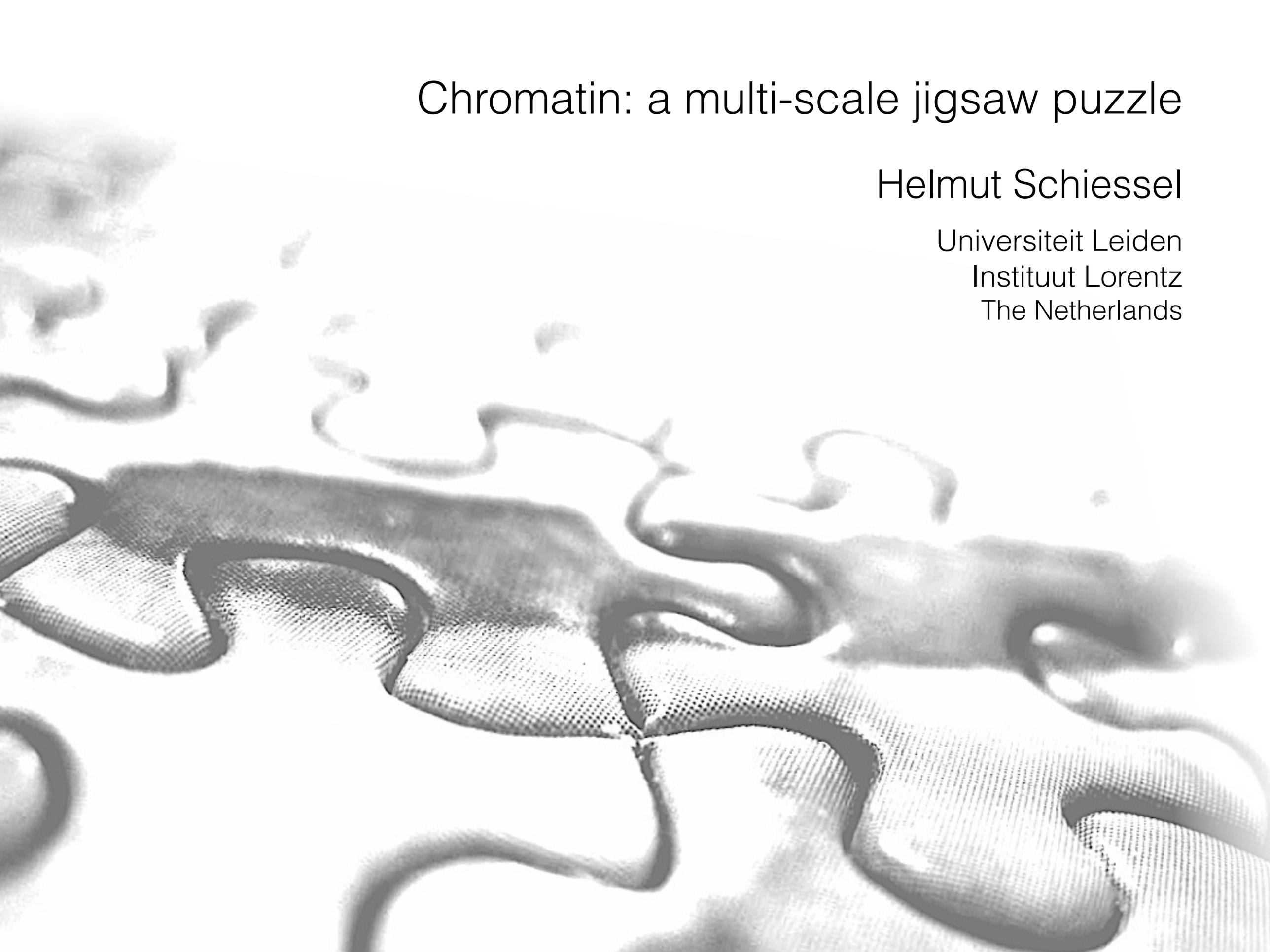
Chromatin: a multi-scale jigsaw puzzle

Helmut Schiessel

Universiteit Leiden
Instituut Lorentz
The Netherlands

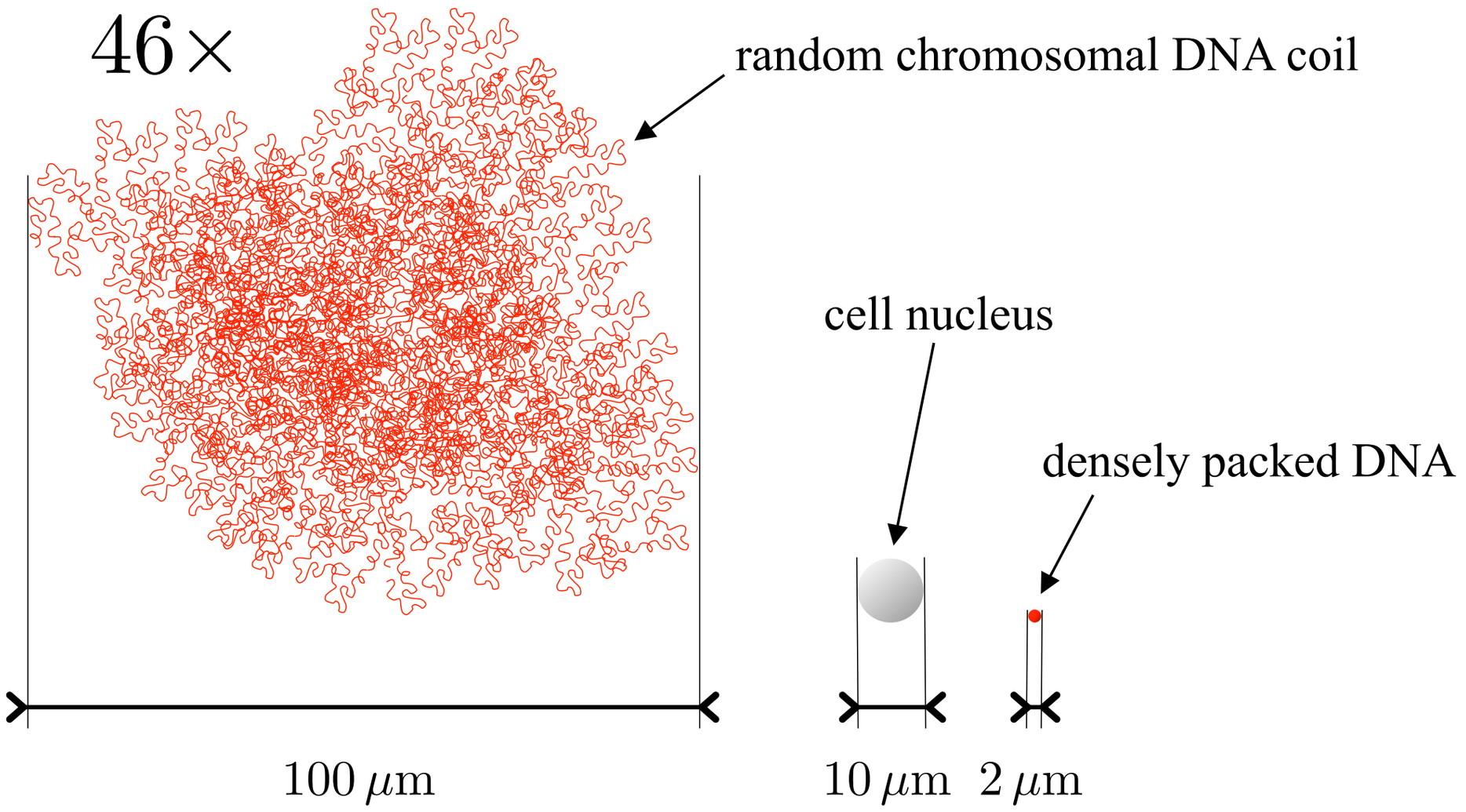


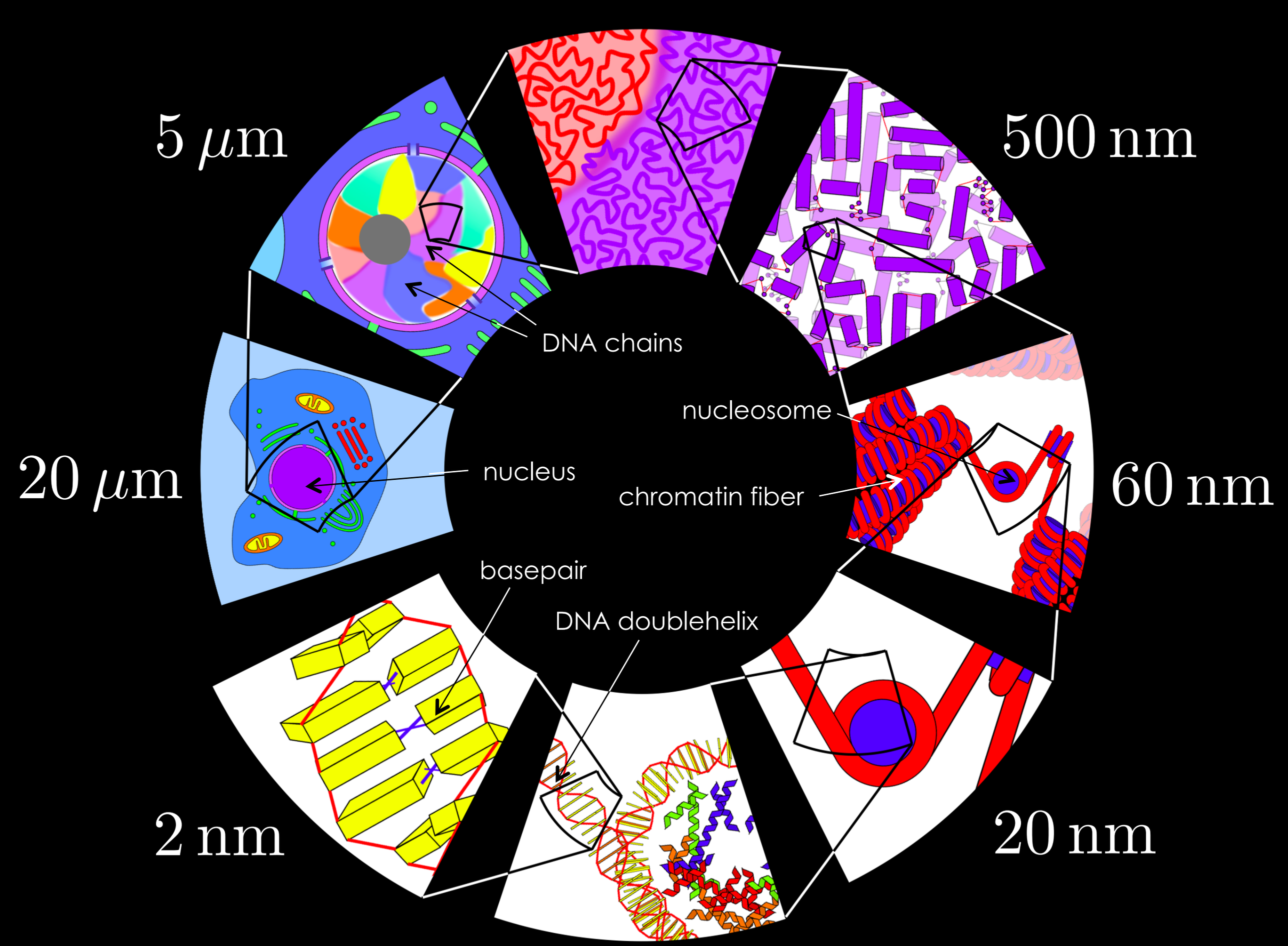
Numbers

human DNA per cell: $2 \times$ human genome (2.9×10^9 bp)

= 2 m DNA in total

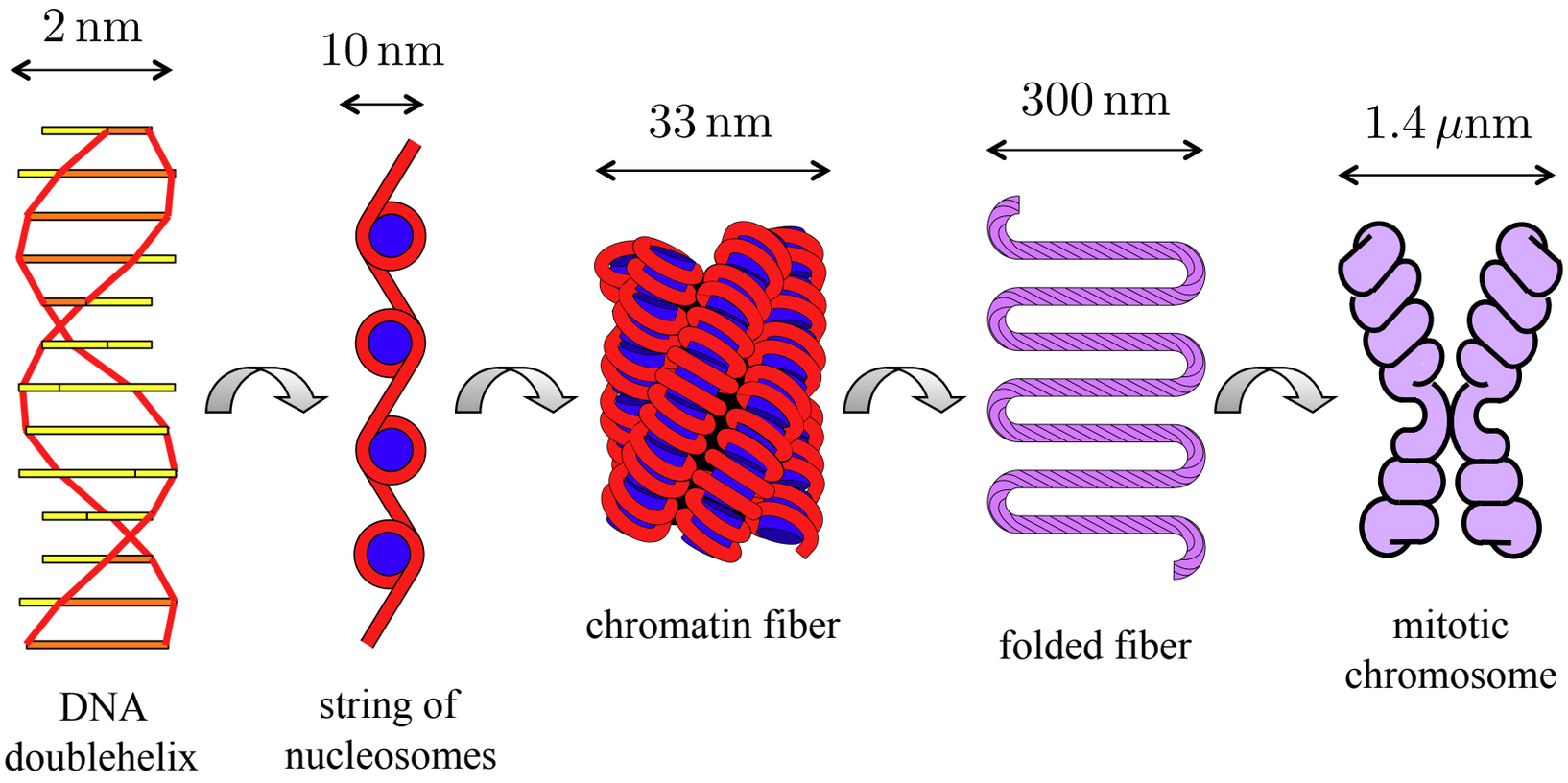
= 46 chromosomes of length ≈ 4 cm



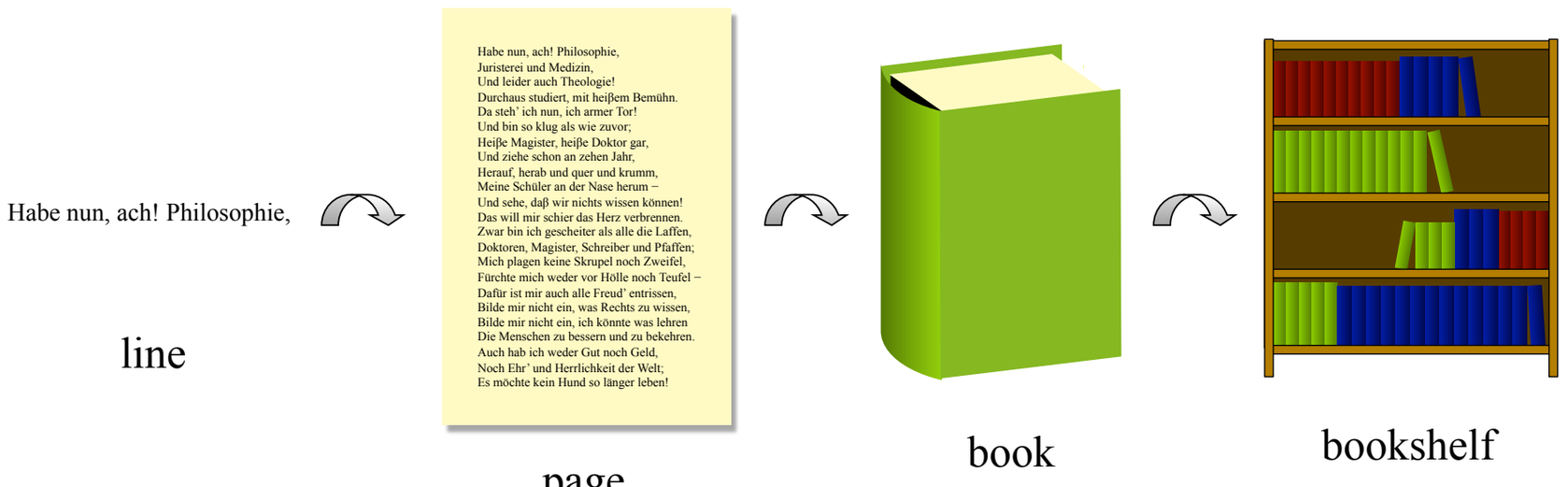

5 μm
500 nm
DNA chains
nucleosome
20 μm
nucleus
chromatin fiber
60 nm
basepair
DNA doublehelix
2 nm
20 nm

Hierarchical folding

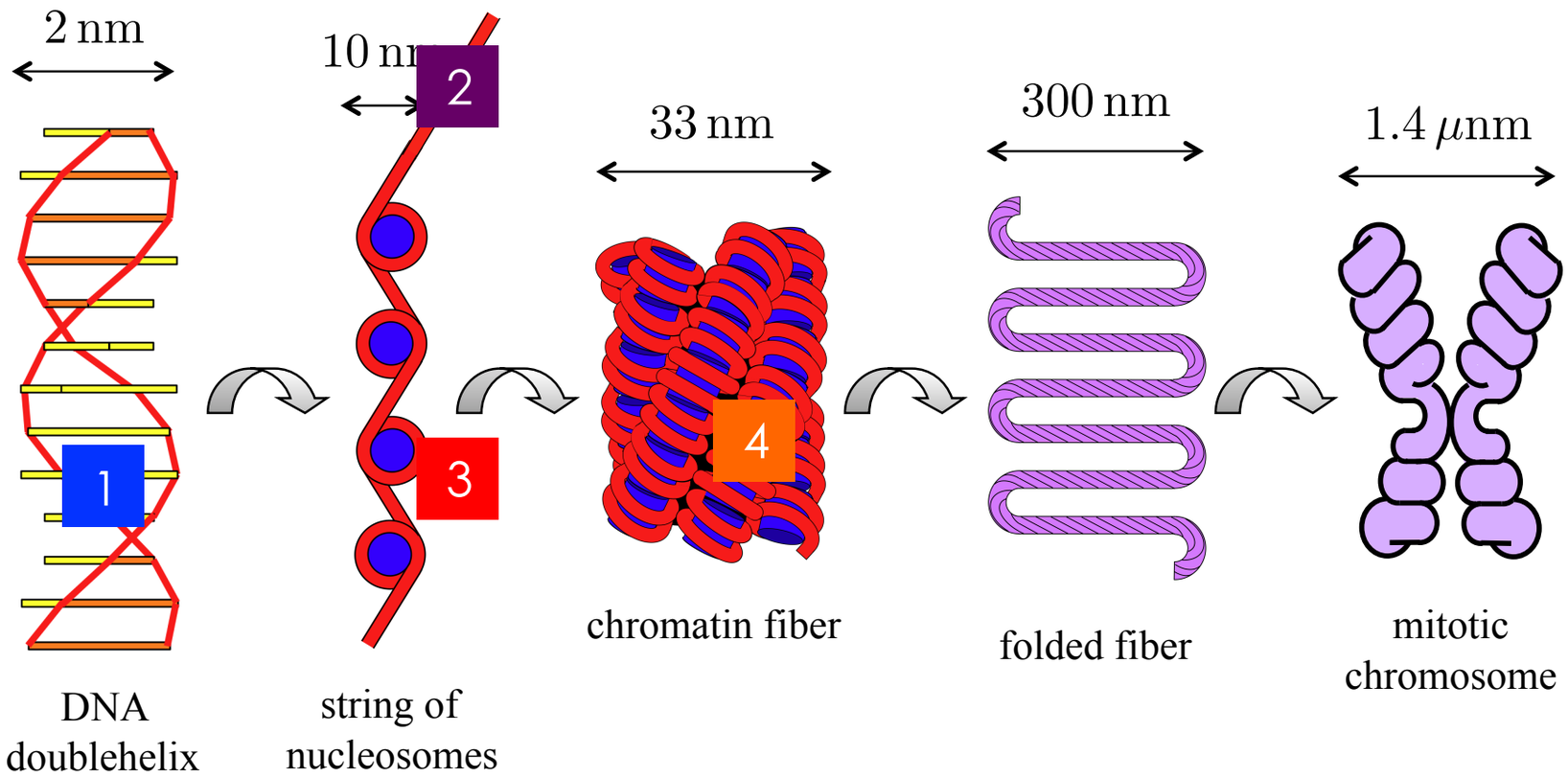
chromatin:



library:



Overview



1 The mechanical genome

2 DNA as a wormlike chain

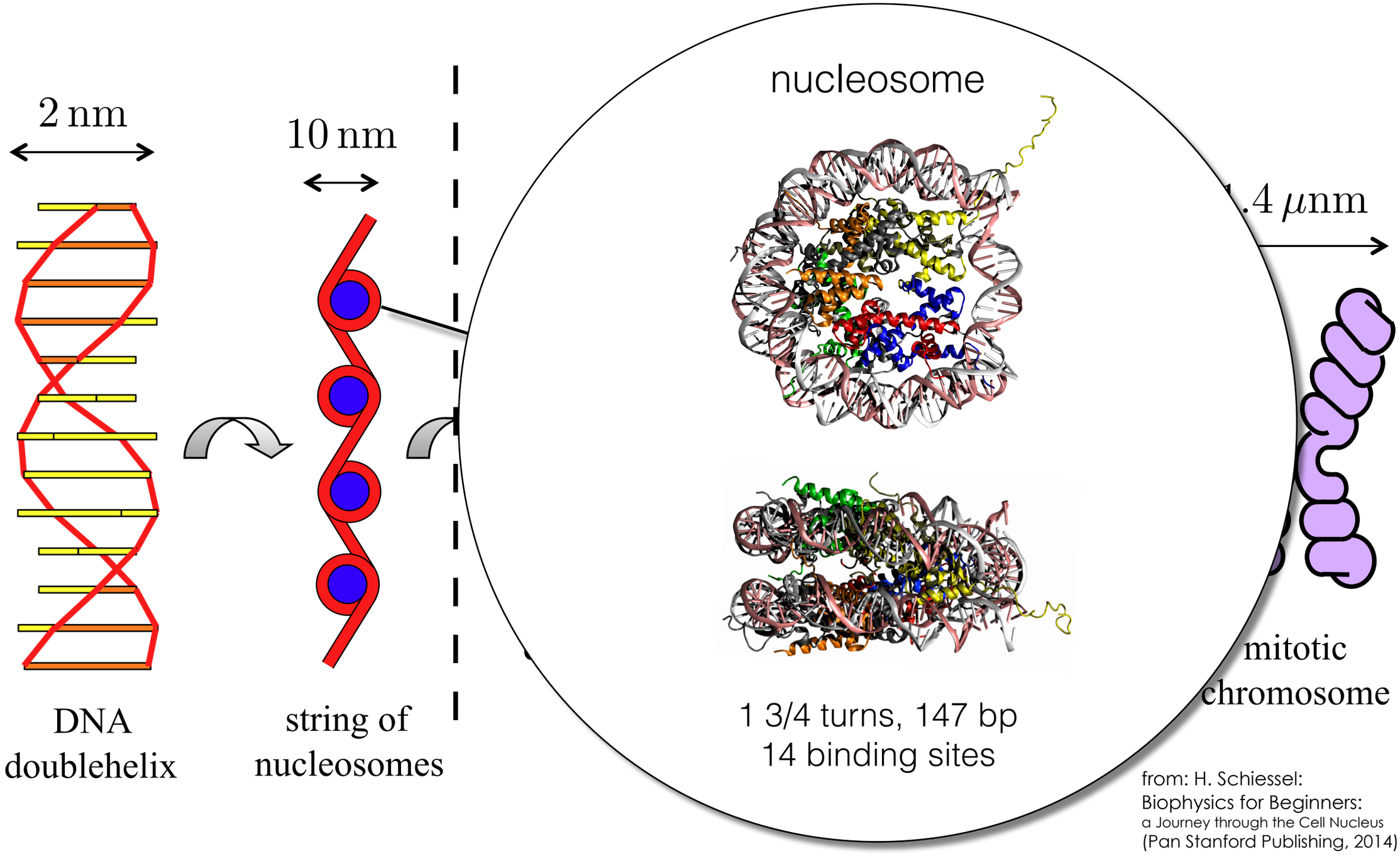
3 Nucleosome unspooling

4 Packing nucleosomes

THE MECHANICAL GENOME

Artwork: Amar van Leewaarde & Marco Tompitak

The chromatin complex



from: H. Schiessel:
Biophysics for Beginners:
a Journey through the Cell Nucleus
(Pan Stanford Publishing, 2014)

nature Vol 442|17 August 2006|doi:10.1038/nature04979

ARTICLES

A genomic code for nucleosome positioning

Eran Segal[1], Yvonne Fondufe-Mittendorf[2], Lingyi Chen[2], AnnChristine Thåström[2], Yair Field[1], Irene K. Moore[2], Ji-Ping Z. Wang[3] & Jonathan Widom[2]

Jonathan Widom

1955-2011

"Genomes care where nucleosomes are on average and so genomes encode explicit information to bias [their positions]."

