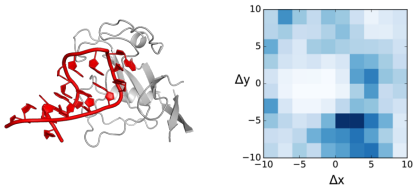
Welcome to CHEM 153B: DNA, RNA, and Protein Synthesis

- This quarter we are going to cover:
  - Nucleic acid structure
  - DNA & RNA polymerases
  - DNA replication
  - DNA repair
  - Transcription
  - RNA processing
  - Translation

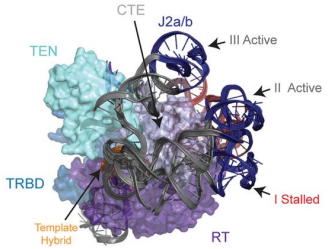
My background: modeling the structures and energetics of RNAs and RNA-protein complexes

RNP-denovo



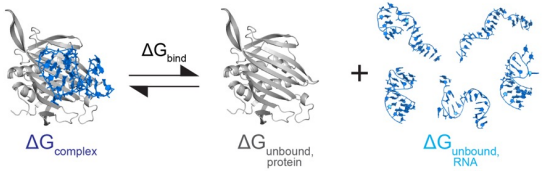
K. Kappel, R. Das. *Structure*, 2019.

smFRET-Rosetta



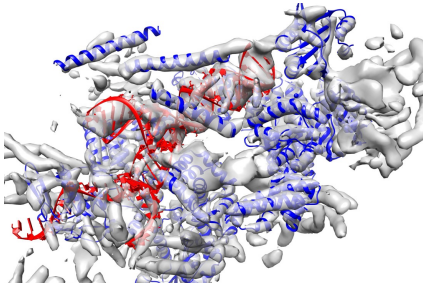
J. Parks*, K. Kappel*, et al. *RNA*, 2017.

RNP-ddG



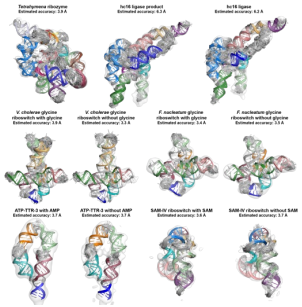
K. Kappel, et al. *PNAS*, 2019.

DRRAFTER



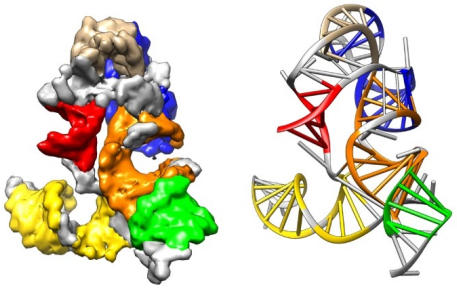
K. Kappel, et al. Nature Methods, 2018.

auto-DRRAFTER

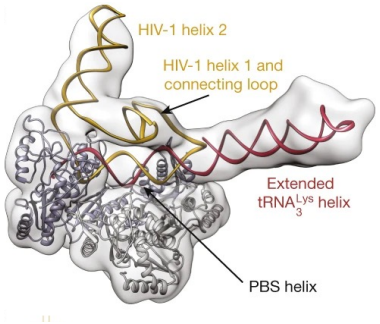


K. Kappel*, K. Zhang*, Z. Su*, et al. Nature Methods, 2020.

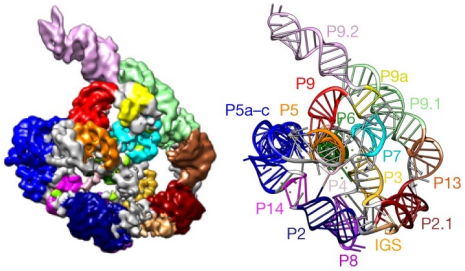
Using cryo-EM and computational modeling to resolve several structures of RNAs and RNA-protein complexes



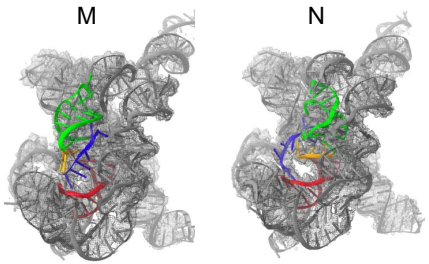
K. Zhang*, S. Li*, K. Kappel*, et al. Nat. Commun. 2018.



K.P. Larsen, Y. Mathiharan, K. Kappel, et al. Nature, 2018.



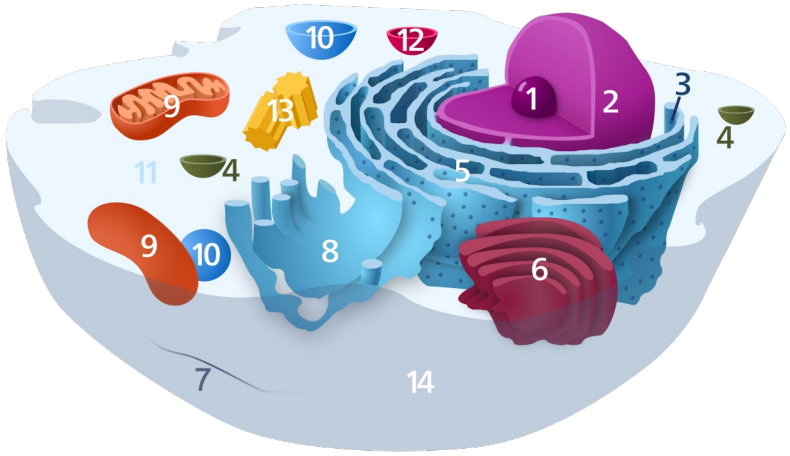
Z. Su*, K. Zhang*, K. Kappel* et al. Nature, 2021.



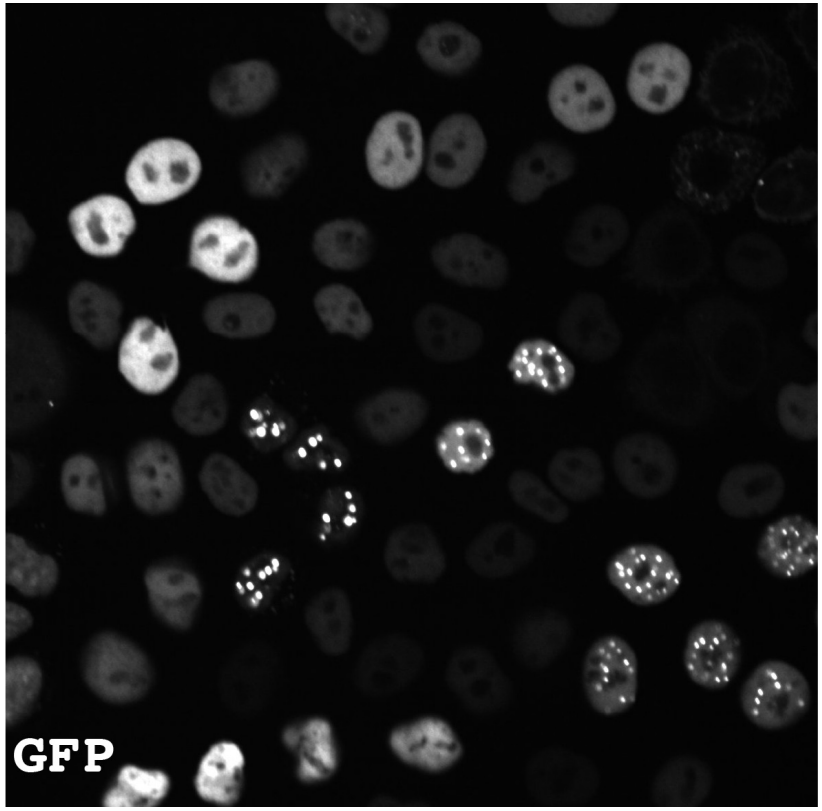
S. Li, et al. PNAS, 2022.

How is the inside of a cell organized?

A human cell contains 5-10 billion protein molecules



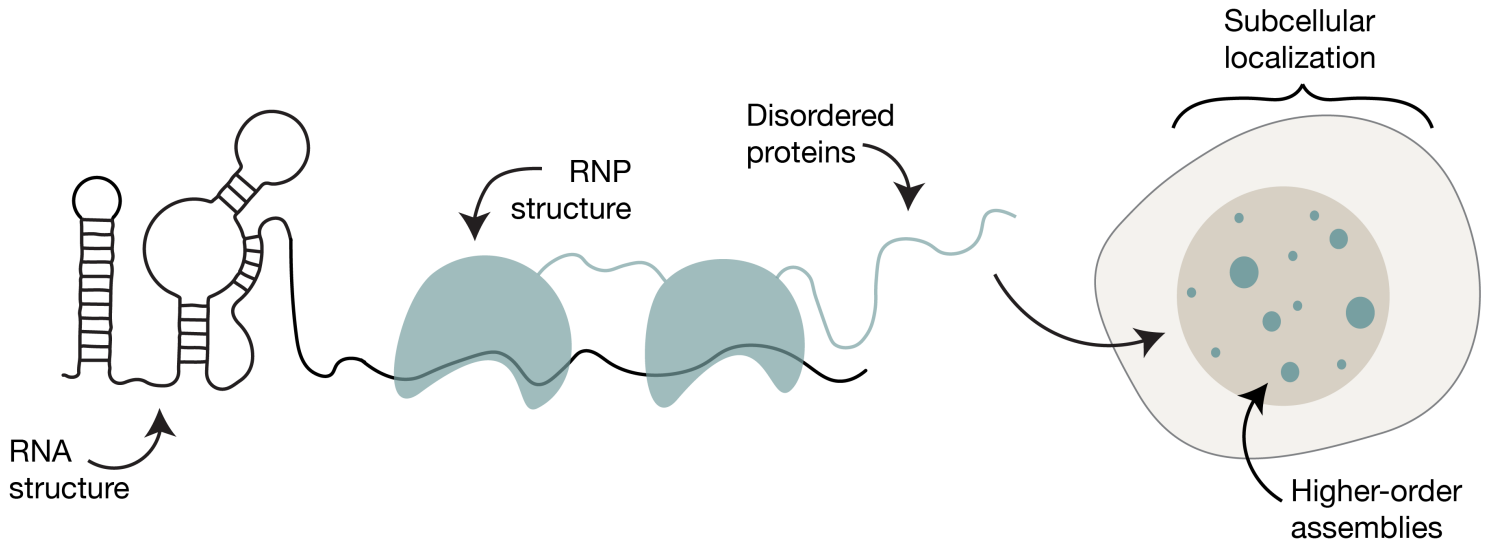
https://en.wikipedia.org/wiki/Organelle



K. Kappel, et al. Nature Methods, 2025.

The Kappel Lab @ UCLA:

Combining high-throughput experiments and computation to decipher sequence-organization-function relationships of proteins, RNA, and their interactions



Chem 153B Syllabus

- The full syllabus is posted on BruinLearn

Course Syllabus
Chem 153B – Winter 2026
Professor Kalli Kappel

Instructor: Professor Kalli Kappel (she/her)

Email: kallikappel@ucla.edu; please use chem153bkappel@gmail.com for course inquiries (see below).

Instructor office hours: 2051 Young Hall, Mondays 2-3 PM and Thursdays 2-3 PM

TAs:
Izaiah Cole (icole@g.ucla.edu)
Jake Cozy (jcozy@g.ucla.edu)
Marcella Mirabelli (mmirabelli@g.ucla.edu)
Vanessa Wiratmo (vanessawiratmo@g.ucla.edu)

TA office hours: Tuesdays 4-5 PM, 4222A Young Hall (except for January 20th, when office hours will be held in 3064 Young Hall)

Course Meetings: M,W,R,F: 1-1:50 PM in Young Hall CS24

Thursday tutorials are optional, except for the days when we have exams.
We will use the tutorials to review course material and go over old exam questions.

Lectures: MWF - Young Hall CS24

Lectures will be in person and will introduce the materials and concepts covered in this course. All materials will be posted on the course BruinLearn page: https://bruinlearn.ucla.edu/courses/220811

Calendar of assignments and exams:

You should submit a homework by 11:59 PM on Monday on all weeks highlighted in Blue.

Lectures YH-CS24

Tutorials YH-CS24

Midterms YH-CS24

No Class

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	1/5		1/7	1/8	1/9
2	1/12		1/14	1/15	1/16
3	1/19		1/21	1/22	1/23
4	1/26		1/28	1/29	1/30
5	2/2		2/4	2/5	2/6
6	2/9		2/11	2/12	2/13
7	2/16		2/18	2/19	2/20
8	2/23		2/25	2/26	2/27
9	3/2		3/4	3/5	3/6
10	3/9		3/11	3/12	3/13

iClickers:

To facilitate active learning and promote engagement with course material, students will answer quizzes shown on some of the lecture slides using iClickers. The iClicker app is now integrated into BruinLearn and you can access it for no additional cost:
https://mhe.my.site.com/iclicker/s/article/Student-Guide-iClicker-Roster-Grade-Sync-Integration
Course Name: Chem 153B
Use this link to register your iClicker for this course (or follow the QR code to the right):
https://join.iclicker.com/WCUO

Please sign up for iClicker!
iClicker participation points will be given for completion regardless of whether or not your answers are correct.

Tutorials (Prof. Kappel):

• Tutorials on January 29th and February 19th will be used for midterm exams in Young Hall CS24.

• The other tutorials will be in person in Young Hall CS24. These sessions typically consist of a brief review of recent course material followed by working through old exam questions in small groups. These sessions are not recorded and attendance is highly recommended since they are mostly used for problem solving in small groups. **I reserve the right to cover exam-relevant material during tutorials if I deem it necessary to achieve the course learning objectives by end of quarter.**

Discussions (TAs):

1A:	T	9:00-9:50 AM	Boelter Hall 4283	Izaiah Cole
1B:	T	1:00-1:50 AM	Young Hall 3069	Izaiah Cole
1C:	W	11:00-11:50 AM	Young Hall 1044	Jake Cozy
1D:	W	3:00-3:50 PM	Young Hall 2200	Jake Cozy
1E:	R	9:00-9:50 AM	Boelter Hall 5252	Marcella Mirabelli
1F:	R	4:00-4:50 PM	Boelter Hall 2444	Marcella Mirabelli

The first discussion section will focus on an introduction to PyMol, and the following discussions will focus on problem sets in preparation of midterms/final exam. Attendance will be taken, and please attend your assigned discussion section.

Grading: Exams, Homework and Participation

Midterms: 30%
Your higher midterm score will count for 20% of your grade, and the lower score as 10%.

Final Exam: 40% **Wednesday, March 18th, 8:00 AM-11:00 AM**
If you have scheduling conflicts regarding the midterms or final exam, please reconsider your enrollment in the course.

