

# Nucleic Acids

Alberto Perez

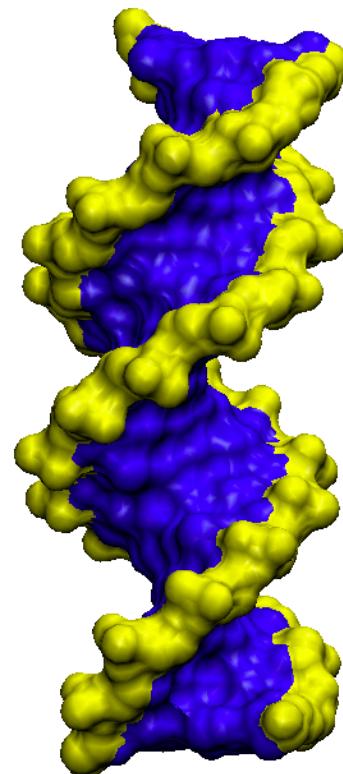
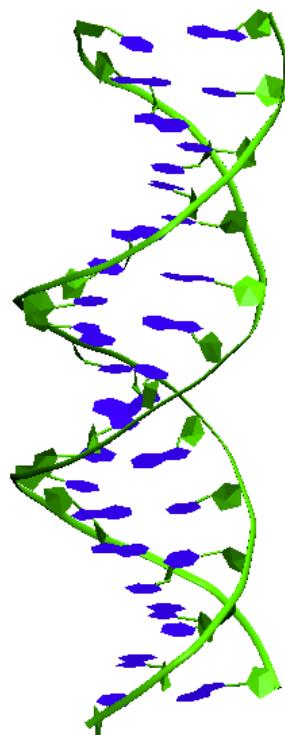
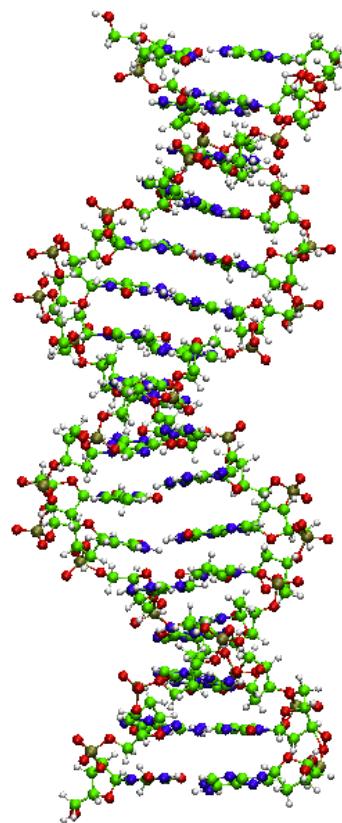
<https://perez.chem.ufl.edu>

Molecular Recognition Group



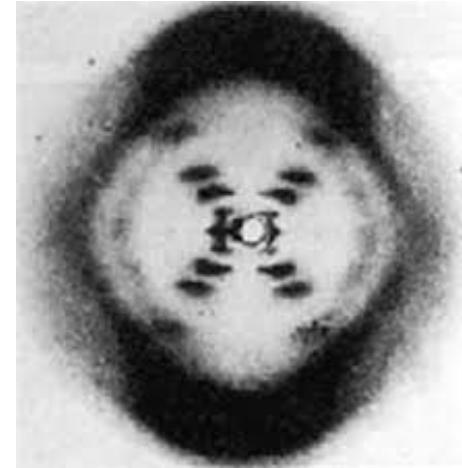
# Classical view

DNA → genes → RNA(m) → Proteins

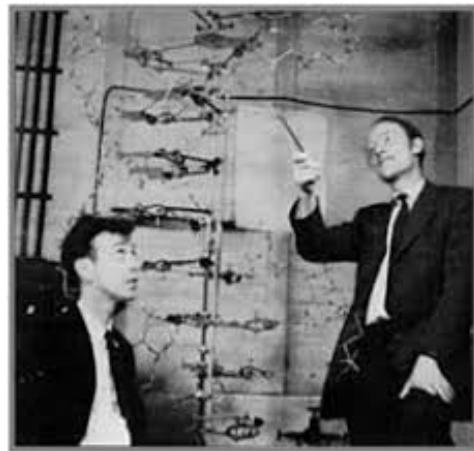




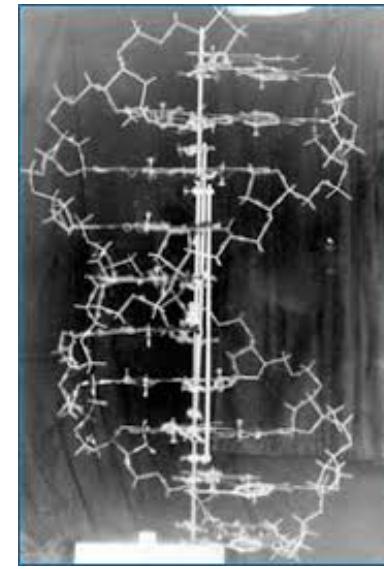
**Rosalind Franklin**



**x-ray structure DNA**



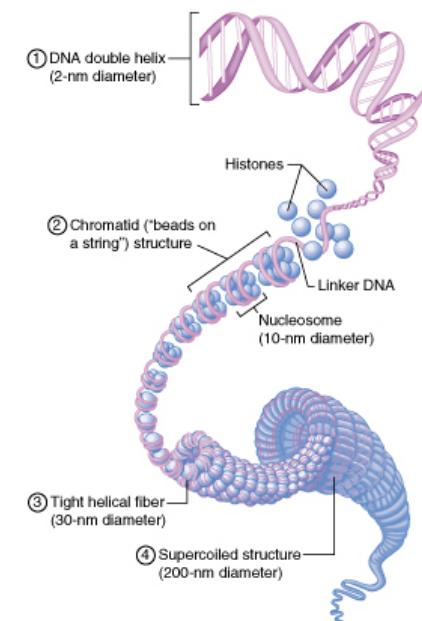
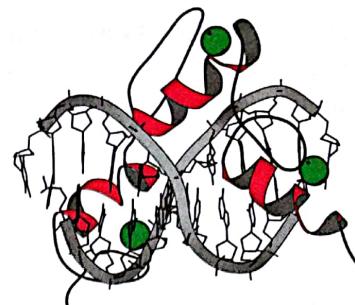
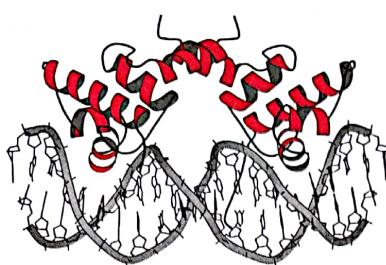
**Watson and Crick**



**DNA model**

# Structural levels

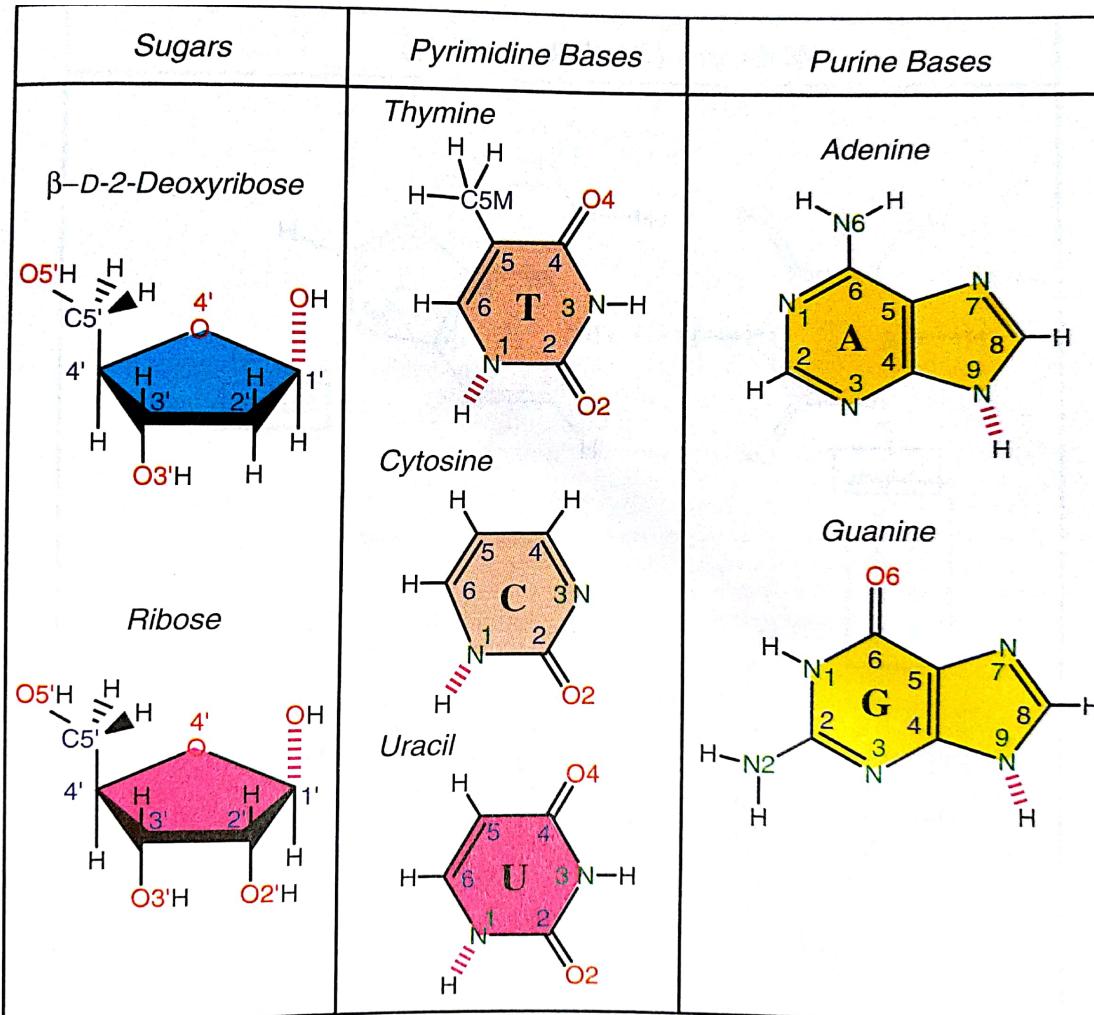
- Primary structure (sequence: ATTACCTGCTGGGAA)
- Secondary structure (helices)
- Tertiary structure (super-helices)
- Quaternary structure (complexes)