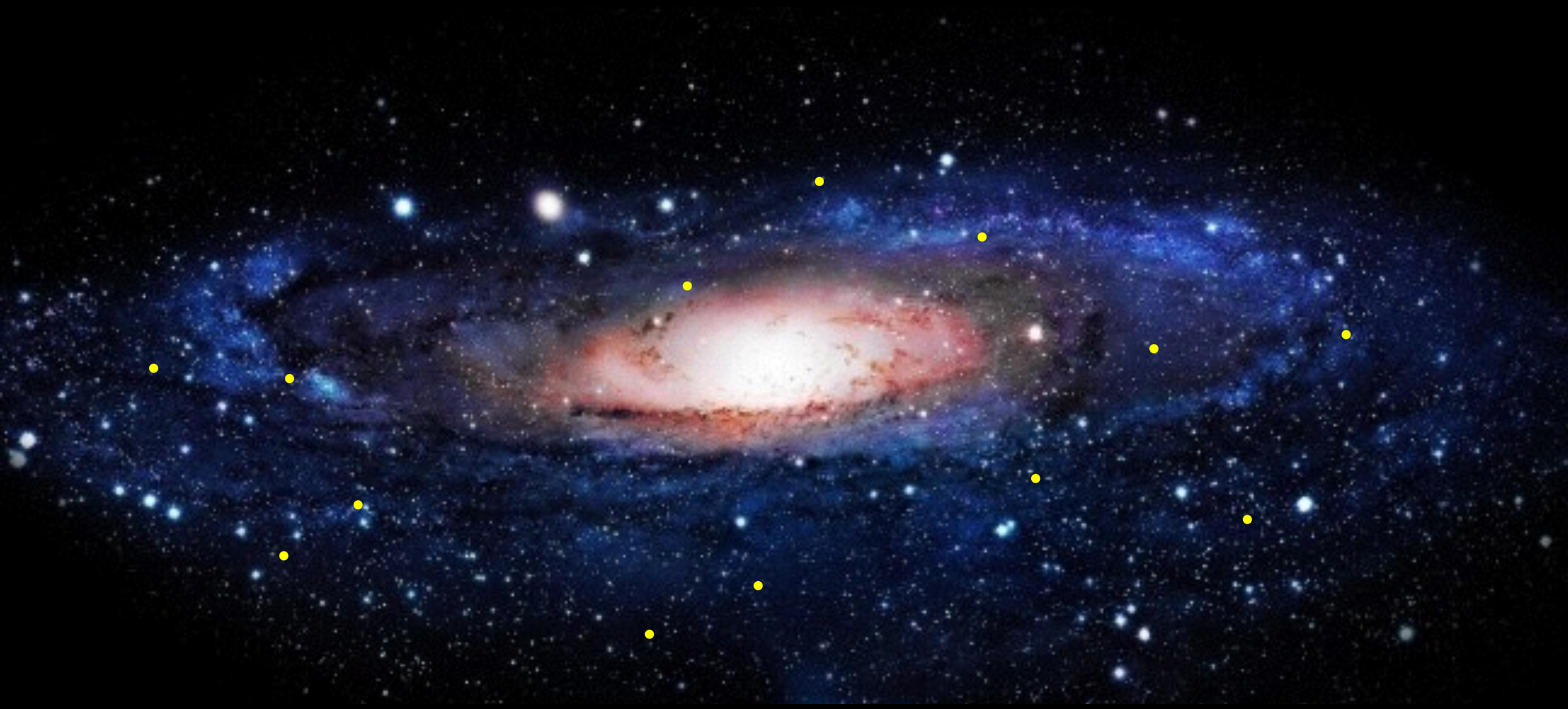
The sequence space

How many sequences can be wrapped around a nucleosome?

$$4^{147}$$

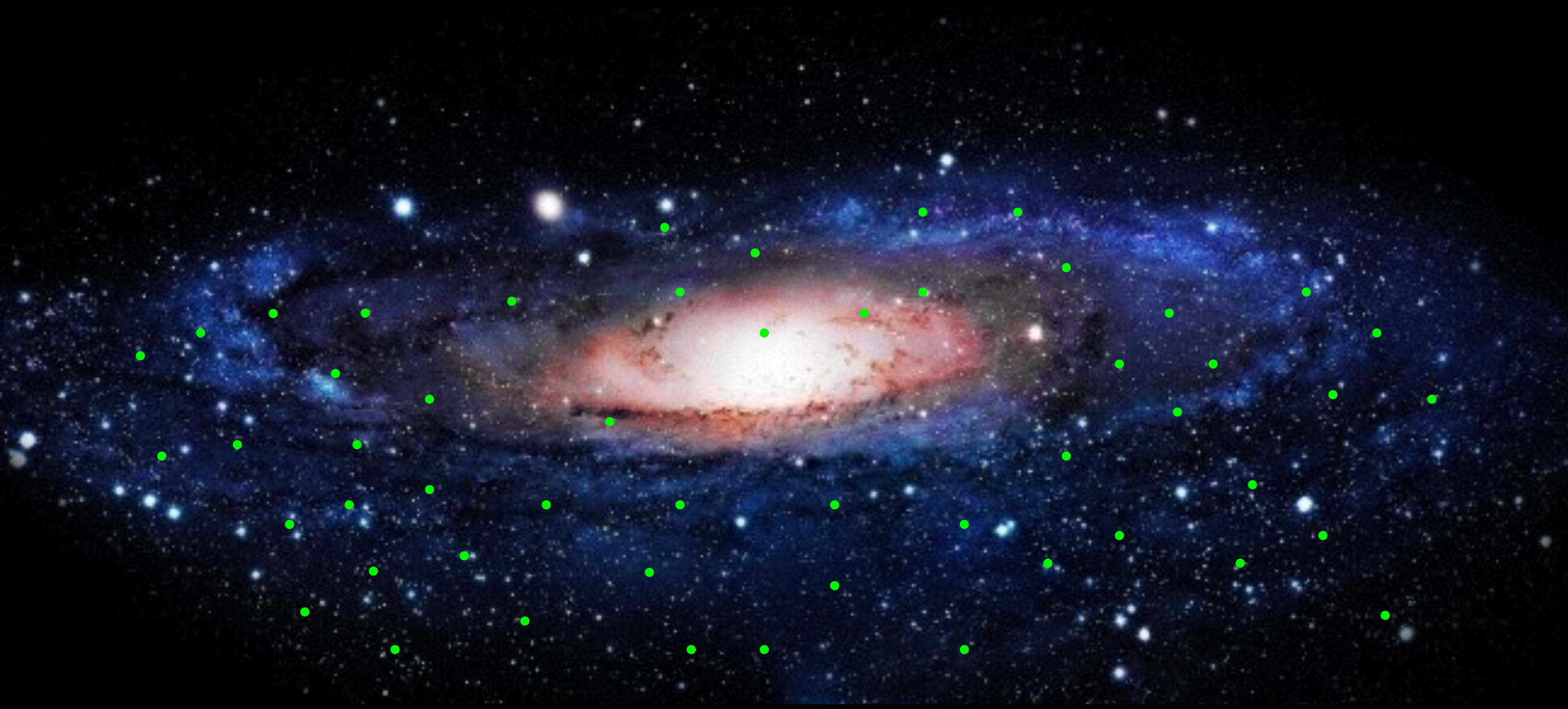
(about 5 times the volume of the Milky Way)

The sequence space



Baker’s Yeast 12 Mb

The sequence space



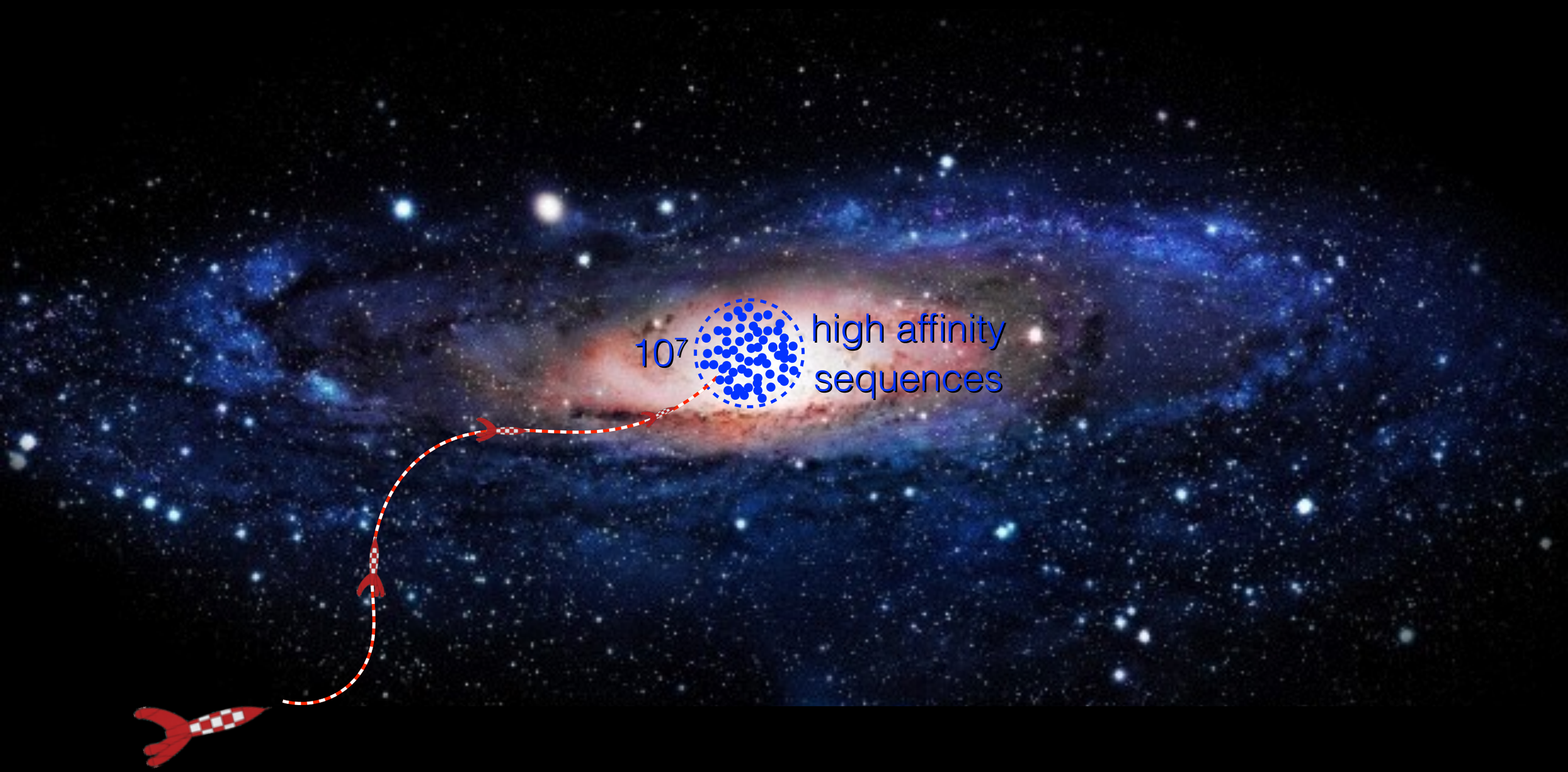
Human 3.2 Gb

The sequence space

Random Pool 5 Tb

Lowary & Widom 1998

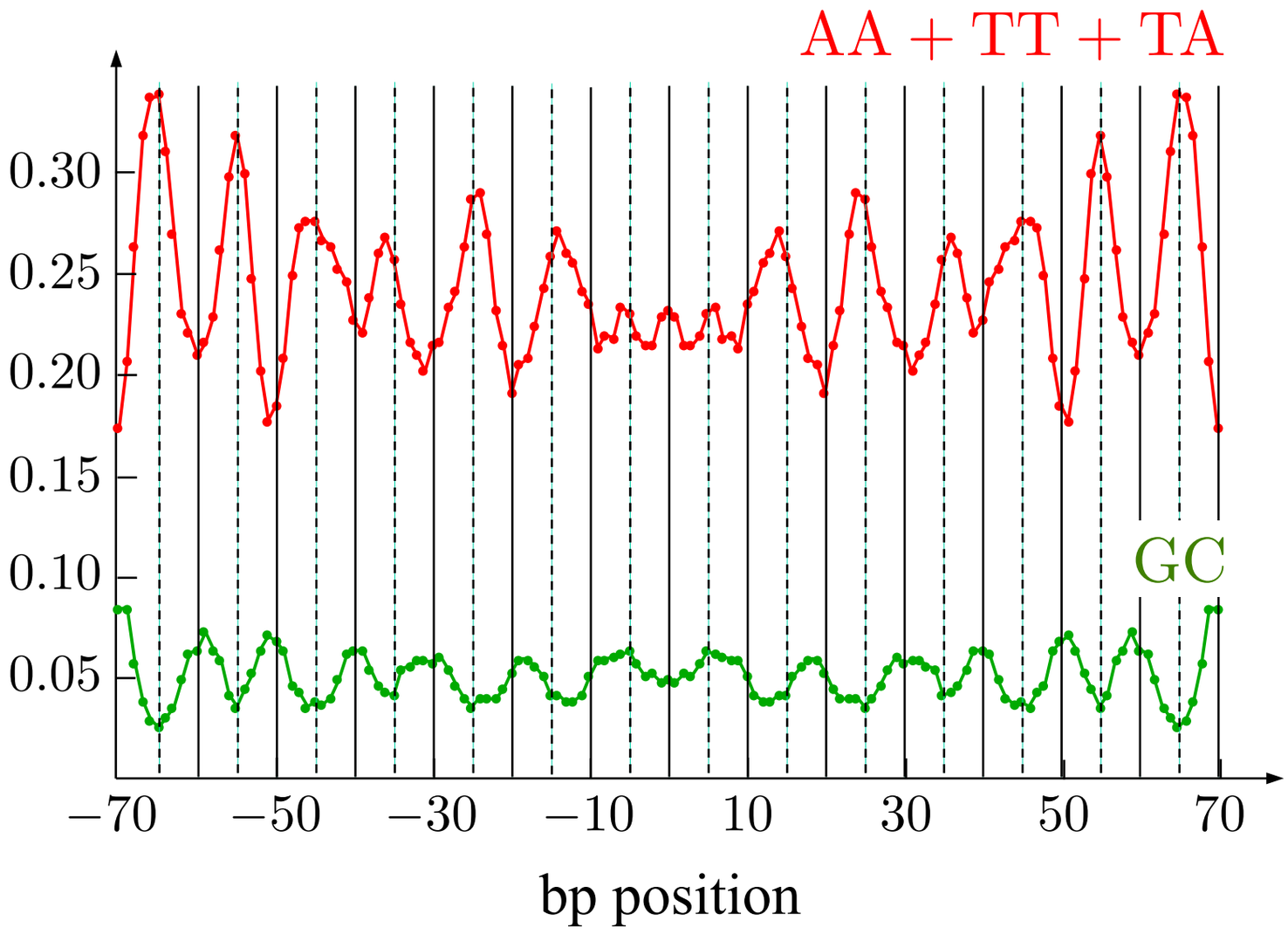
The sequence space

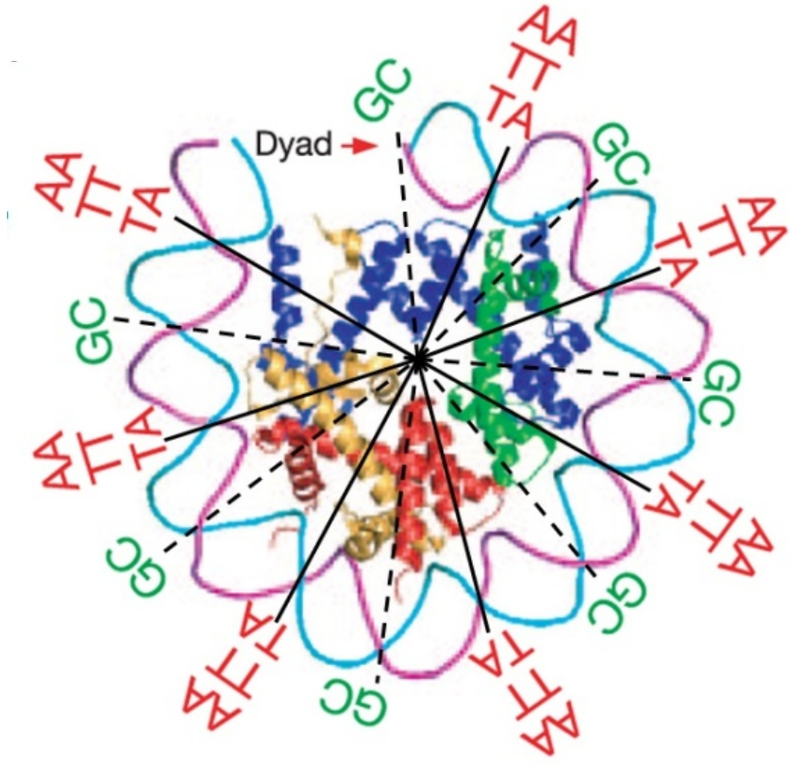


Mutation Monte Carlo method

The genomic code for nucleosome positioning

Satchwell, Drew & Travers, J. Mol. Biol. **191** (1986) 659





Segal et al., Nature, 2006

THE MECHANICAL GENOME

Our questions:

- Can the positioning rules be explained by a purely mechanical model?
- Can mechanical information be multiplexed with classical genetic information?

see also Seok-Cheol Hong's lecture from yesterday

Historical background

What is a gene ?

Somehow properties are inherited by the daughter cells...

Via DNA? No, boring molecule

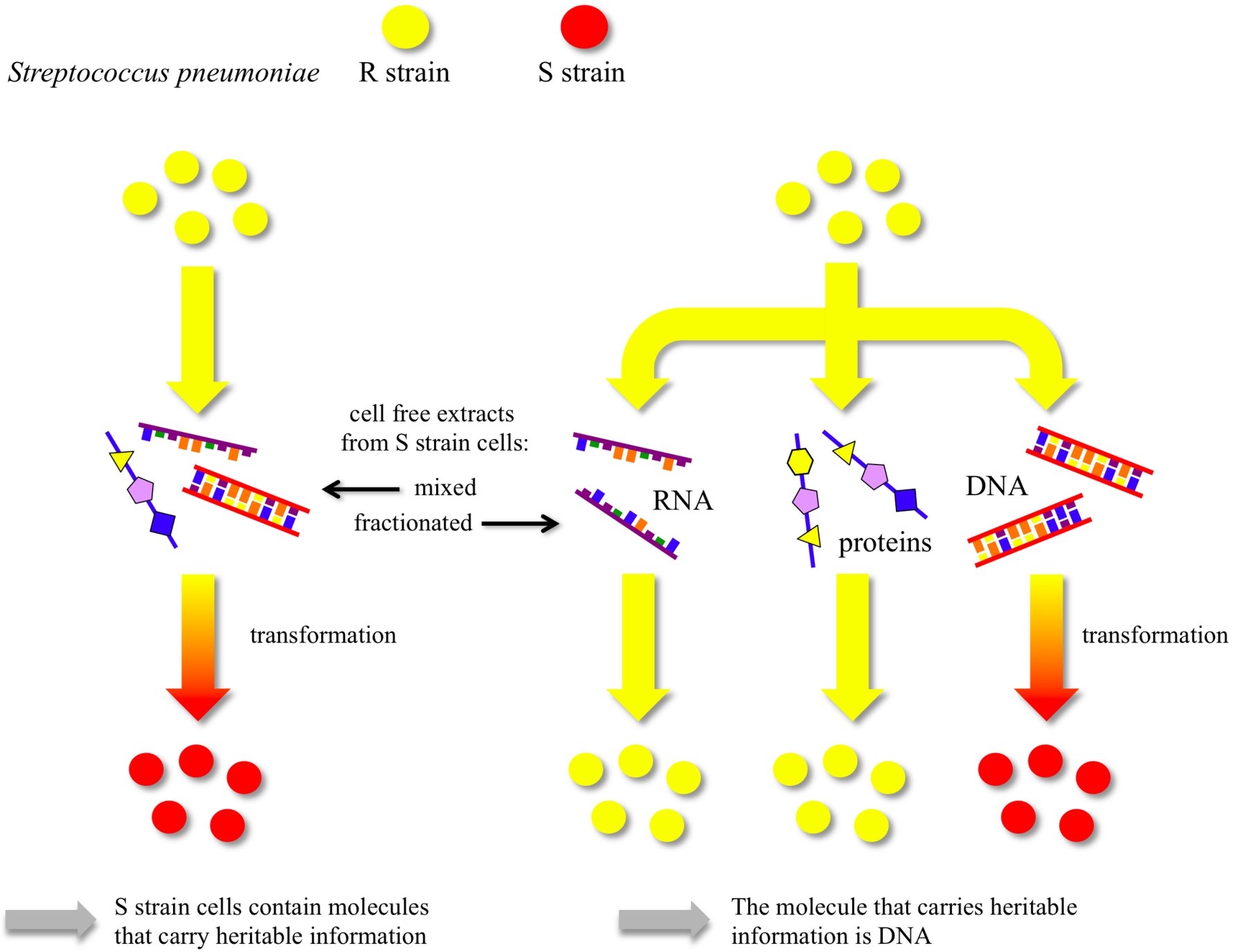
Via Proteins? Yes! Very rich class of extremely diverse molecules!

Genes are special sets of protein molecules!

How to test this idea?