Homework: 10%
To encourage you to explore the problems presented in the weekly discussions ahead of time, you will be required to complete and submit one (out of the four possible) discussion problems on Monday prior to your discussion section (11:59PM deadline) for completion credit. The question you will answer will be the same as your group number, which will be assigned to you during the first week. Your submission should be a screenshot of a digital version or a photo of a paper version of your single question (not a PDF or word document) to help reduce formatting errors that could arise from various document types. You are welcome to work in groups to answer these questions, but every student should write their own answers. If you need clarification, please ask your TAs to help. You will get 1.25% credit for each completed homework, with a maximum possible percentage of 10%. *This equates to being able to miss one problem set during the quarter without penalty.*

Attendance and Participation Points: 10%
- iClicker lecture questions = 10%. While I expect that you will attend every class, you will receive full credit if you submit all iClicker responses in at least 20 lectures throughout the quarter (≥ 20 would be 10%, 19 would be 9.5%, etc). *In short, this equates to 8 skips that you can use throughout the quarter for mandatory classes without penalty.* iClicker participation points are given for completion regardless of whether or not the answer is correct. Since the class is not on Zoom and we will be using active learning techniques**, you must be physically in the classroom to be eligible for iClicker points on any given day**. If you have trouble with iClicker during any given lecture, please speak with me after class.

Discussions Participations = 10%
Please go to your assigned discussion section. While we expect that you will attend all discussions, you must attend at least 8 for full credit (less than 8 is 1.25% per discussion attended). *This equates to 2 skips during the quarter for mandatory discussion sections without penalty.*

Extra Credit: 1% for submitting an online evaluation for the course by the campus deadline.

Grading Scale: Minimum scores for particular grades are listed below. Scores are based on exceeding a cutoff without rounding (for example, 97.9 is an A, not an A+).

A+ ≥ 98%	A ≥ 90%	A- ≥ 85%	B+ ≥ 80%	B ≥ 75%	B- ≥ 70%
C+ ≥ 66%	C ≥ 63%	C- ≥ 60%	D ≥ 55%	F <55%	

Course Policies:

- **Questions and Course Logistics:** If you need me to answer specific questions related to the materials, **please use the BruinLearn discussion forum rather than emailing me or a TA** as it is beneficial for all students to know the answer to this question. If you have a question about course logistics, **please send an email to chem153bkappel@gmail.com. Emails should only be sent directly to the instructor in rare cases that involve private or personal matters.**
- **Technology Policy**: Students are welcome to take class notes however they prefer: pen and paper, using an iPad, or on a laptop. However, students should remain engaged in the class and refrain from using social media, web browsing, online shopping, etc. during class.
- **Sick Policy**: ***In general, do not come to class sick***. If you suspect you have COVID-19, please stay home until you have a negative rapid antigen test. If you feel well enough to attend class, please wear a mask.
- **Participation**: All students are expected to be active participants in the course. This includes attending every class meeting, asking questions, and being present and involved in discussions.
- **Regrade policy**: You are allowed two regrade challenges for the entire term-use them wisely. Like in tennis or football, your available challenges do not decrease if your challenge is found to be valid. However, once you use a challenge and it is not upheld, it counts toward your two. Before submitting a regrade request, carefully ask yourself:
    - Does my answer contain all the information requested by the question?
    - Does my answer contain any incorrect or unsupported claims?

  Historically, valid regrades fall into two categories:
    - Clerical errors: Examples include arithmetic mistakes in point totals, overlooked responses, or issues with automated grading (e.g., multiple-choice answers not captured by Gradescope).
    - Fundamental flaws in the question: If a flaw in the question affects multiple students, a broader correction will be made, and no challenge will be deducted from your total.

  Please note that we do not negotiate changes in partial credit. Regrade requests should be submitted on GradeScope, along with a clear explanation that directly addresses the questions above.
- **Course Evaluations**: All students are strongly encouraged to fill out course evaluations.

Textbook:
There is no required textbook for the course as none of the current textbooks cover all the information I give during lectures. You may find it useful to have one of the following texts as a reference to supplement the lecture:

Lehninger Principles of Biochemistry - David Nelson and Michael M. Cox

Many of you may already own the eBook or the hard copy book. For example, if you purchased Achieve/Lehninger for 153A, you already have 4-year access to the eBook. If you can still access your Chem 153A Achieve page, you should be able to read it through the Vital Source website or to download it to your computer for offline viewing (downloading for offline viewing requires that you first install the MacMillan Learning eBook or Bookshelf app on your device).
If you are no longer able to access the previous course, the eBook can be found at macmillan.vitalsource.com. The same email address and password that you used for Achieve must be used and you should then be able to gain free access. If you have any issues with this, contact Macmillan Support at (800) 936-6899 or via their website.

If you already own the book, there is no advantage to purchasing Achieve/Lehninger for this course since I will not be assigning homework in Achieve, and so you should opt out of Inclusive Access before the end of the second week of the course. Access the opt out option from the UCLA Store Course Materials tool in the menu on the left or the Inclusive Access materials list link on program emails or by sending an email to inclusiveaccess@asucla.ucla.edu.

Discussion information week 1

- Please attend your assigned discussion section
- Download PyMol *before* your discussion section this week
- Bring your laptop to your discussion section this week
- Homework: if you are assigned group “1A” – please submit solutions for all of question 1

iClicker

Chem 153B

Go to

join.iClicker.com

WCUO

Share this join link with students

https://join.iclicker.com/WCUO

Copy

What is your major?

Unit 1: Nucleic Acid Structure

1-Nucleosides and Nucleotides

2- A, B and Z conformations of nucleic acids double helices

3- Principles of DNA recognition by sequence-specific DNA binding proteins nucleic acids/protein interactions

4- Principles of nucleic acids
Denaturation

5- DNA tertiary structures
G-quadruplex

6- RNA secondary and tertiary structures

Learning outcomes:
What you need to know/understand after this unit

Memorize structures of nucleosides and understand their chemical specificities

Be able to recognize A/B/Z and quadruplex conformations of DNA/RNA and understand the main structural features in these conformations; recognize variations and differences

Understand and recognize chemical groups in proteins that interact with distinct components of nucleic acids

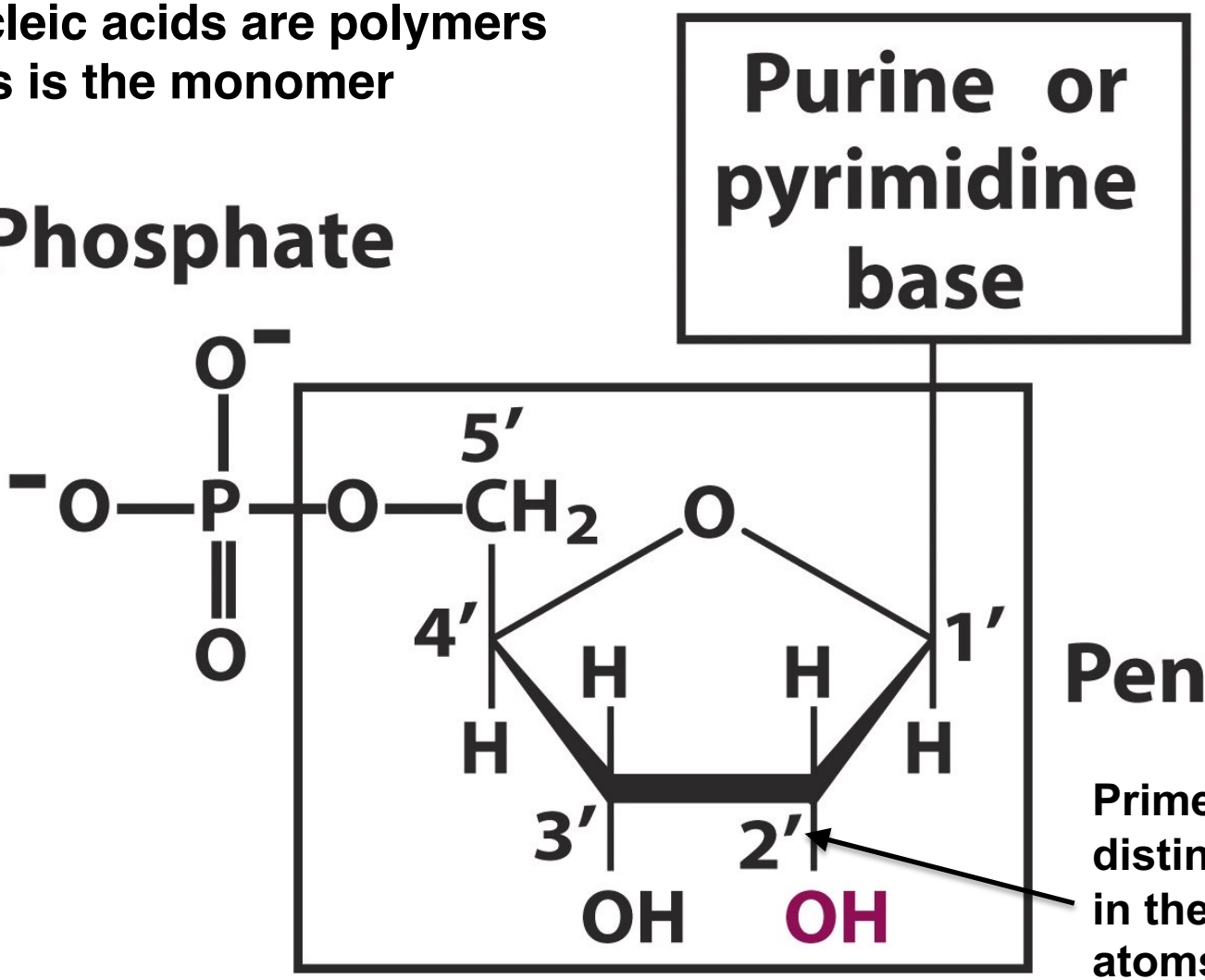
Understand how proteins interact with DNA in a sequence-specific or non-sequence specific manner; recognize sequence-specific vs. non-sequence specific interactions

Understand the main parameters that stabilize each conformation, as well as the transition from double-stranded to single stranded

Recognize basic elements of secondary and tertiary structure in RNA

Chemical definitions/nomenclature:

Nucleic acids are polymers
This is the monomer



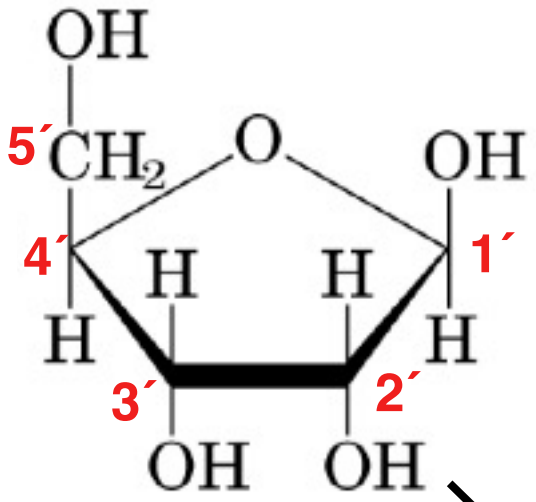
Nucleoside = Pentose + base

Nucleotide = Pentose + base + phosphate

Sugar

5' end

Ribose: (RNA)



3' end

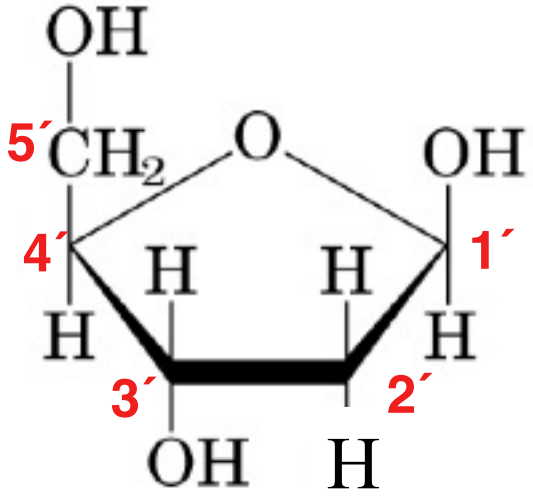
Numbering of carbons: C1´, C2´…
(´ used to prevent confusion with the numbering of atoms in bases).

The base is connected at the 1' position.

The presence of the 2´OH confers special chemical reactivity and structural properties to RNA compared to DNA

or

Deoxyribose: (DNA)



The sugar is not actually flat!

PyMol: G_C3'endo.pse

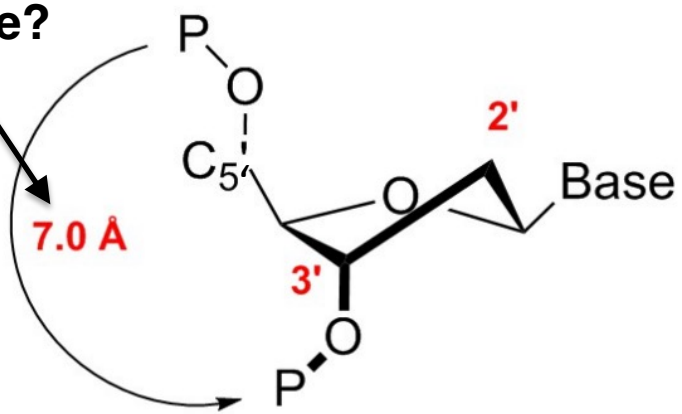
Numbering is the same in DNA and RNA

Sugar puckering: C2´ endo or C3´ endo

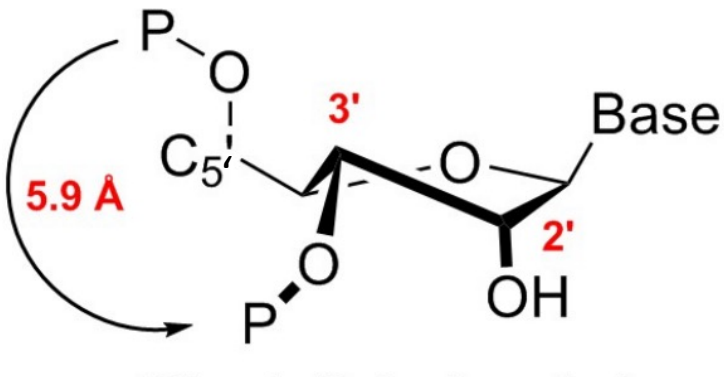
Planar Conformations of ribose/deoxyriboses are not stable → Puckering

Why might this be energetically favorable?

2 Major Conformations:



C2'-endo



C3'-endo

Ribose in polymers are constrained in the C3´ endo conformation for steric reasons (in 2' endo, the 2' OH would clash with the phosphate) --> RNA is nearly always found as C3´ endo

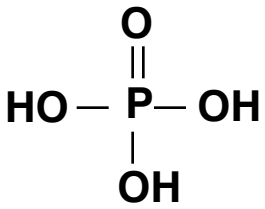
Deoxyriboses in DNA are in the C2´ or C3´ endo conformation (C2' preferred)

Differences in Sugar Pucker impact distances between consecutive phosphate groups! → important to remember for polymer sizes (spacing between 5' and 3'

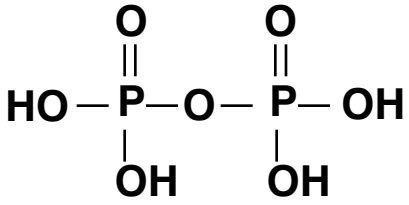
https://www.ibiology.org/biochemistry/rna-structure/

Phosphates

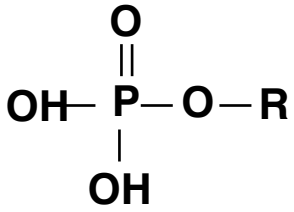
Phosphoric Acid



Pyrophosphate

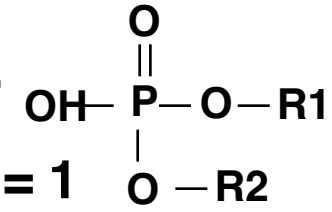


Phosphomonoester



pKas = 1, 6

Phosphodiester

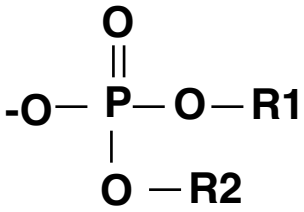


pKa = 1

pKa important to remember for ionization of polymers at physiological pH

An atom is protonated at pH < pKa
An atom is unprotonated at pH > pKa

Which amino acids in proteins would typically interact with phosphate groups in nucleic acids?