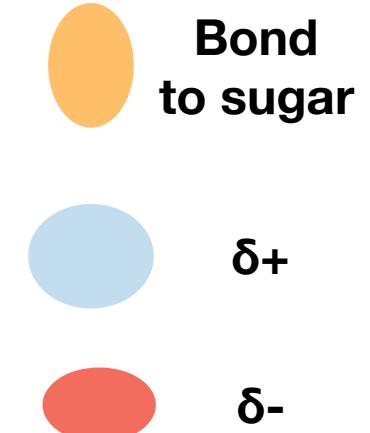
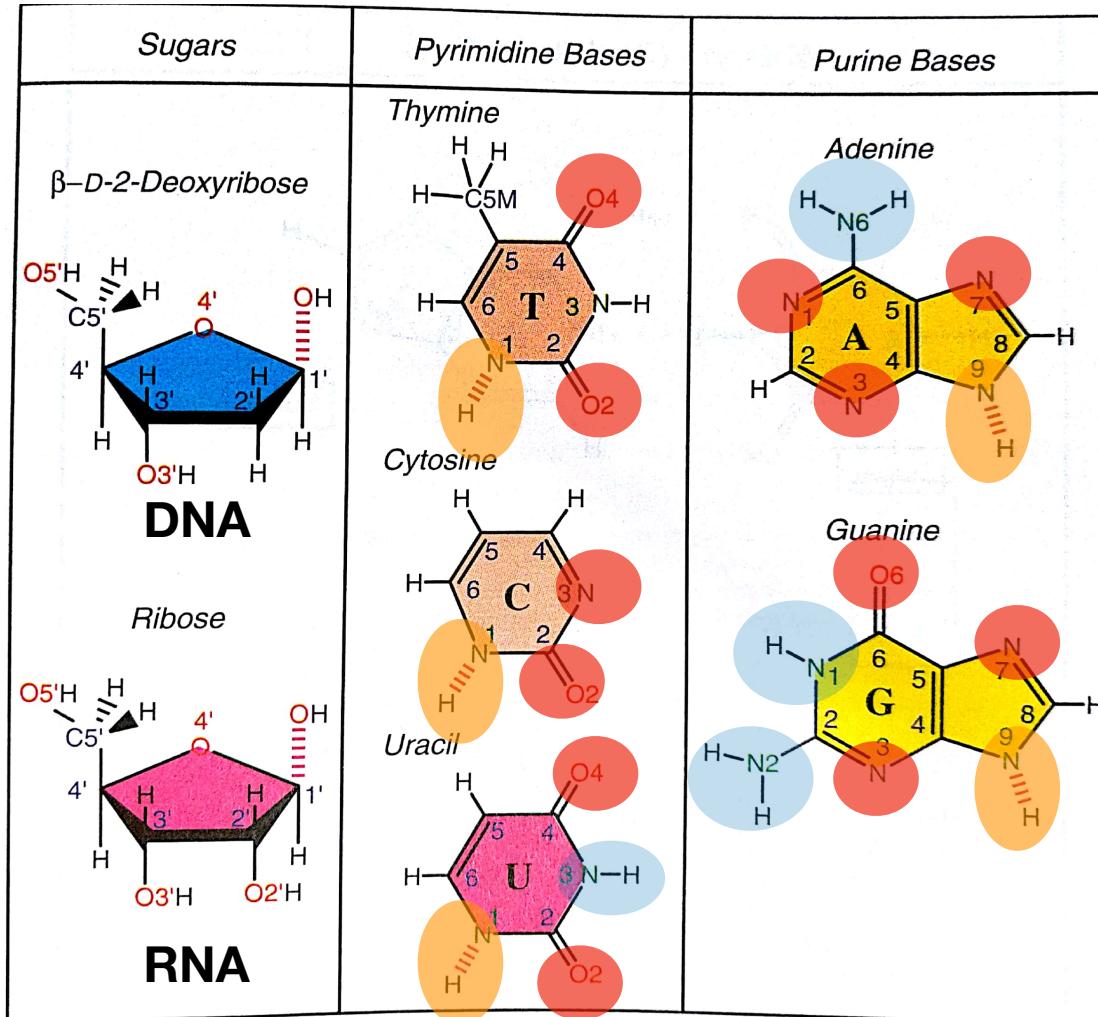
Copyright © 2001 Benjamin Cummings, an imprint of Addison Wesley Longman, Inc.