DNA is the carrier of the genetic information

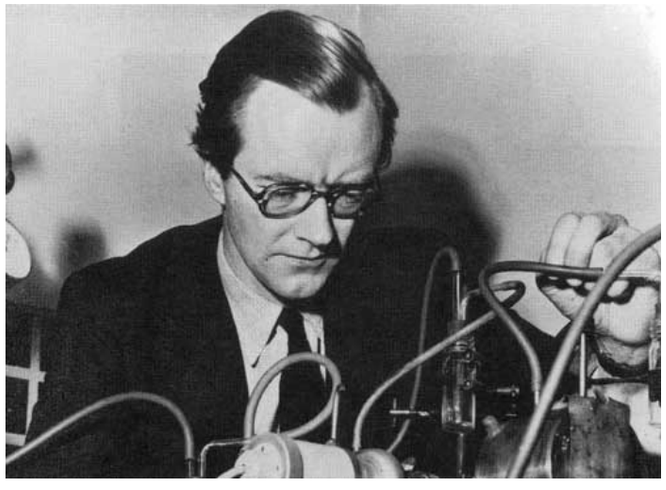
Oswald T. Avery, 1944



S strain cells contain molecules that carry heritable information

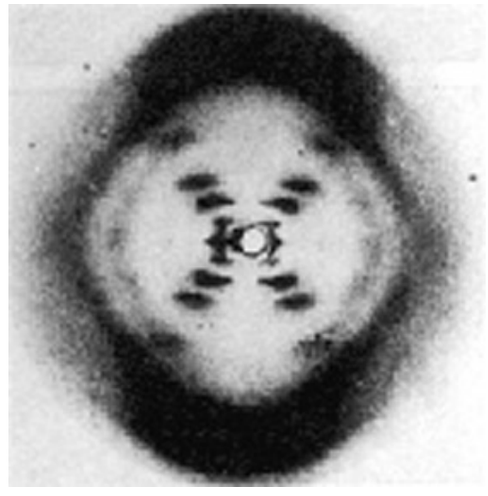
The molecule that carries heritable information is DNA

The discovery of the DNA double helix



Maurice Wilkins

Rosalind Franklin



X ray diffraction

DNA forms helix

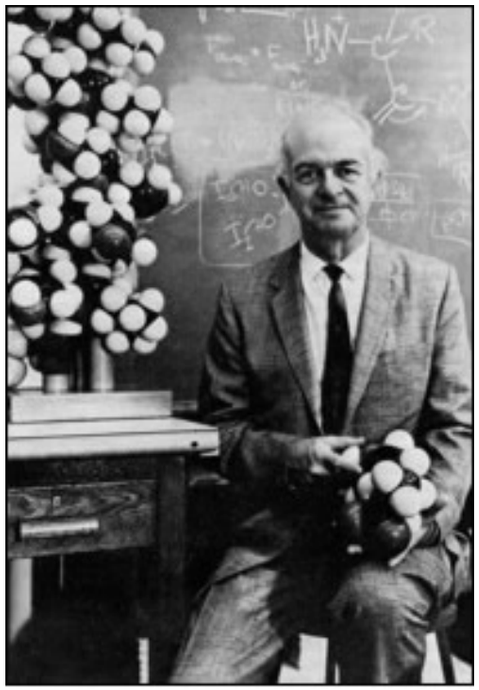
James D. Watson

Francis Crick

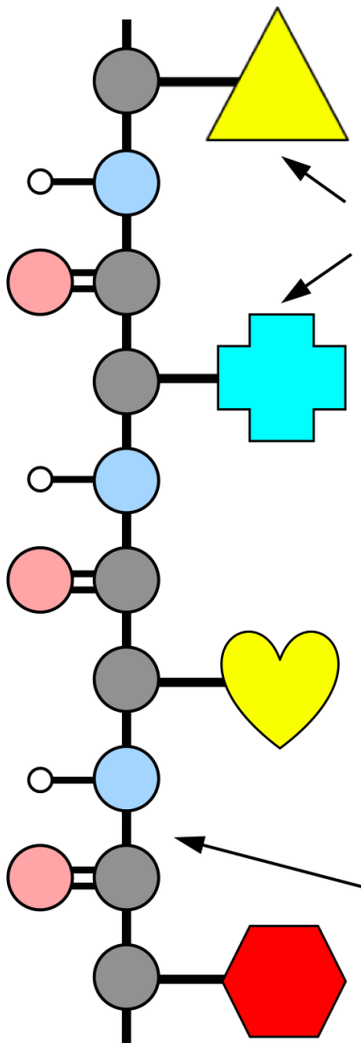
model building

The alpha helix in proteins

Pauling & Corey, 1951

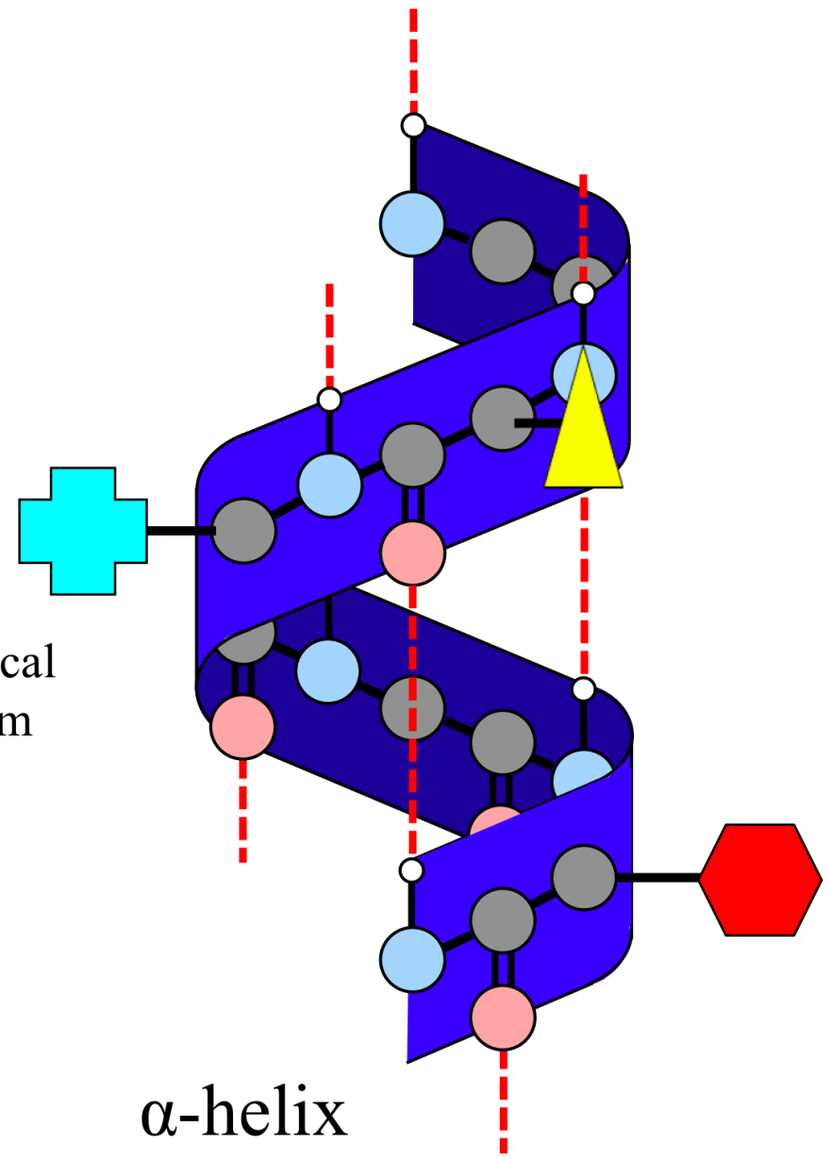


Linus Pauling



carbon
nitrogen
oxygen
hydrogen

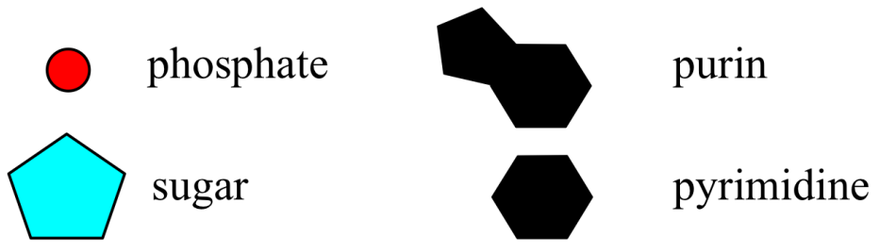
Assumption: same chemical environment for each atom of regular backbone

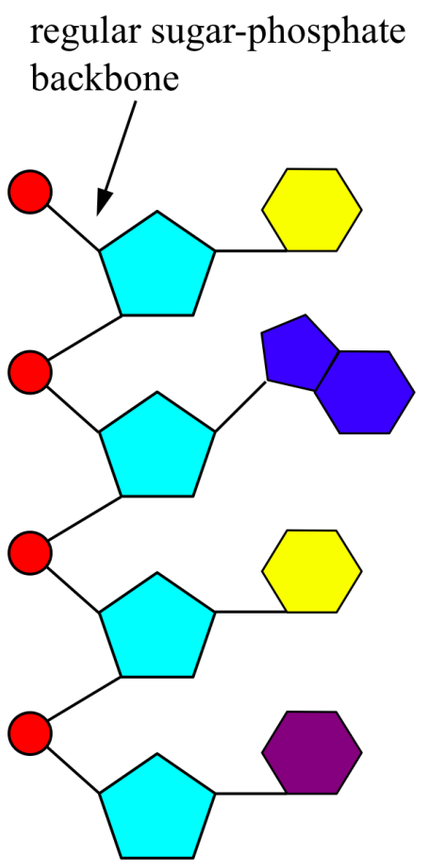


α-helix

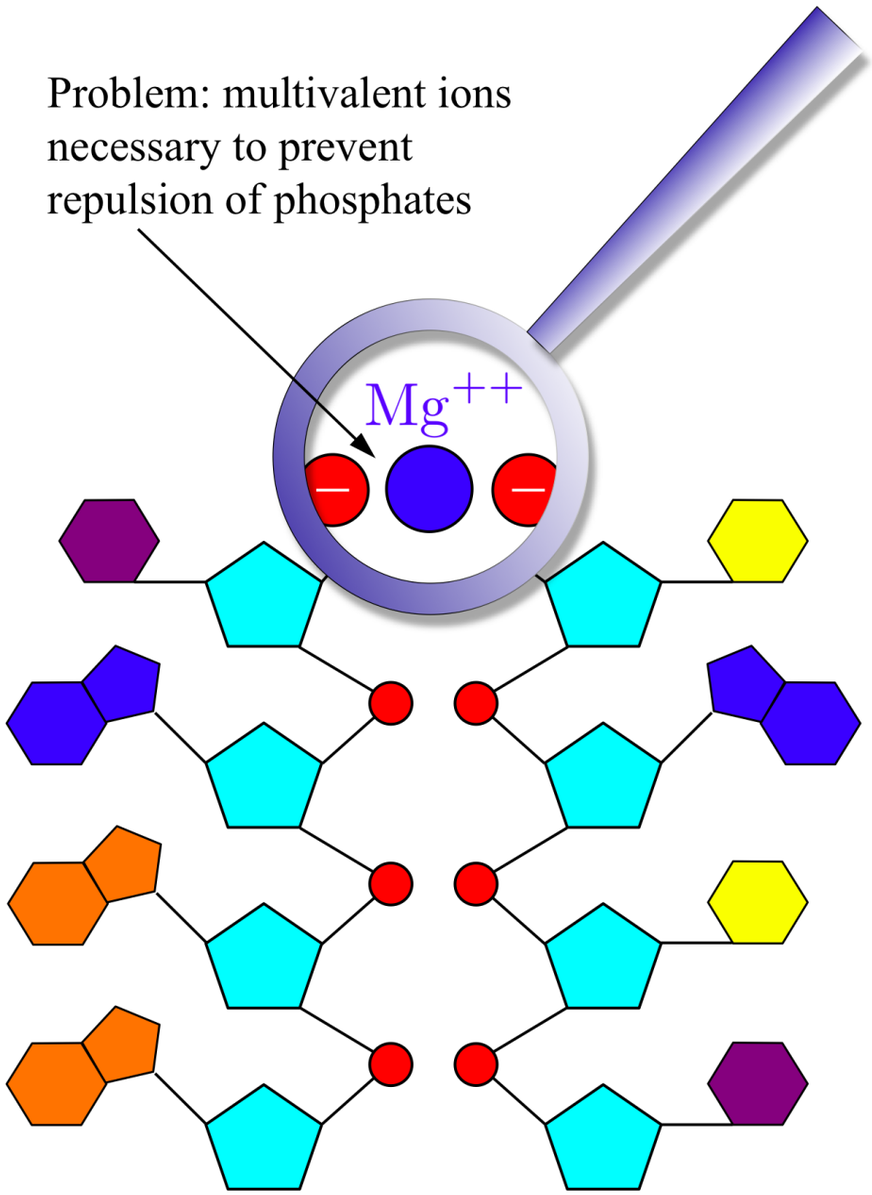
A wrong DNA double helix

Watson & Crick, 1951, not published





Assumption: same chemical environment for each atom of regular backbone AND 2 or 3 intertwined chains (X-ray)



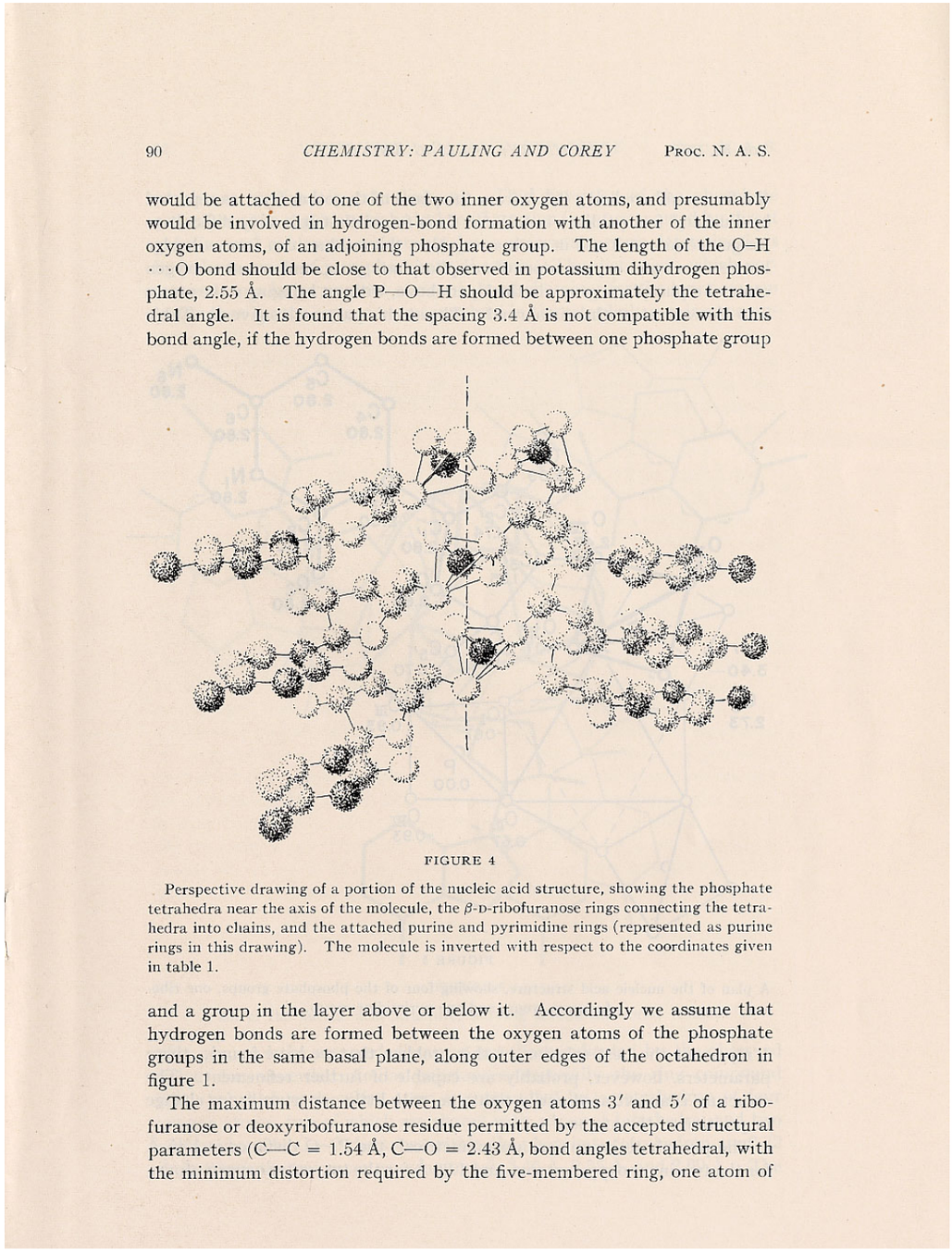
A wrong DNA triple helix

Pauling & Corey, Feb. 1953

three intertwined chains with the sugar-phosphate backbones in the middle

major mistake: phosphate groups assumed to be uncharged!

Published in PNAS
(cited 86 times…)



would be attached to one of the two inner oxygen atoms, and presumably would be involved in hydrogen-bond formation with another of the inner oxygen atoms, of an adjoining phosphate group. The length of the O–H ···O bond should be close to that observed in potassium dihydrogen phosphate, 2.55 Å. The angle P—O—H should be approximately the tetrahedral angle. It is found that the spacing 3.4 Å is not compatible with this bond angle, if the hydrogen bonds are formed between one phosphate group

FIGURE 4

Perspective drawing of a portion of the nucleic acid structure, showing the phosphate tetrahedra near the axis of the molecule, the β-D-ribofuranose rings connecting the tetrahedra into chains, and the attached purine and pyrimidine rings (represented as purine rings in this drawing). The molecule is inverted with respect to the coordinates given in table 1.

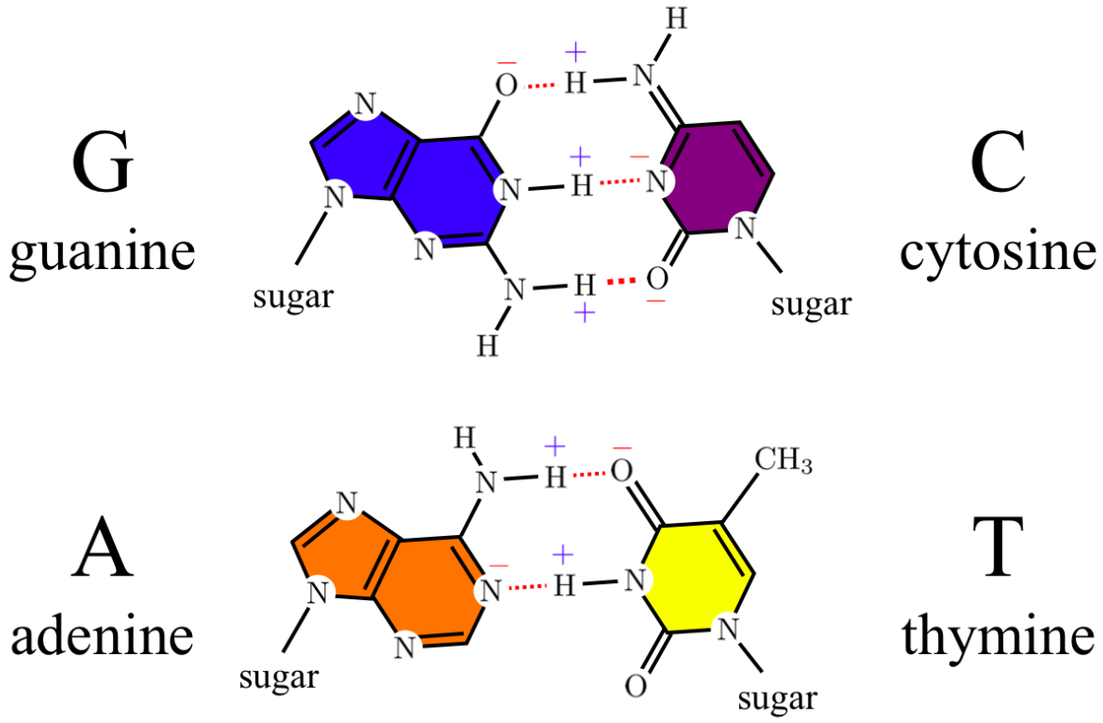
and a group in the layer above or below it. Accordingly we assume that hydrogen bonds are formed between the oxygen atoms of the phosphate groups in the same basal plane, along outer edges of the octahedron in figure 1.

The maximum distance between the oxygen atoms 3′ and 5′ of a ribofuranose or deoxyribofuranose residue permitted by the accepted structural parameters (C—C = 1.54 Å, C—O = 2.43 Å, bond angles tetrahedral, with the minimum distortion required by the five-membered ring, one atom of

Base pairing

How to get the charged backbone to the outside?
Somehow the backbone should be of similar size.