(pKa = 1)

A: Hydrophobic amino acids (e.g. Leucine)

B: Amino acids with Aromatic side chains (eg Tryptophan)

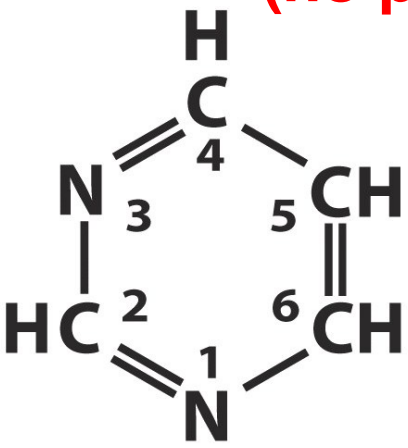
C: Positively charged amino acids (e.g. Lysine)

D: you cannot say in general - it depends whether it´s RNA or DNA

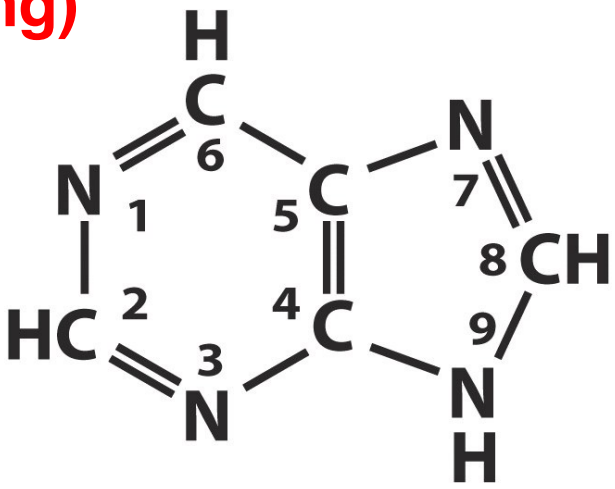
Bases

Aromaticity of bases and consequences:

Bases are planar
(no puckering)



Pyrimidine



Purine

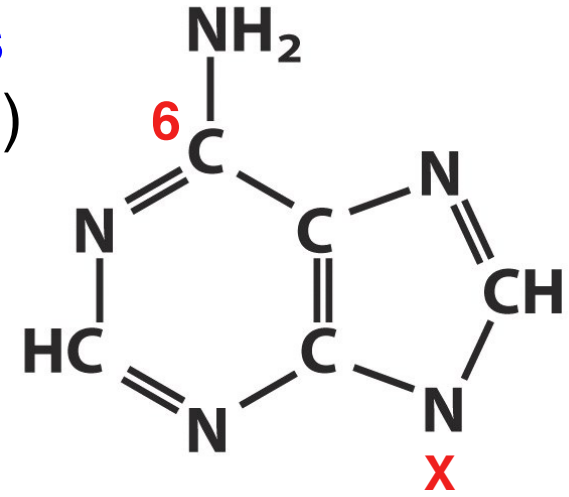
Bases have hydrophobic aromatic rings which promotes aromatic stacking (π stacking), an important factor in stabilizing nucleic acids in helical structures

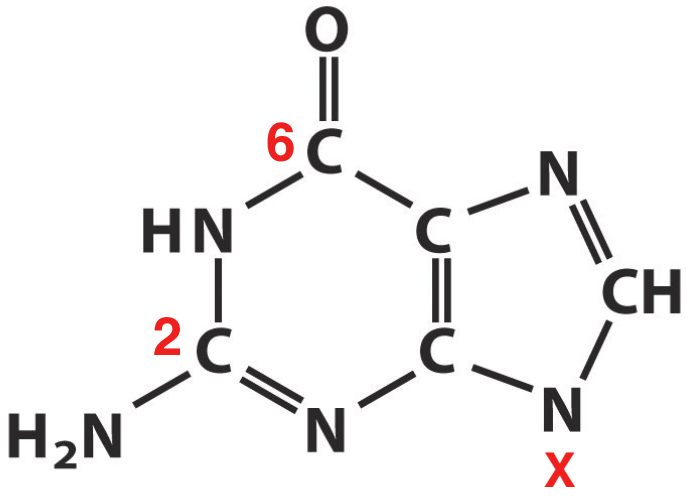
Bases are planar
PyMol: G_C3'endo.pse

Purine Bases and Nucleosides

(abbreviated as R)

Structures need to be memorized



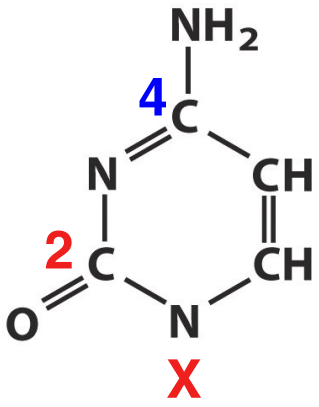


Base (X=H)	**Adenine**	**Guanine**
Nucleoside (X=pentose)	**Adenosine** A/dA/rA	**Guanosine** G/dG/rG
Nucleotide (X=pentose phosphate)	**Adenosine monophosphate (AMP)**	**Guanosine monophosphate GMP**

Base (X=H)

Nucleoside (X=pentose)

Nucleotide (X=pentose phosphate)

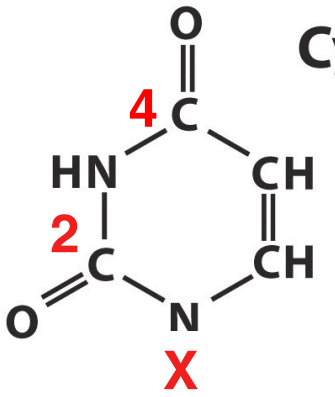


Cytosine

Cytidine
C/dC/rC

Cytidine monophospate
CMP

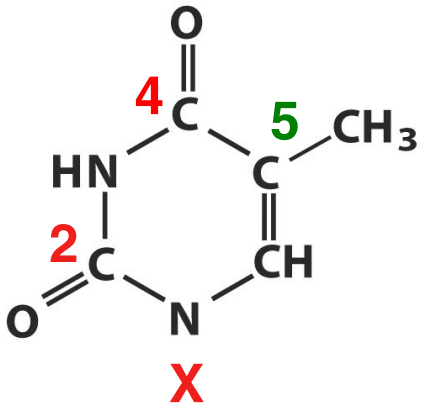
Structures need to be memorized



Uracil (RNA)

Uridine
U/rU

Uridine monophospate
UMP



Thymine (DNA)

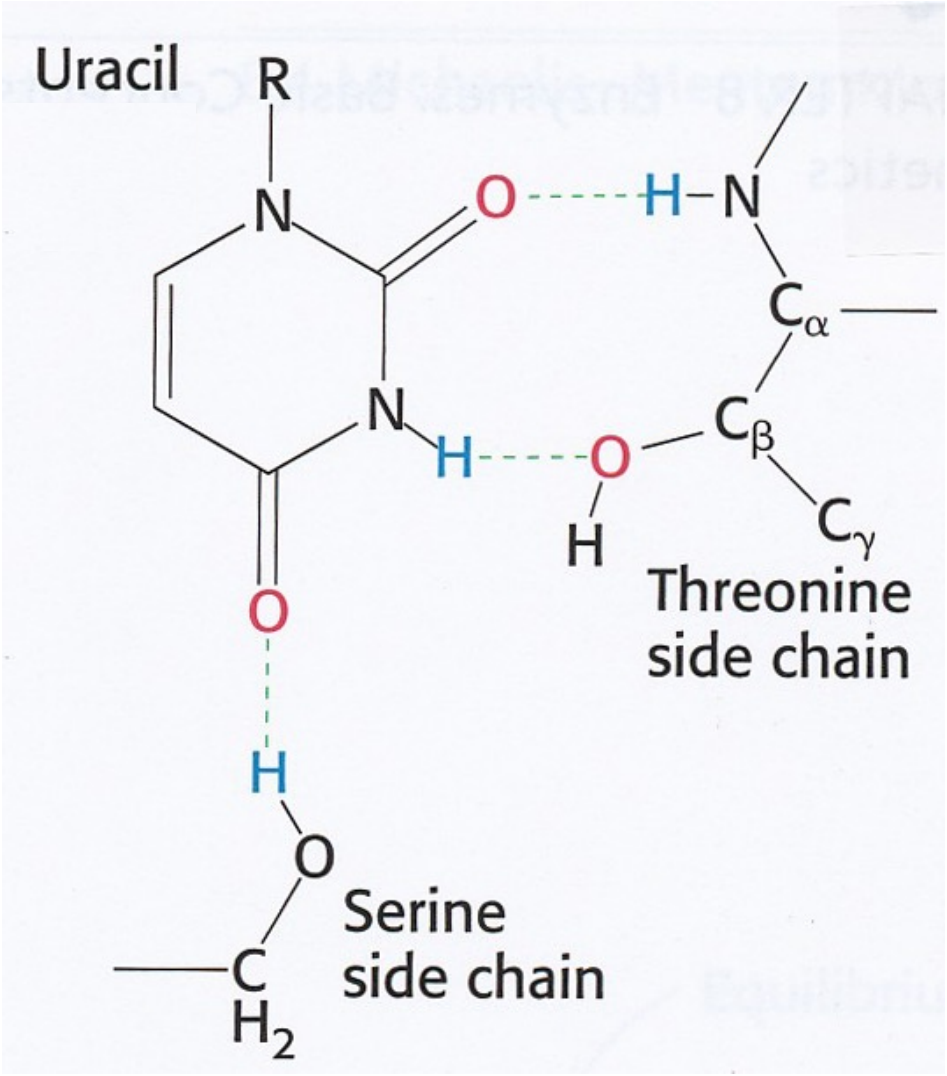
Thymidine
T/dT

Thymidine monophospate
TMP (dTMP)

Pyrimidine Bases and Nucleosides
(abbreviated as **Y**)

Molecular groups in bases can serve as Hydrogen bond donors or acceptors:

Hydrogen-bond donor	Hydrogen-bond acceptor
N—H	N
N—H	O
O—H	N
O—H	O



These groups provide the chemical basis for interaction between two nucleic acid strands

• Also provide some key elements of interactions between protein and nucleic acids

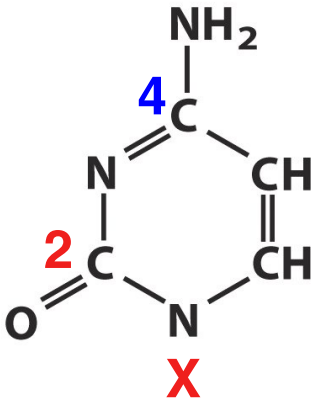
Which chemical group in pyrimidine bases cannot be used as H-bond donor or acceptor?

A: carbonyl at position 2

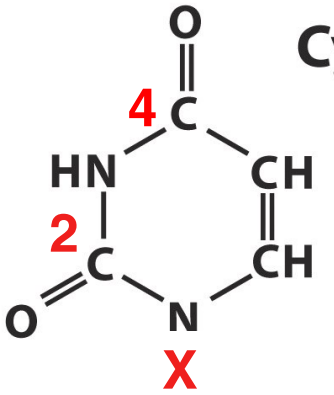
B: amine at position 4 of C

C: methyl at position 5 of T

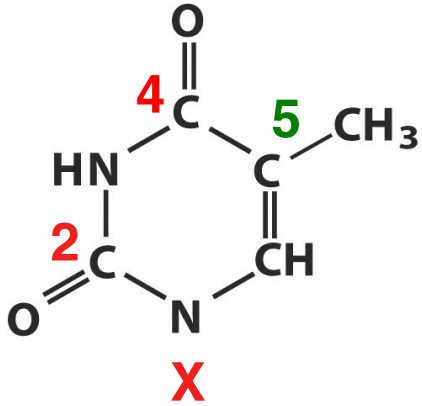
D: none – they can all be used



Cytosine

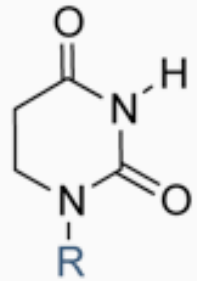


Uracil
(RNA)

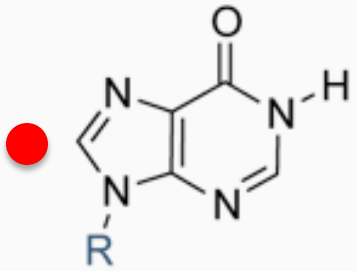


Thymine
(DNA)

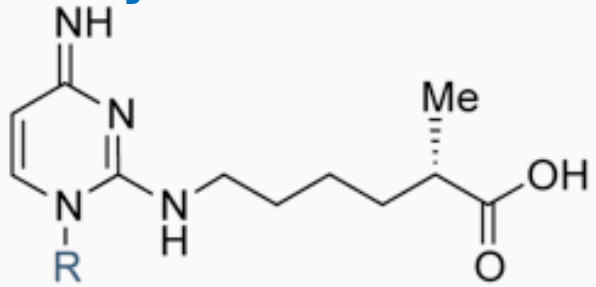
RNA (and to some extent DNA) can contain chemically modified bases which can impact the ability to form H-bonds