# **Structural parts in nucleic acids**

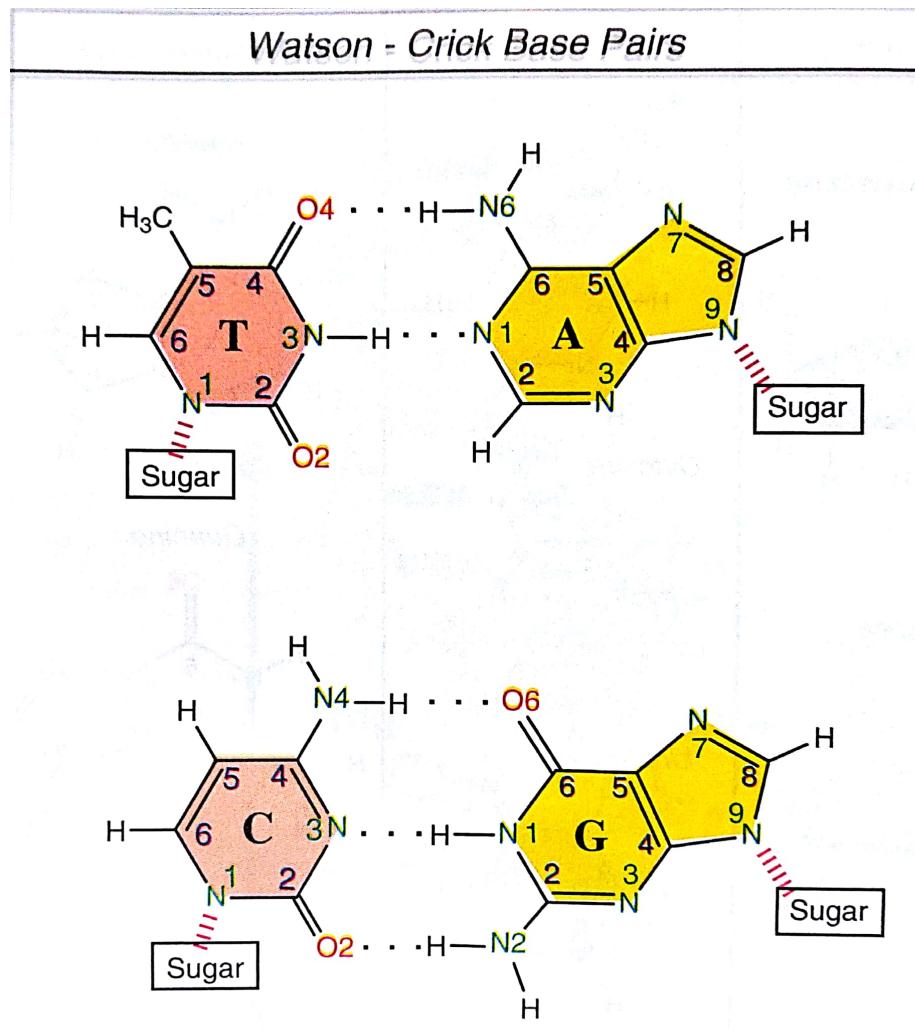
# Nitrogenous bases and sugars



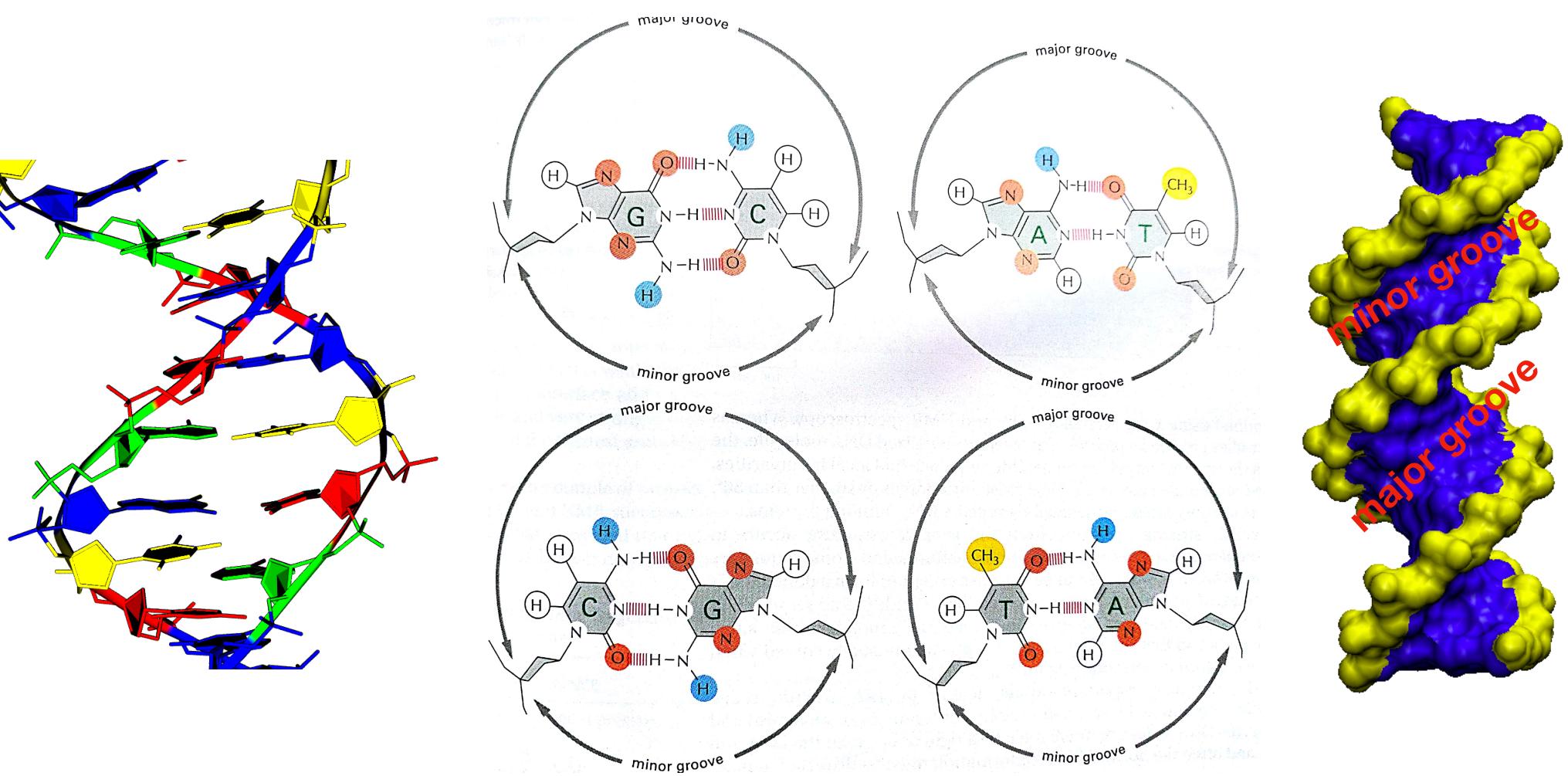
# Bases give a unique H-bond pattern



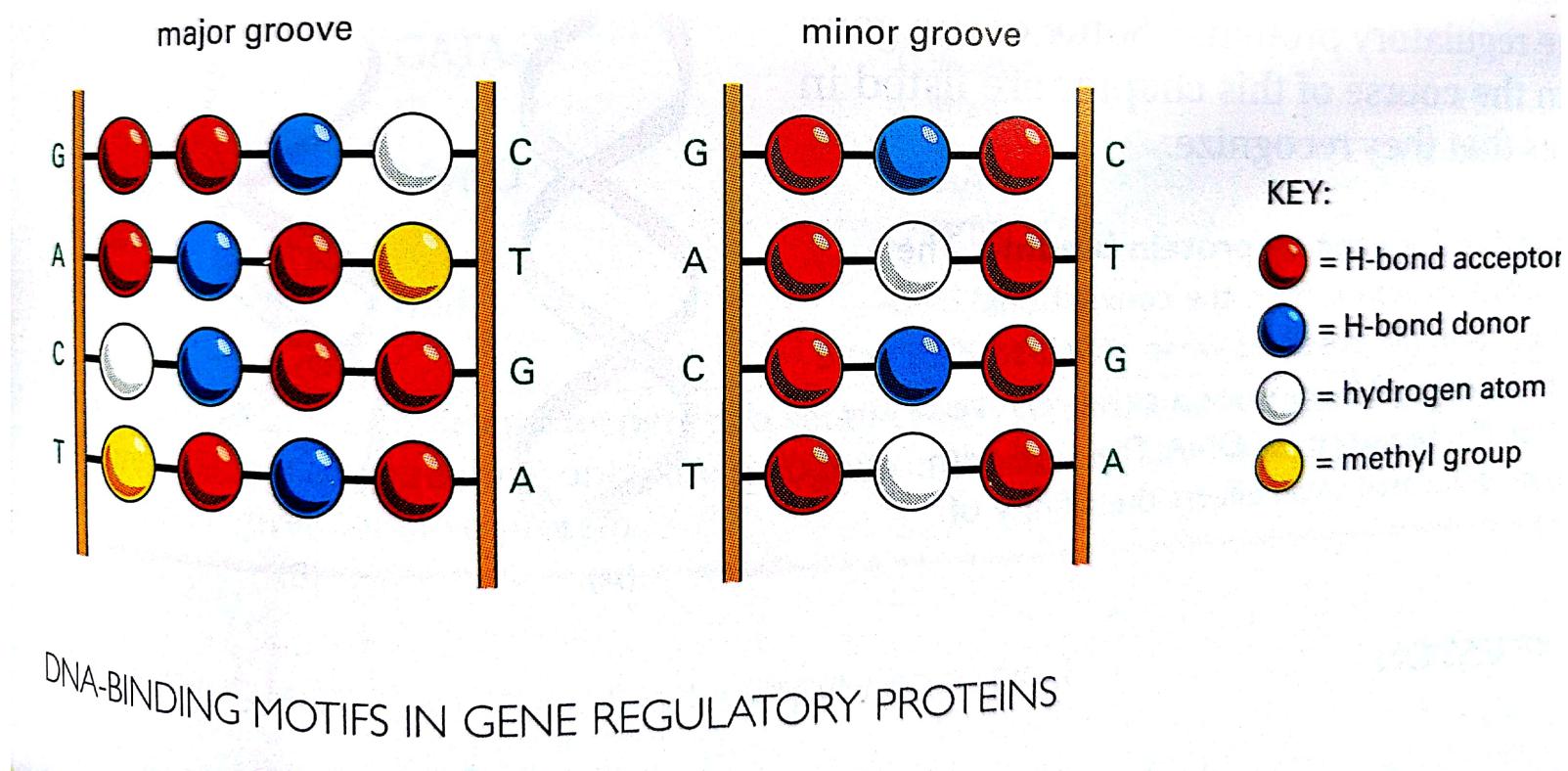
# Canonical hydrogen bond pattern



# Canonical hydrogen bond pattern

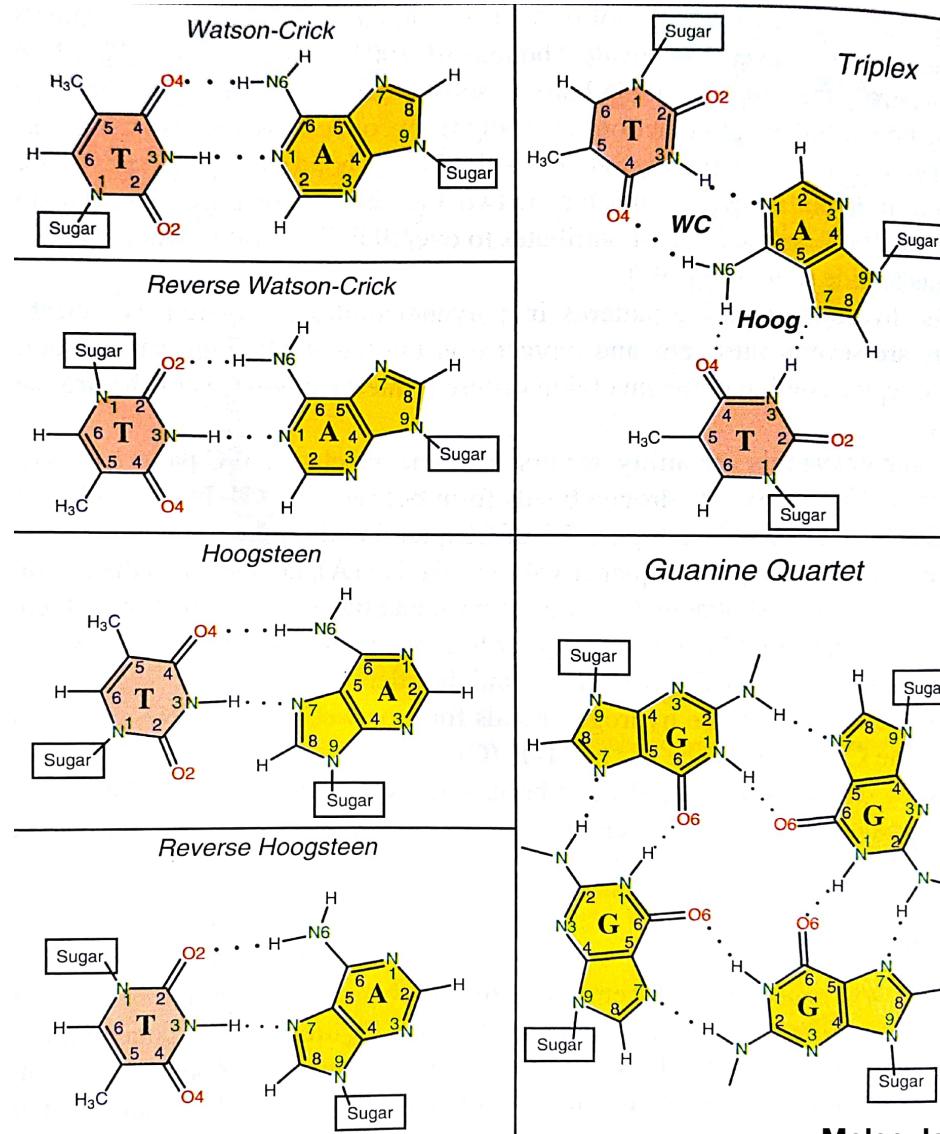


# How other molecules can interact DNA



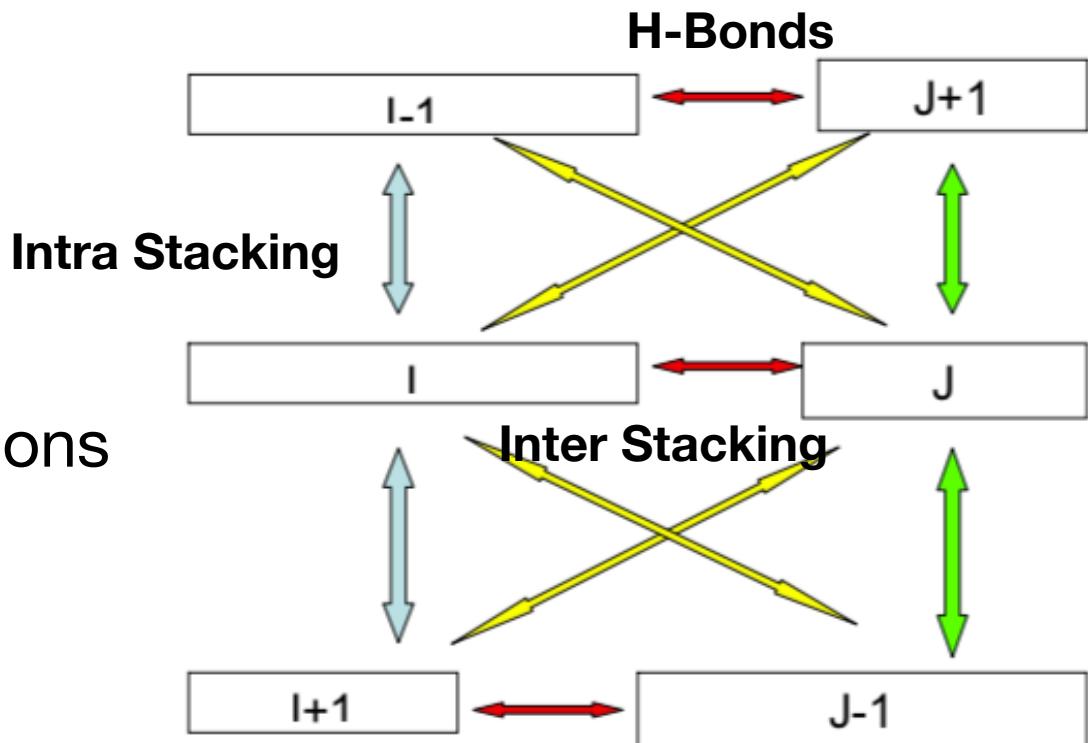
**Not all hydrogen bond donors/acceptors are used  
Possibility to interact with:  
Drugs  
Proteins**

# Many possible non-canonical pairings!

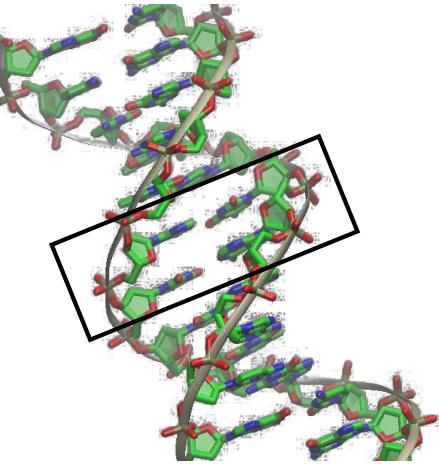


# Base interactions

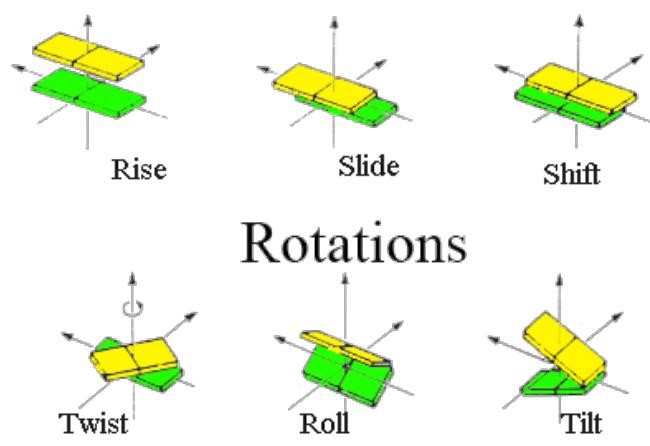
- Stacking interactions
- Hydrogen Bond interactions
- Solvation



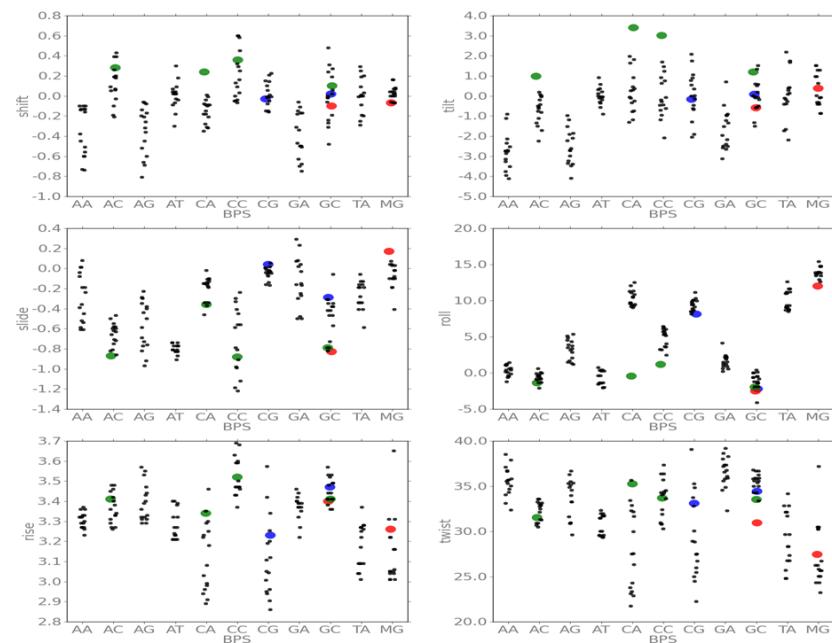
# DNA exhibits DNA sequence dependent properties in terms of structure and flexibility