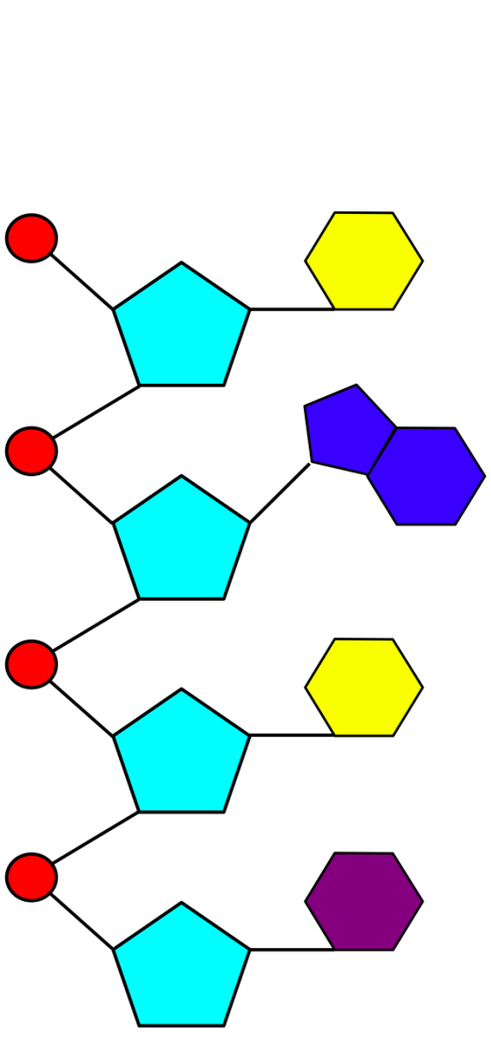
The solution: Watson-Crick base-pairs (found by playing with paper cutouts of the bases)



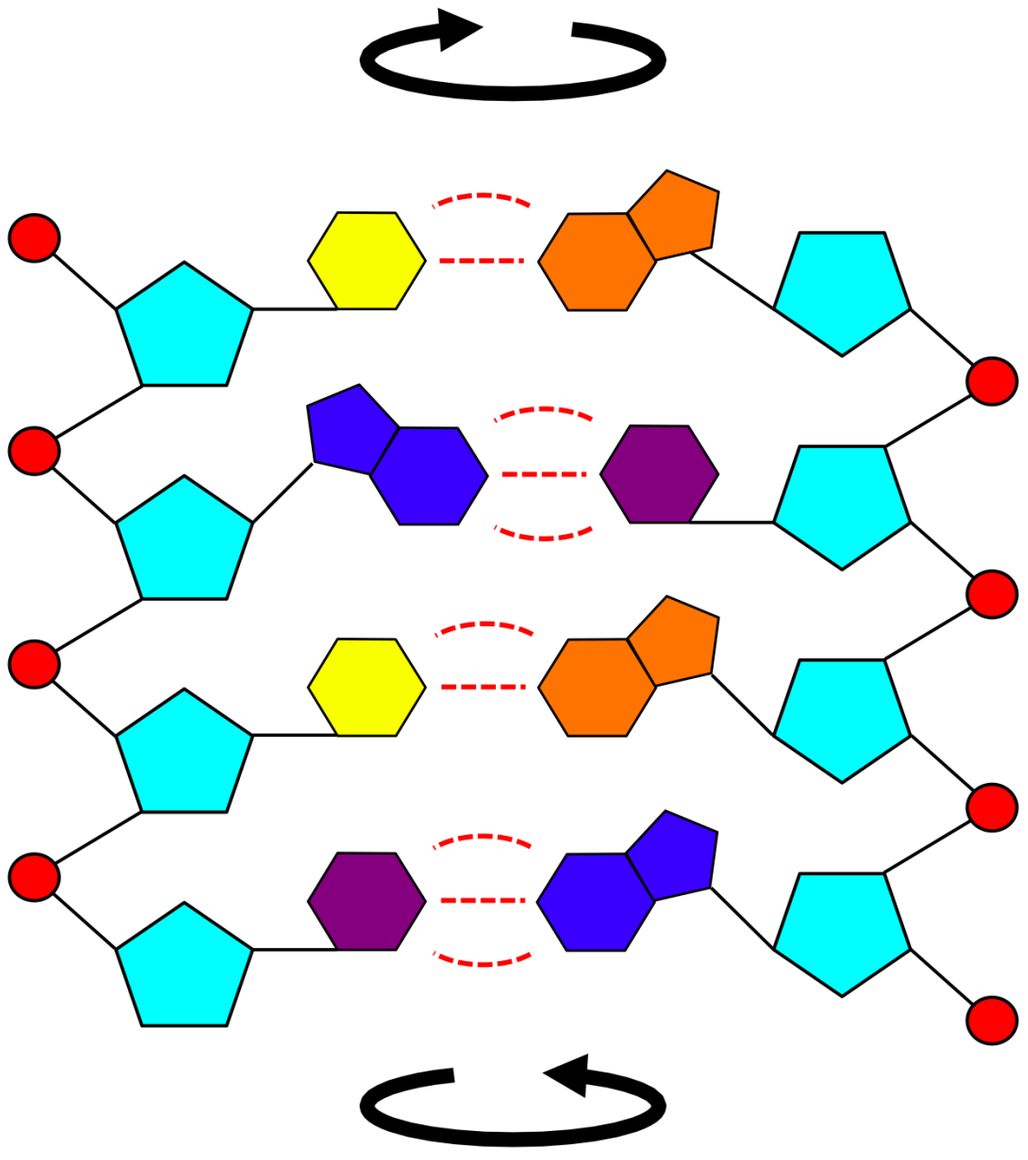
Consistent with the Chargaff rules:
for any DNA sample: #A = #T and #C = #G

The correct double helix

Watson & Crick, 1953

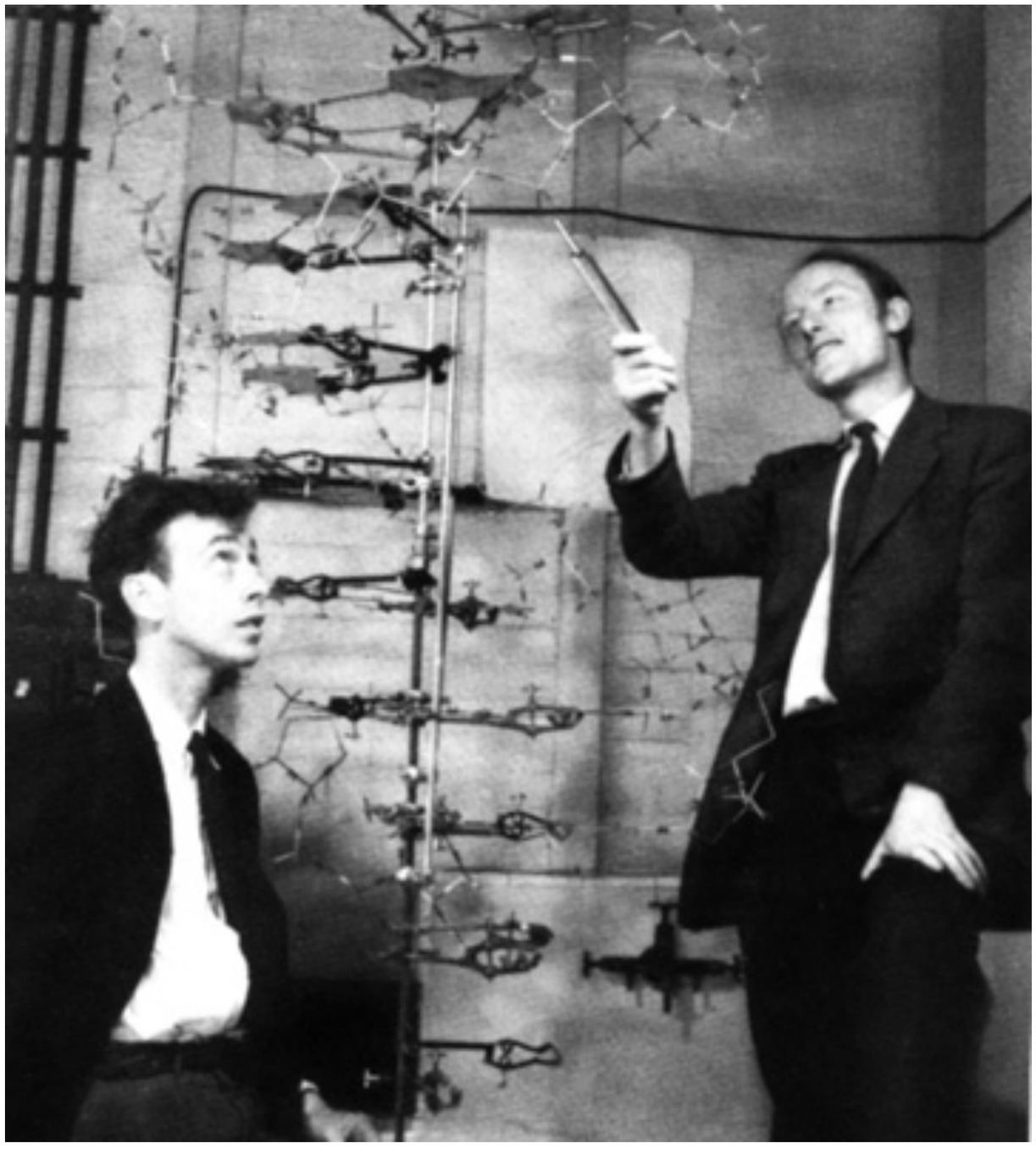


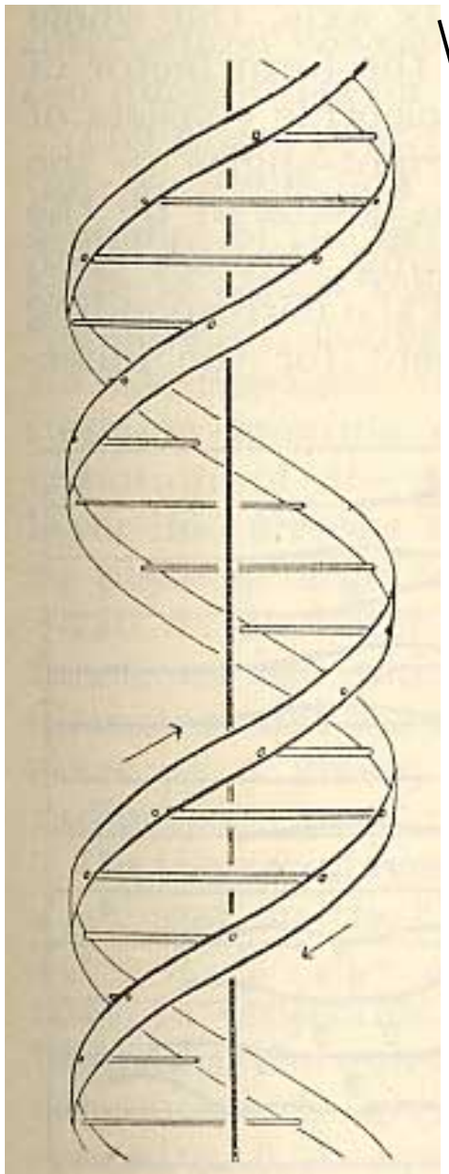
Assumption: baired bases
stacked in the center



Model building

Watson & Crick, 1953





This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (*Stockholm*), **2** (11) (1905).

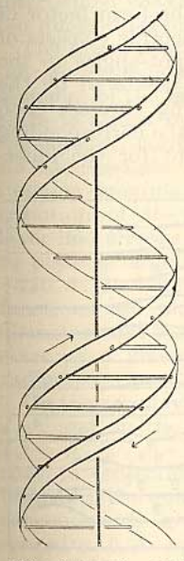
MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.



This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furcular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

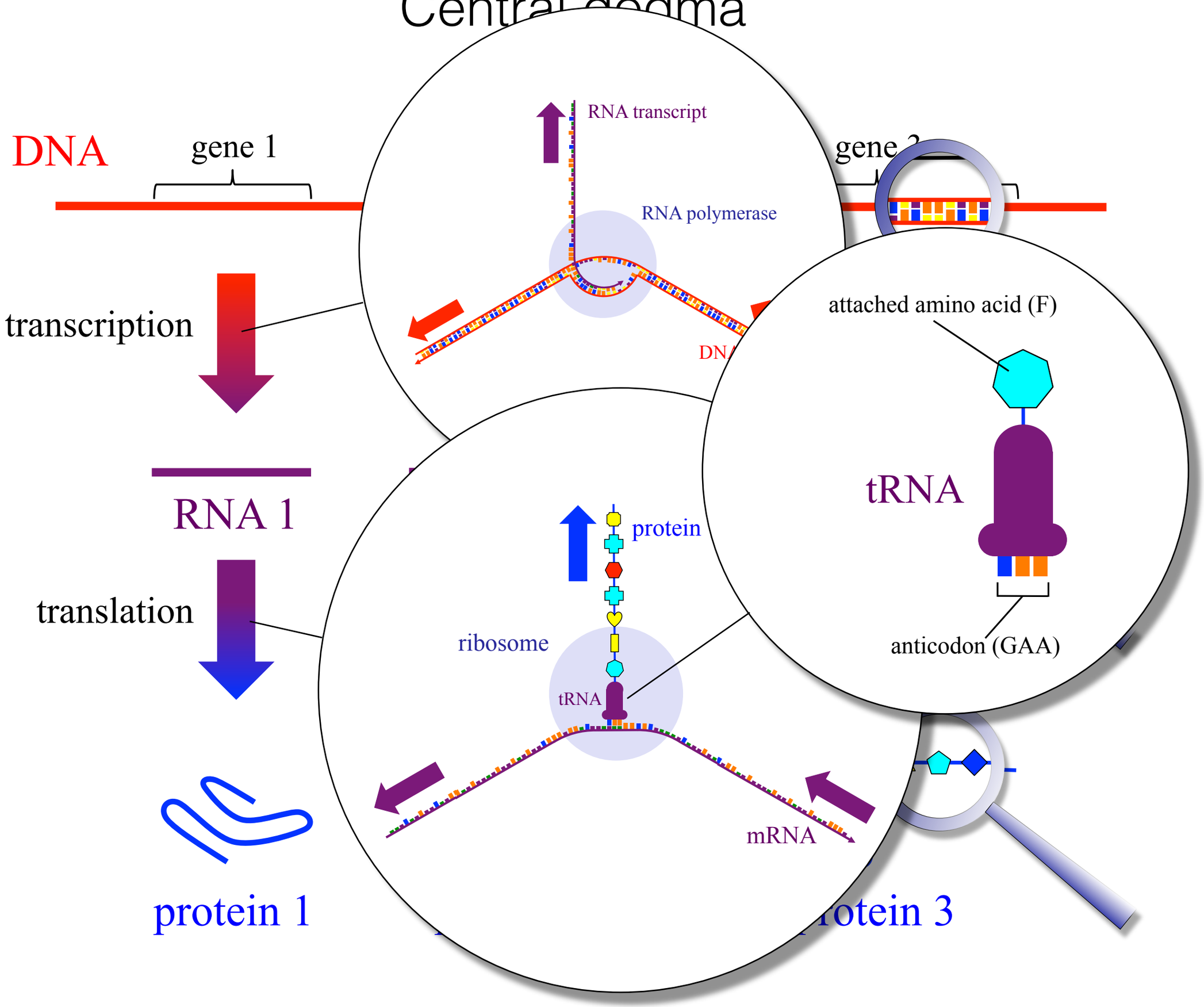
It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the con-

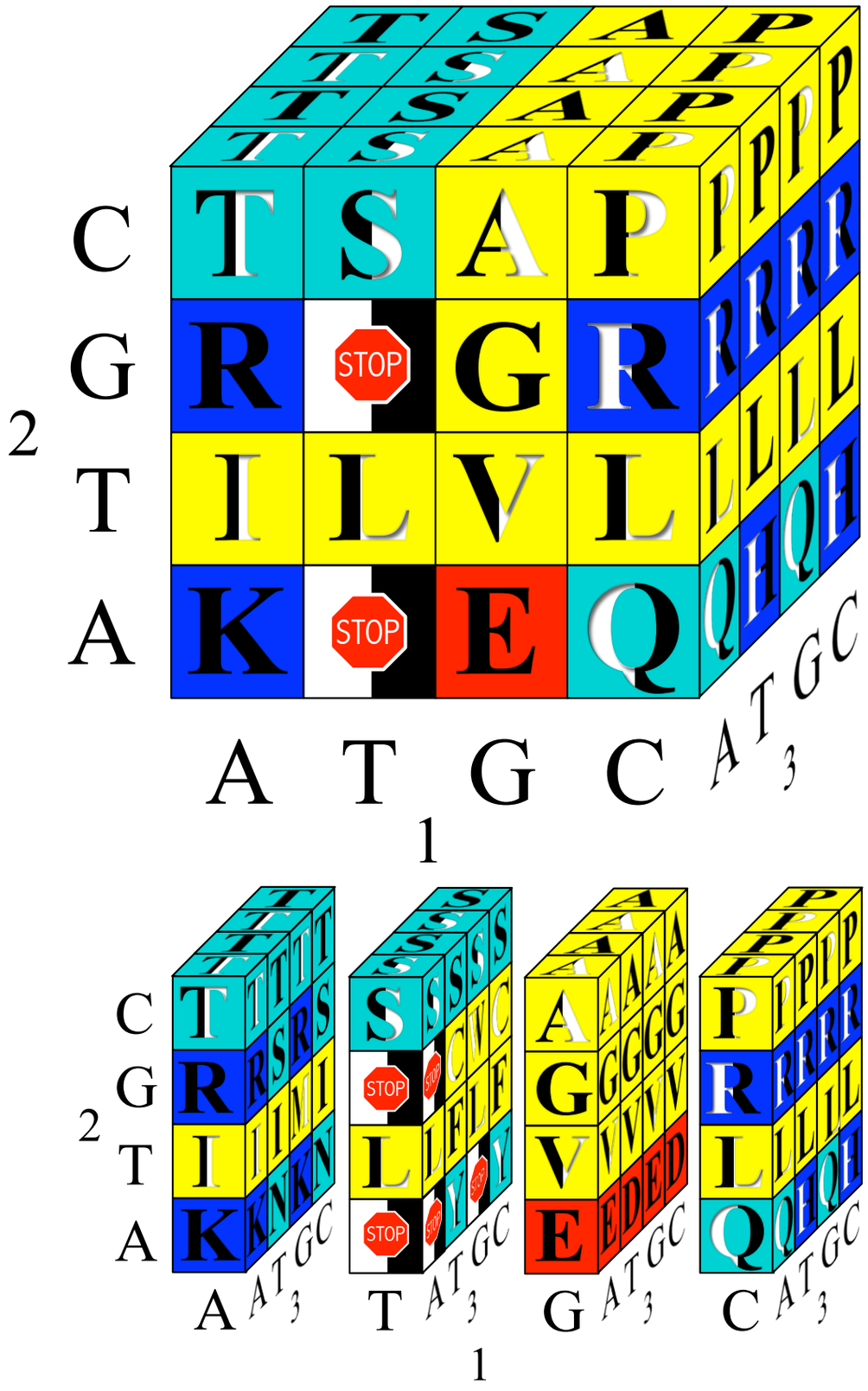
It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Wilkins, Dr. R. E. Franklin and their co-workers at

Central dogma



codon cube:



negatively charged aa	
D:	aspartic acid (Asp)
E:	glutamic acid (Glu)

positively charged aa	
H:	histidine (His)
K:	lysine (Lys)
R:	arginine (Arg)

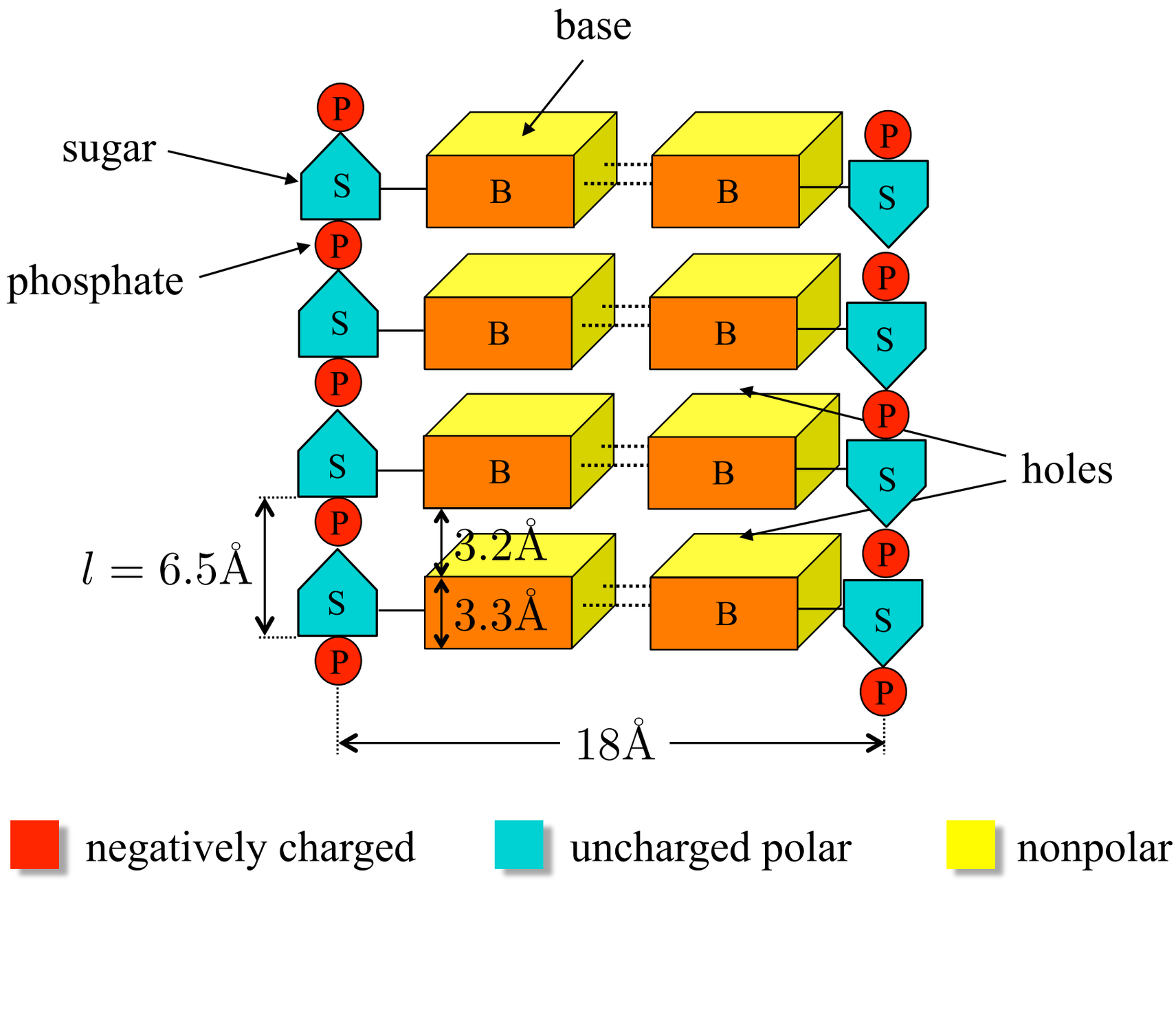
uncharged polar aa	
N:	asparagine (Asn)
Q:	glutamine (Gln)
S:	serine (Ser)
T:	threonine (Thr)
Y:	tyrosine (Tyr)

nonpolar aa	
A:	alanine (Ala)
C:	cysteine (Cys)
F:	phenylalanine (Phe)
G:	glycine (Gly)
I:	isoleucine (Ile)
L:	leucine (Leu)
M:	methionine (Met)
V:	valine (Val)
P:	proline (Pro)
W:	tryptophan (Trp)

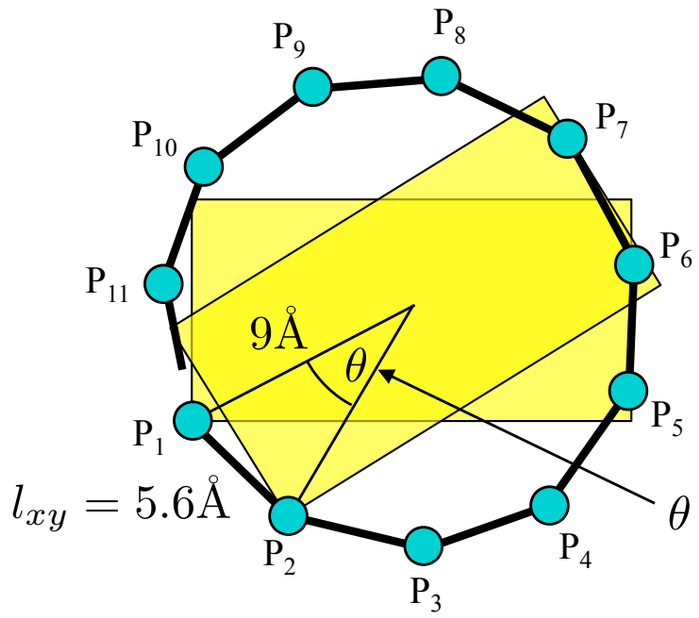
from: H. Schiessel, Biophysics for Beginners: a Journey though the Cell Nucleus

Why does DNA form a double helix?

