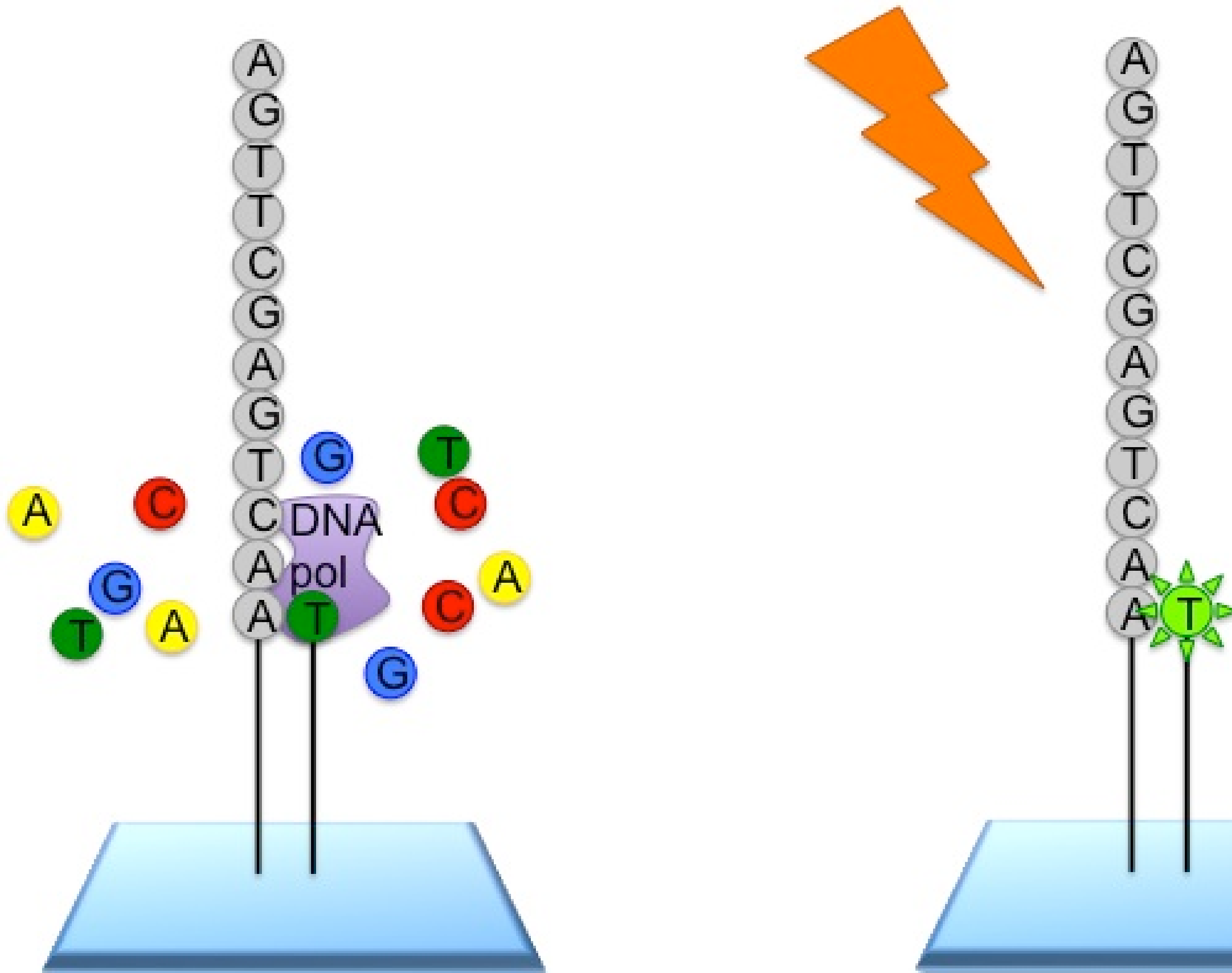
Scheme


mRNA Extracted from Cell
Reverse Transcription
cDNA
Transcription
Biotin-labeled cRNA
Fragmentation
Fragmented, Biotin-labeled cRNA
GeneChip Expression Array
Hybridize
Wash and Stain
Scan and Quantitate
LONG