Structure



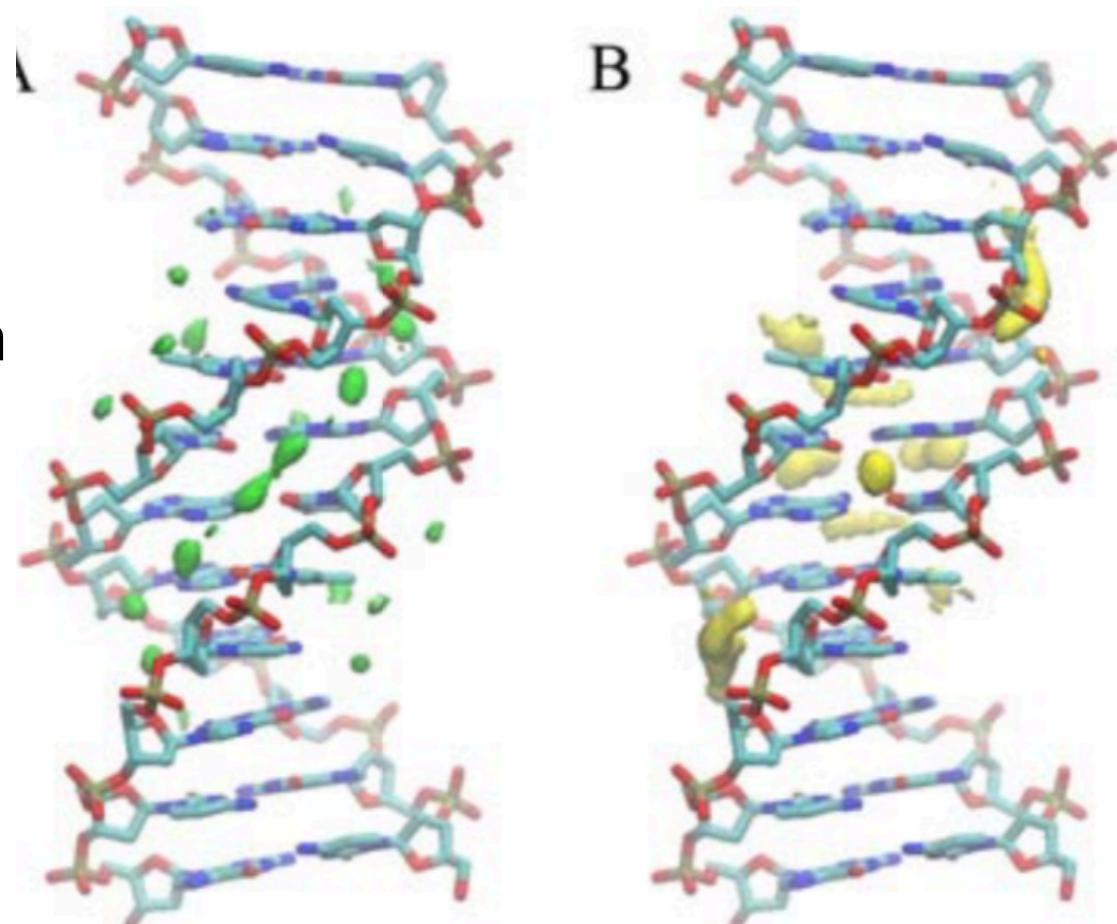
Measure



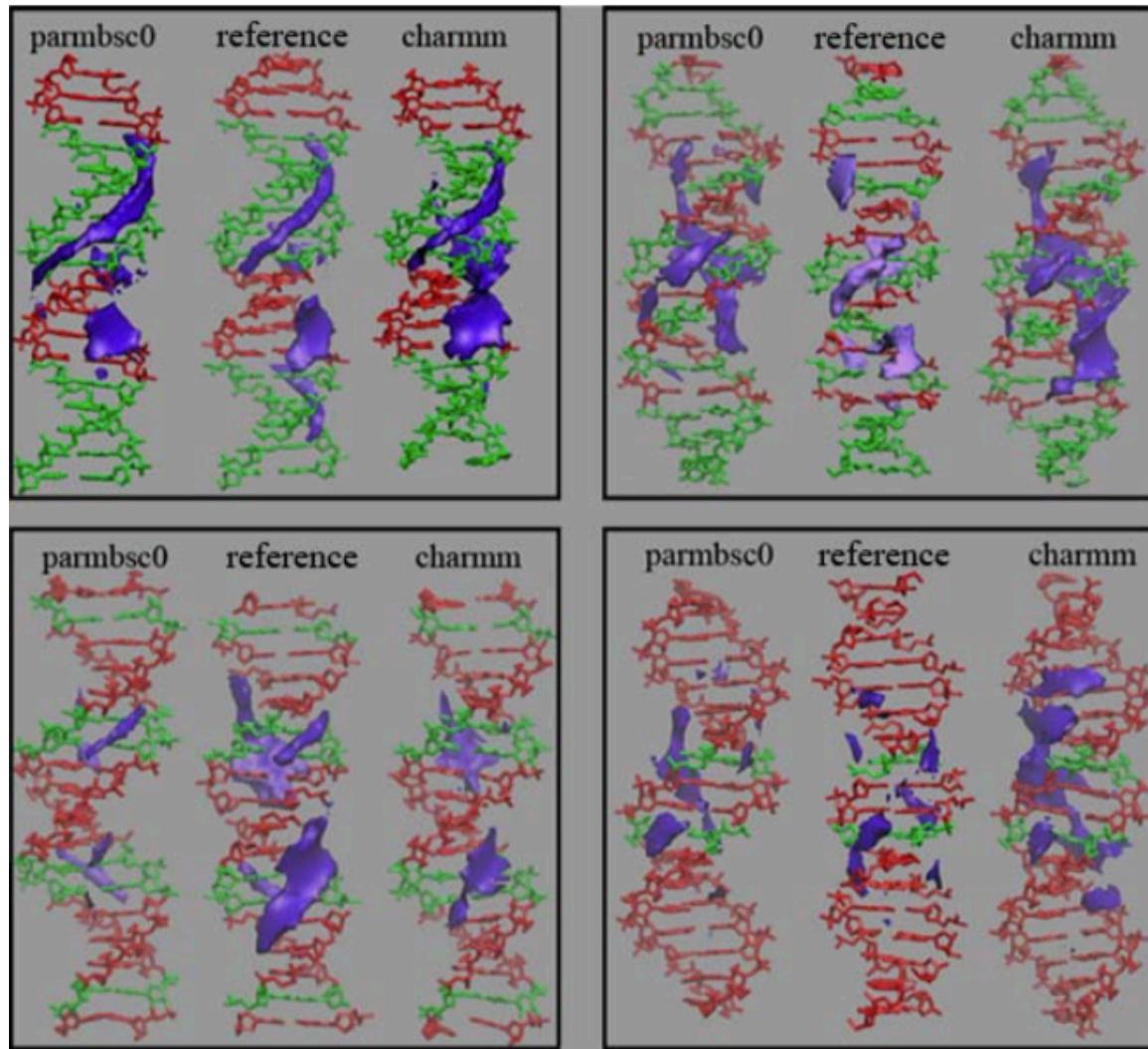
Sequence dependent behavior

# Solvation

- The double chain and arrangement of phosphates (minor groove width and depth allow for regions of preferred interaction with water / ions).
- Seen here are solvation maps where the water density is 2.5 times higher than bulk (green) and ion densities are greater than 2 times bulk (yellow).

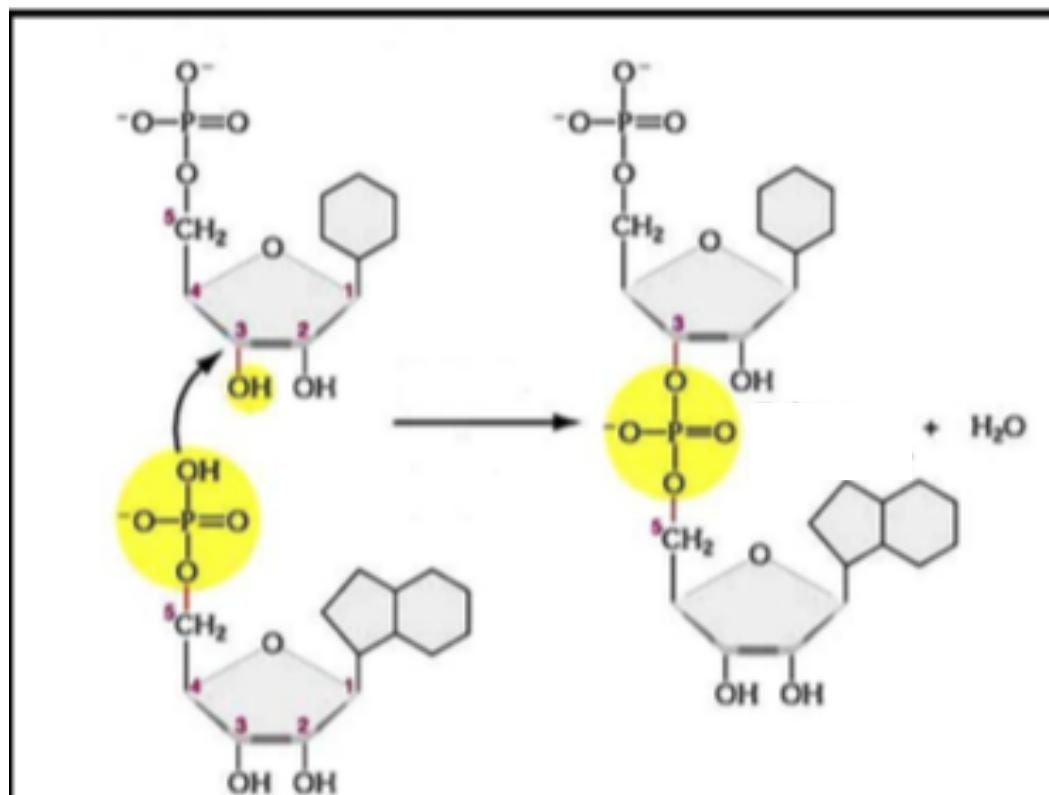


# Properties are sequence dependent



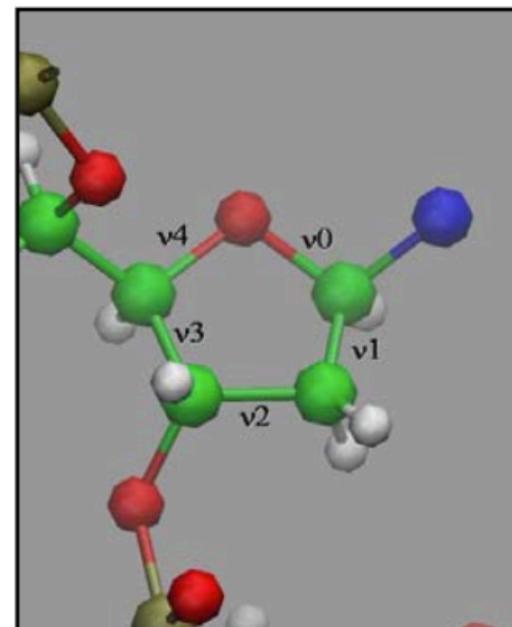
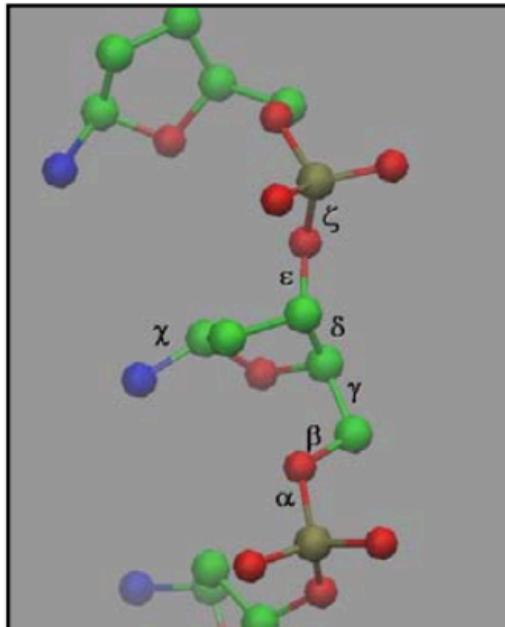
C:G pairs  
A:T pairs