Calladine, Drew, Luisi, Travers, Understanding DNA

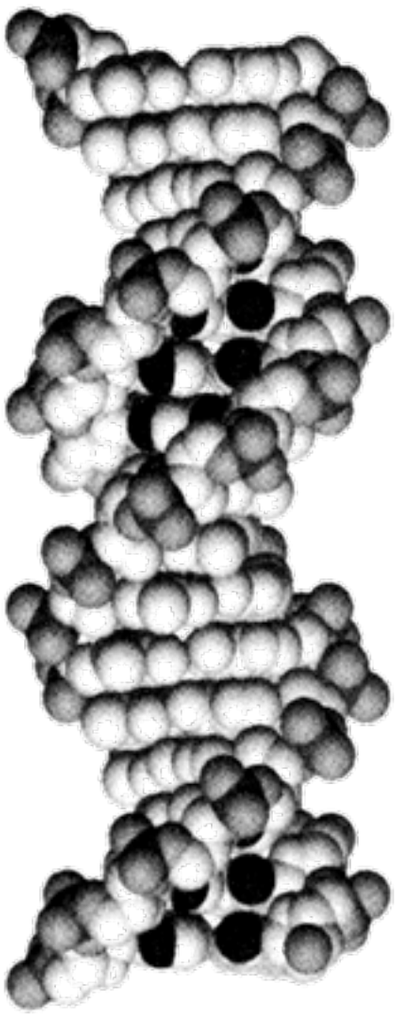


The DNA double helix:

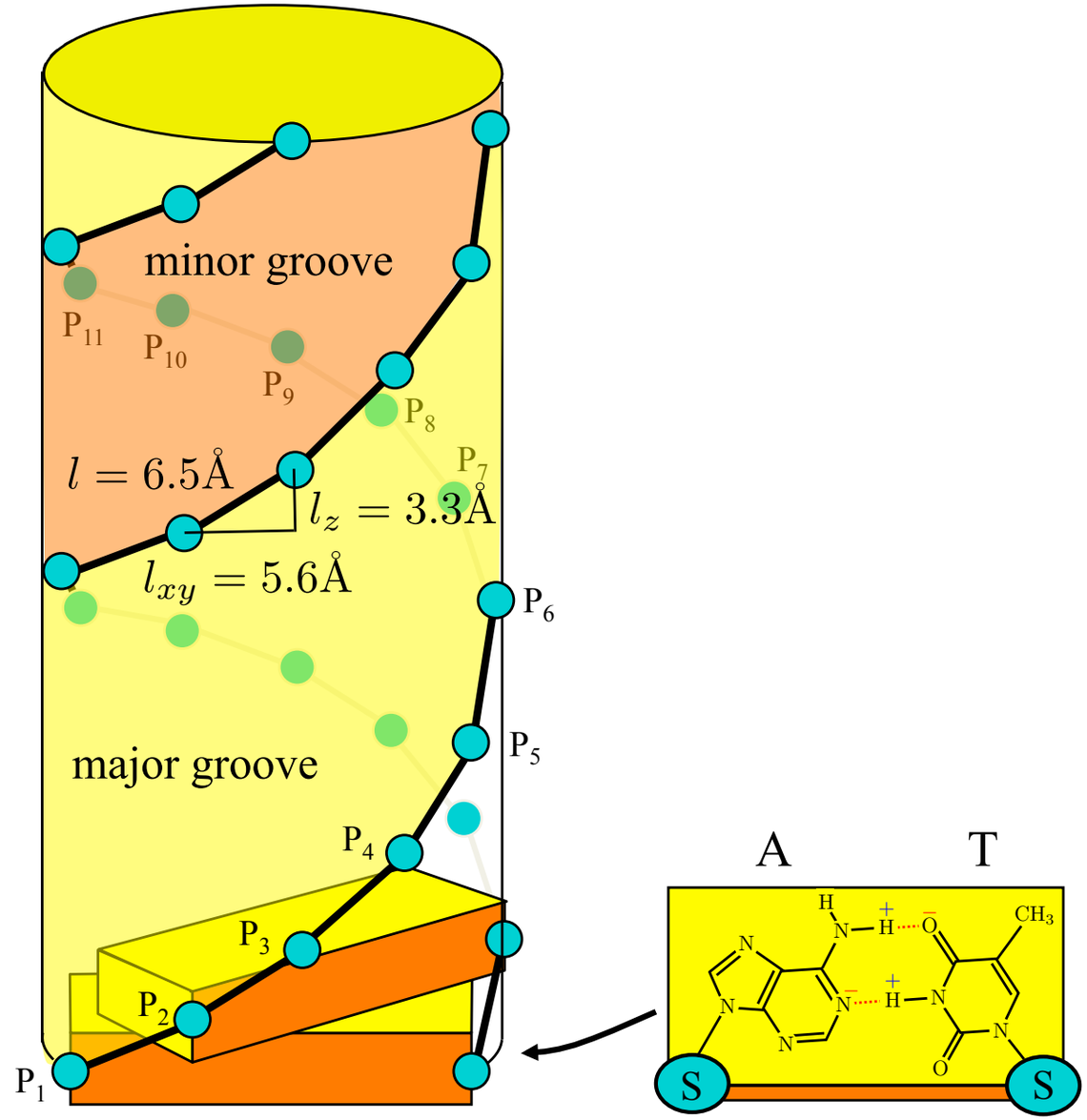


$$l_{xy} = 5.6\text{Å}$$

$$\theta = 2\arcsin\left(\frac{l_{xy}}{2 \times 9\ \text{Å}}\right) \approx 36^\circ$$

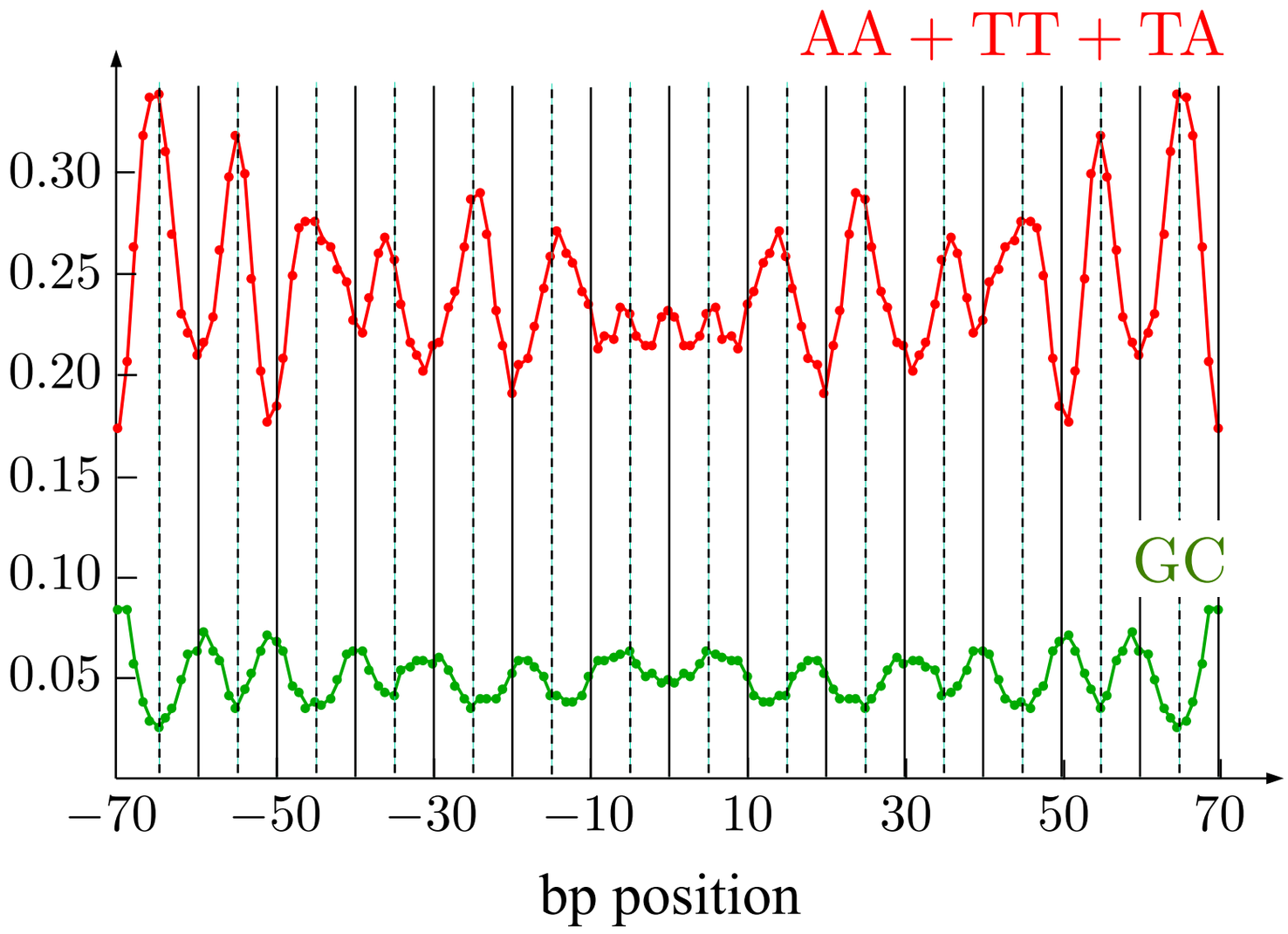


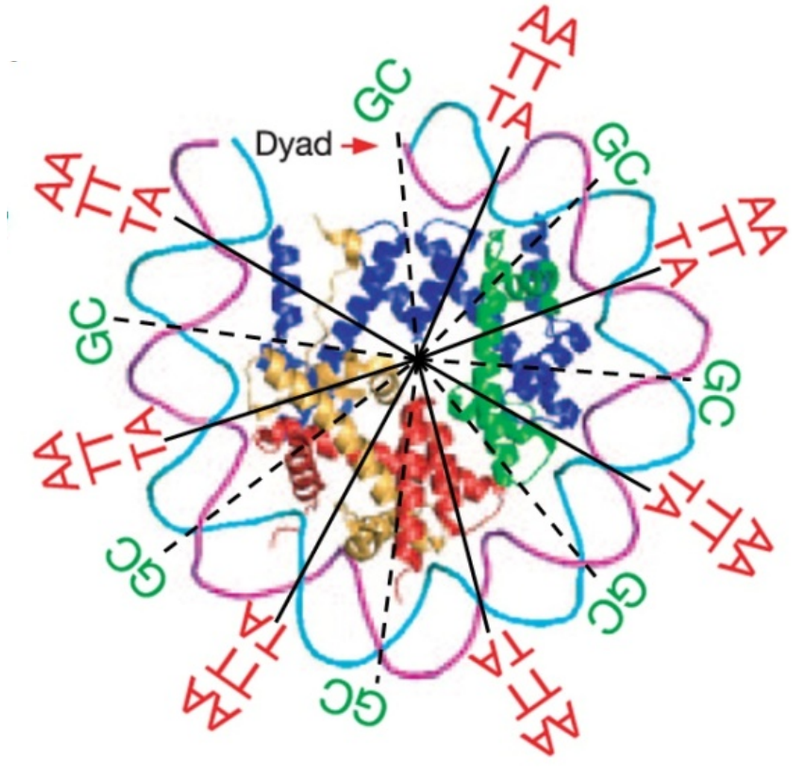
B-DNA



The genomic code for nucleosome positioning

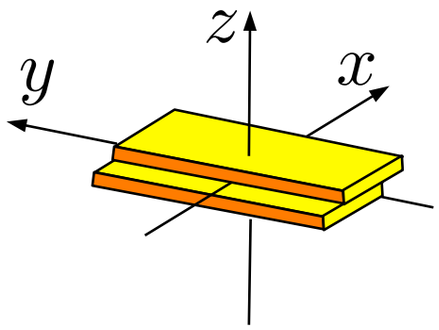
Satchwell, Drew & Travers, J. Mol. Biol. **191** (1986) 659