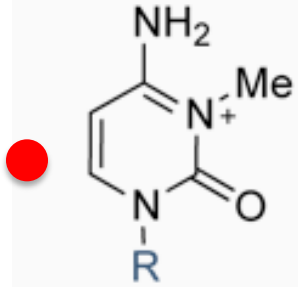
dihydrouridine



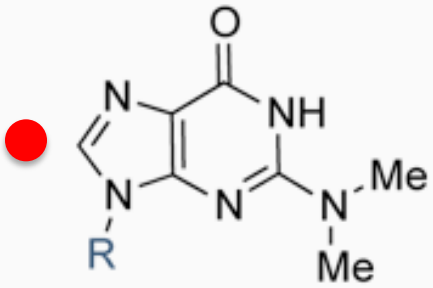
inosine



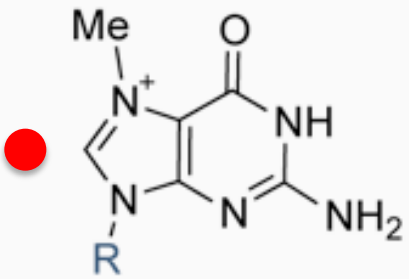
lysidine



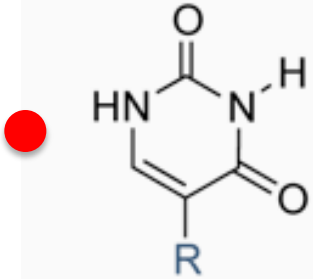
3-methylcytidine



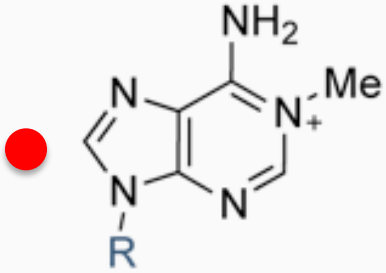
N(2)-dimethylguanosine



7-methylguanosine



pseudouridine

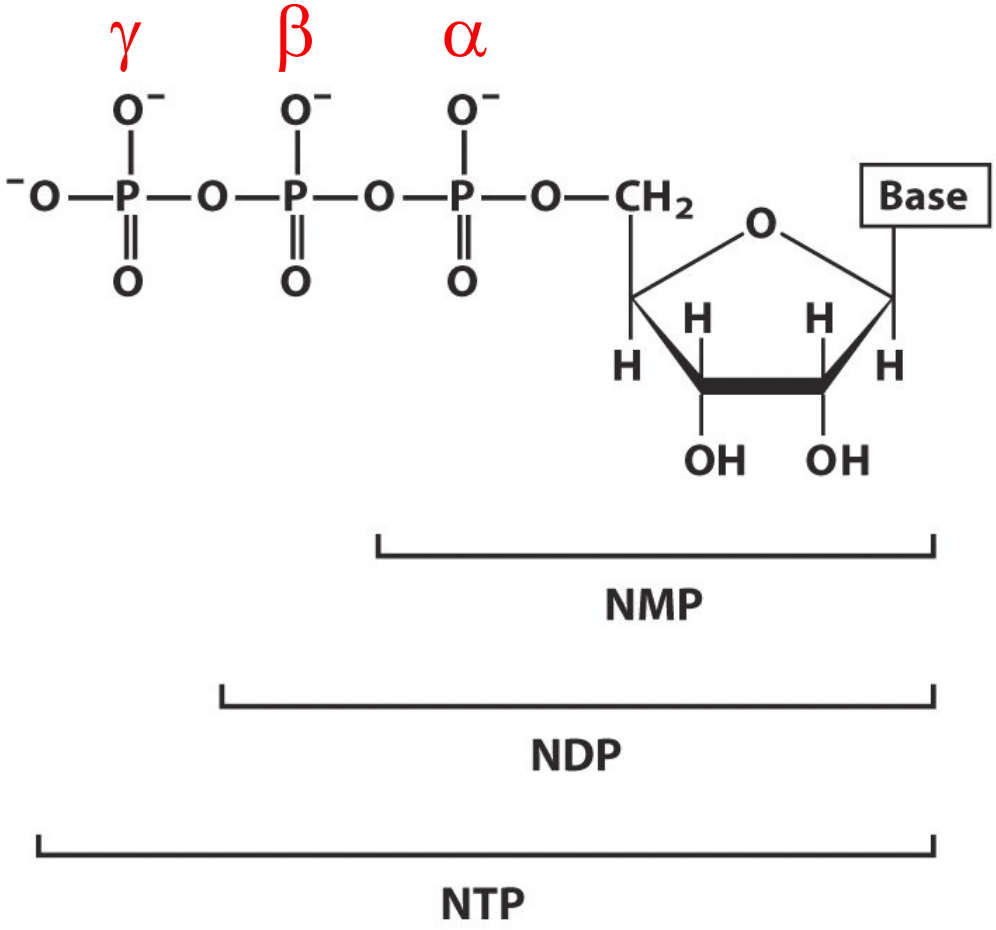


1-methyladenosine

Modifications generated by modifying enzymes on the polymers

● **= structures need to be known**

R = ribose

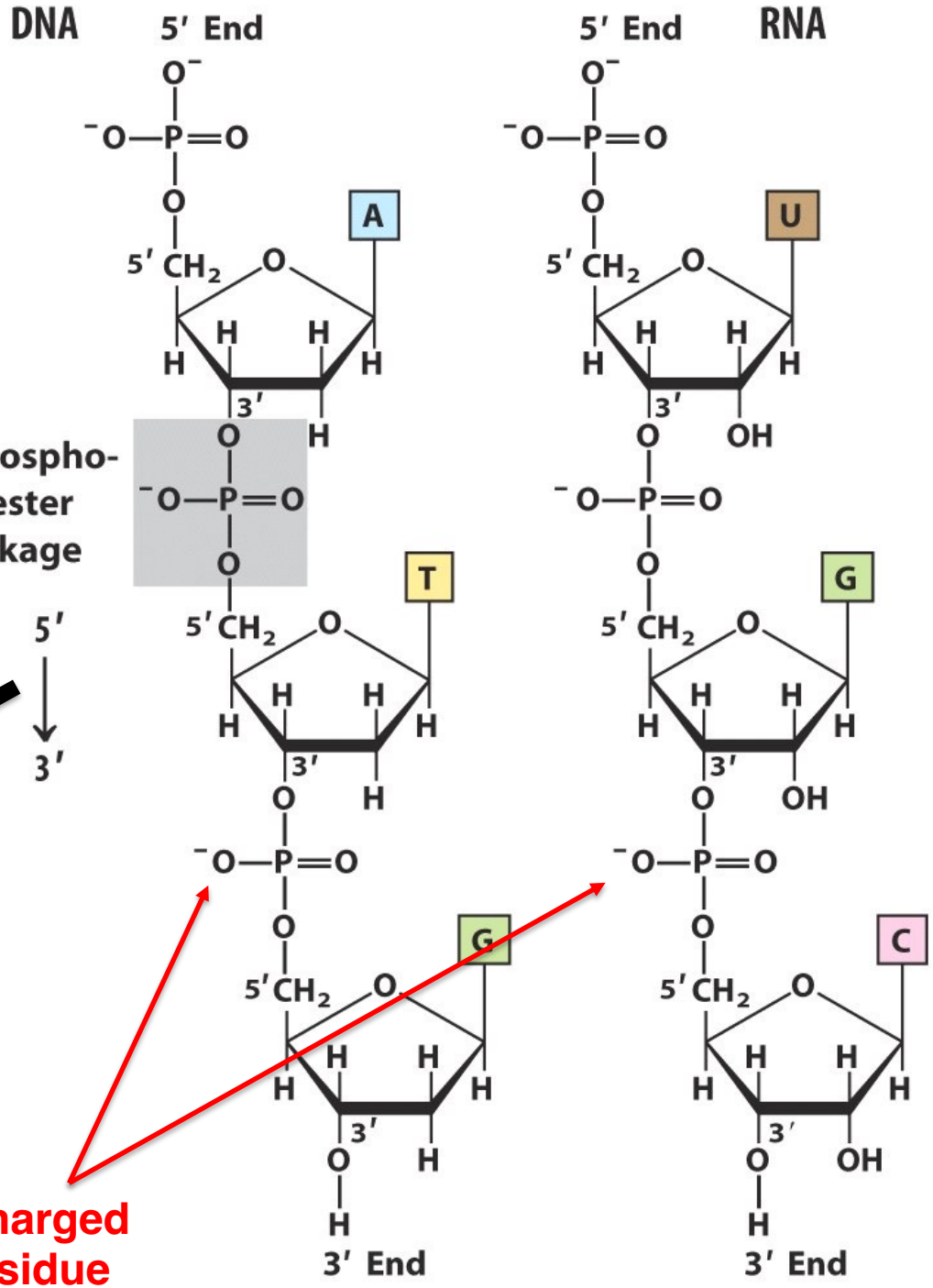


Common Bases and Nucleosides/ Nucleotides Nomenclature and abbreviations

Abbreviations of ribonucleoside 5′-phosphates			
Base	**Mono-**	**Di-**	**Tri-**
Adenine	AMP	ADP	ATP
Guanine	GMP	GDP	GTP
Cytosine	CMP	CDP	CTP
Uracil	UMP	UDP	UTP

Abbreviations of deoxyribonucleoside 5′-phosphates			
Base	**Mono-**	**Di-**	**Tri-**
Adenine	dAMP	dADP	dATP
Guanine	dGMP	dGDP	dGTP
Cytosine	dCMP	dCDP	dCTP
Thymine	dTMP	dTDP	dTTP

Polymeric Structure of Nucleic Acids



Links 3´ C of preceding nucleoside to 5´ C of next one

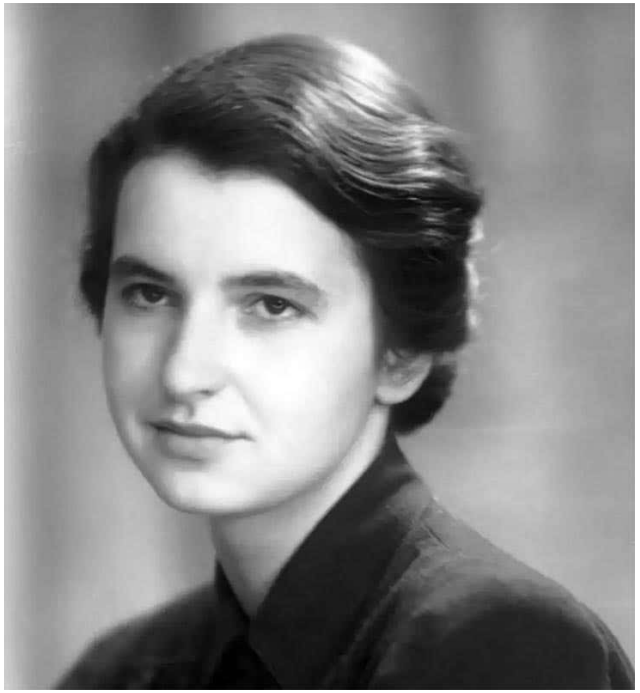
Indicates 5´-3´polarity
Polarity as in "direction" of the polymer
(not in the chemical sense)

- **1 negatively charged Oxygen per residue**

The discovery of the DNA double helix

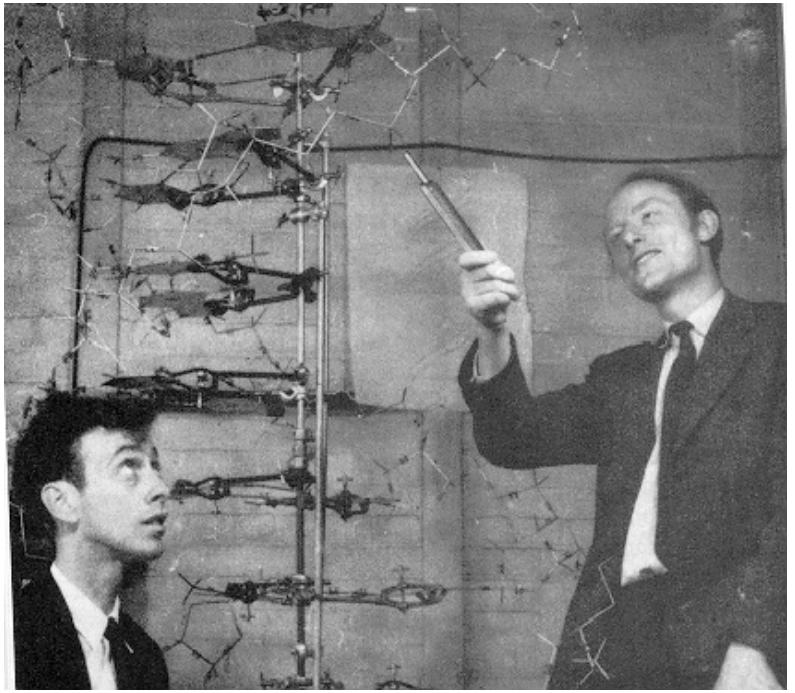
Setting the stage:

- DNA was first isolated and characterized in 1869 (Miescher)
- 1940s: first compelling evidence that DNA is the genetic material (Avery, MacLeod, McCarty)
- After that, lots of interest in solving its structure
- X-ray crystallography and biochemistry not yet advanced enough to solve the structure



Rosalind Franklin (1950-1953)

- **Physical chemist @ King's College London**
- **Used fiber diffraction to study DNA structure**
- **Died at age 37**



Watson and Crick (1953)

- **Theorists at Cambridge in the protein structure group**
- **Their model may have required Franklin's data…**

<u>Watson and Crick (1953)</u>: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material".

The discovery of the DNA double helix

3 Side-by-Side papers in Nature in 1953....
the authors of only two got the Nobel Prize in 1962

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 inter-twined 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 di-ester groups joining β-D-deoxy-ribofuranose 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 Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular 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 deoxy-ribose 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 stereo-chemical 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 conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the poly-nucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all specimens (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleo-tides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid ('structure *B*' in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the inter-nucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[4] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through

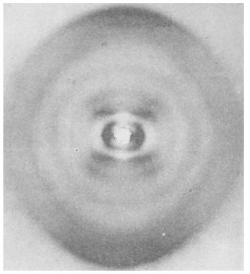


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion (J_n^2) on the nth layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[5] being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to C in Fig. 2.

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

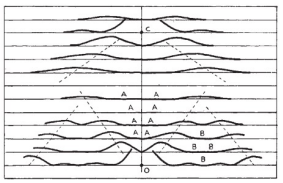


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About C on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

We wish to thank Prof. J. T. Randall for encouragement; Profs. E. Chargaff, R. Signer, J. A. V. Butler and Drs. J. D. Watson, J. D. Smith, L. Hamilton, J. C. White and G. R. Wyatt for supplying material without which this work would have been impossible; also Drs. J. D. Watson and Mr. F. H. C. Crick for stimulation, and our colleagues R. E. Franklin, R. G. Gosling, G. L. Brown and W. E. Seeds for discussion. One of us (H. R. W.) wishes to acknowledge the award of a University of Wales Fellowship.

M. H. F. Wilkins
Medical Research Council Biophysics Research Unit,
A. R. Stokes
H. R. Wilson
Wheatstone Physics Laboratory,
King's College, London.
April 2.

[1] Astbury, W. T., Symp. Soc. Exp. Biol., 1, Nucleic Acid (Cambridge Univ. Press, 1947).
[2] Riley, D. P., and Oster, G., *Biochim. et Biophys. Acta*, **7**, 526 (1951).
[3] Wilkins, M. H. F., Gosling, R. G., and Seeds, W. E., *Nature*, **167**, 759 (1951).
[4] Astbury, W. T., and Bell, F. O., Cold Spring Harb. Symp. Quant. Biol., **6**, 109 (1938).
[5] Cochran, W., Crick, F. H. C., and Vand, V., *Acta Cryst.*, **5**, 581 (1952).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

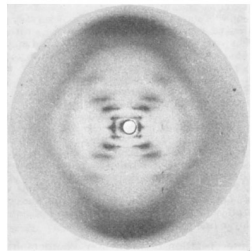
Molecular Configuration in Sodium Thymonucleate

Sodium thymonucleate fibres give two distinct types of X-ray diagram. The first corresponds to a crystalline form, structure *A*, obtained at about 75 per cent relative humidity; a study of this is described in detail elsewhere[1]. At higher humidities a different structure, structure *B*, showing a lower degree of order, appears and persists over a wide range of ambient humidity. The change from *A* to *B* is reversible. The water content of structure *B* fibres which undergo this reversible change may vary from 40–50 per cent to several hundred per cent of the dry weight. Moreover, some fibres never show structure *A*, and in these structure *B* can be obtained with an even lower water content.