cDNA Microarray



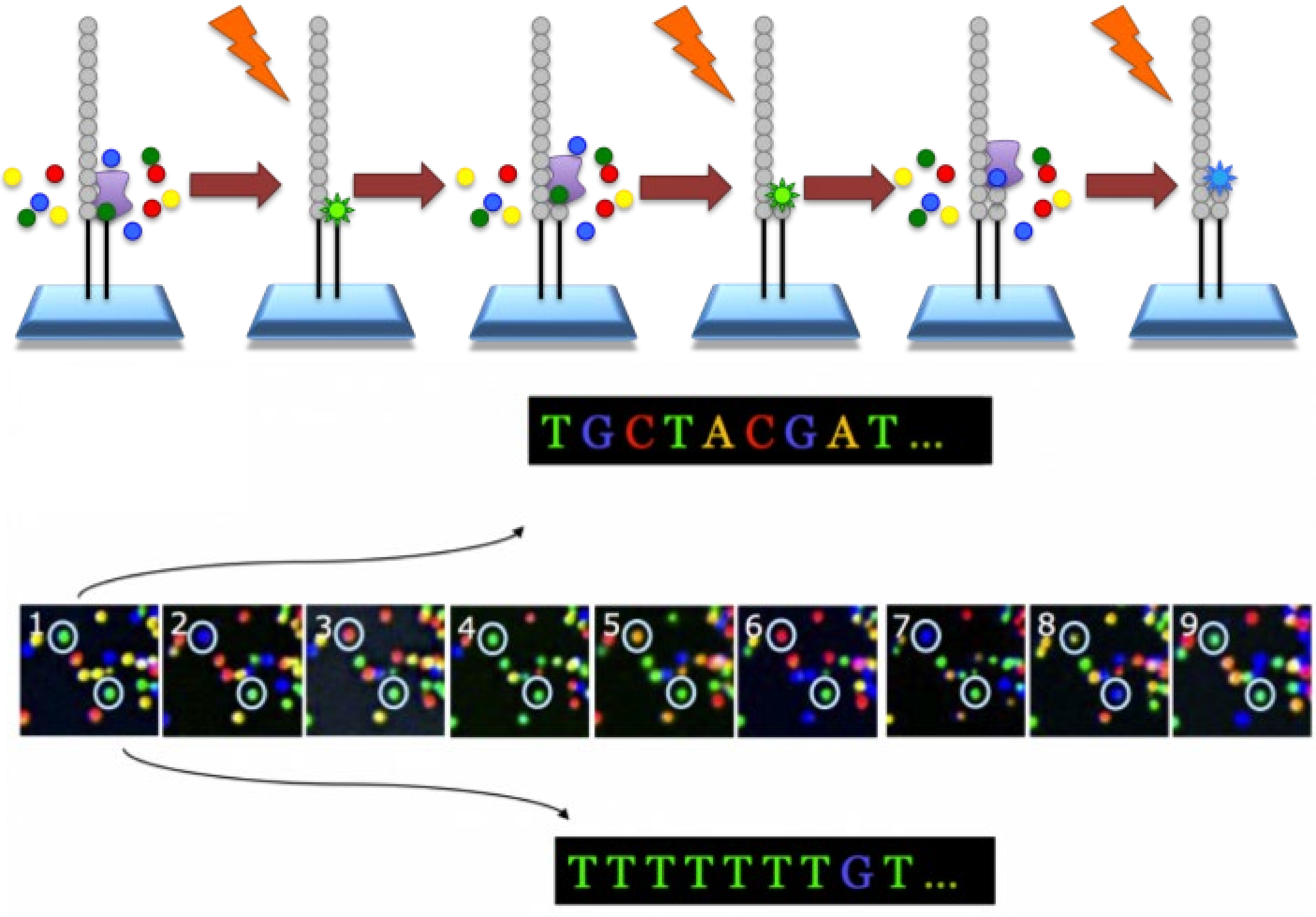
NGS Illumina



100-150 bp

NGS Illumina

https://www.youtube.com/watch?v=fCd6B5HRaZ8



The current version of the human genome reference assembly, GRCh38.p14 (GRCh38), has millions of bases represented by the letter "N," which means that the actual base residing at that location is unknown.