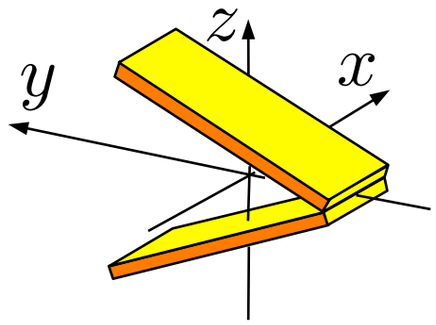


Segal et al., Nature, 2006

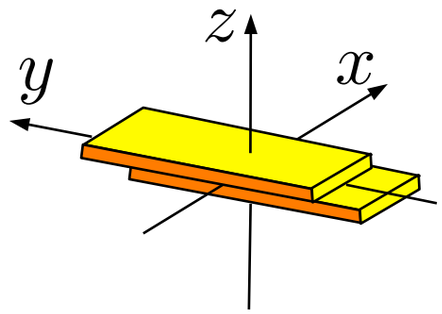
The rigid base-pair representation



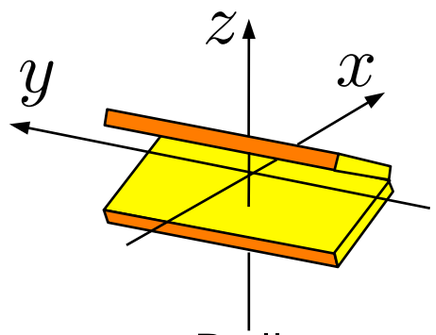
Shift



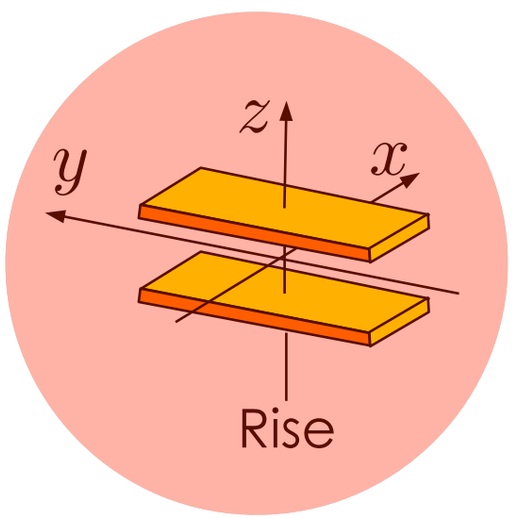
Tilt



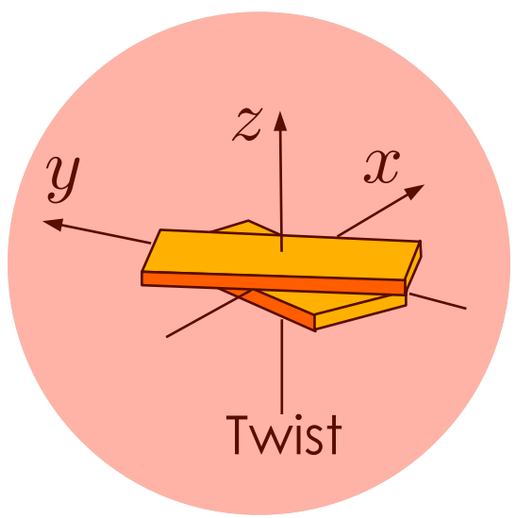
Slide



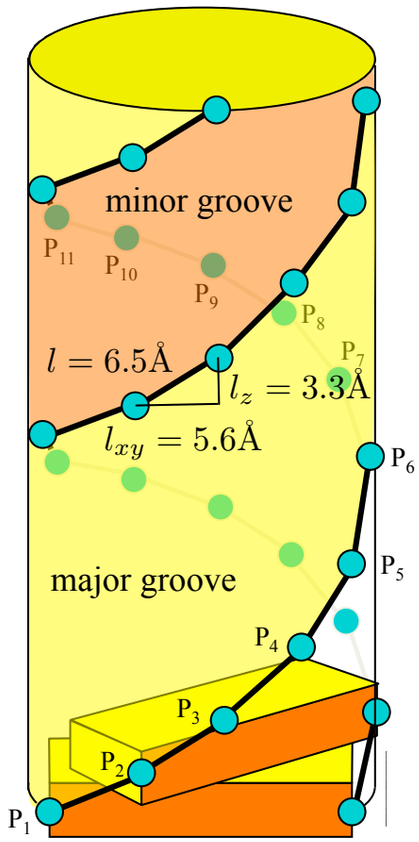
Roll



Rise

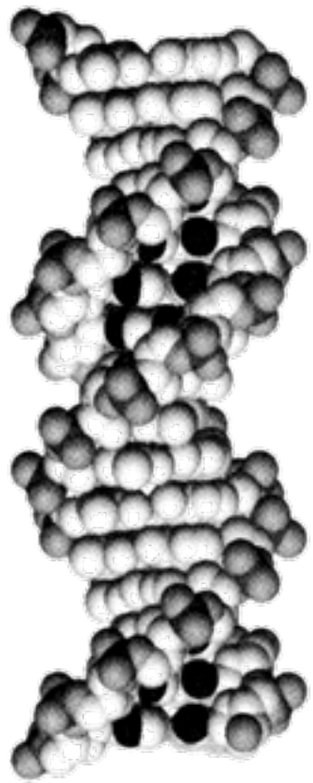


Twist



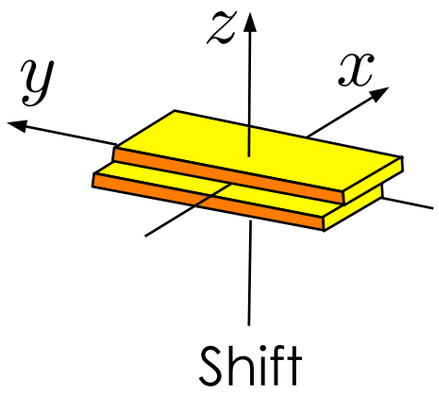
$$\text{Rise} = 0.34\,\text{nm}$$

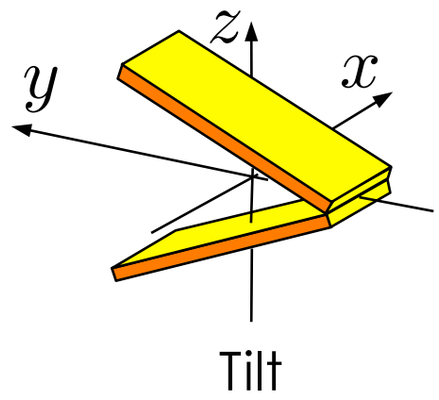
$$\text{Twist} = 36^\circ$$

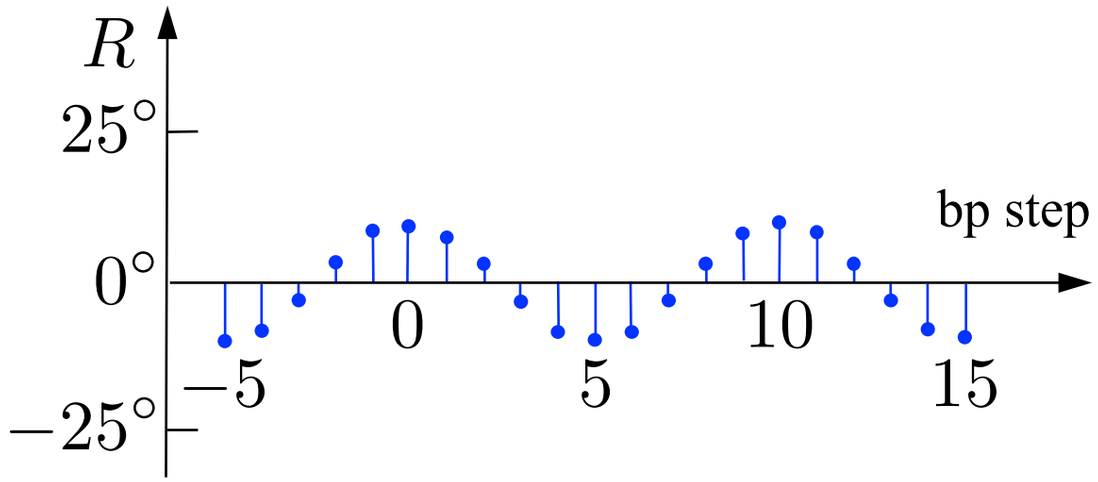


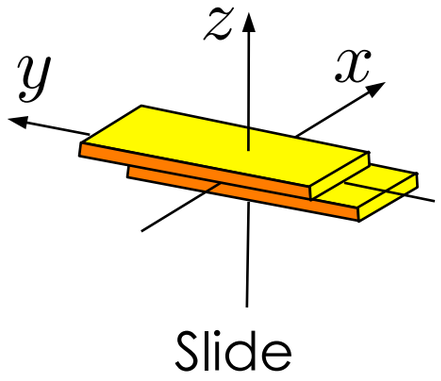
B-DNA

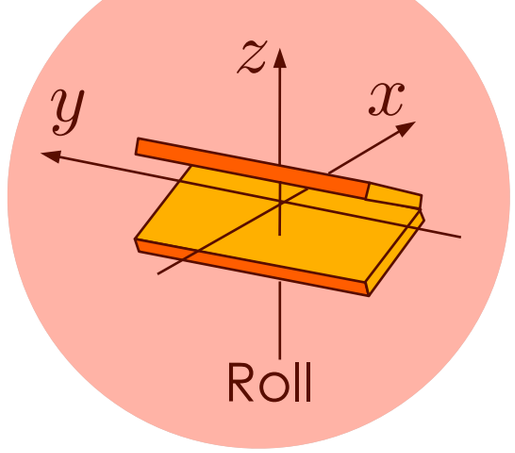
The rigid base-pair representation

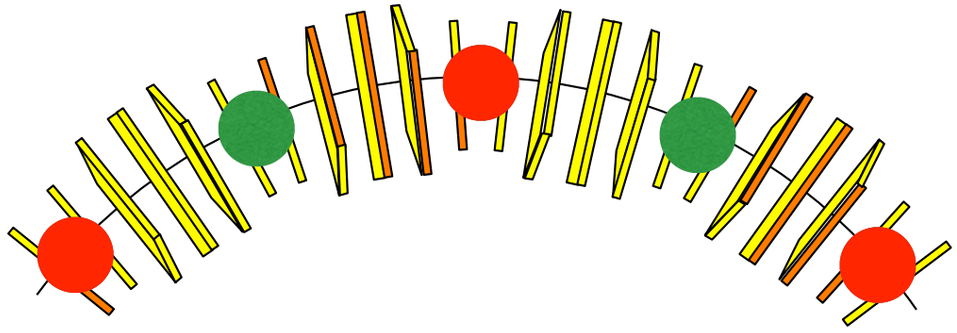


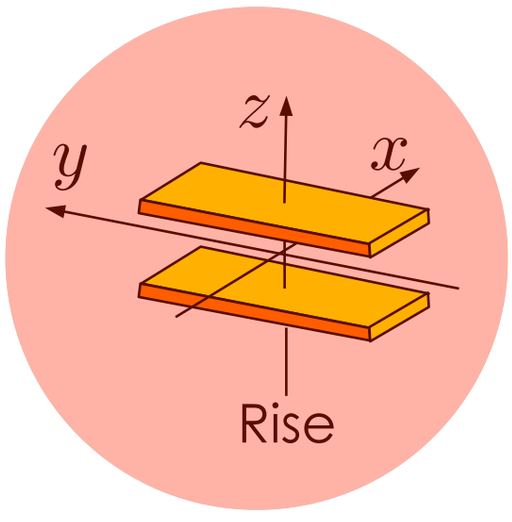


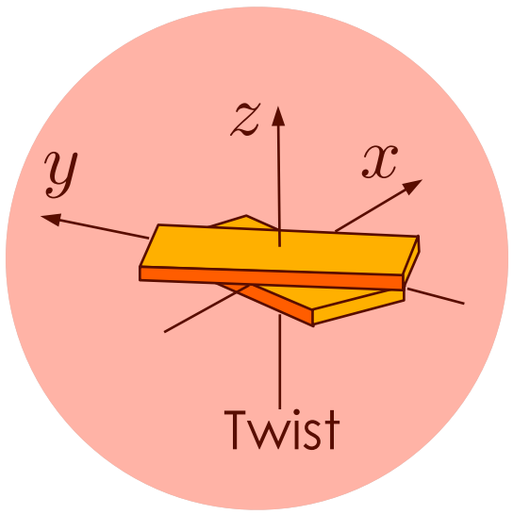


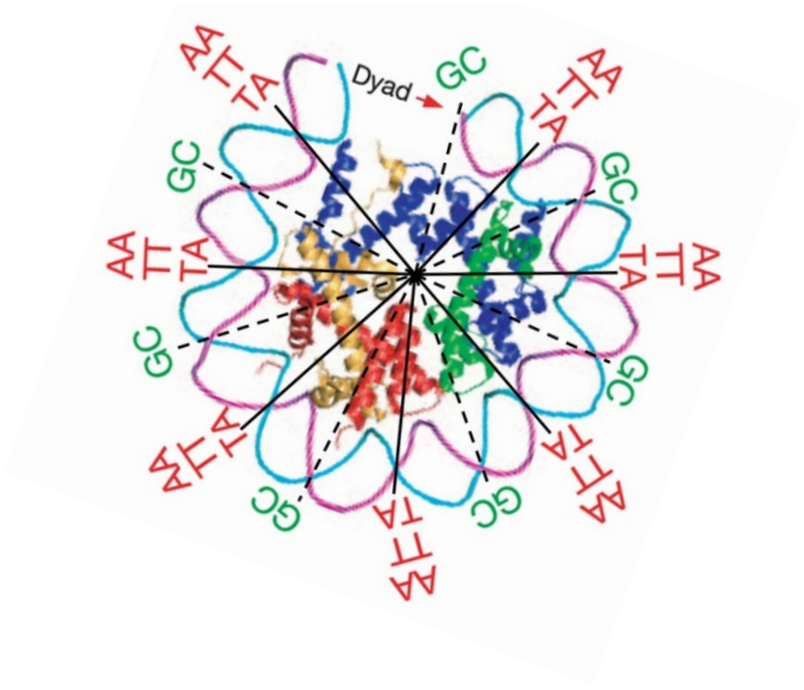




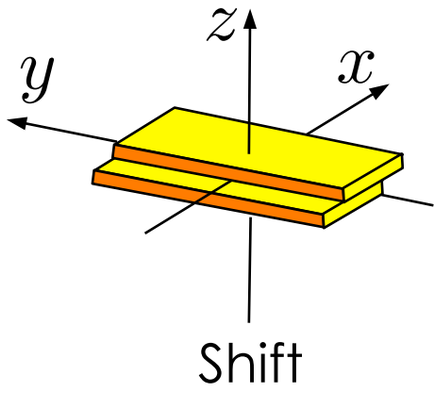


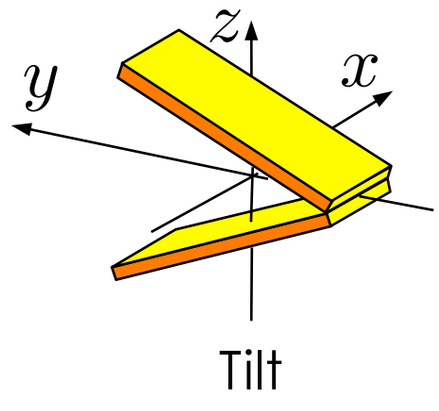


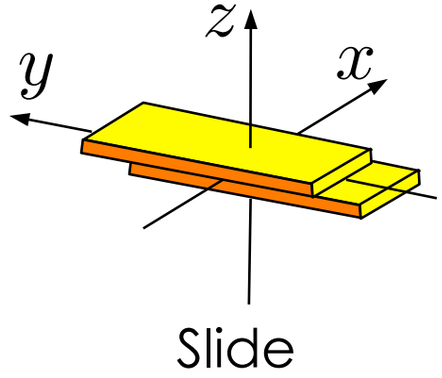


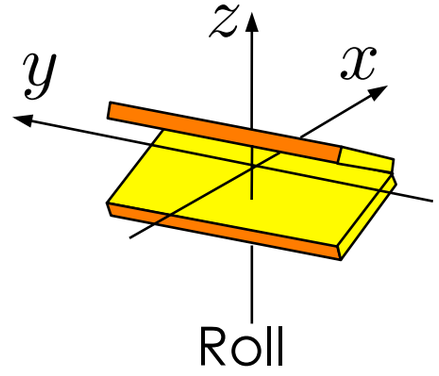


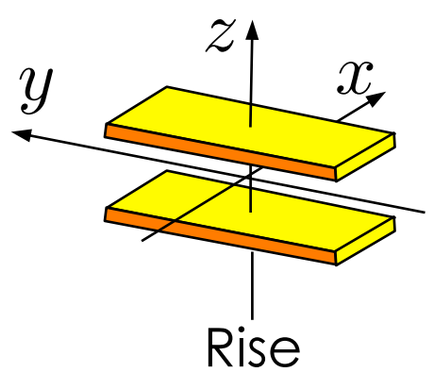
A local harmonic model

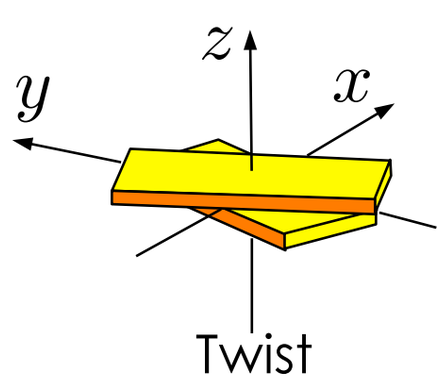












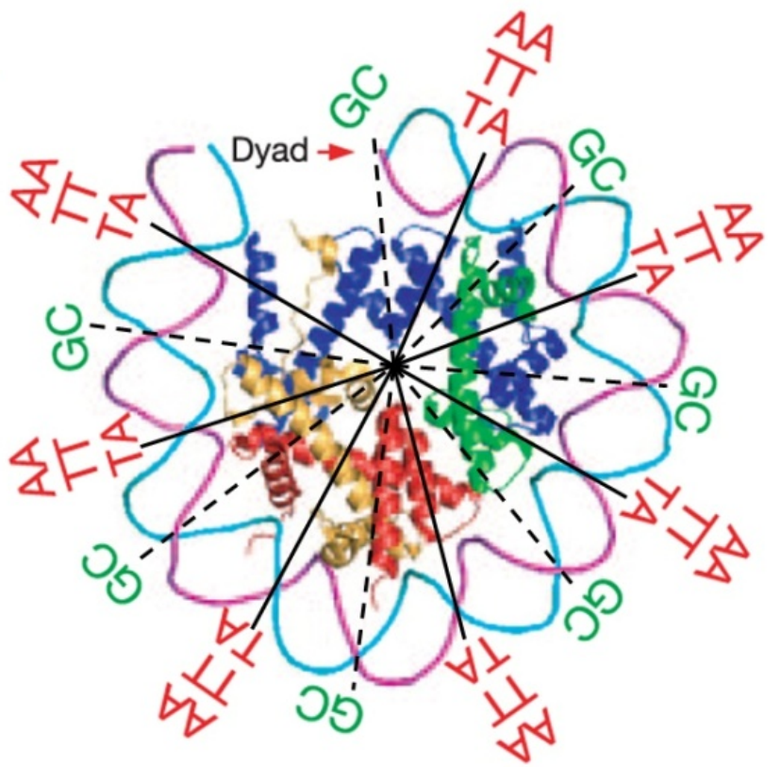
$$\vec{x} = [\text{Tilt},\ \text{Roll},\ \text{Twist},\ \text{Shift},\ \text{Slide},\ \text{Rise}]$$

$$E_{\text{step}} = \frac{1}{2}\left(\vec{x} - \vec{x}^{0}\right)^{T} \mathbf{Q} \left(\vec{x} - \vec{x}^{0}\right)$$

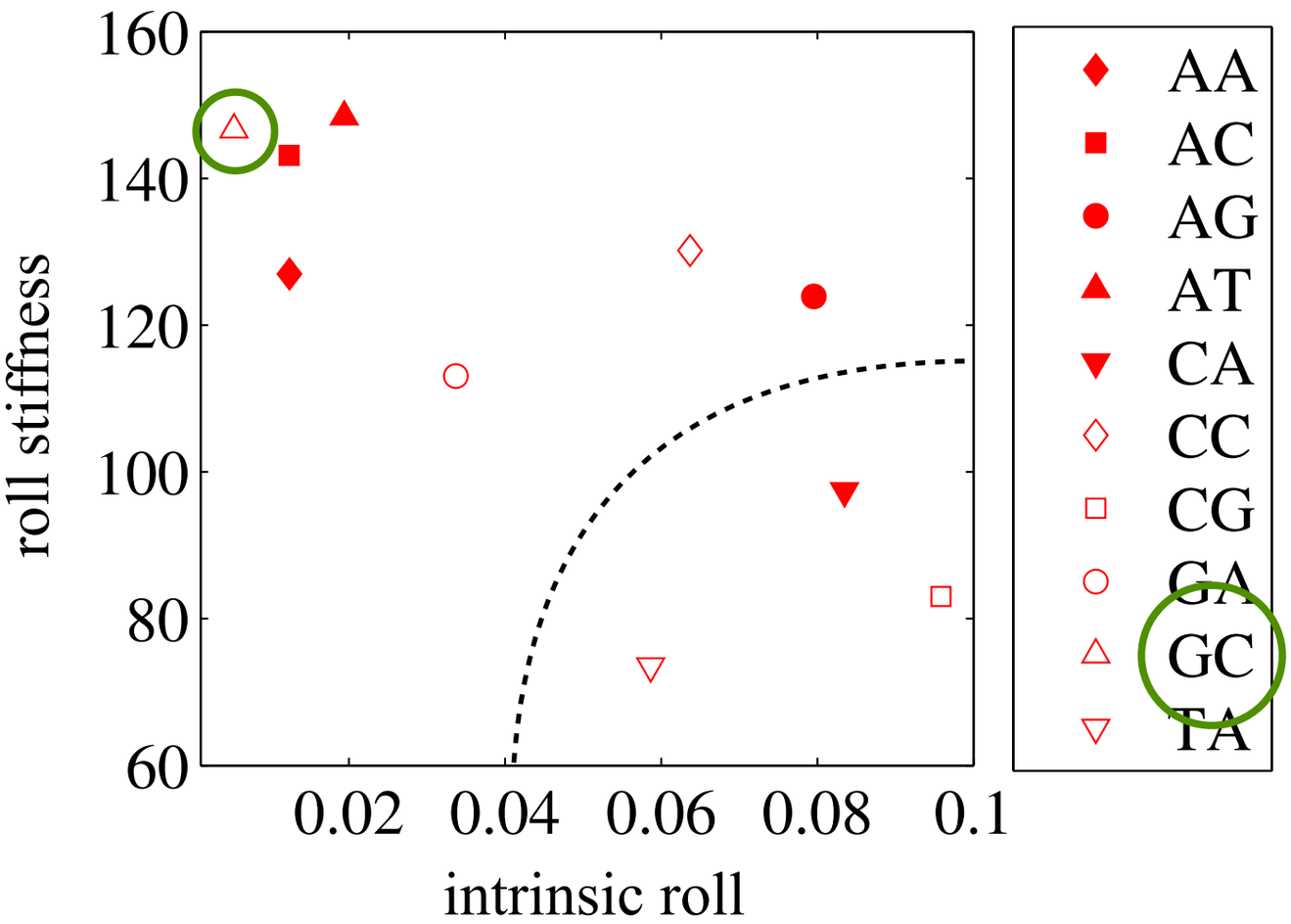
protein-DNA cocrystals
Olson, Gorin, Lu, Hock & Zhurkin
PNAS **95** (1998) 11163

all atom MD simulations
Lavery et al.
Nucleid Acids Res. **38** (2010) 299

The mysterious GC step

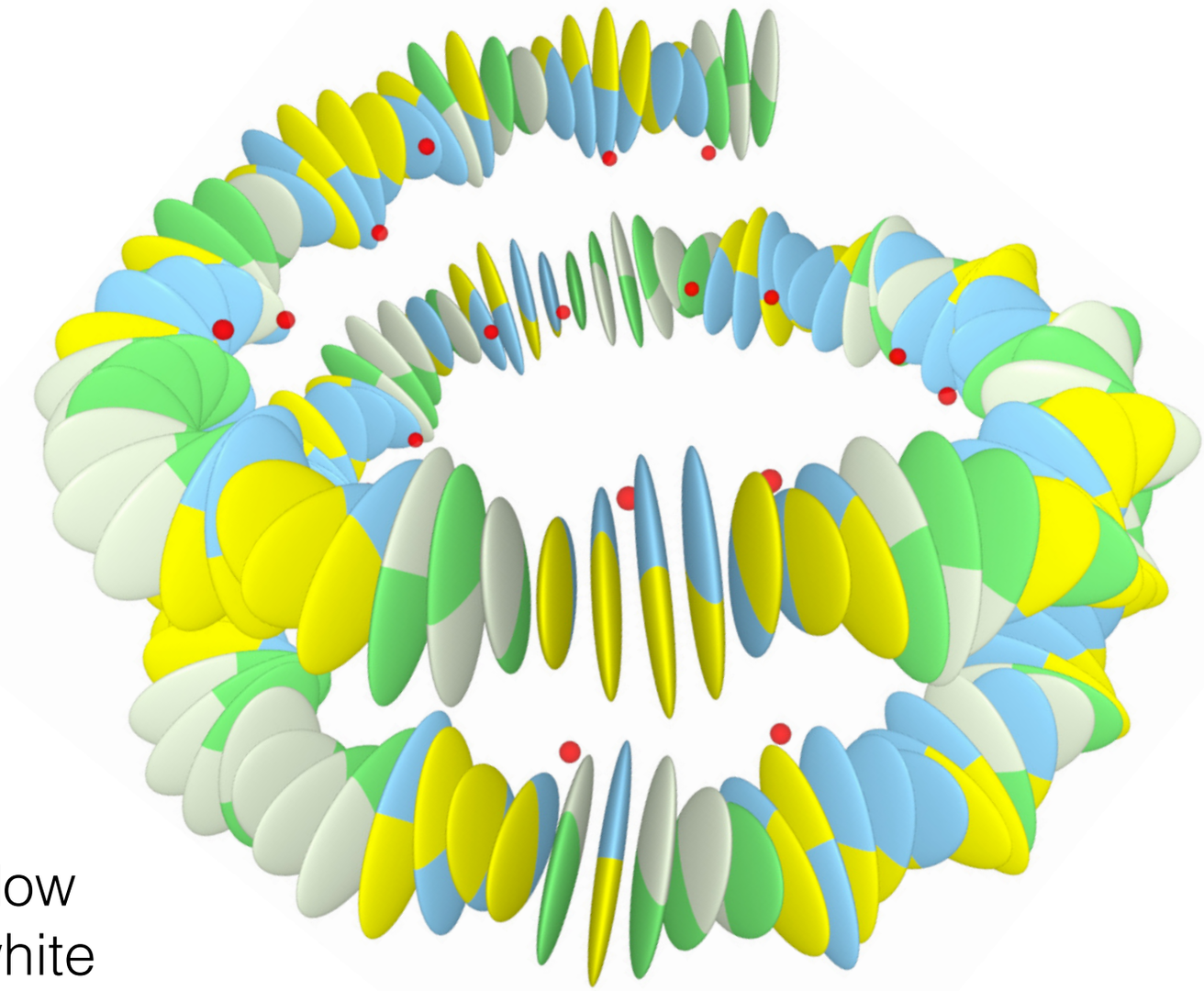


GC peaks at positive roll



Our nucleosome model

Eslami-Mossalam, Schram, Tompitak, van Noort & Schiessel

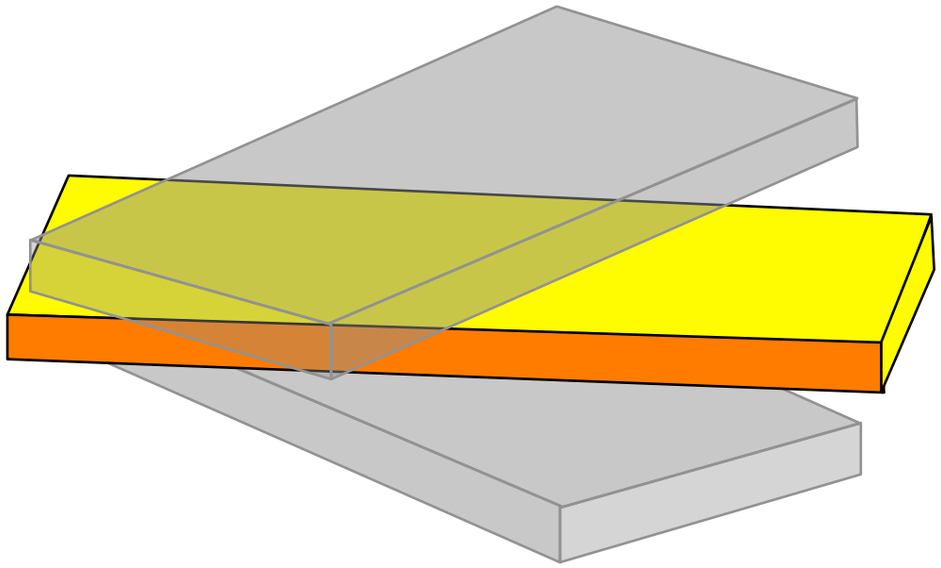


AT: blue/yellow
GC: green/white

similar models:

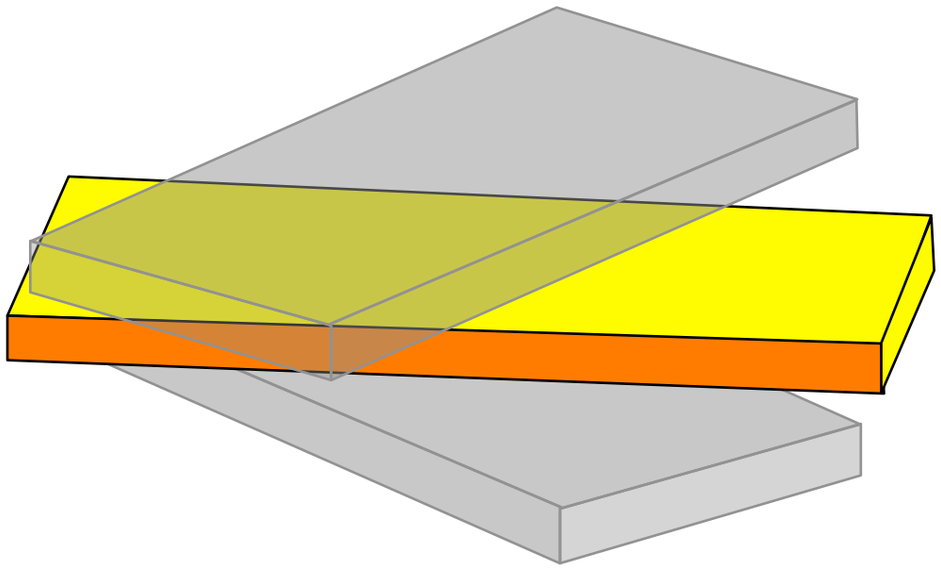
Tolstorukov et al., J. Mol. Biol. 2007
Vaillant, Audit & Arneodo, 2007
Morozov et al., Nucl. Acids Res. 2009
Becker & Everaers, Structure 2009
Fathizadeh, Besya, Ejtehadi & Schiessel, EPJE 2013