# Phosphodiester bonds form the NA backbone

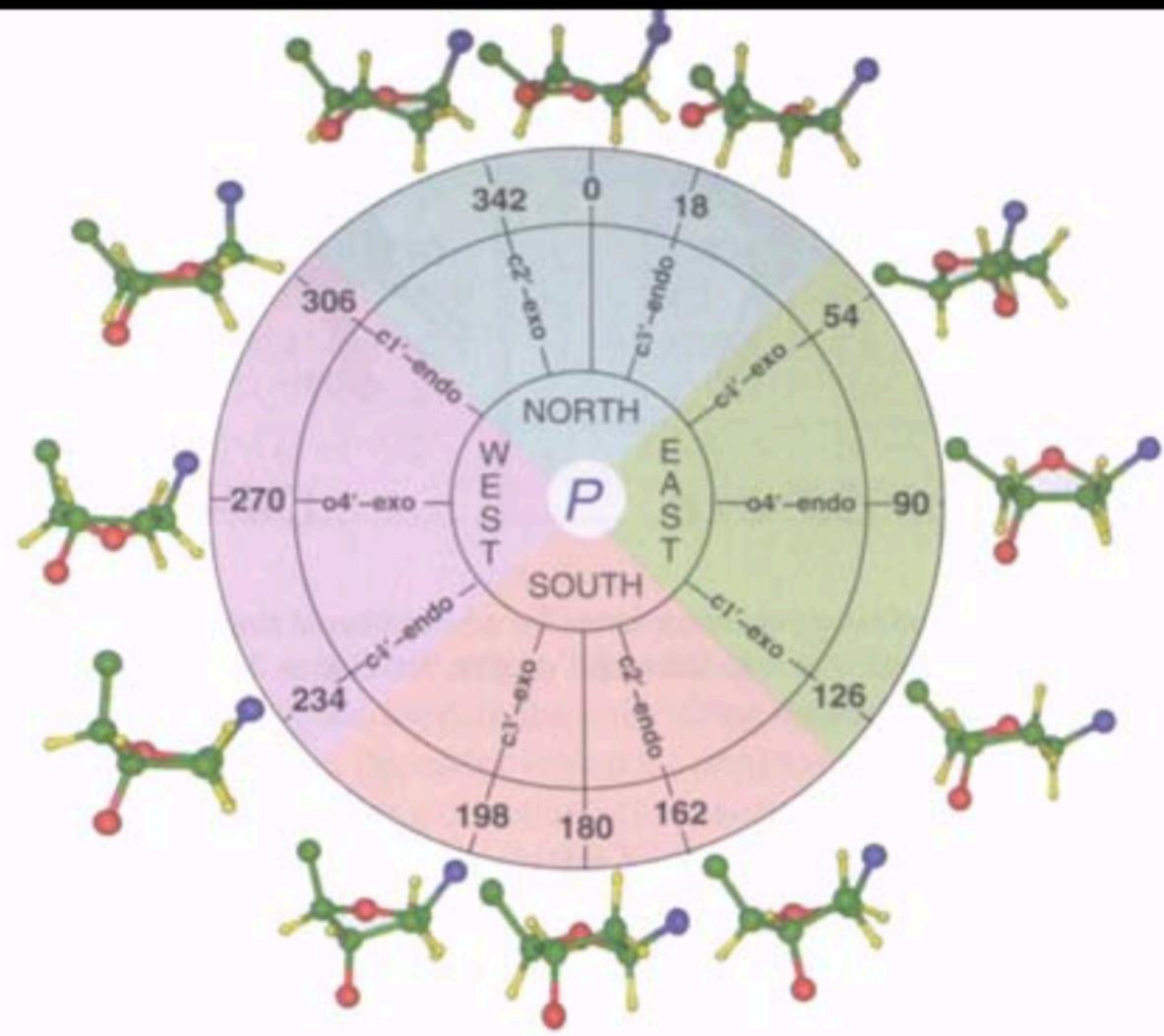


# The sugar

- Highly flexible. Links between the phosphates. RNA has a ribose. DNA has a 2-deoxyRibose. Ribose has more limited conformational preferences than 2-deoxyRibose in this environment.
- If bases give the recognition properties for NA (direct readout), the sugars and phosphodiesters give the conformational flexibility (indirect readout)

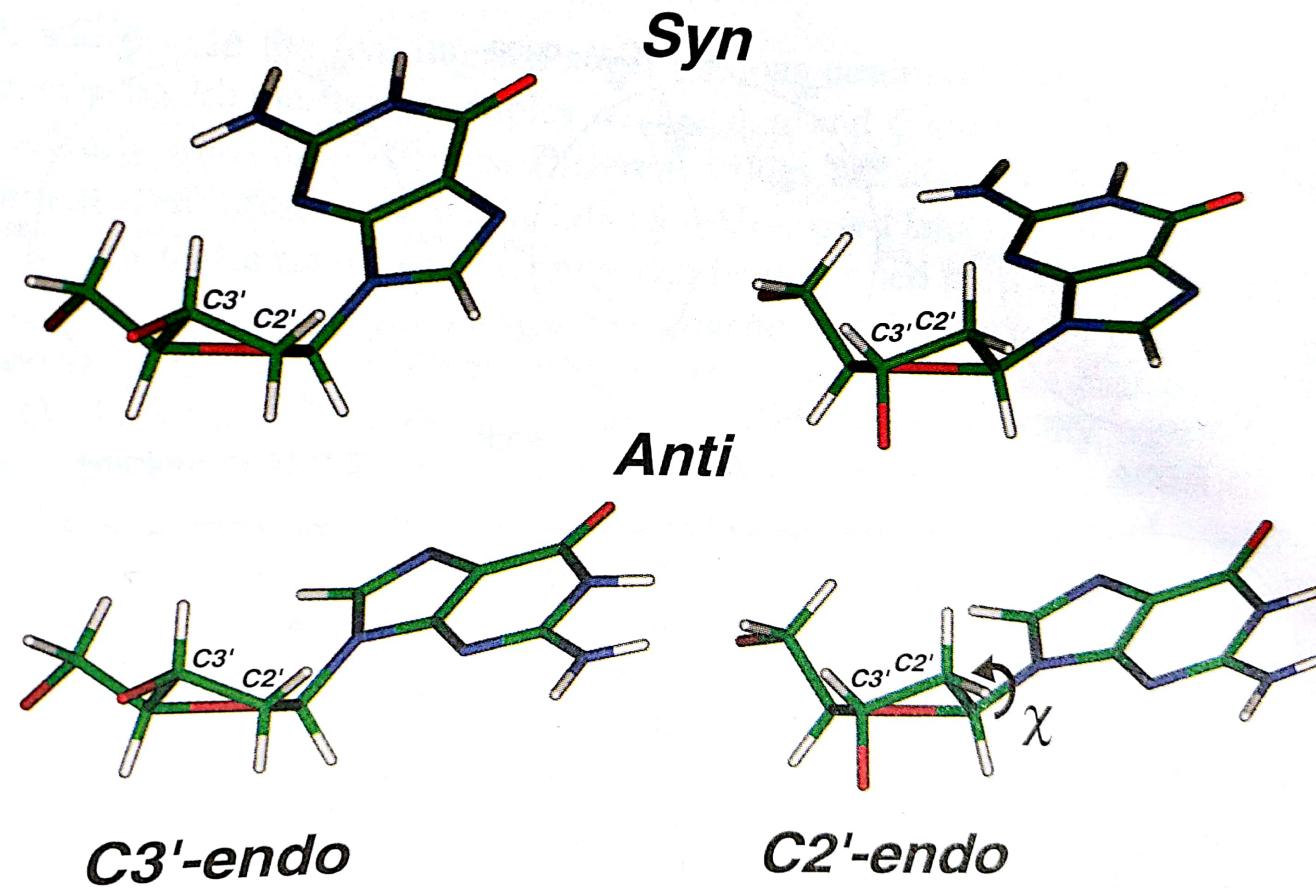


# Sugar puckering



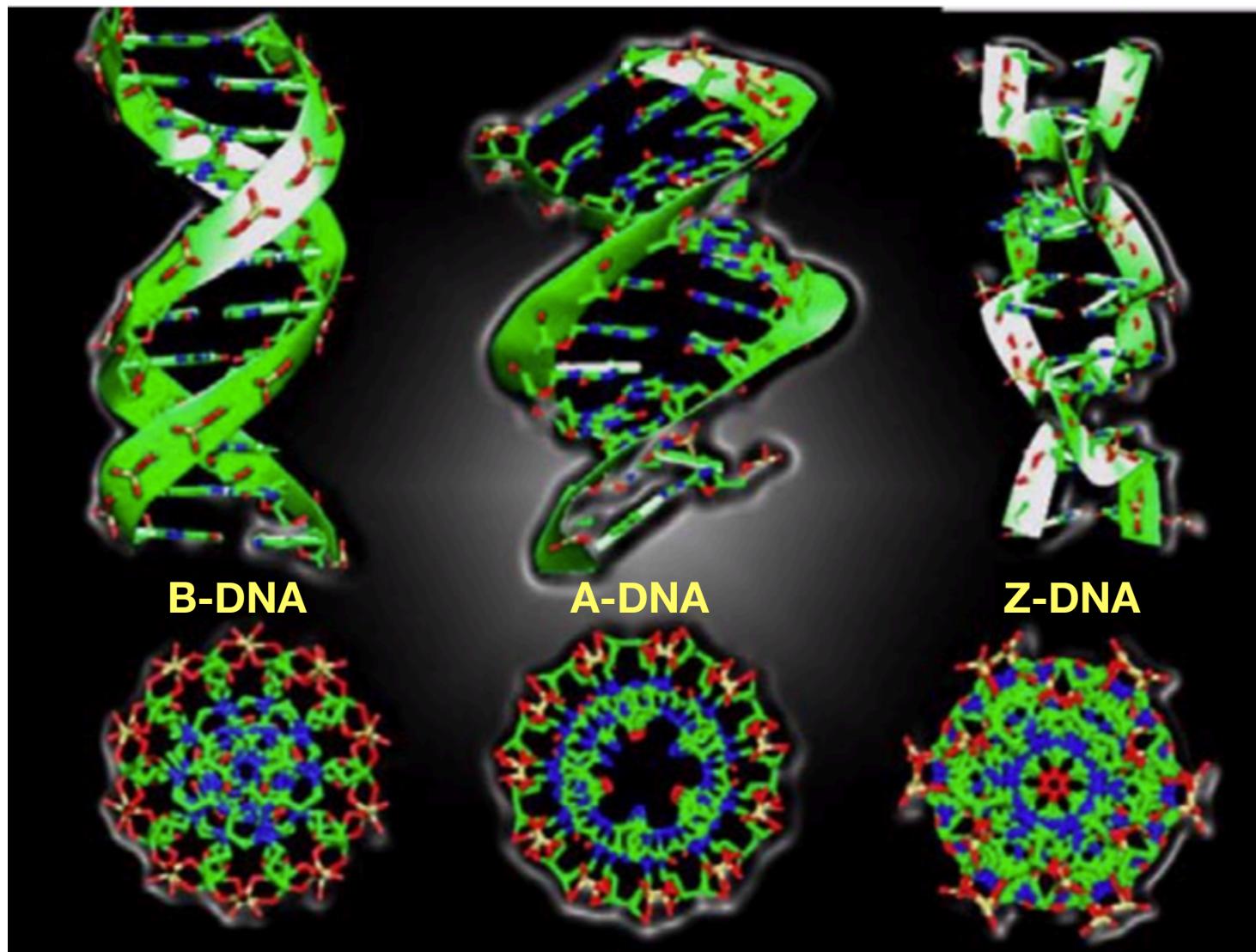
- North (C3' endo) and South (C2' endo) are the most common
- W is never accessible
- RNA, A-DNA → N puckering
- B-DNA: mostly S, some E.

# Base with respect to sugar can be either syn or anti



Anti is the most common. Guanine is the only nucleoside that sometimes goes to Syn

# Solvation preferences, puckering, backbone determine DNA structure

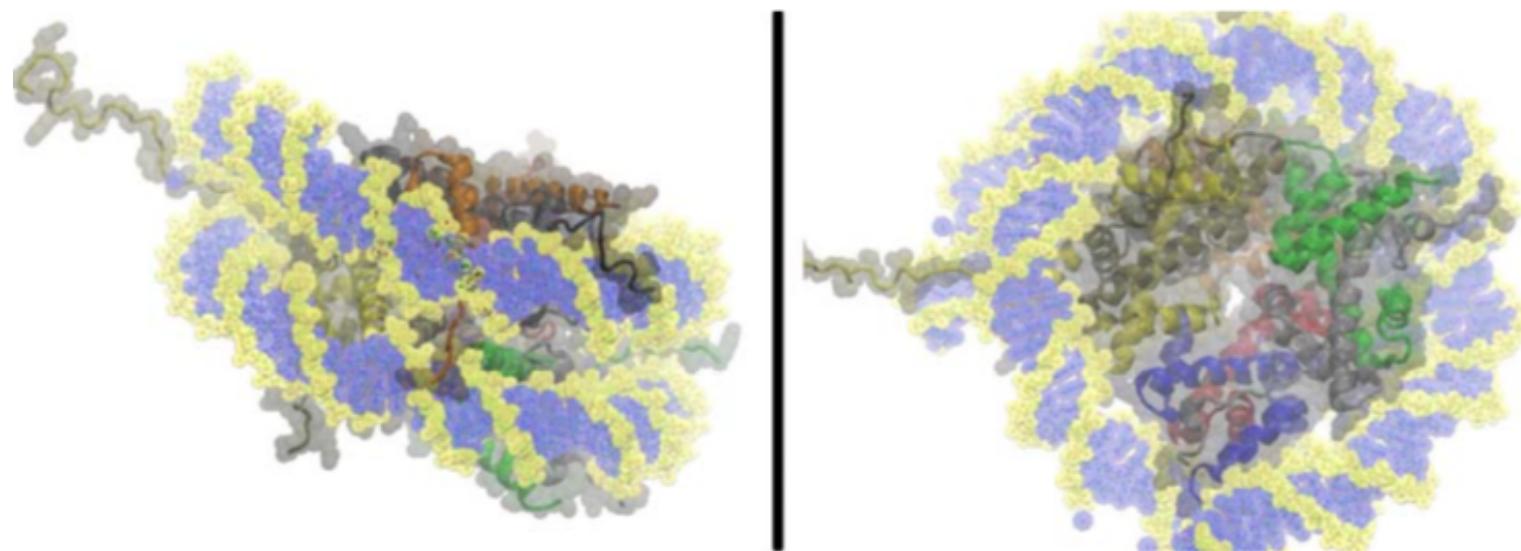


# Higher order DNA sequence

- We have 2 meters of DNA (6 feet) in every cell in our bodies. They have to fit in a nucleus that is  $5\mu\text{m}$  in length.
  - Need to pack DNA
  - ~2% of the DNA codes for proteins. Other regions are regulatory, repressors, enhancers,... involved in gene regulation.
  - Need to keep DNA accessible