The X-ray diagram of structure *B* (see photograph) shows in striking manner the features characteristic of helical structures, first worked out in this laboratory by Stokes (unpublished) and by Crick, Cochran and Vand[2]. Stokes and Wilkins were the first to propose such structures for nucleic acid as a result of direct studies of nucleic acid fibres, although a helical structure had been previously suggested by Furberg (thesis, London, 1949) on the basis of X-ray studies of nucleosides and nucleotides.

While the X-ray evidence cannot, at present, be taken as direct proof that the structure is helical, other considerations discussed below make the existence of a helical structure highly probable.

Structure *B* is derived from the crystalline structure *A* when the sodium thymonucleate fibres take up quantities of water in excess of about 40 per cent of their weight. The change is accompanied by an increase of about 30 per cent in the length of the fibre, and by a substantial re-arrangement of the molecule. It therefore seems reasonable to suppose that in structure *B* the structural units of sodium thymonucleate (molecules or groups of molecules) are relatively free from the influence of neighbouring



Sodium deoxyribose nucleate from calf thymus. Structure *B*

molecules, each unit being shielded by a sheath of water. Each unit is then free to take up its least-energy configuration independently of its neighbours and, in view of the nature of the long-chain molecules involved, it is highly likely that the general form will be helical[3]. If we adopt the hypothesis of a helical structure, it is immediately possible, from the X-ray diagram of structure *B*, to make certain deductions as to the nature and dimensions of the helix.

The innermost maxima on the first, second, third and fifth layer lines lie approximately on straight lines radiating from the origin. For a smooth single-strand helix the structure factor on the nth layer line is given by:

$$F_n = J_n(2\pi rR)\exp i\,n(\psi + \tfrac{1}{2}\pi),$$

where $J_n(u)$ is the nth-order Bessel function of u, r is the radius of the helix, and R and ψ are the radial and azimuthal co-ordinates in reciprocal space[2]; this expression leads to an approximately linear array of intensity maxima of the type observed, corresponding to the first maxima in the functions J_1, J_2, J_3, etc.

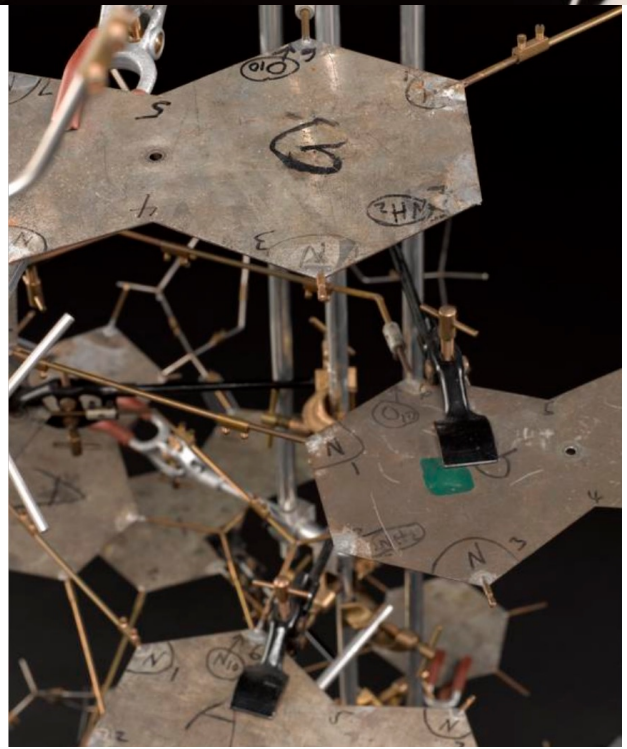
If, instead of a smooth helix, we consider a series of residues equally spaced along the helix, the transform in the general case treated by Crick, Cochran and Vand is more complicated. But if there is a whole number, m, of residues per turn, the form of the transform is as for a smooth helix with the addition, only, of the same pattern repeated with its origin at heights mc^*, $2mc^*$ . . . etc. (c is the fibre-axis period).

In the present case the fibre-axis period is 34 A. and the very strong reflexion at 3·4 A. lies on the tenth layer line. Moreover, lines of maxima radiating from the 3·4-A. reflexion as from the origin are visible on the fifth and lower layer lines, having a J_4 maximum coincident with that of the origin series on the fifth layer line. (The strong outer streaks which apparently radiate from the 3·4-A. maximum are not, however, so easily explained.) This suggests strongly that there are exactly 10 residues per turn of the helix. If this is so, then from a measurement of R_R the position of the first maximum on the nth layer line (for $n \gg$), the radius of the helix, can be obtained. In the present instance, measurements of R_1, R_2, R_3 and R_5 all lead to values of r of about 10 A.

Watson & Crick

Wilkins, Stokes & Wilson

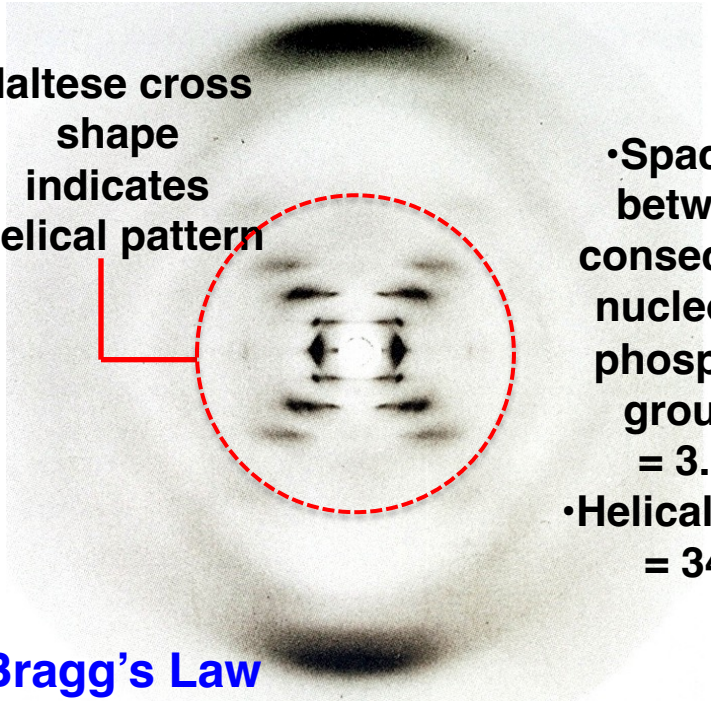
Franklin (died in 1958) & Gosling



The (almost original) model of DNA structure by Watson & Crick (Science Museum London)

Information that Watson and Crick used to propose the double helix model:

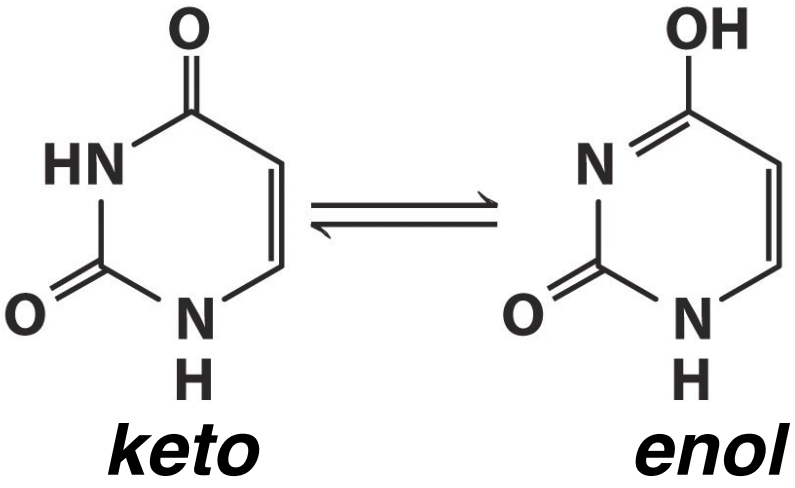
1) R. Franklin DNA fibers
X-ray diffraction data helped to define shape and dimensions



- Spacing between consecutive nucleotide phosphate groups = 3.4A
- Helical Pitch = 34A

Bragg's Law
used to interpret X-ray diffraction pictures

2) bases are in the keto conformation



keto ***enol***

4) Density measurements:
~2 polymers/helix

5) C2´ endo sugar pucker conformation

3) Chargaff´s rules:
(G+C)/(A+T) can vary but (G+A)/(C+T) = G/C = A/T =1

Chargaff. 1950: "It is, however, noteworthy-whether this is more than accidental, cannot yet be said-that in all deoxypentose nucleic acids examined thus far the molar ratios of total purines to total pyrimidines, and also of adenine to thymine and of guanine to cytosine, were not far from 1".

The original model for DNA structure
Watson and Crick (1953)
Nature 171, 964-967

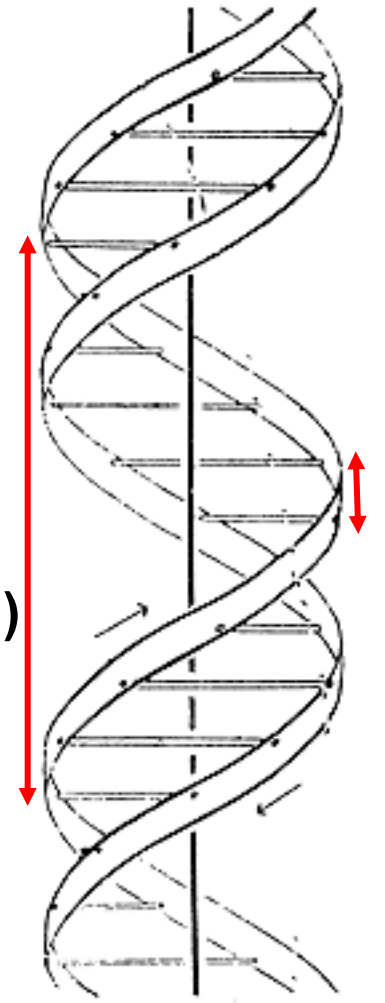
- See posted PDF of original paper

Essential features of the model that proved correct:

1) *Antiparallel* right-handed double helix

2) Strands are linked by complementary sets of hydrogen bond donors and acceptor groups on bases

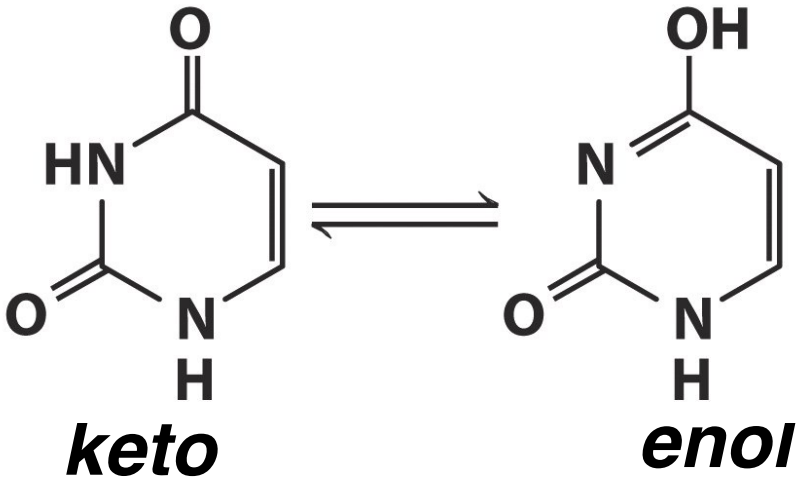
Helical Pitch = 34 A (10 residues/turn)



Rise/ residue = 3.4 A

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

Why is it important to know the exact keto/enol conformation of bases ?



keto ***enol***

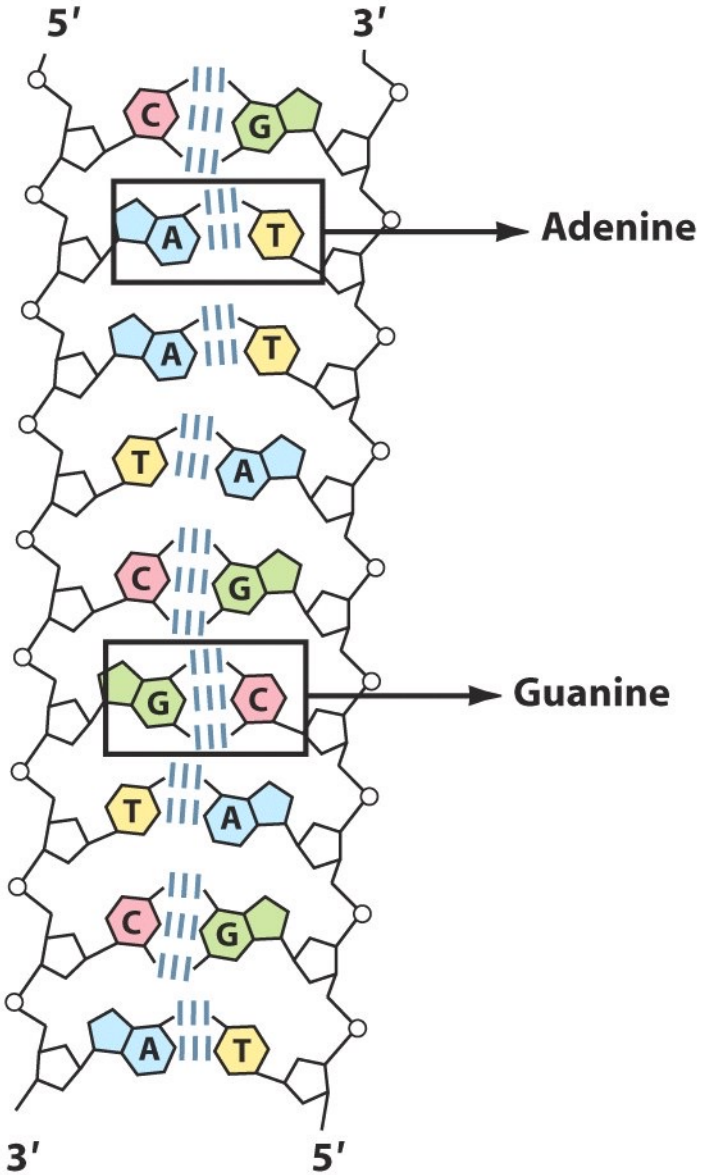
A: The keto/enol conformations influence electrons in the π orbital system and therefore base stacking geometry