The two Mutation Monte Carlo moves

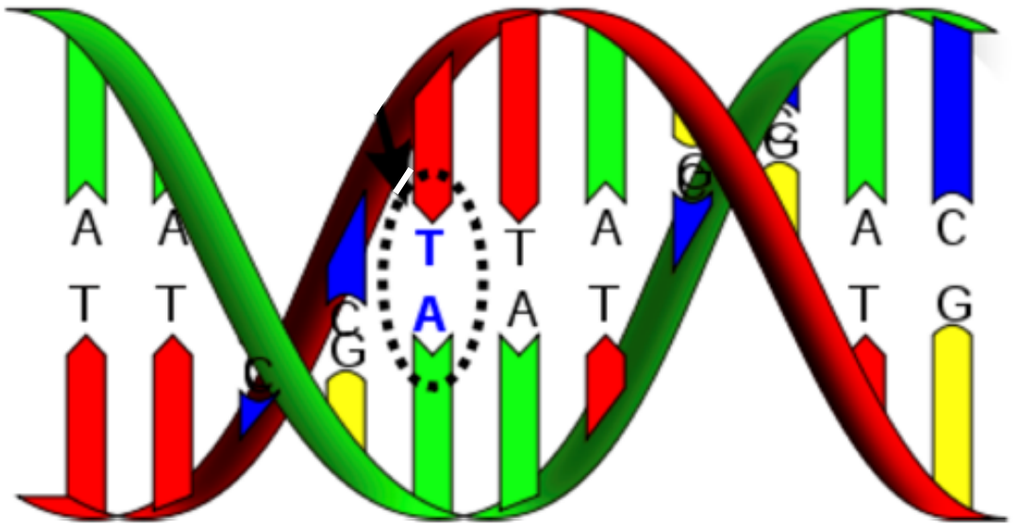


configurational move

The two Mutation Monte Carlo moves

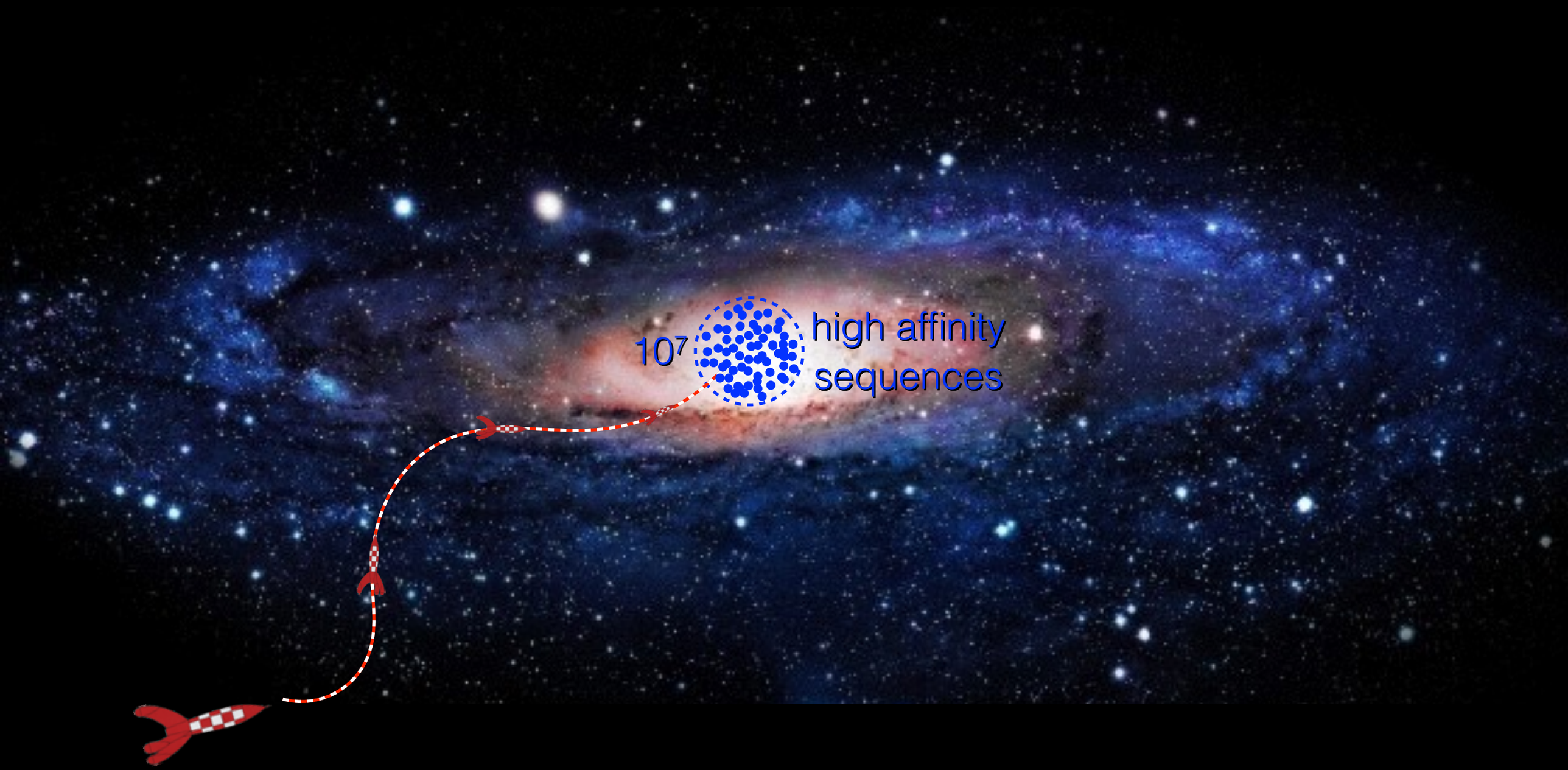


configurational move



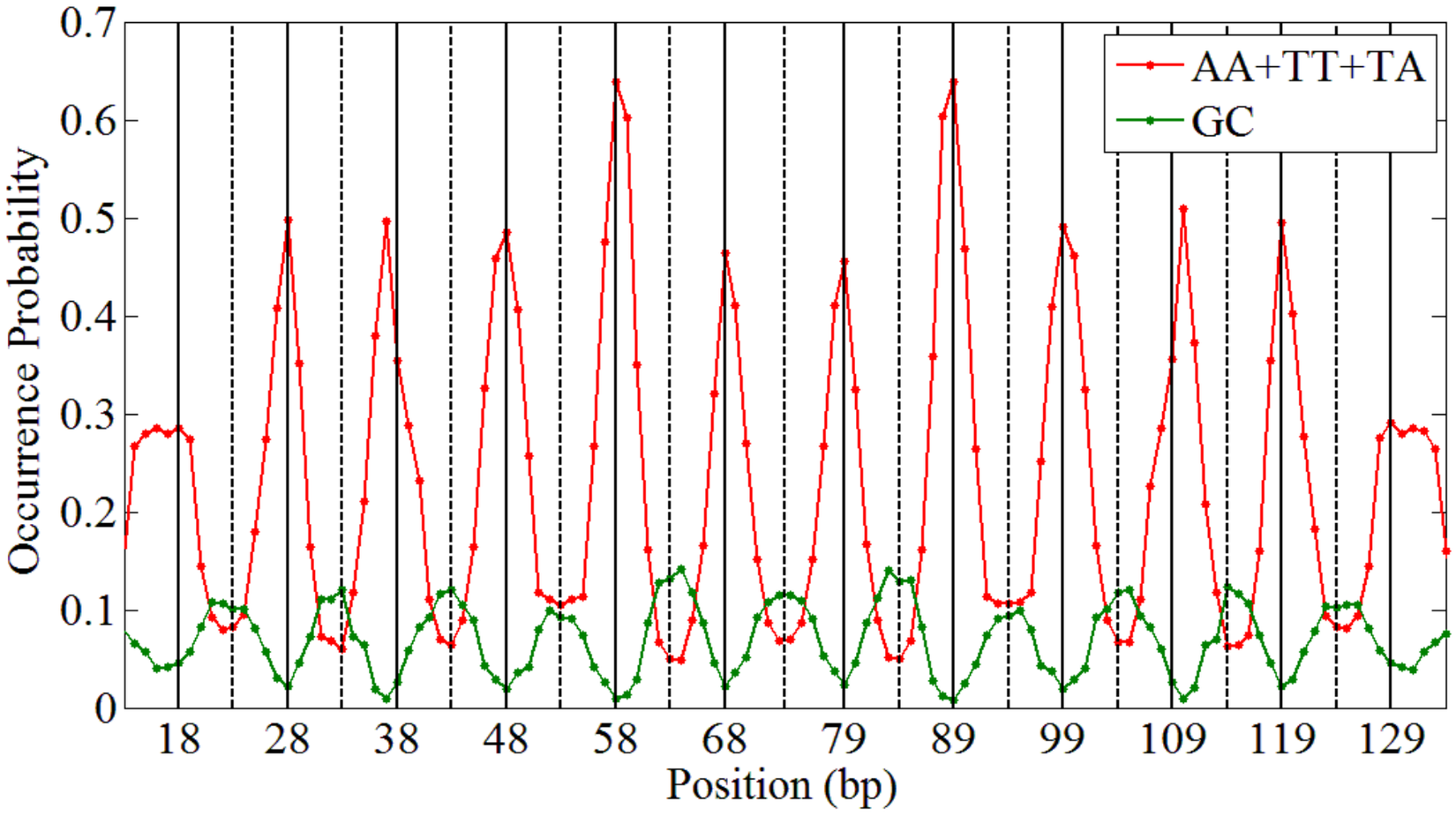
point mutation

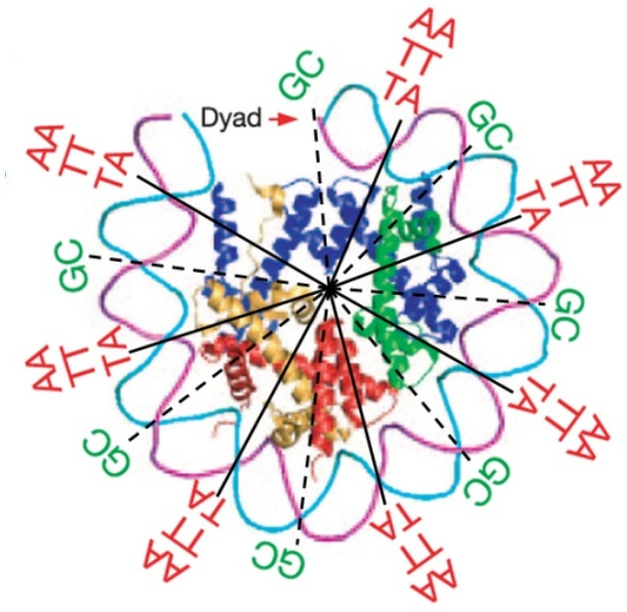
Mutation Monte Carlo method



Recovering the positioning rules

10 million sequences at 100 K





GC steps peak at their least favorite positions. Why?

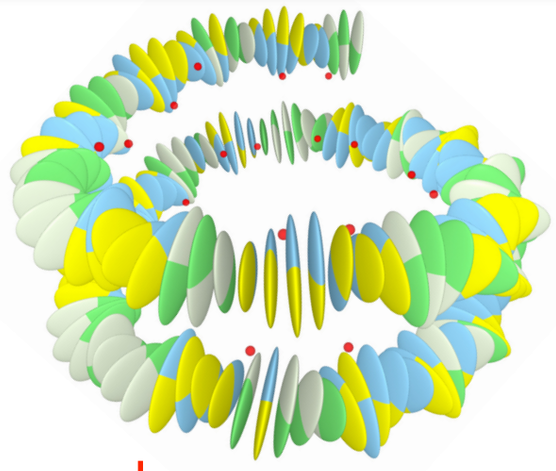
GC brings in good neighbors, e.g. AGCT

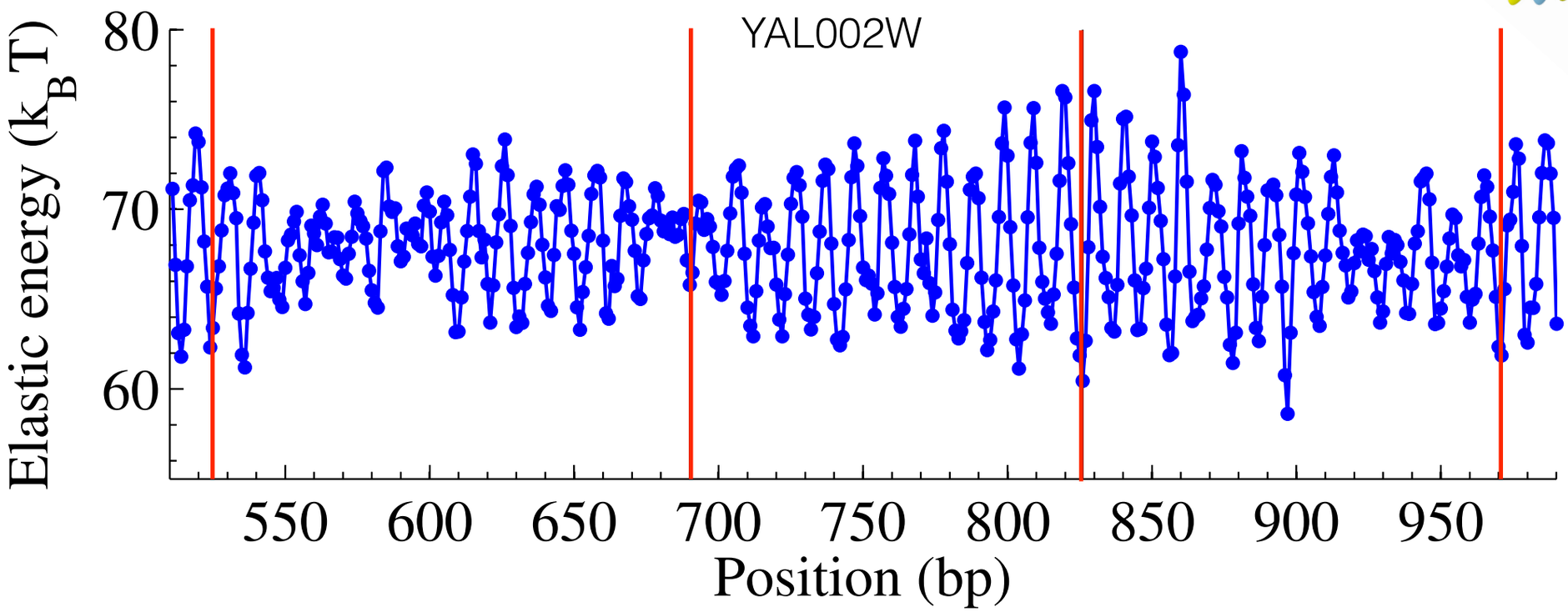
THE MECHANICAL GENOME

Our questions:

- Yes, the positioning rules can be explained by a purely mechanical model.
- Can mechanical information be multiplexed with classical genetic information?

The energy landscape of a gene





Nucleosome mapping at basepair resolution in *Saccharomyces cerevisiae*

Brogaard, Xi, Wang & Widom

Nature **486** (2012) 496

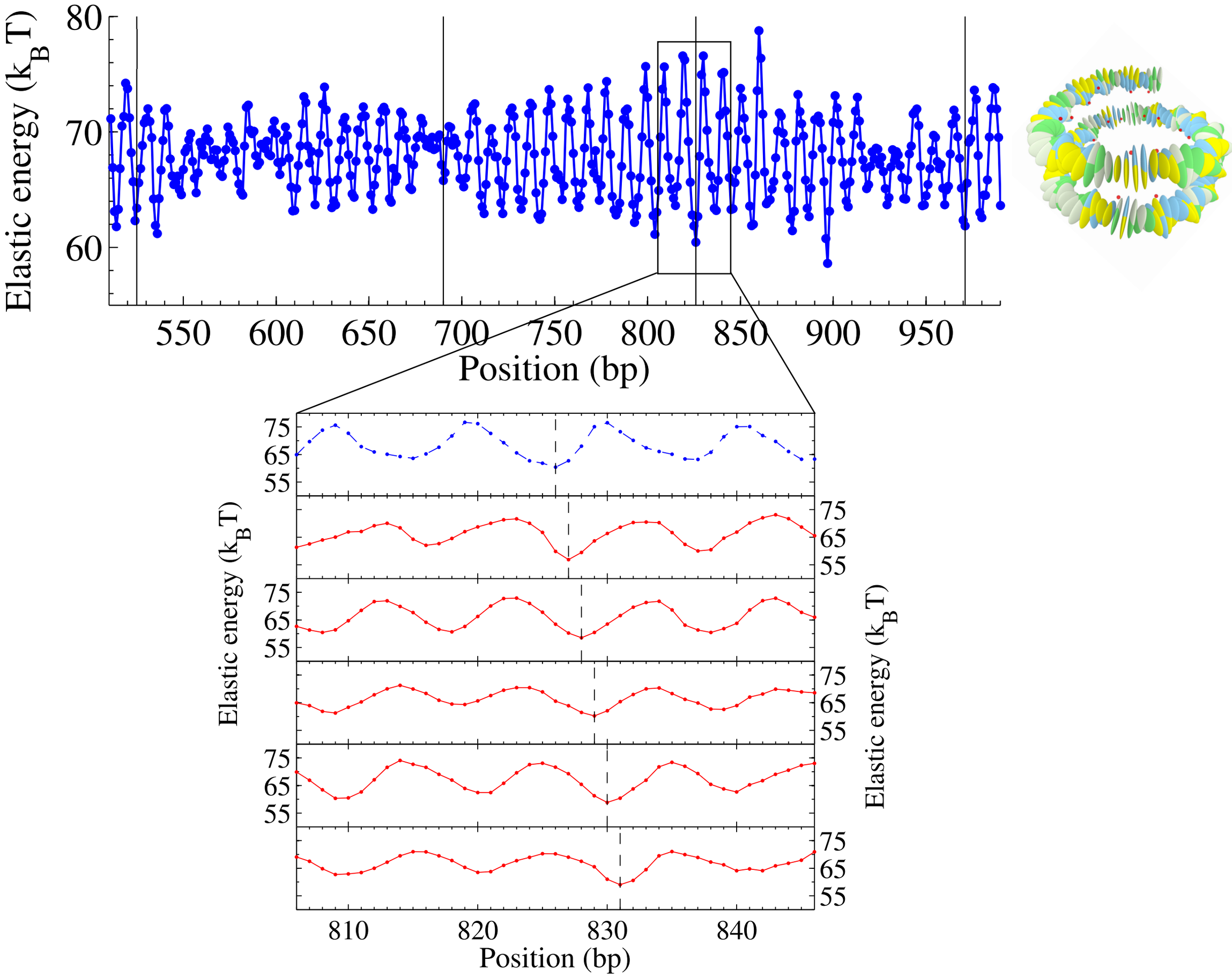
Synonymous Mutation Monte Carlo

… GTA CTC ACA ACT ACA CAT TTT GCC CTT ATT …

… Val Leu Thr Thr Thr His Phe Ala Leu Ile …

allowed mutations

GTA	TTA	ACA	ACA	ACA	CAT	TTT	GCA	TTA	ATA
GTT	TTG	ACT	ACT	ACT	CAC	TTC	GCT	TTG	ATT
GTG	CTA	ACG	ACG	ACG			GCG	CTA	ATC
GTC	CTT	ACC	ACC	ACC			GCC	CTT	
	CTG							CTG	
	CTC							CTC	


Elastic energy (k_BT)
80
70
60
550
600
650
700
750
800
850
900
950
Position (bp)
75
65
55
Elastic energy (k_BT)
810
820
830
840
Position (bp)

THE MECHANICAL GENOME

Our questions:

- Yes, the positioning rules can be explained by a purely mechanical model.
- Yes, mechanical and genetic information can be multiplexed together.

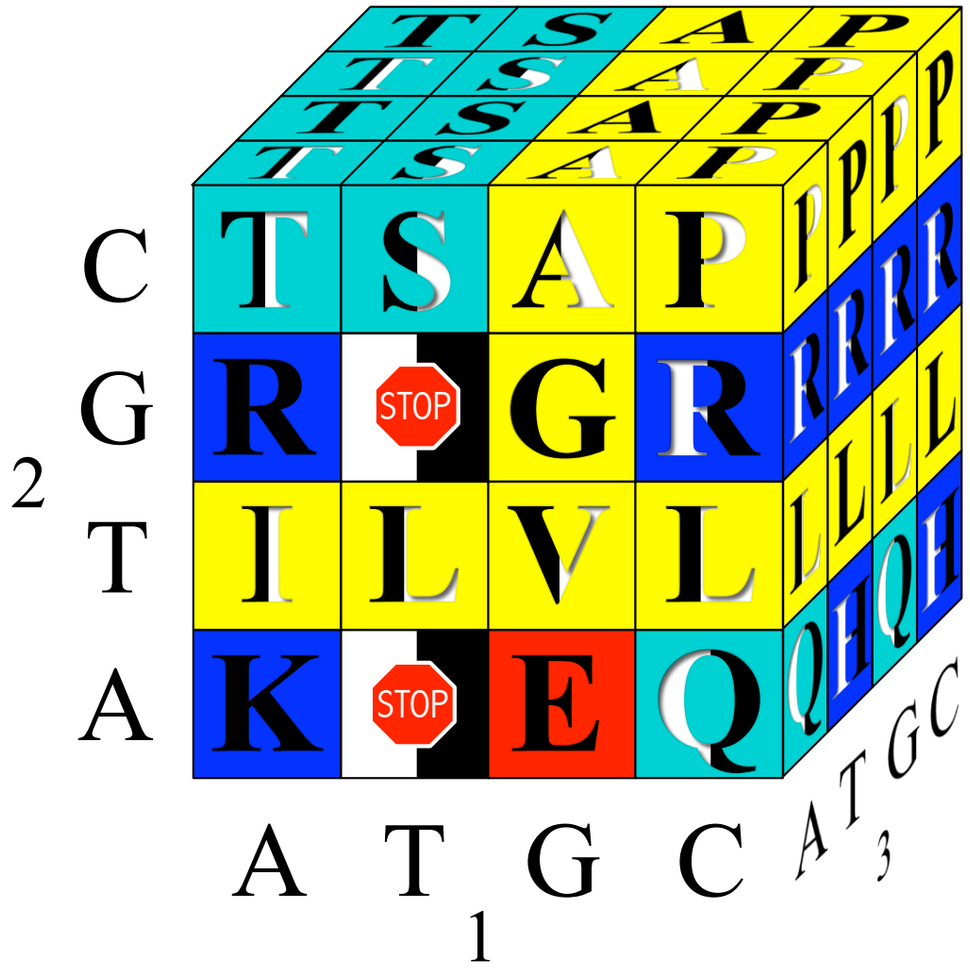
THE MECHANICAL GENOME

New question:

How do you multiplex those two types of genetic information?

Three mechanisms allow for multiplexing

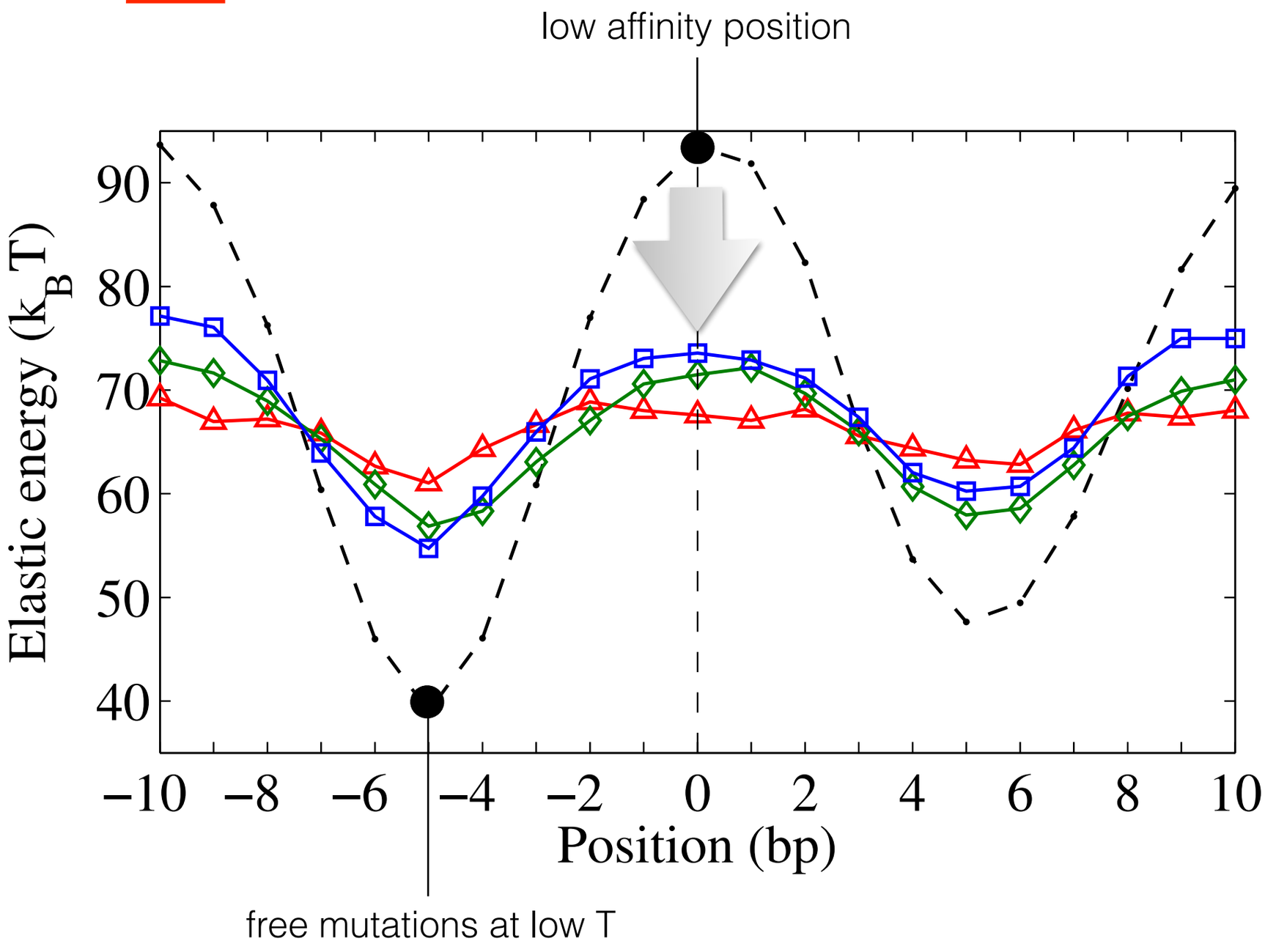
1 protein sequences highly degenerate



from: H. Schiessel:
Biophysics for Beginners:
a Journey through the Cell Nucleus
(Pan Stanford Publishing, 2014)

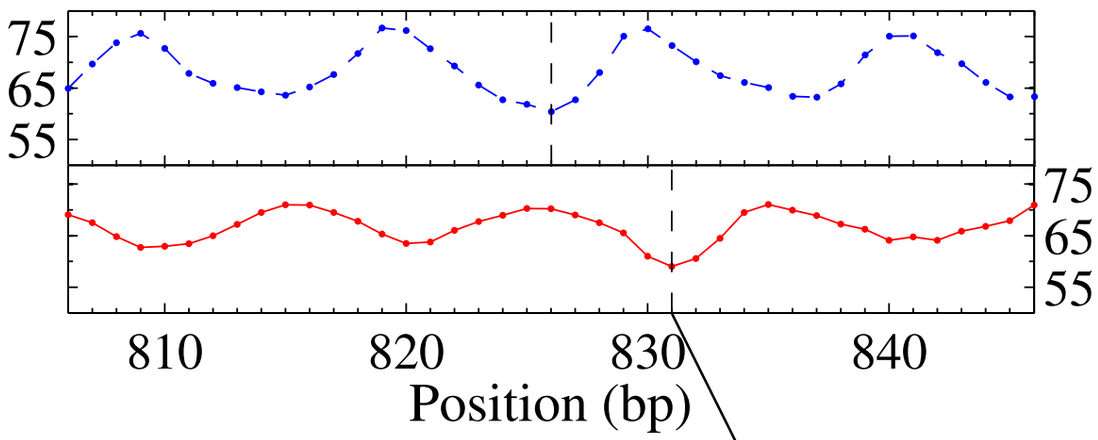
Three mechanisms allow for multiplexing

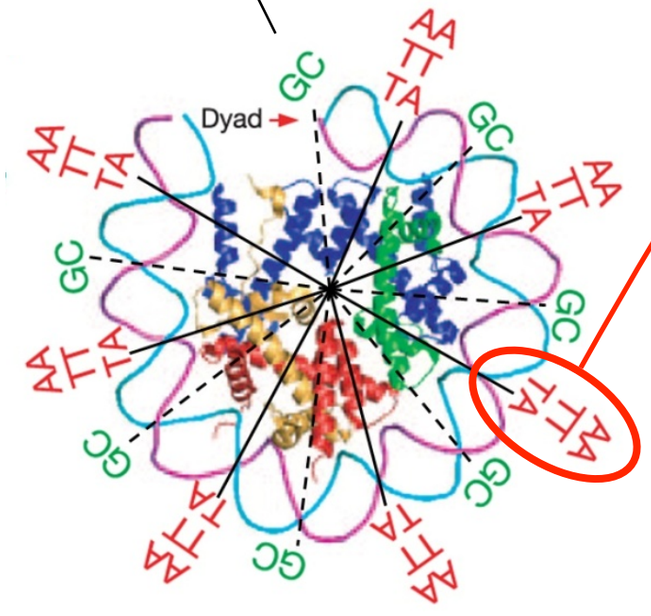
2 genomes not selected for highest affinity



Three mechanisms allow for multiplexing

3 plasticity due to mechanical nature of DNA readout





… GAA ACT …

… Glu Thr …

GAA ACA
GAG ACT
ACG
ACC

suboptimal location of a good basepair step

Three mechanisms allow for multiplexing

1 protein sequences highly degenerate

2 genomes not selected for highest affinity

3 plasticity due to mechanical nature of DNA readout

Jonathan Widom
“nucleosome positioning” at the KITP conference
“Soft Matter Physics Approaches to Biology”
Santa Barbara, May 23rd 2011

THE MECHANICAL GENOME

Summary of what we found so far:

The nucleosome positioning rules are mechanical in nature.

Classical genetic and mechanical information can be multiplexed, based on three different mechanisms.

Does this mean that we have proven
the existence of a mechanical genome, which
would be the result of the mechanical evolution of
DNA molecules?

No!