There are also 169 sequences that cannot confidently be ordered or oriented within the assembly, typically owing to their repetitive nature

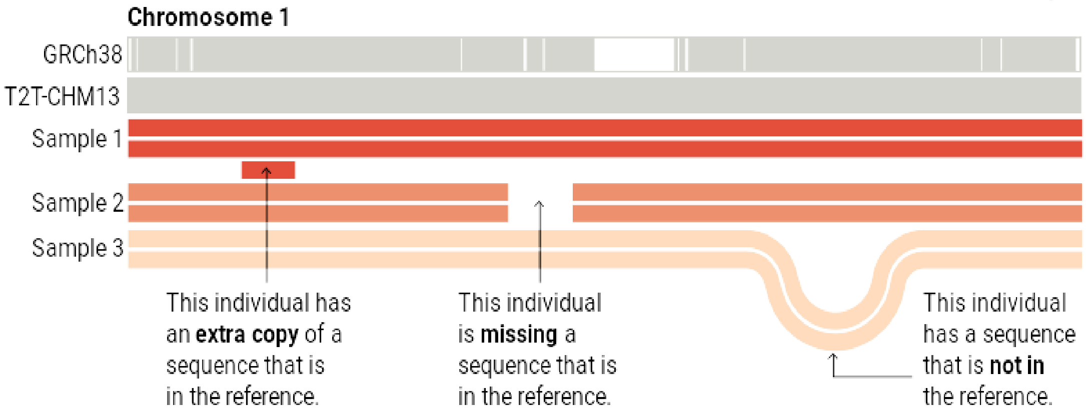
Until recently, limitations of sequencing technology, primarily that the sequencers could read no more than about 1000 bases at a time,

The HGP opted for a more structured approach. This involved cloning genomic DNA into pieces that could be grown in bacteria (clones) and indexed in 96-well plates. Clones from these libraries were first mapped to chromosome

An important attribute of the human reference assembly is that the source DNA was derived from multiple individuals.

A more complete reference

The new human genome assembly, T2T-CHM13 from the Telomere-to-Telomere Consortium, includes complex and repetitive regions of chromosomes that had not been included in previous versions of the human genome assembly (GRCh38). Although the Y chromosome remains to be completed, this new reference could be annotated with regulatory regions, variants, and sequence diversity to give a fuller picture of human genomic variation.



when two clones from different haplotypes of an individual are adjacent in the reference assembly, this can create sequence representations that are not normally found in the population

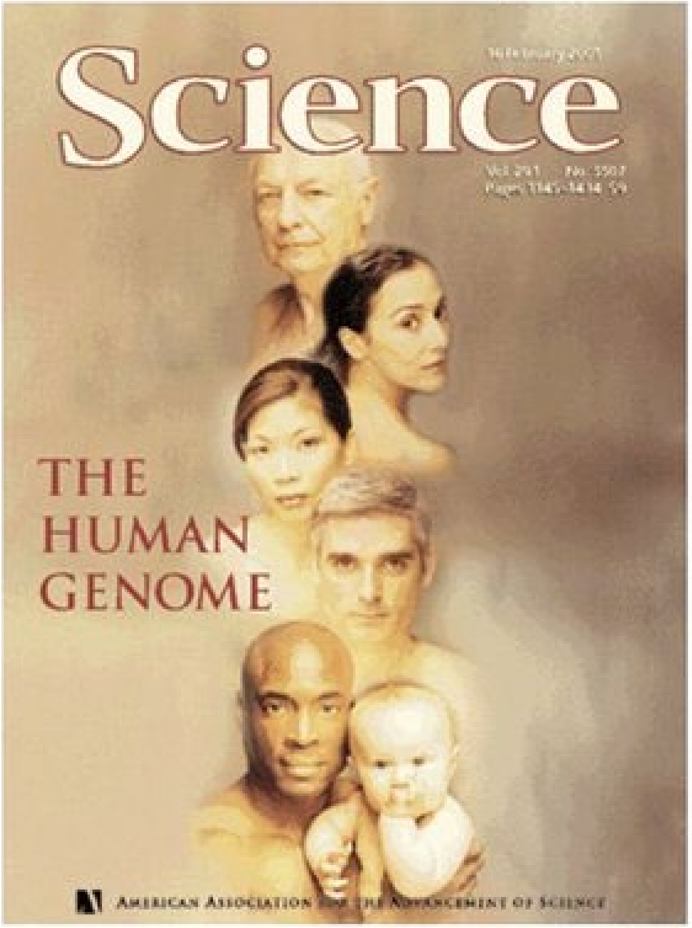
T2T Consortium, eliminated the problem of allelic diversity by sequencing the genome of a cell line derived from a complete hydatidiform mole (CHM).