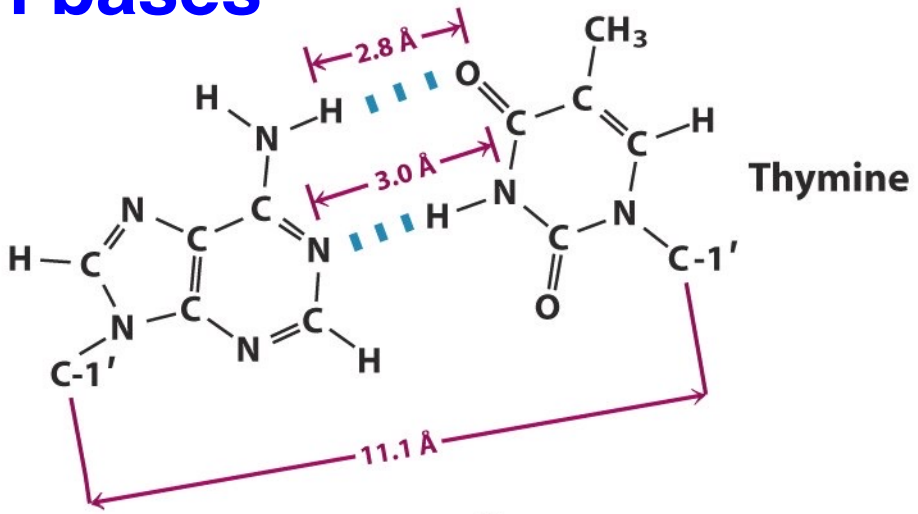
B: The keto/enol conformation indirectly impact sugar puckering because of potential steric hindrances with C2´ endo sugar pucker

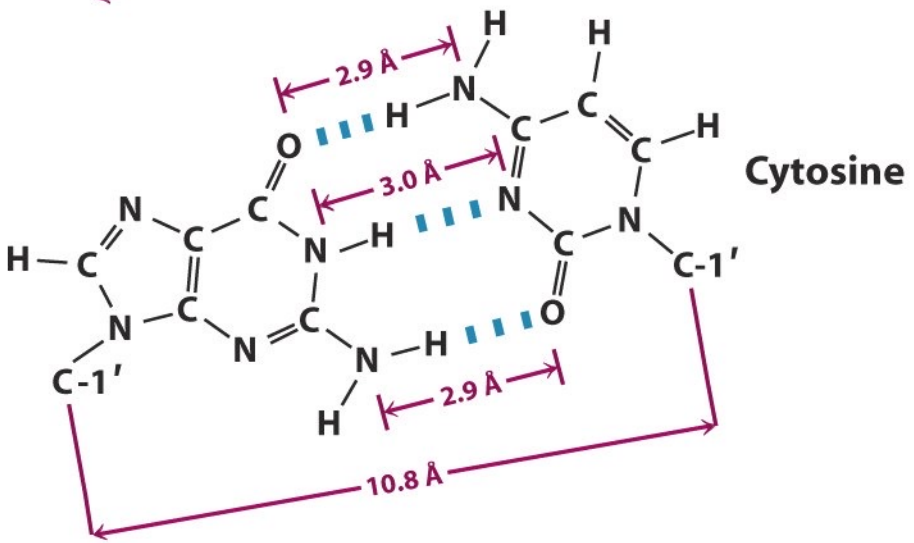
C: Because bases in RNA and DNA have different keto/enol conformations

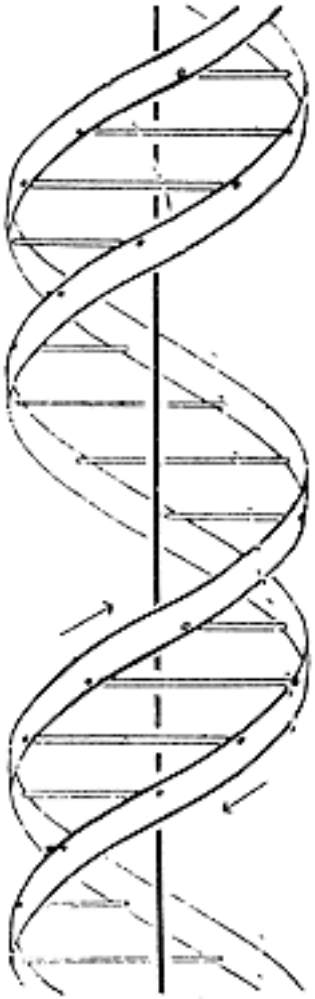
D: The keto/enol conformations affect the ability of specific bases substituents to donate/accept H-bonds and therefore base pairing partners

complementary sets of donors and acceptor groups on bases

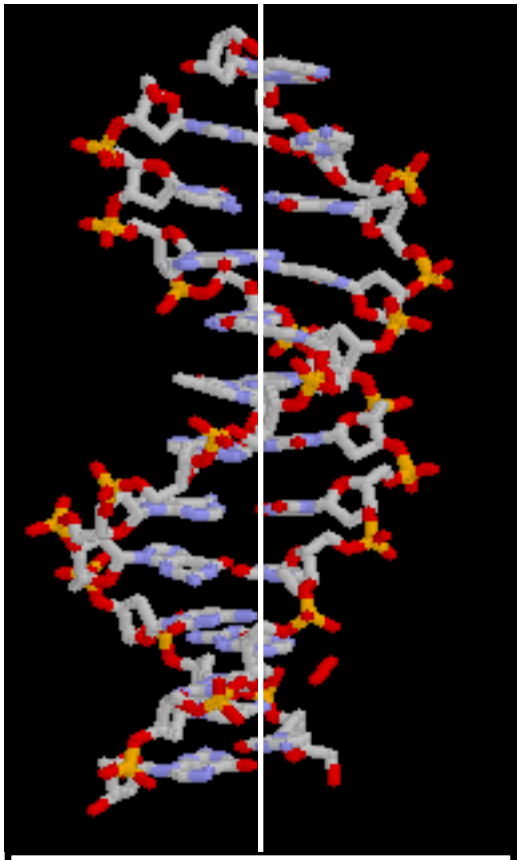








Watson-Crick Model



The Dickerson Dodecamer
X-ray structure
(CGCGAATTCGCG)

A comparison of the Watson-Crick model (1953) and of the first B-DNA structure solved (1980)

PDB ID:
1BNA

PyMol: DickersonDodecamer.pse

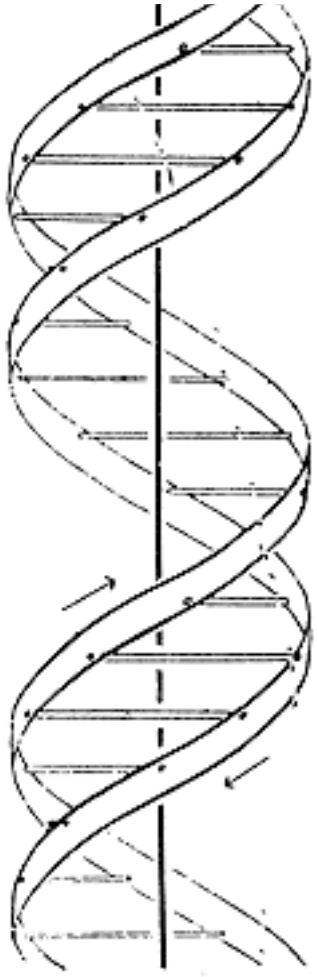
Differences between the WC model and the actual DNA structure solved ?

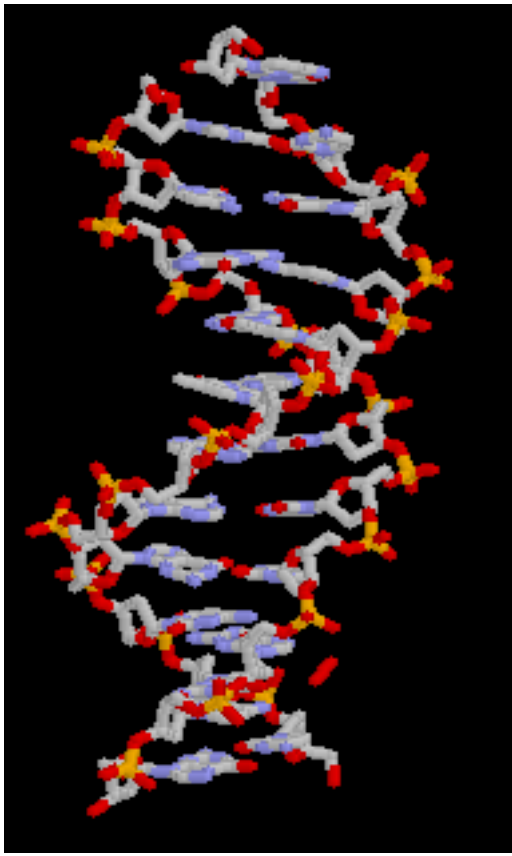
A: the two bases forming each base pair are not aligned in a plane in the actual structure

B: Base pairs cross the helical axis in the structure while they´re shifted away from the axis in the WC model

C: No major difference

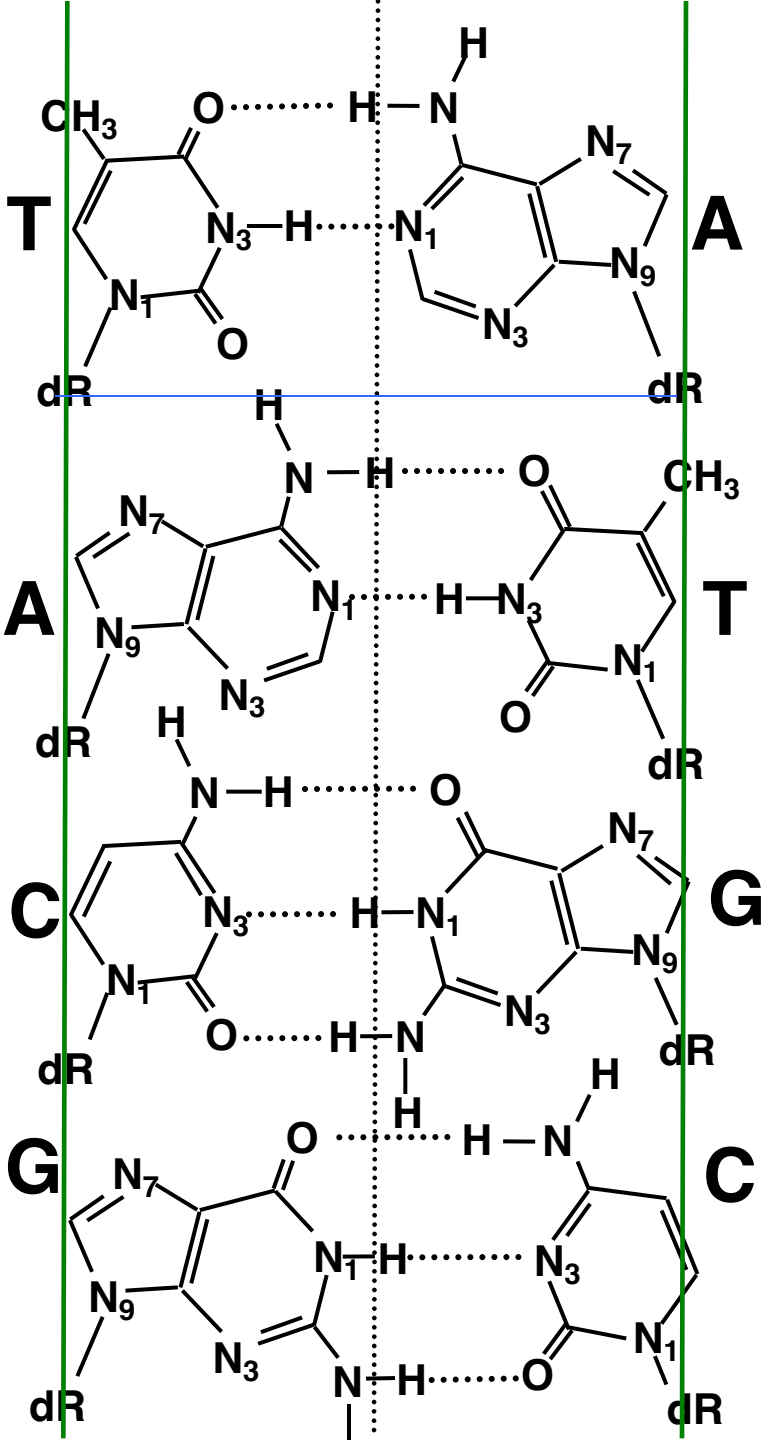
D: Base pairs are not perpendicular to the helical axis in the actual structure, as opposed to the WC model



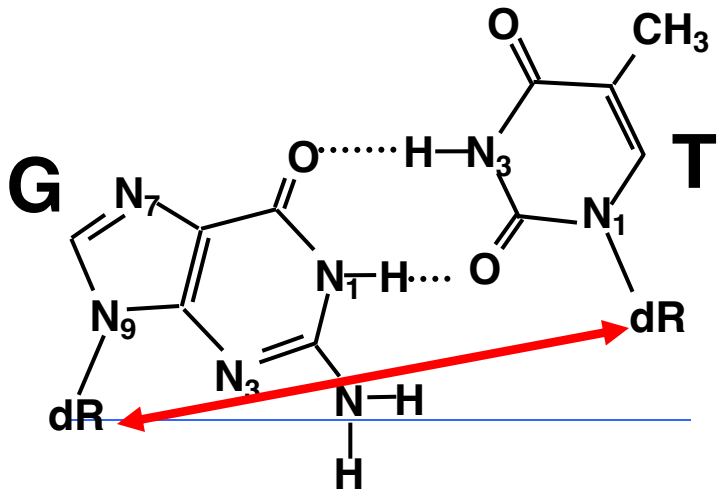


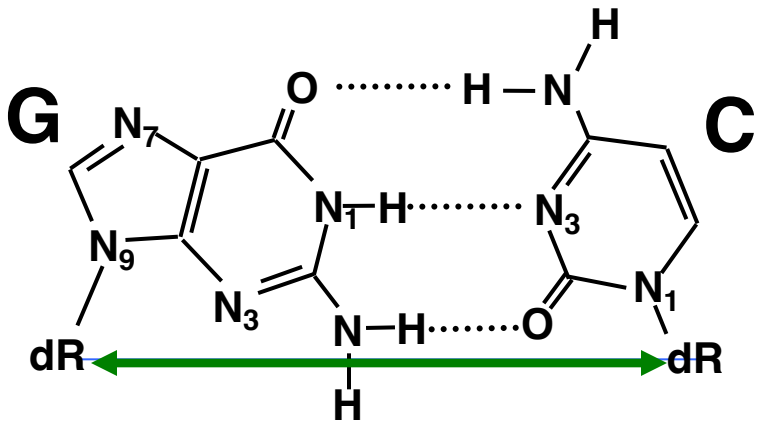
Isostericity of Watson-Crick Base Pairs

(and non isostericity of non WC base pairs)