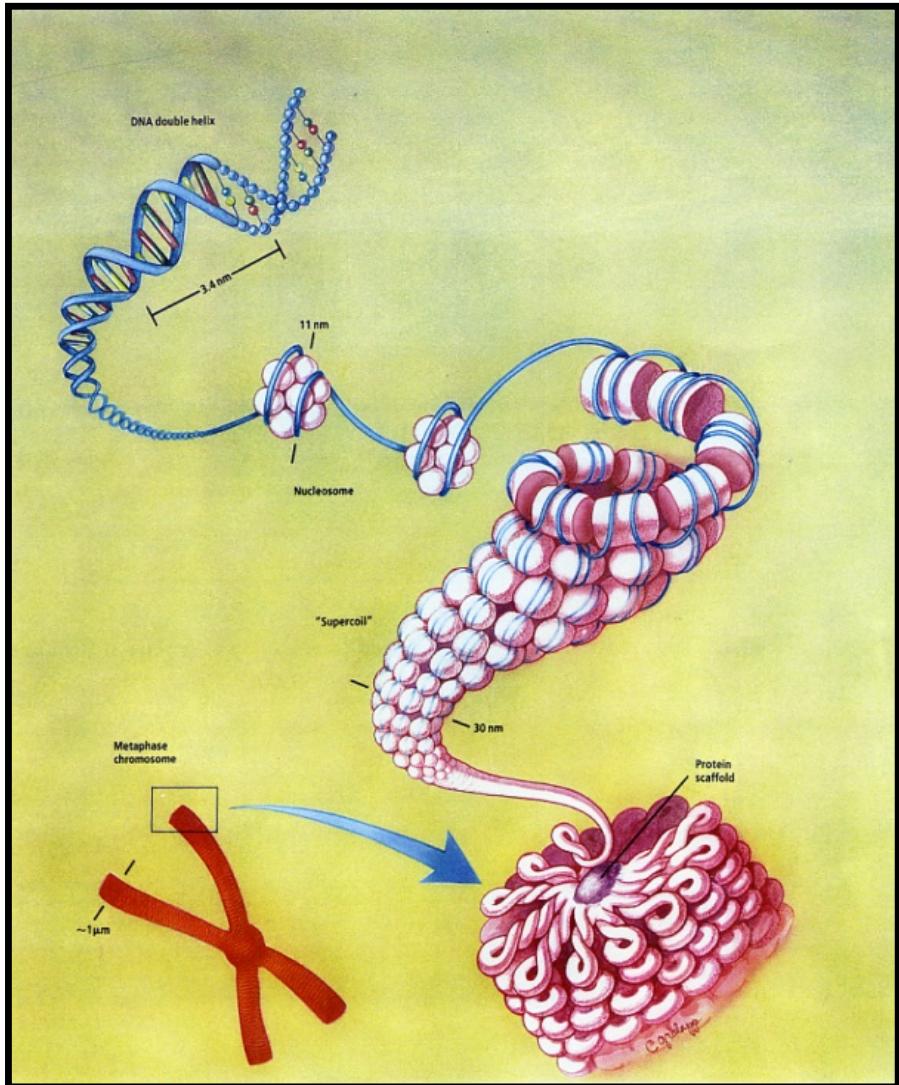


# The nucleosome is the core particle for packing

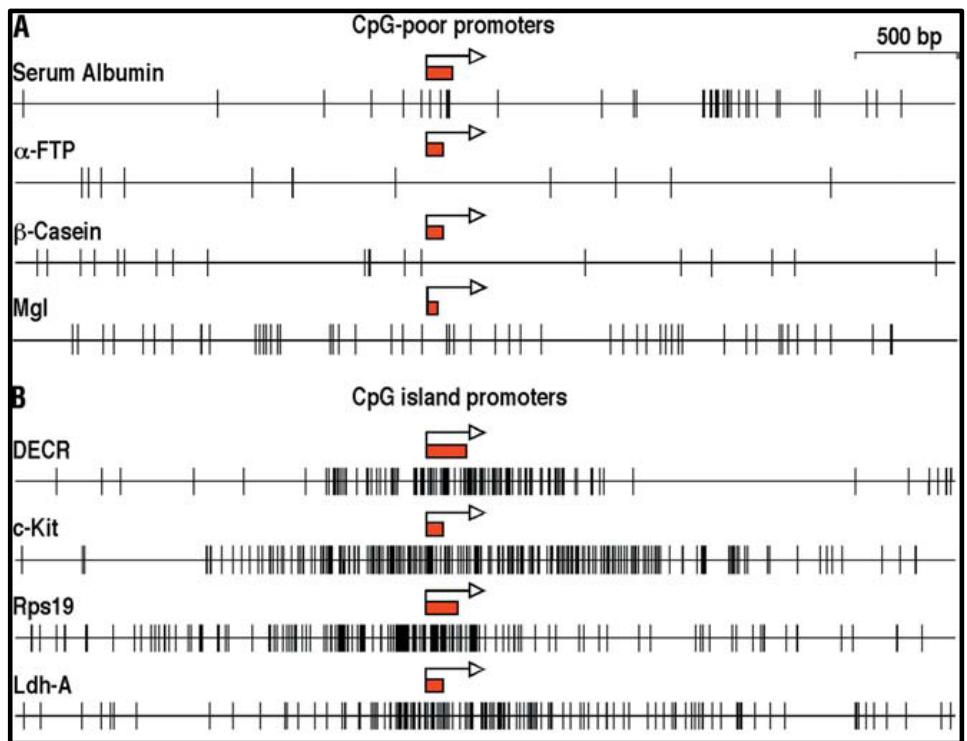


- 147 bases of DNA wrap themselves 1.75 times around the histone octamer (H2A, H2B, H3 and H4).
- Resulting structure is 220Å (10x compression).
- Linker DNA (~100 base pairs) is used to connect nucleosomes

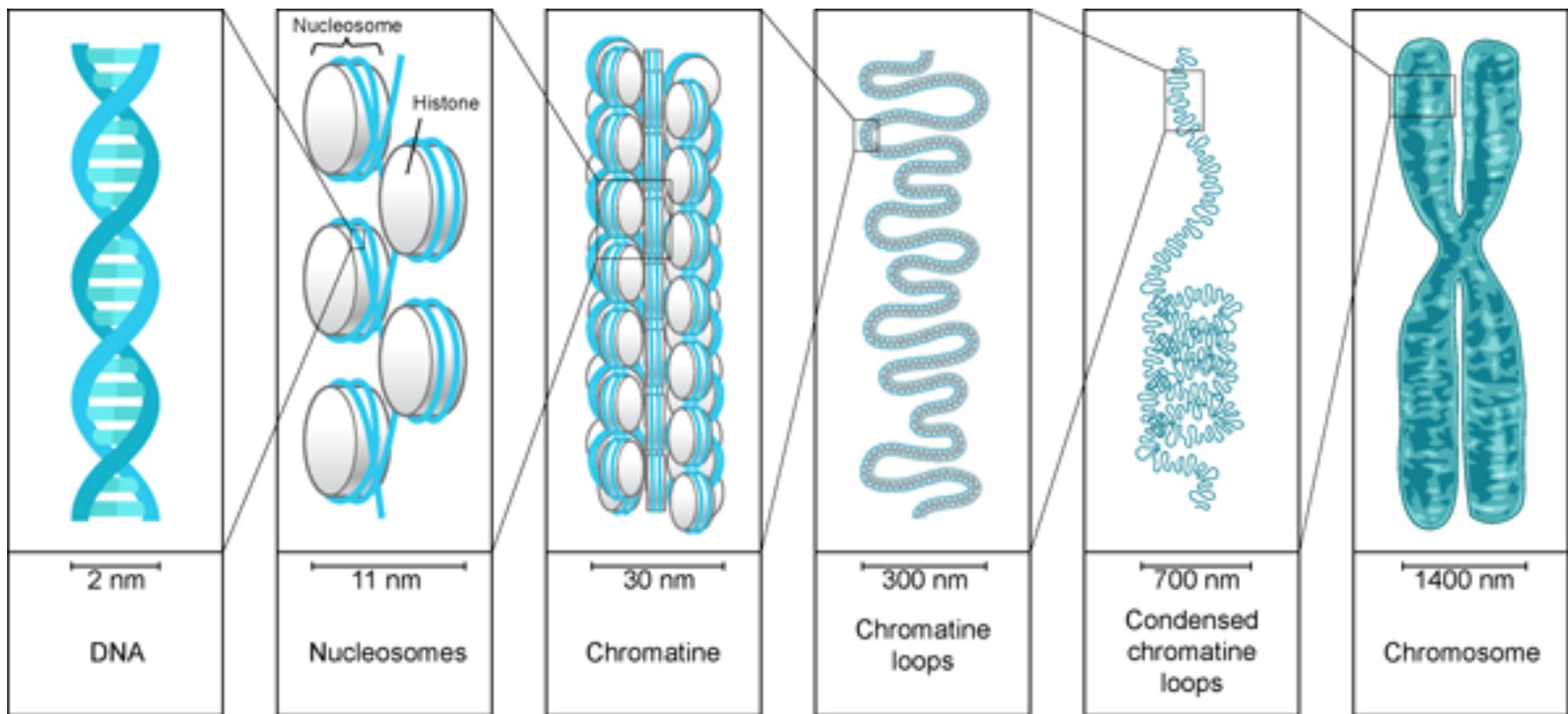
# The nucleosome is the core particle for packing



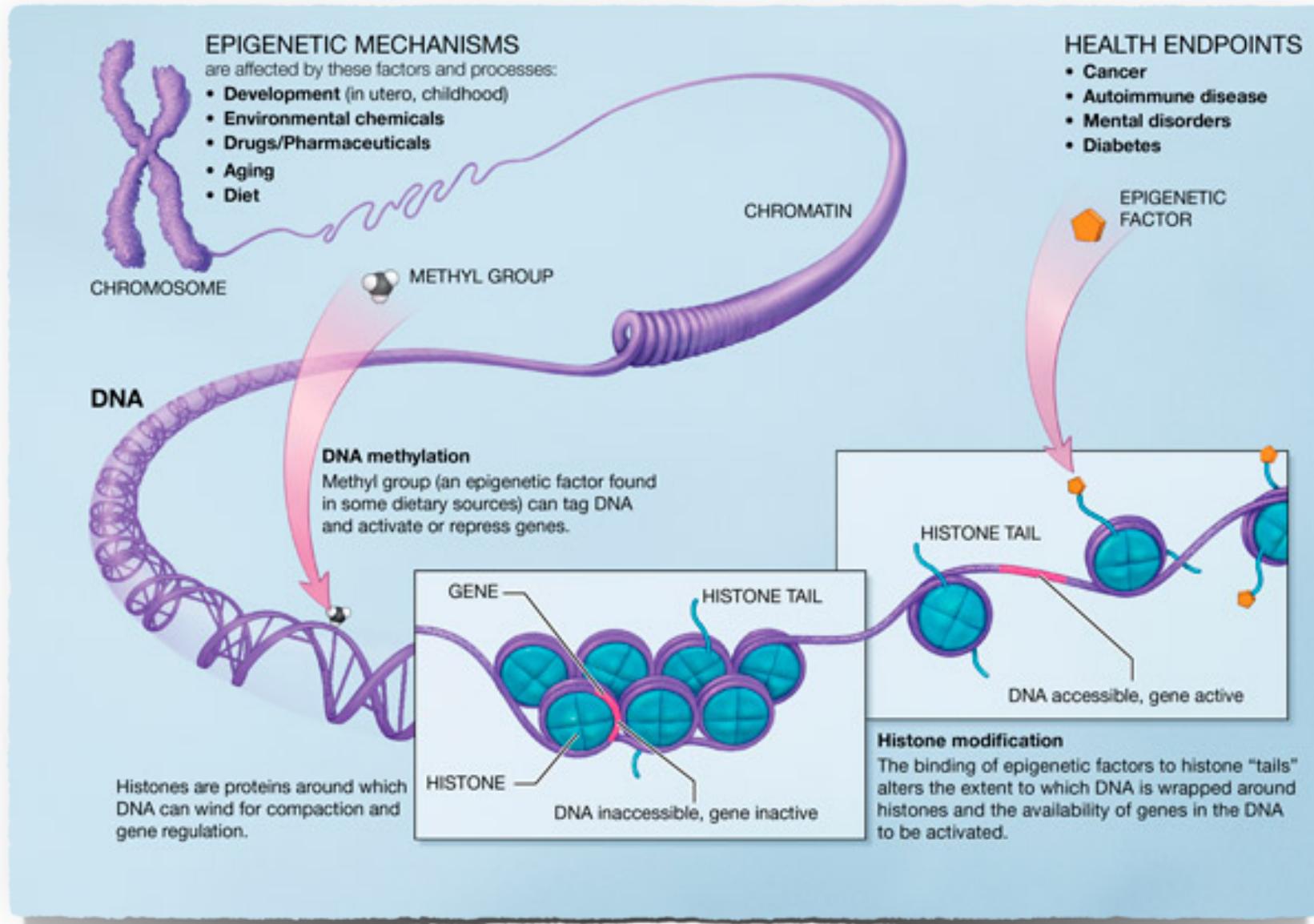
- Packing affects which information can be used