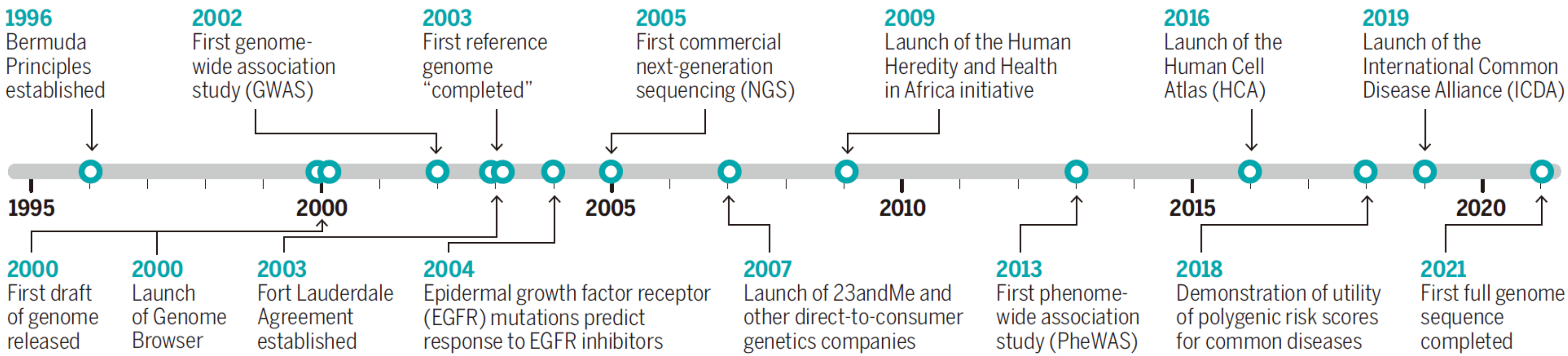
This is duplicated so that the cell contains two copies of the same parental genome

A next-generation human genome sequence, Volume: 376, Issue: 6588, Pages: 34-35, DOI: (10.1126/science.abo5367)

Genome sequencing

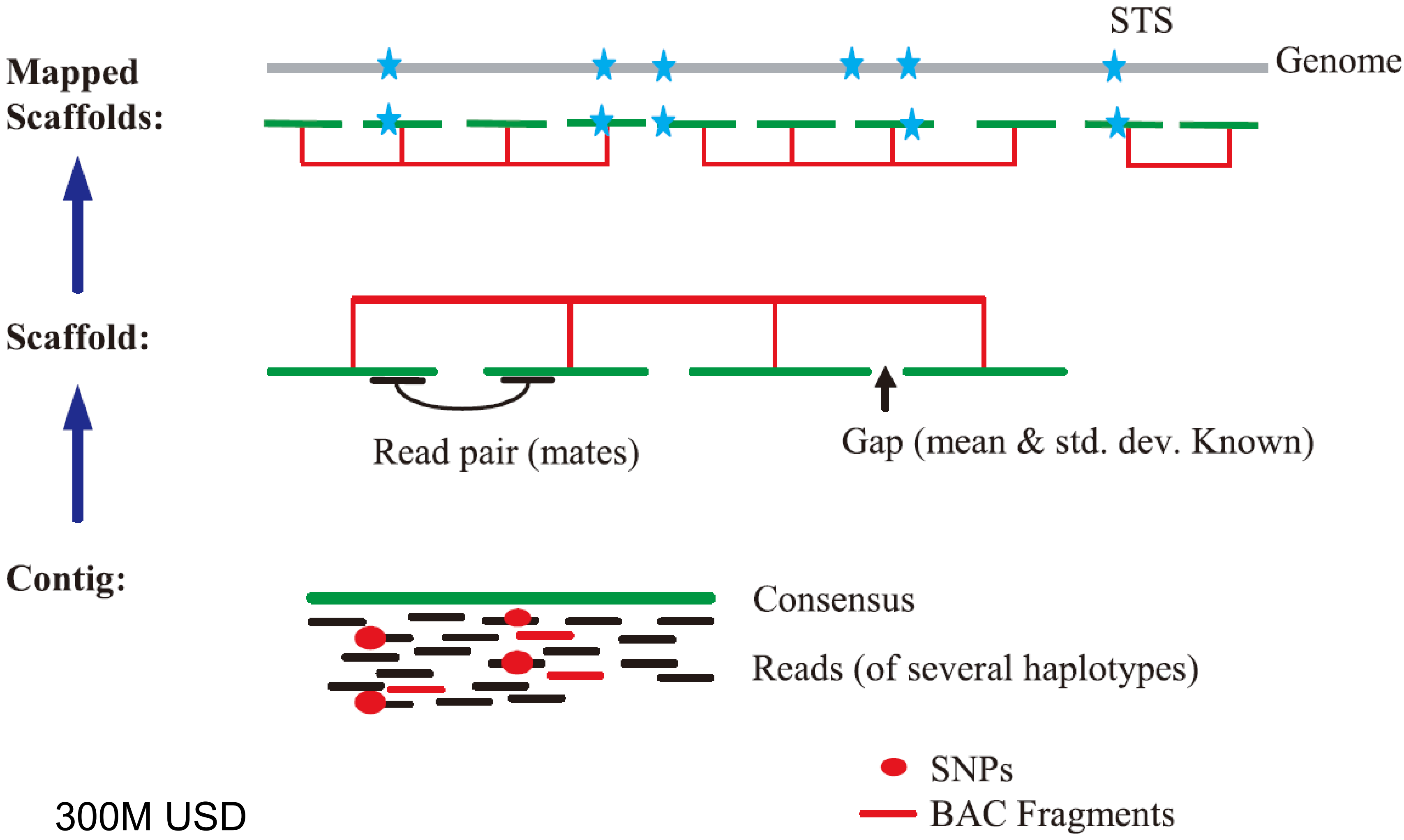
February 2001 - Publication of the first draft of the human genome