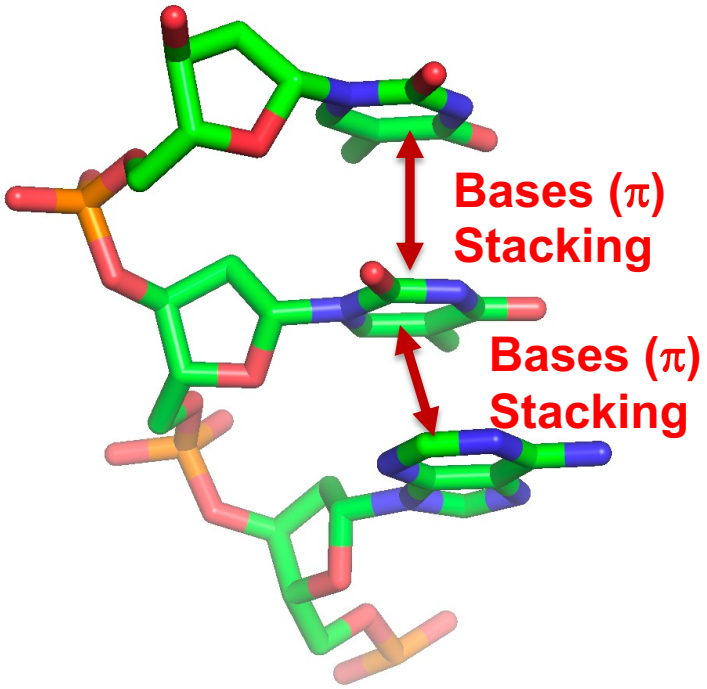


Example of a G-T non WC base pair





Bases of one strand are stacked onto each other:



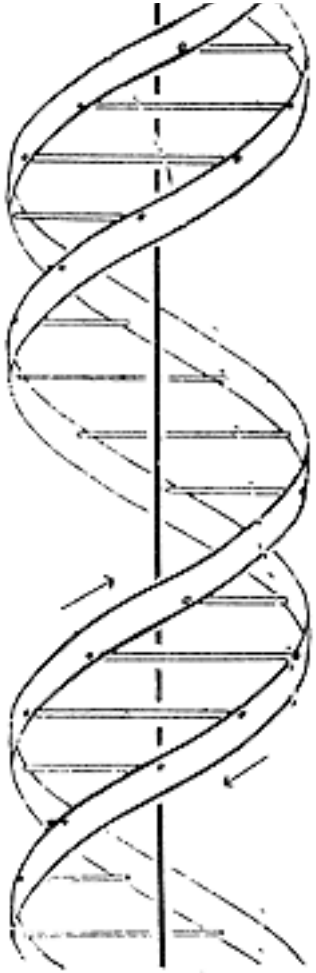
DNA double helix structure is primarily stabilized by stacking, not hydrogen bonding

Right-handed double helix

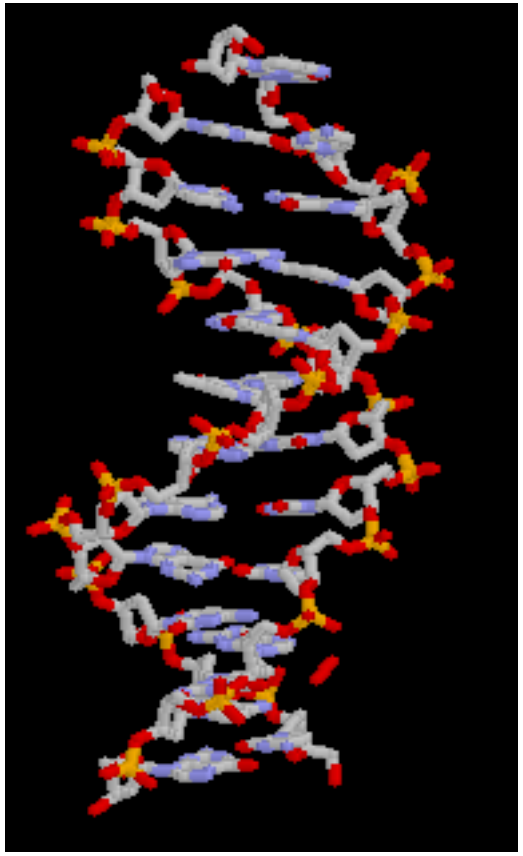
Right-handed

Left-handed

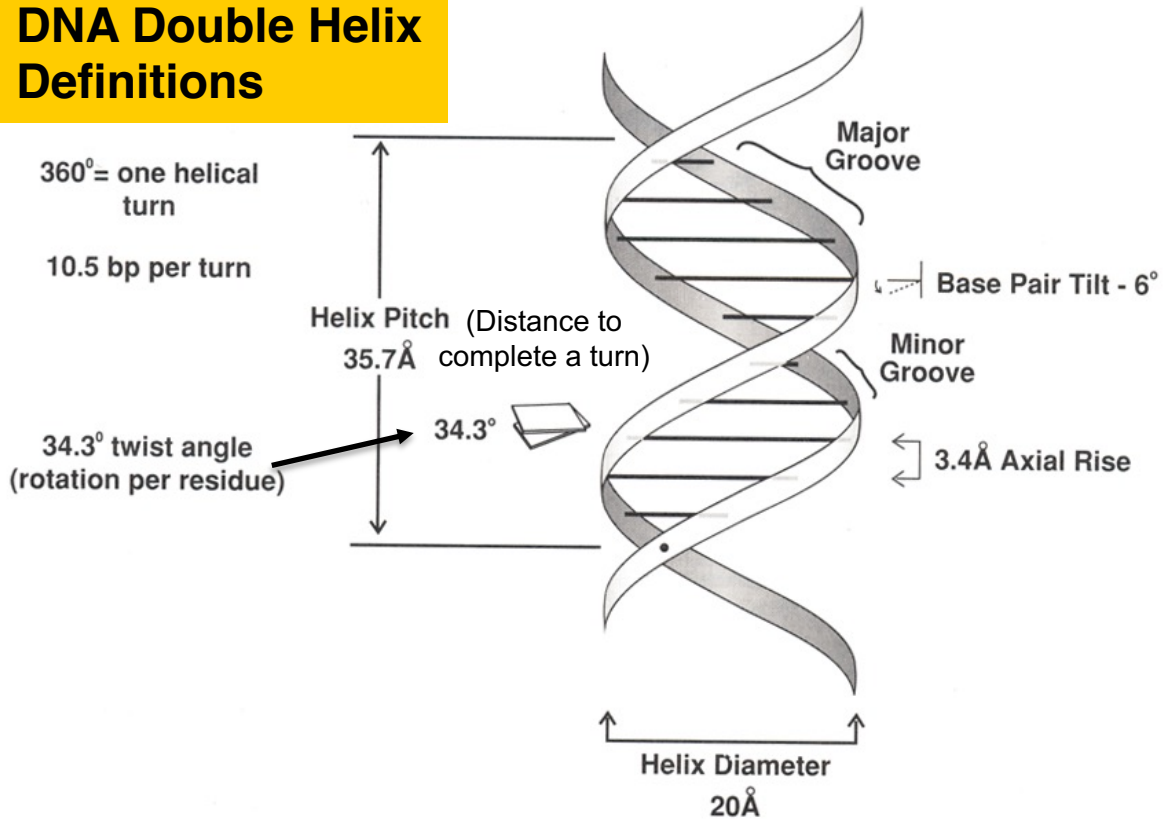
Watson-Crick model

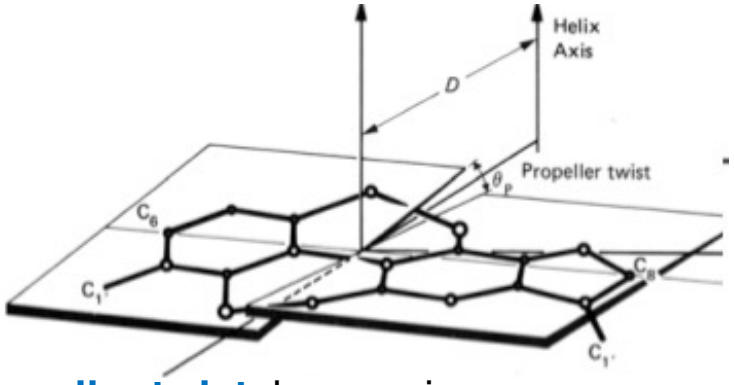


Dickerson dodecamer



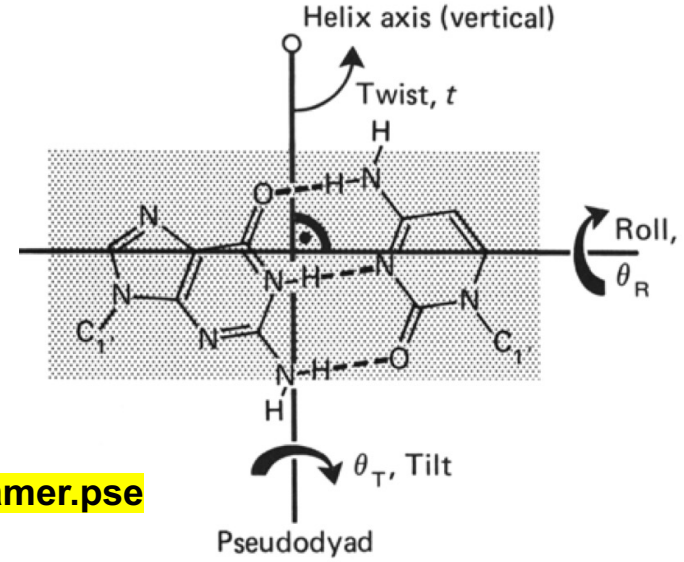
DNA Double Helix Definitions





Propeller twist: base pairs are not planar. The angle between the two planes of the bases defines the propeller twist angle

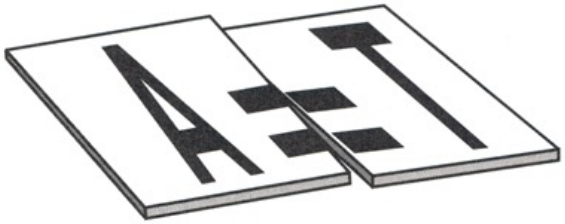
PyMol: DickersonDodecamer.pse
PyMol: GCbasepair.pse



Tilt

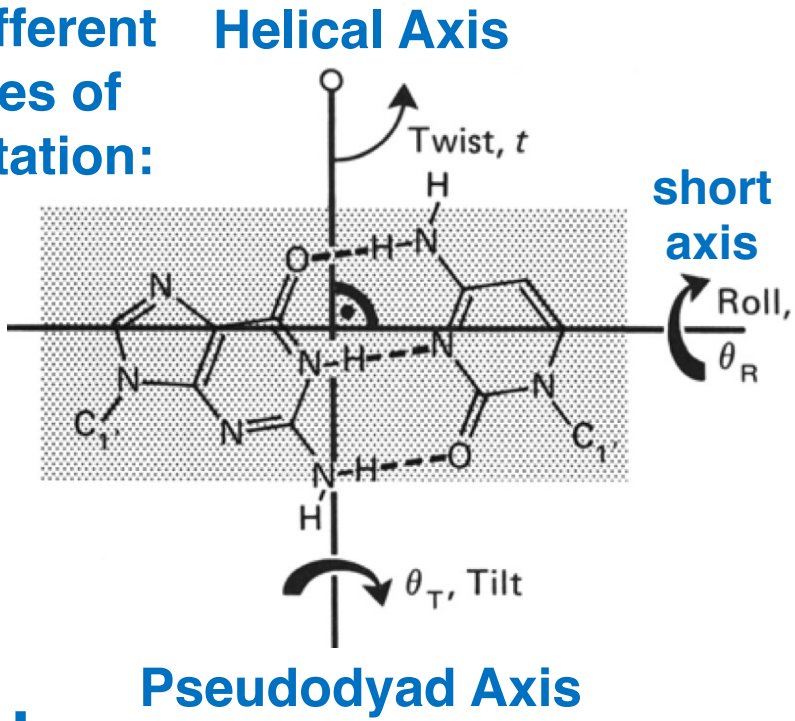
Roll

Twist



Propeller Twist

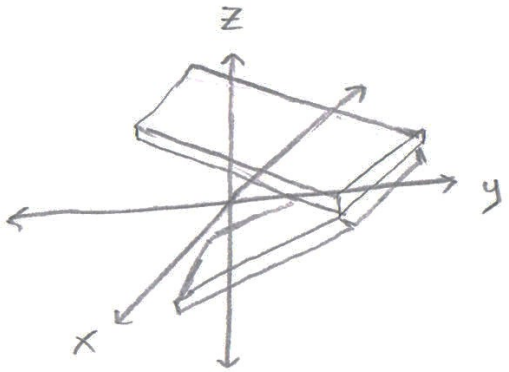
Different axes of rotation:



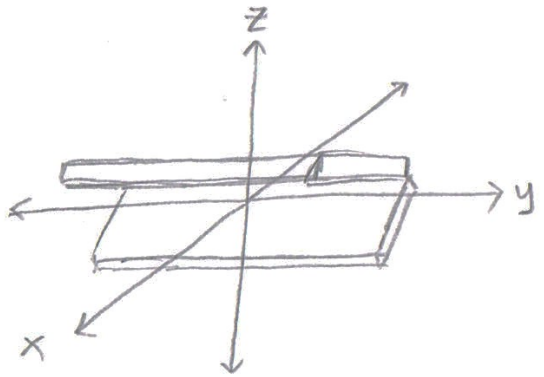
Base pair tilt: An average mean plane of the base pair can be defined (ignoring the propeller twist) – see gray area above. The rotation of this plane about the pseudo-dyad axis defines base pair tilt.

Base pair roll: the degree of departure of the mean plane of the base pairs from the perpendicular helix axis on the short axis of the base pairs.

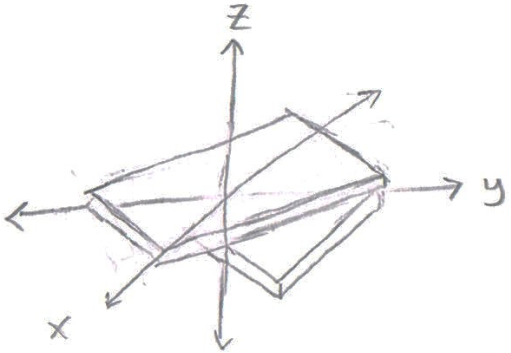
Helical twist. Defines the degree of rotation when moving vertically from one base pair to the next.