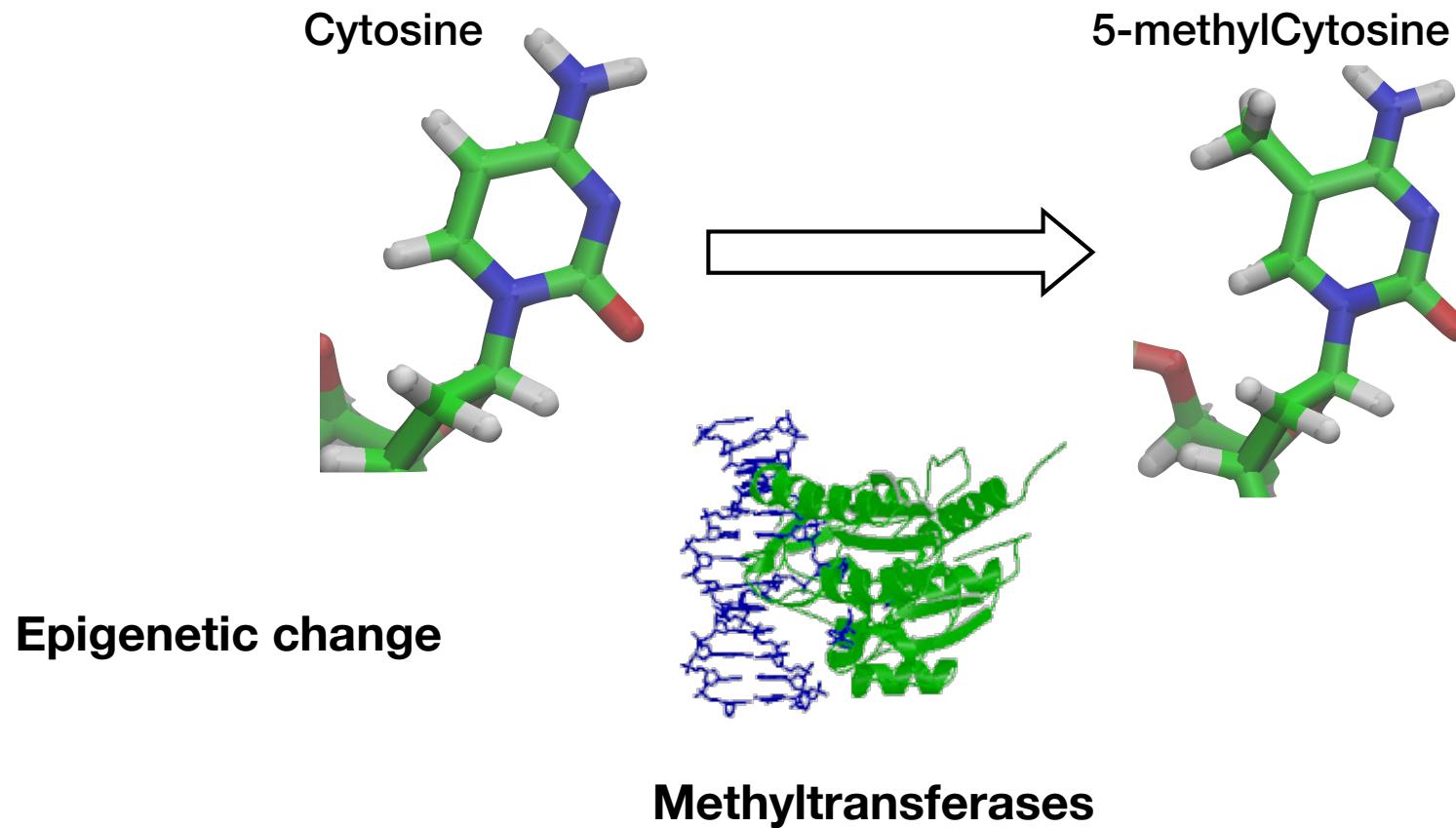
There is no clear picture of the steps going from nucleosome to chromosome—  
HiC experiments could change that



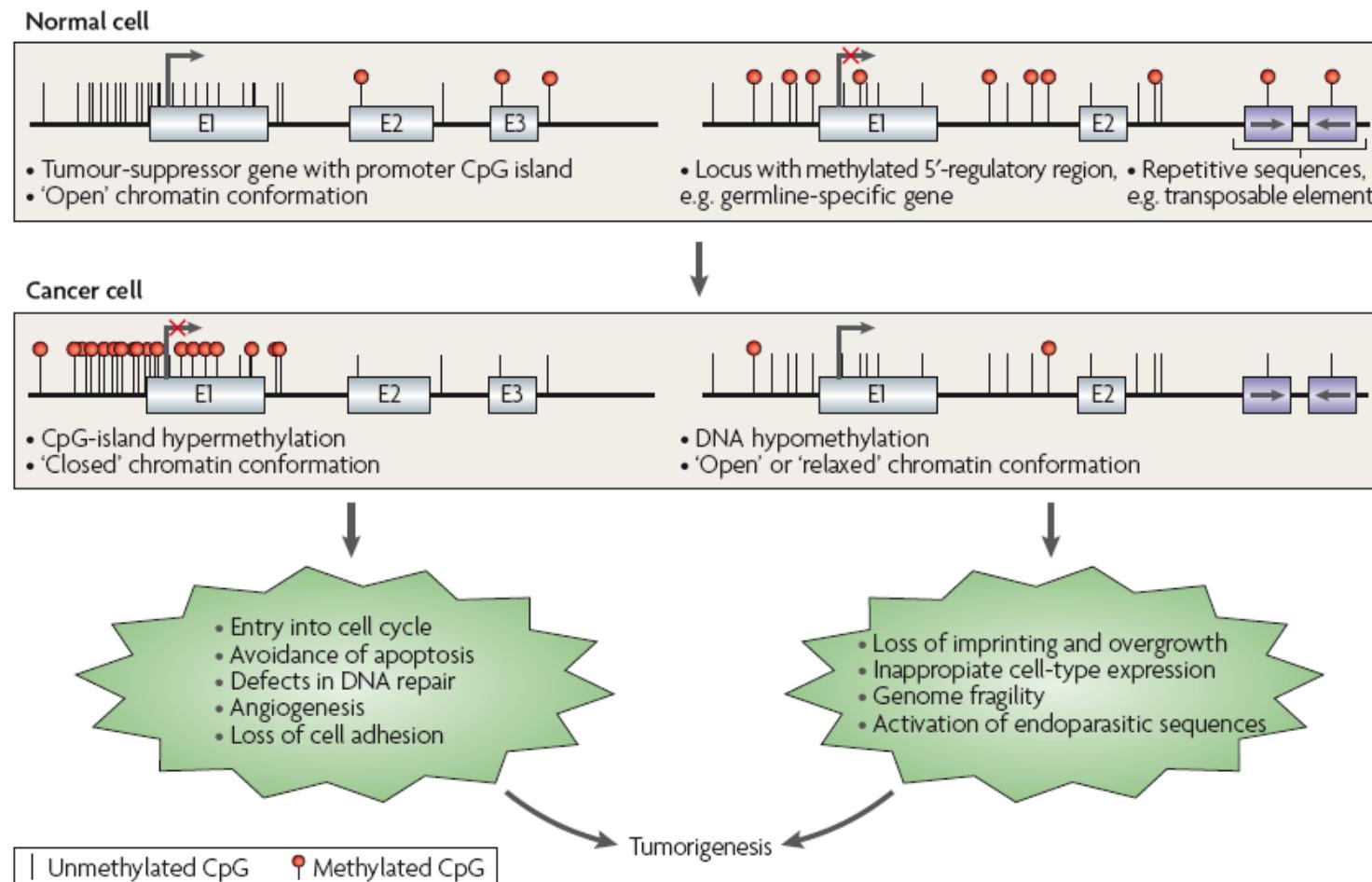
# Epigenetic changes affect packing preferences (cells in different tissues express different proteins)



# DNA methylation happens in mammals at 5'-CpG-3'



# Abnormal methylation patterns lead to cancer



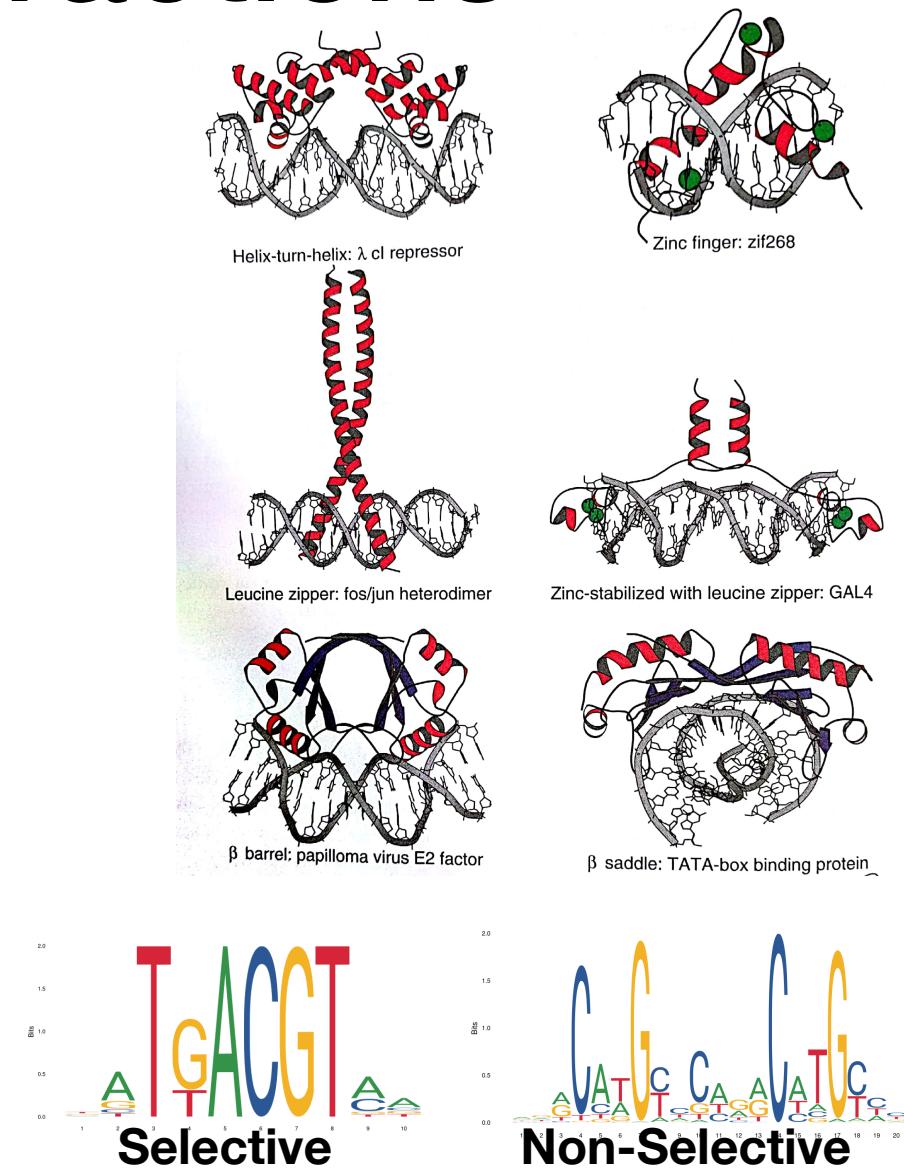
# A few known cancers due to hypermethylation