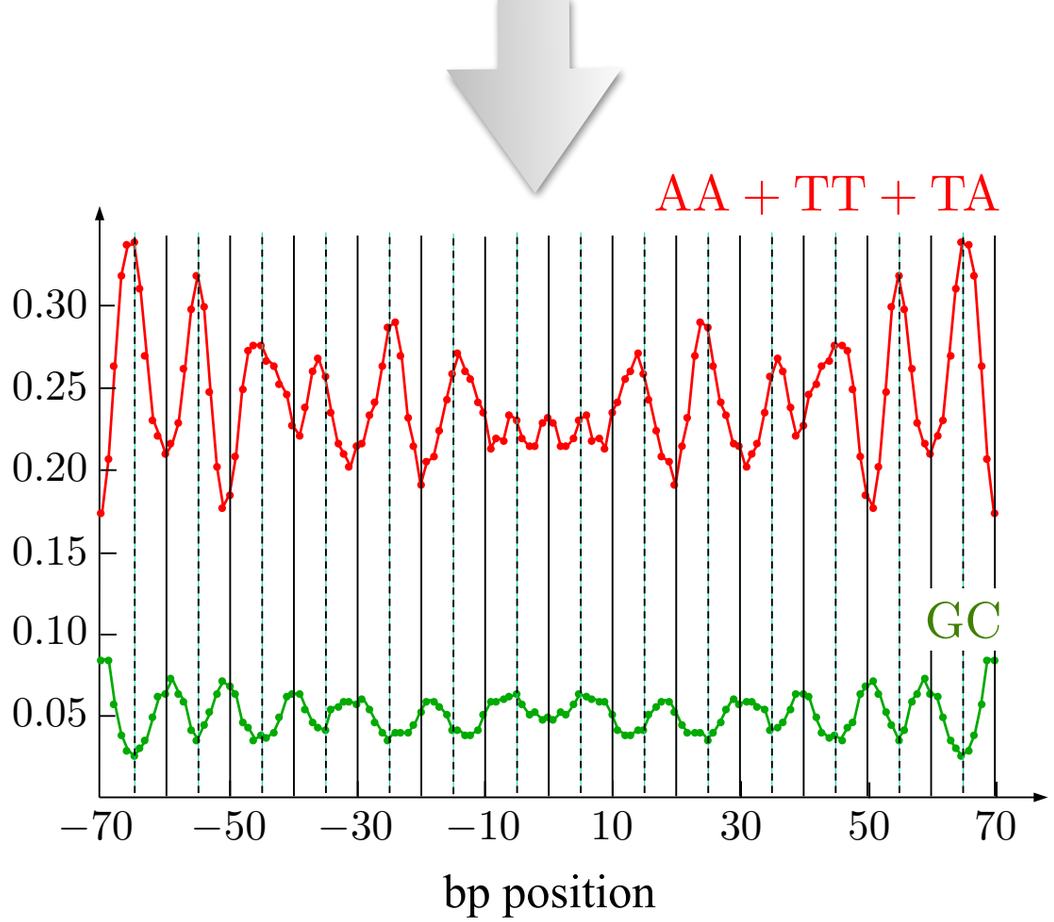
Unfortunately not!

Alternative scenario

Genomes do not care where nucleosomes are.

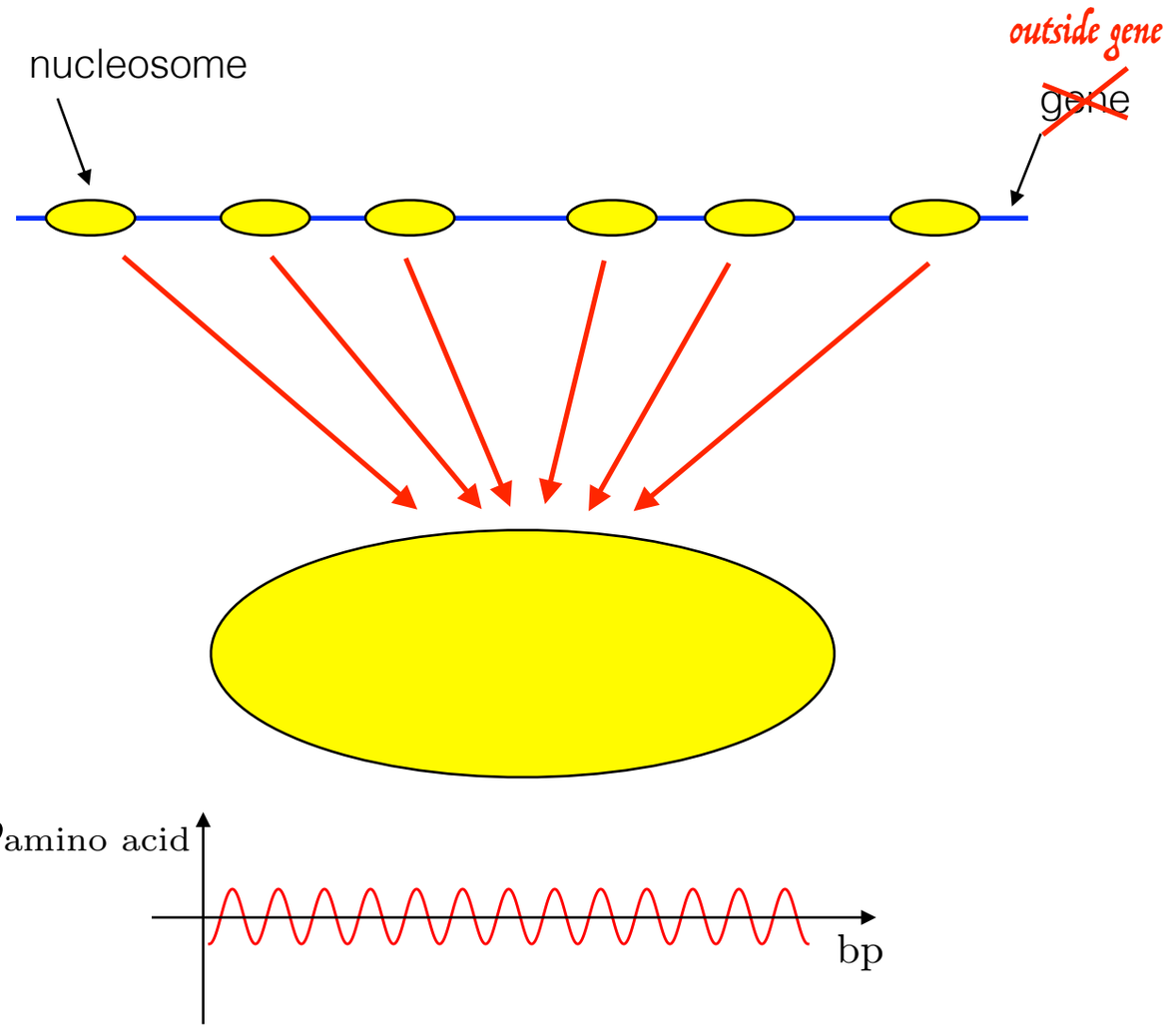
The mechanical variations along DNA molecules are
just random site products of given sequences.

Nucleosomes bind where the mechanical energy is lowest
and experimentalists map these positions.

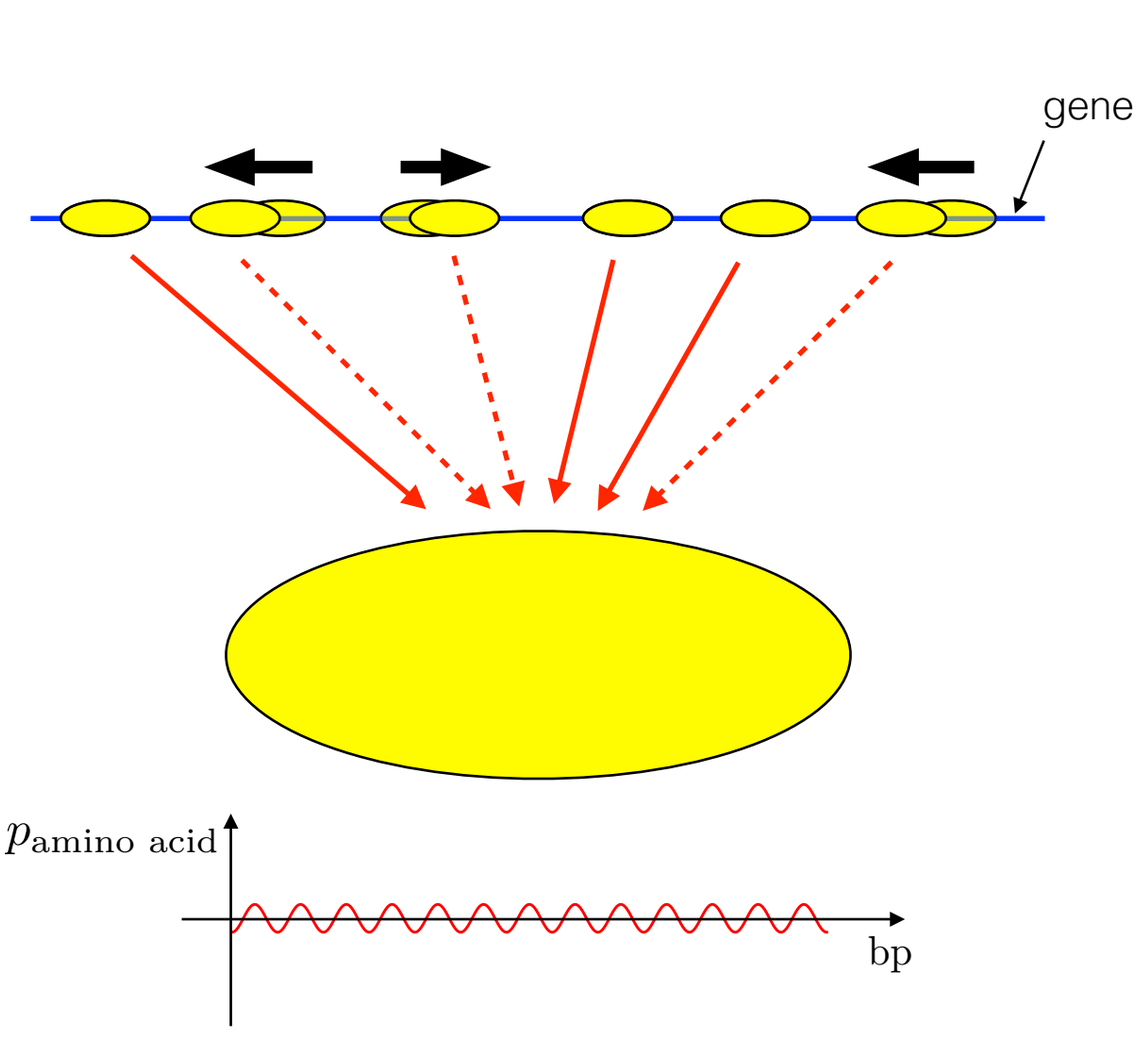


The two scenarios

no multiplexing

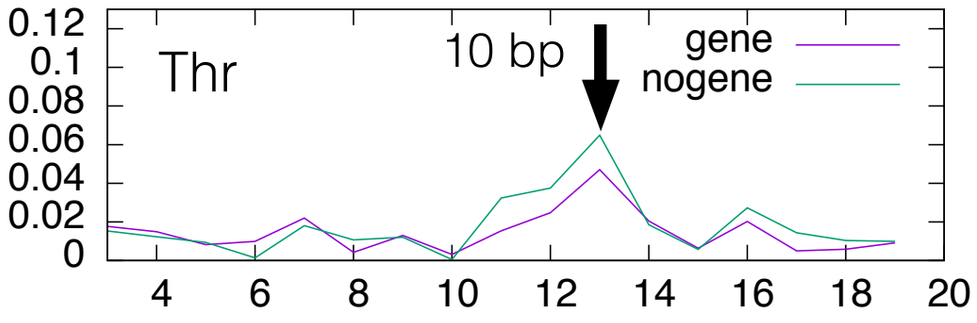


multiplexing

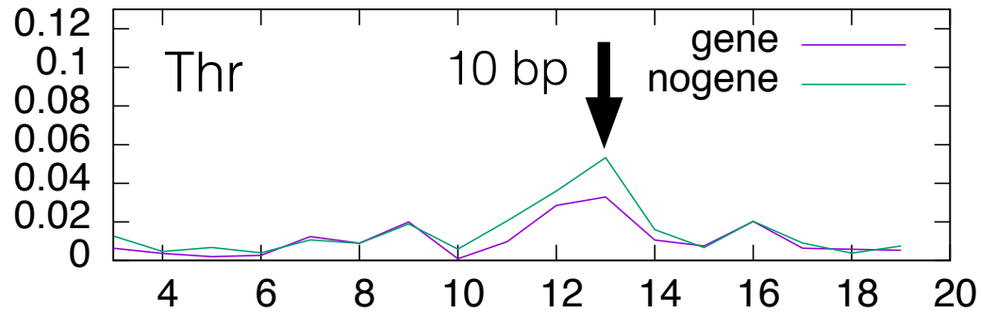


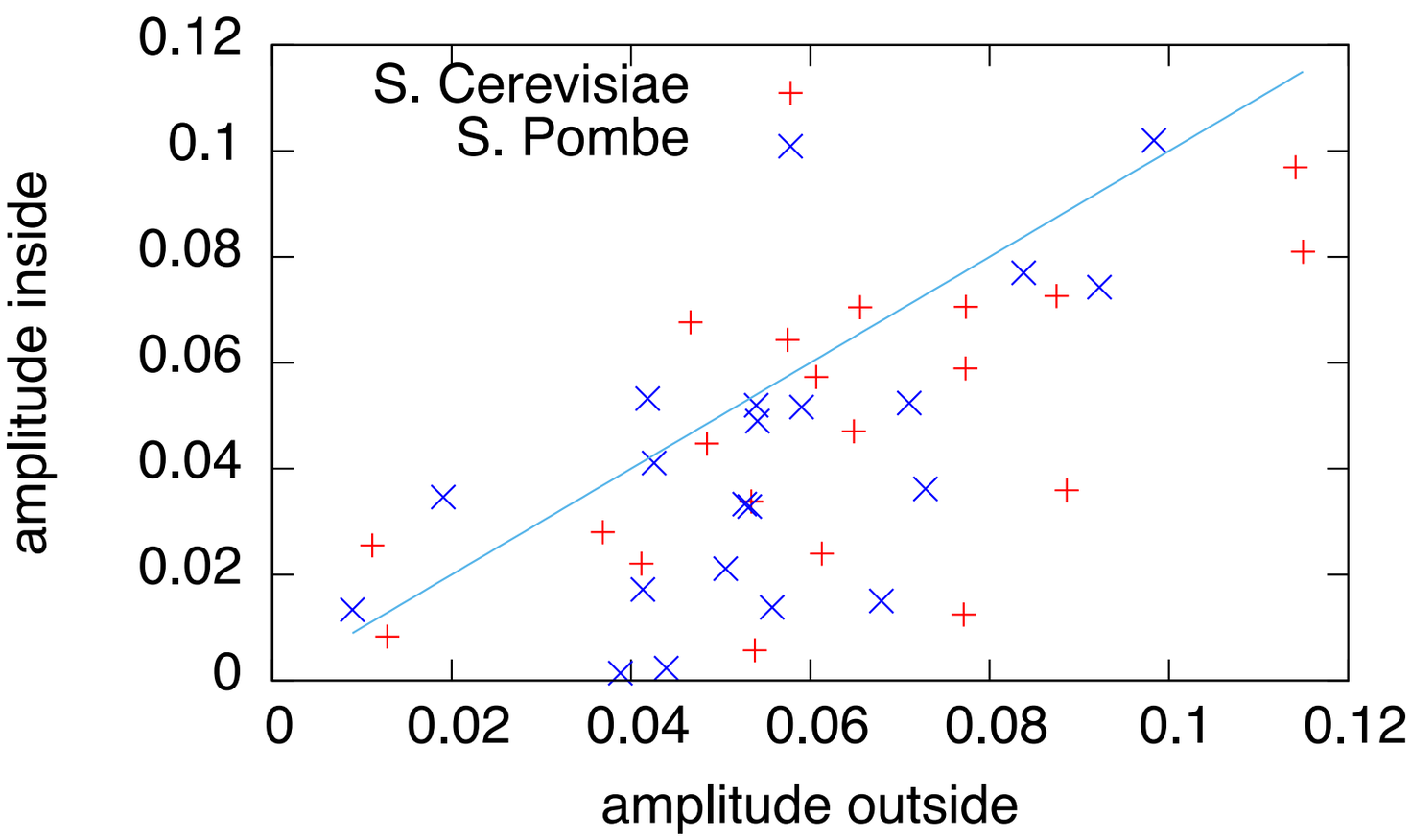
Direct test for multiplexing

S. cerevisiae



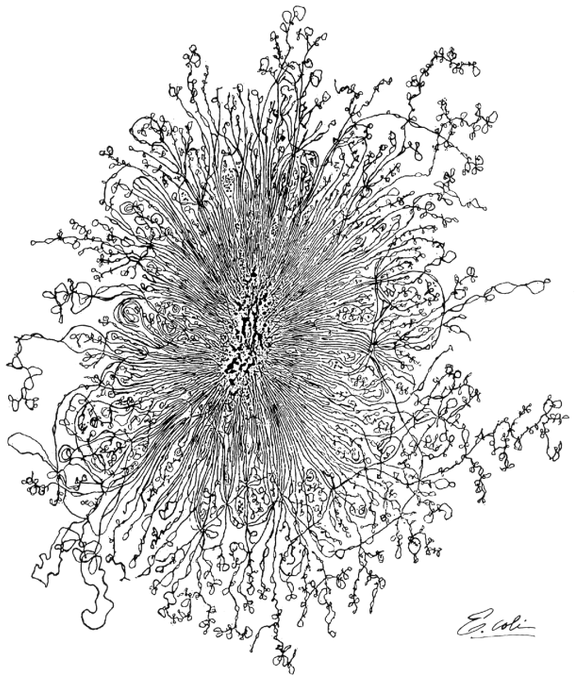
S. pombe





The three domains of life

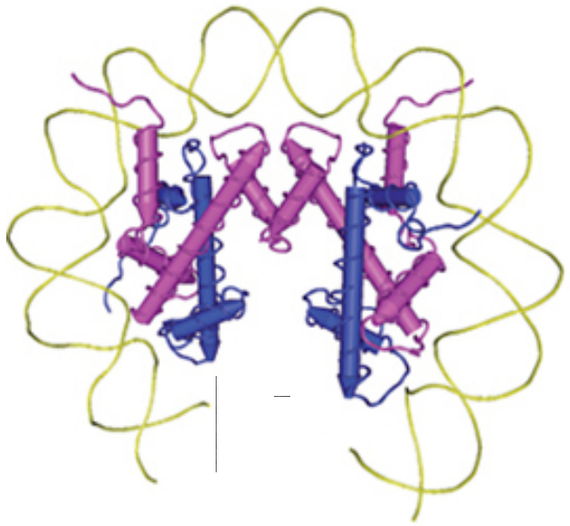
bacteria



E. coli after lysis

http://www.pitt.edu/~mcs2/ecoli.html

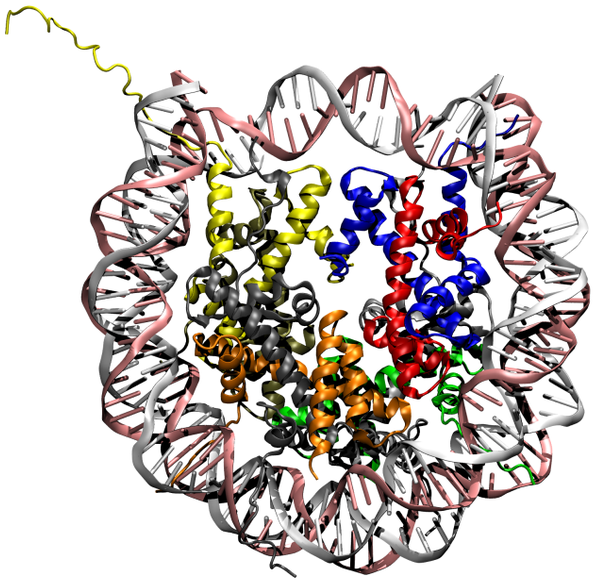
archaea



“proto-nucleosome”

Talbet & Henikoff, 2010

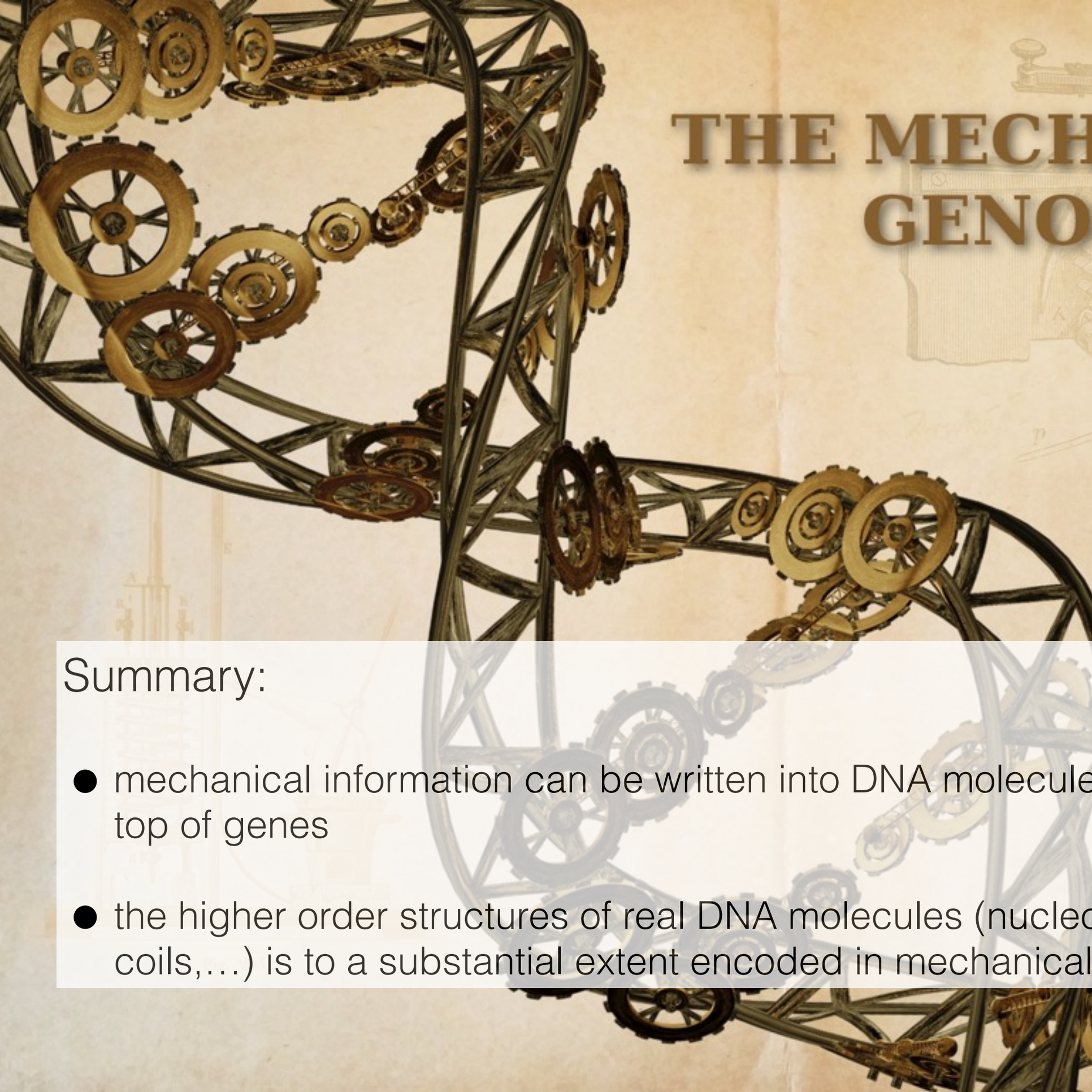
eukaryotes



nucleosome

Luger et al., 1997

THE MECHANICAL GENOME



Summary:

- mechanical information can be written into DNA molecules, even on top of genes
- the higher order structures of real DNA molecules (nucleosomes, super-coils,…) is to a substantial extent encoded in mechanical genomes