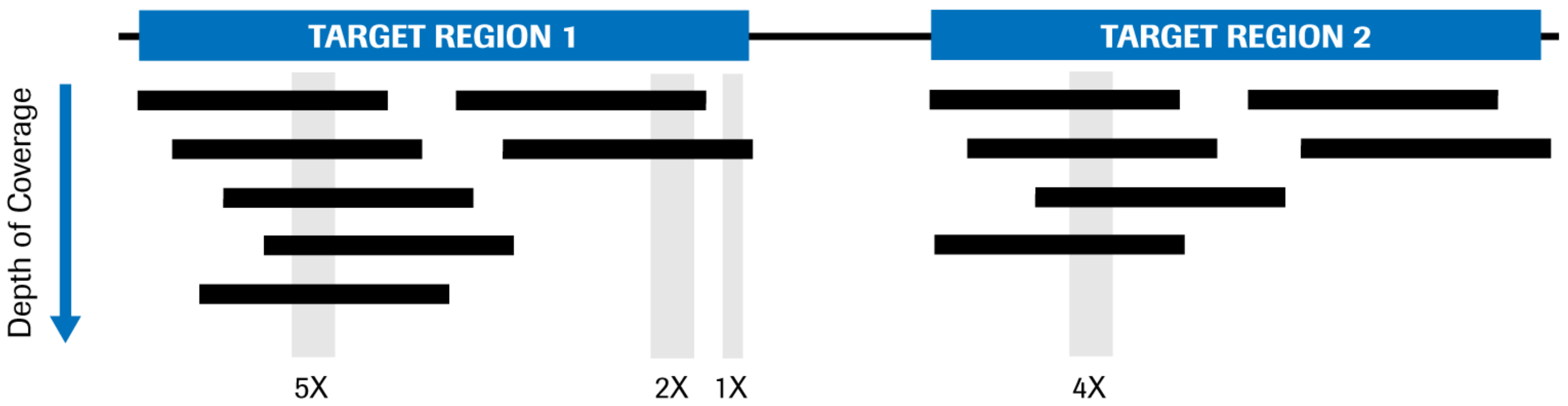
Bermuda Principles

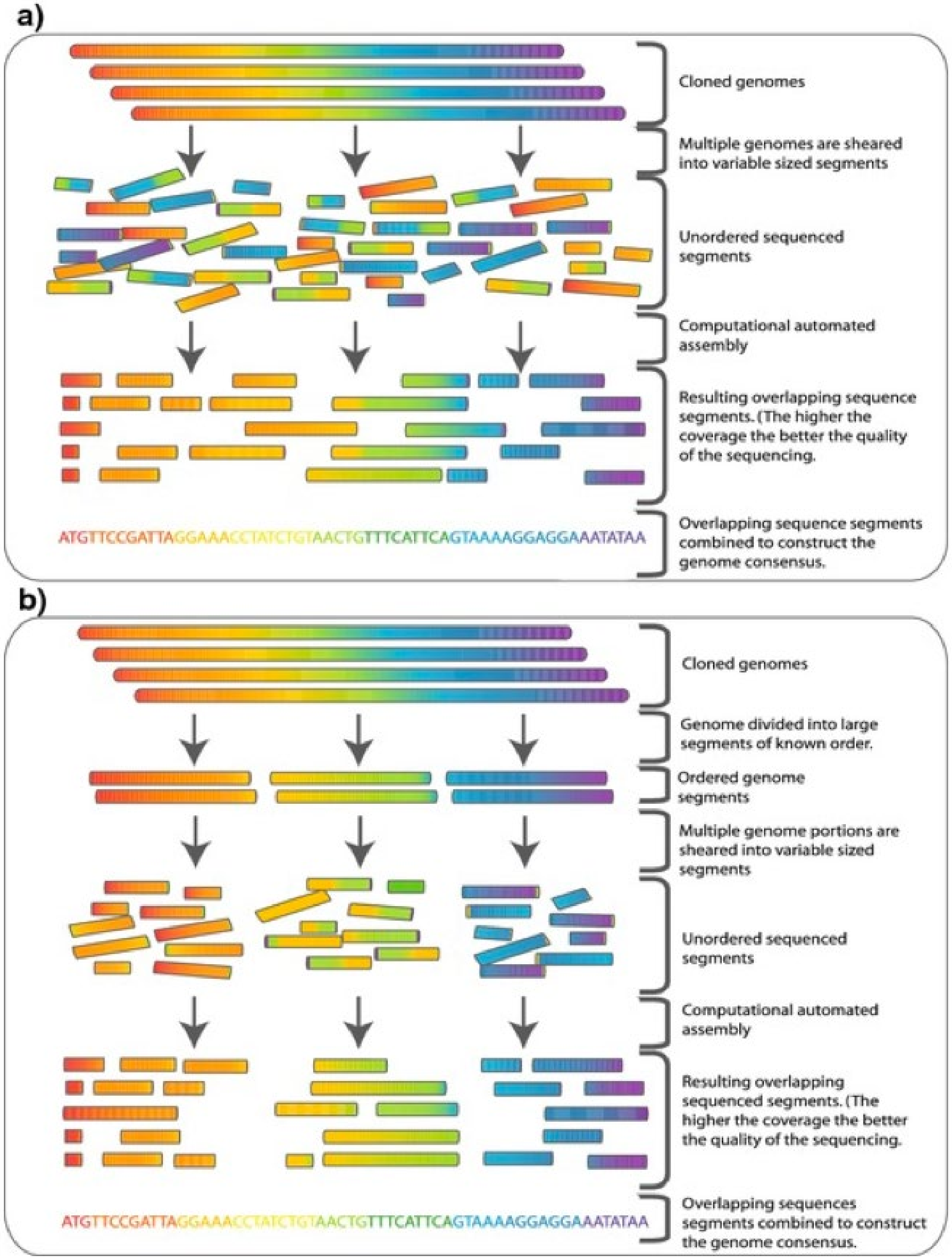
•Automatic release of sequence assemblies larger than 1 kb (preferably within 24 hours).
•Immediate publication of finished annotated sequences.
•Aim to make the entire sequence freely available in the public domain for both research and development in order to maximise benefits to society.



300M USD
2.9 bbp
9 months
5 donors
5.1 folds
Whole-genome shotgun

coverage depth




a)
Cloned genomes
Multiple genomes are sheared into variable sized segments
Unordered sequenced segments
Computational automated assembly
Resulting overlapping sequence segments. (The higher the coverage the better the quality of the sequencing.
ATGTTCCGATTAGGAAACCTATCTGTAACTGTTTCATTCAGTAAAAGGAGGAAATATAA
Overlapping sequence segments combined to construct the genome consensus.
b)
Cloned genomes
Genome divided into large segments of known order.
Ordered genome segments
Multiple genome portions are sheared into variable sized segments
Unordered sequenced segments
Computational automated assembly
Resulting overlapping sequenced segments. (The higher the coverage the better the quality of the sequencing.
ATGTTCCGATTAGGAAACCTATCTGTAACTGTTTCATTCAGTAAAAGGAGGAAATATAA
Overlapping sequences segments combined to construct the genome consensus.