Table 1 | A catalogue of genes silenced by CpG Island promoter hypermethylation in human cancer

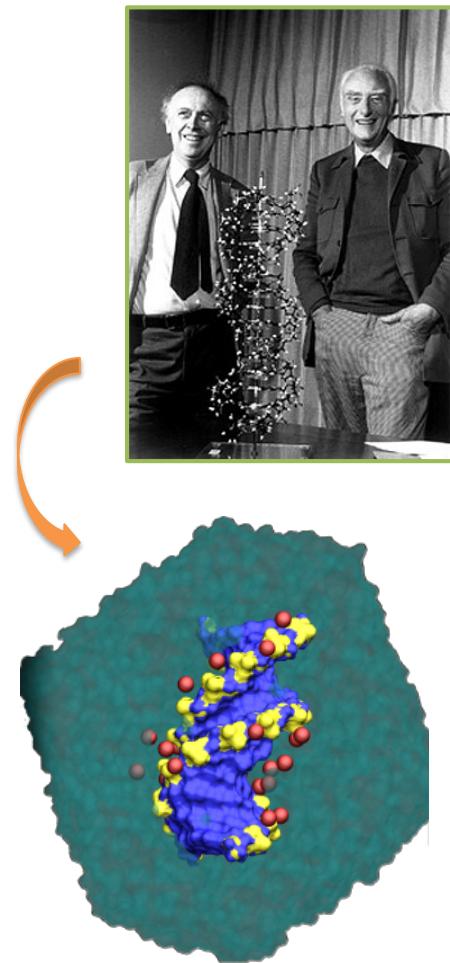
Gene	Function	Location	Tumour type	Consequences	FAT	Cadherin, tumour suppressor	4q34–35	Colon	Dissemination?
MLH1	DNA mismatch repair	3p21.3	Colon, endometrium, stomach	Frameshift mutations	HIC1	Transcription factor	17p13.3	Multiple types	Unknown
BRCA1	DNA repair, transcription	17q21	Breast, ovary	Double-strand breaks?	APC	Inhibitor of β-catenin	5q21	Aerodigestive tract	Activation β-catenin route
p16 <sup>INK4a</sup>	Cyclin-dependent kinase inhibitor	9p21	Multiple types	Entrance in cell cycle	SFRP1	Secreted frizzled-related protein 1	8p12–p11	Colon	Activation Wnt signalling
p14 <sup>ARF</sup>	MDM2 inhibitor	9p21	Colon, stomach, kidney	Degradation of p53	DKK1	Extracellular Wnt inhibitor	10q11.2	Colon	Activation Wnt signalling
p15 <sup>INK4b</sup>	Cyclin-dependent kinase inhibitor	9p21	Leukaemia	Entrance into cell cycle	WIF1	Wnt inhibitory factor	12q14.3	Colon, lung	Activation Wnt signalling
MGMT	DNA repair of 06-alkyl-guanine	10q26	Multiple types	Mutations, chemosensitivity	COK2	Cyclooxygenase-2	1q25	Colon, stomach	Anti-inflammatory resistance?
GSTM1	Conjugation to glutathione	11q13	Prostate, breast, kidney	Adduct accumulation?	SOCS1	Inhibitor of JAK–STAT pathway	16p13.13	Liver, myeloma	JAK2 activation
p73	p53 homologue	1p36	Lymphoma	Unknown	SOCS3	Inhibitor of JAK–STAT pathway	17q25	Lung	JAK2 activation
LKB1/STK11	Serine–threonine kinase	19p13.3	Colon, breast, lung	Unknown	GATA4	Transcription factor	8p23–p22	Colon, stomach	Silencing of target genes
ER	Oestrogen receptor	6q25.1	Breast	Hormone insensitivity	GATA5	Transcription factor	20q13	Colon, stomach	Silencing of target genes
PR	Progesterone receptor	11q22	Breast	Hormone insensitivity	ID4	Transcription factor	6p22–p21.3	Leukaemia, stomach	Unknown
AR	Androgen receptor	Xq11	Prostate	Hormone insensitivity	SRBC	BRCA1-binding protein	1p15	Breast, lung	Unknown
PRLR	Prolactin receptor	5p13–p12	Breast	Hormone insensitivity	SYK	Tyrosine kinase	9q22	Breast	Unknown
TSHR	Thyroid-stimulating hormone receptor	14q31	Thyroid	Hormone insensitivity	RIZ1	Histone/protein methyltransferase	1p36	Breast, liver	Aberrant gene expression?
RARβ2	Retinoic acid receptor-β2	3p24	Colon, lung, head and neck	Vitamin insensitivity?	DAPK	Pro-apoptotic	9q34.1	Lymphoma, lung, colon	Resistance to apoptosis
CRBP1	Retinol-binding protein	3q21–q22	Colon, stomach, lymphoma	Vitamin insensitivity?	TMS1	Pro-apoptotic	16p11	Breast	Resistance to apoptosis
RASSF1A	Ras effector homologue	3p21.3	Multiple types	Unknown	IGFBP3	Growth-factor-binding protein	7p14–p12	Lung, skin	Resistance to apoptosis
NORE1A	Ras effector homologue	1q32	Lung	Unknown	TPEF/HPP1	Transmembrane protein	2q33	Colon, bladder	Unknown
VHL	Ubiquitin ligase component	3p25	Kidney, haemangioblastoma	Loss of hypoxic response?	SLC5A8	Sodium transporter	12q23	Glioma, colon	Unknown
Rb	Cell-cycle inhibitor	13q14	Retinoblastoma	Entrance into cell cycle	HOXA9	Homeobox protein	7p15–p14	Neuroblastoma	Unknown
THBS1	Thrombospondin-1, Anti-angiogenic	15q15	Glioma	Neovascularization	EXT1	Heparan sulphate synthesis	8q24	Leukaemia, skin	Cellular detachment
CDH1	E-cadherin, cell adhesion	16q22.1	Breast, stomach, Leukaemia	Dissemination	LaminA/C	Nuclear intermediate filament	1q21.2	Lymphoma, leukaemia	Unknown
CDH13	H-cadherin, cell adhesion	16q24	Breast, lung	Dissemination?	WRN	DNA repair	8p12–p11.2	Colon, stomach, sarcoma	DNA breakage, chemosensitivity

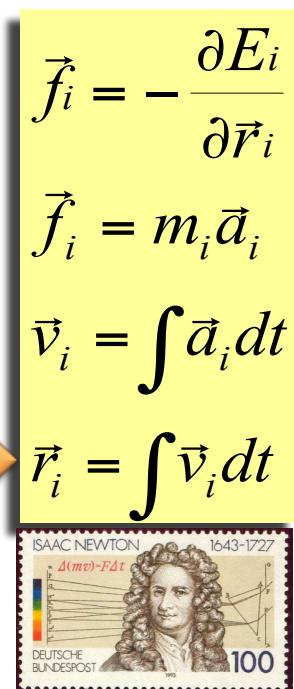
# Gene regulation: protein-DNA interactions

- Transcription factors (proteins) bind specific sequences of DNA.
- They bind a few bases of DNA (6 to 10).
  - There are  $4^{**}10$  possible sequences of DNA 10mers
- Need to differentiate affinity and selectivity:
  - Affinity: strength of the interaction protein-DNA
  - Selectivity: how well the protein differentiates between different DNA sequences
- Some TF are very selective — some are very promiscuous
  - Rules for understanding TF binding are an area of hot research



# Atomistic models of DNA





$\vec{f}_i = - \frac{\partial E_i}{\partial \vec{r}_i}$

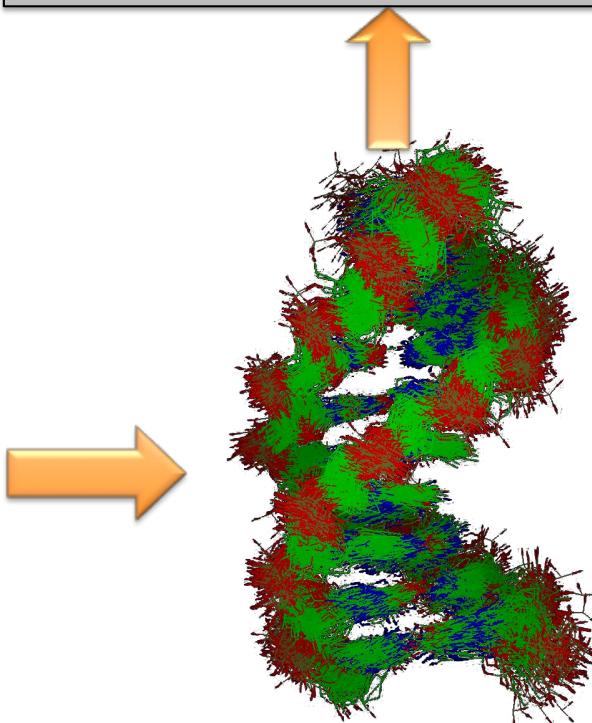
$\vec{f}_i = m_i \vec{a}_i$

$\vec{v}_i = \int \vec{a}_i dt$

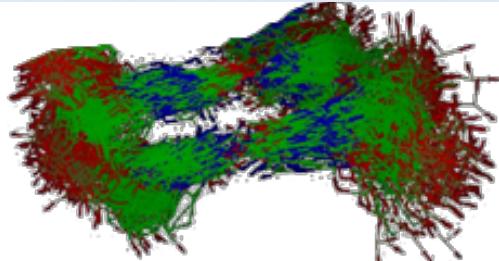
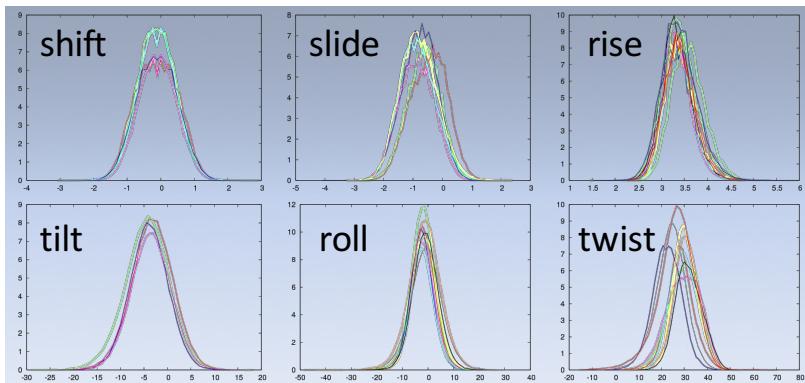
$\vec{r}_i = \int \vec{v}_i dt$

ISAAC NEWTON 1643-1727  
DEUTSCHE BUNDESPOST 100

Flexibility  
Physical properties  
Atomistic detail interactions

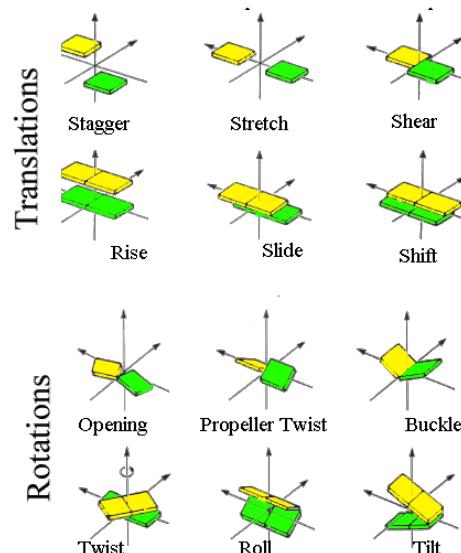


# DNA is flexible



$$K = kTC^{-1}$$

$$E_{\text{def}} = \sum_i \frac{K_{ii}}{2} (X_i - X_{i_0})^2 + \sum_{i \neq j} \frac{K_{ij}}{2} (X_i - X_{i_0})(X_j - X_{j_0})$$



- Structure and flexibility dictate DNA's ability to interact with proteins, form nucleosomes, ...