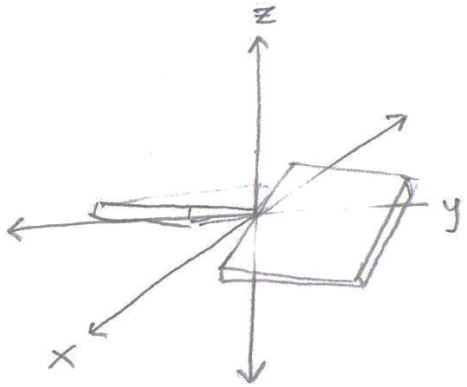
Tilt



Roll



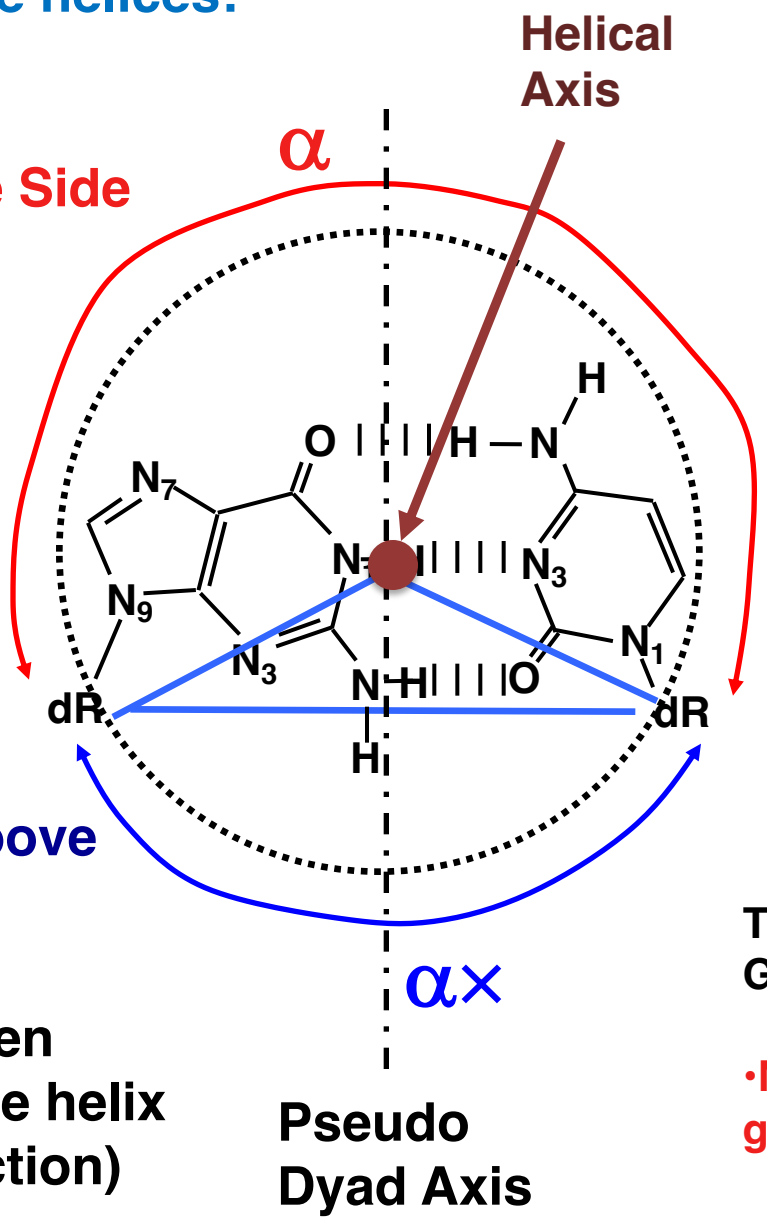
Twist

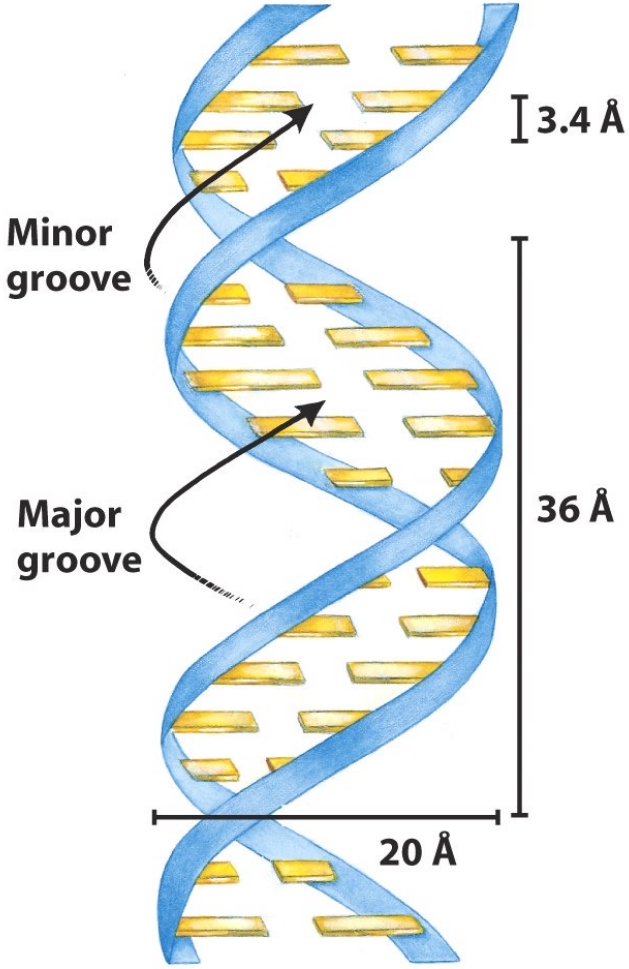


Propeller Twist

https://en.wikipedia.org/wiki/Nucleic_acid_double_helix

Major Groove and Minor Groove in B-form double helices:



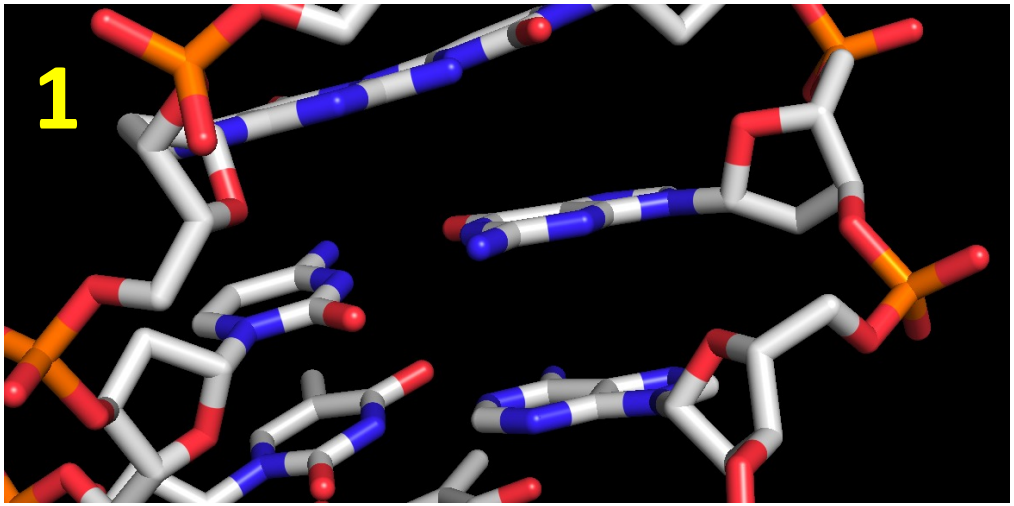


Tip to visualize Major and Minor Groove sides in structures:

- **N7 of Purines always on major groove side**
- **N3 of Purines always on minor groove side**

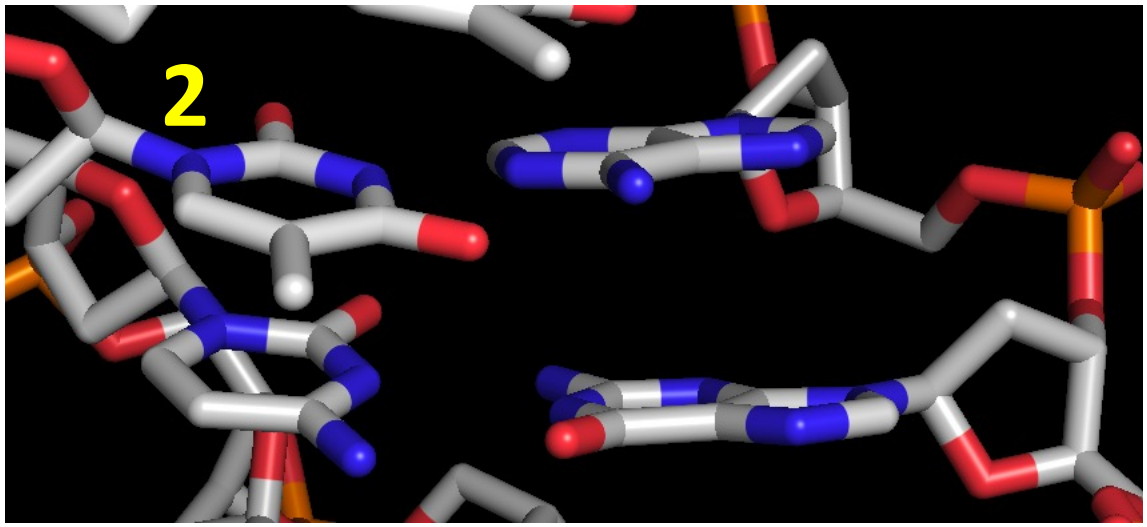
PyMol: DickersonDodecamer.pse

Which grooves are we looking at?



A: 1-Major 2-Major

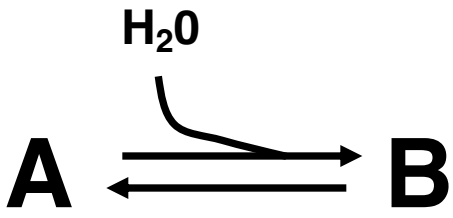
B: 1-Minor 2-Minor



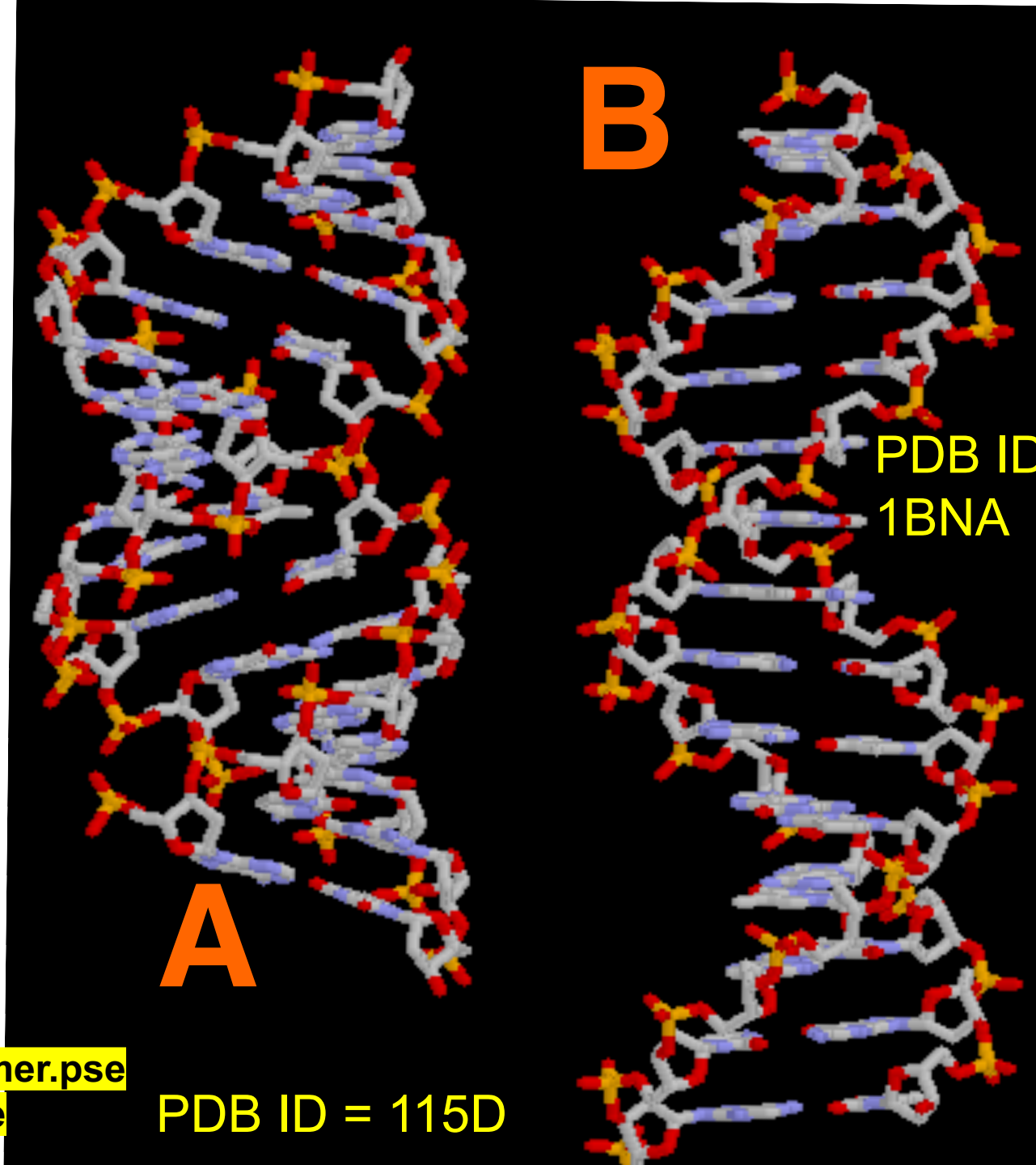
C: 1-Major 2-Minor

D: 1-Minor 2-Major

A vs B DNA



A-dsRNA = 2KYD

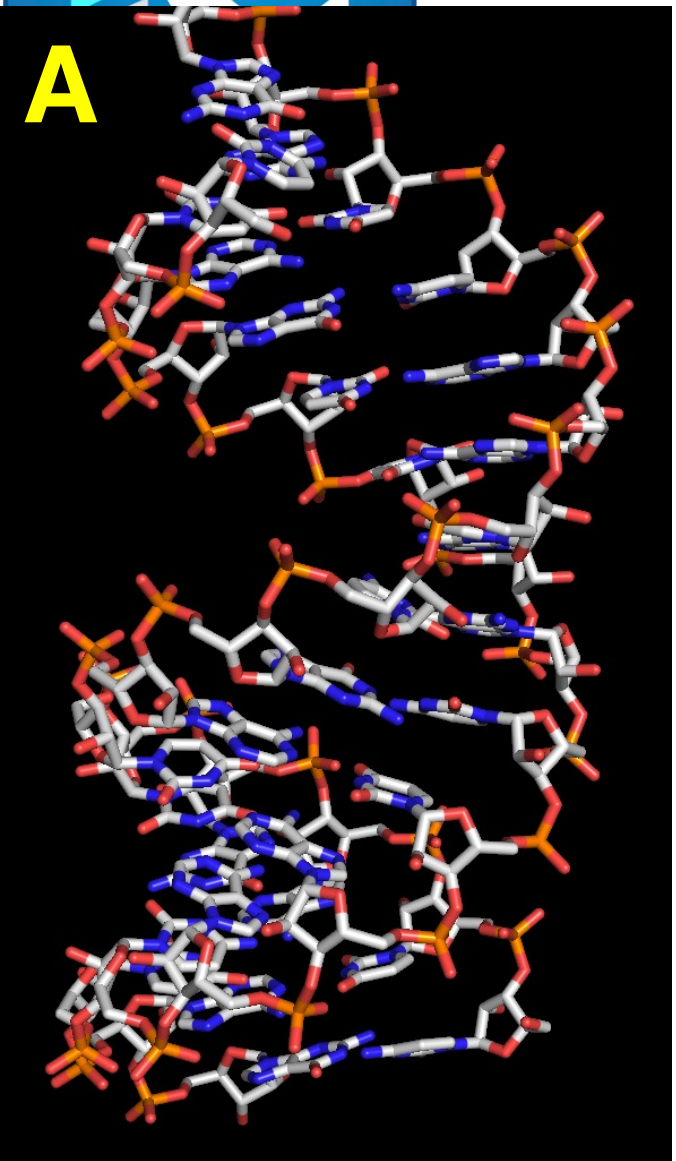


PyMol: DickersonDodecamer.pse
PyMol: A-form_dsRNA.pse

PDB ID = 115D

What differences can you see between A and B forms?

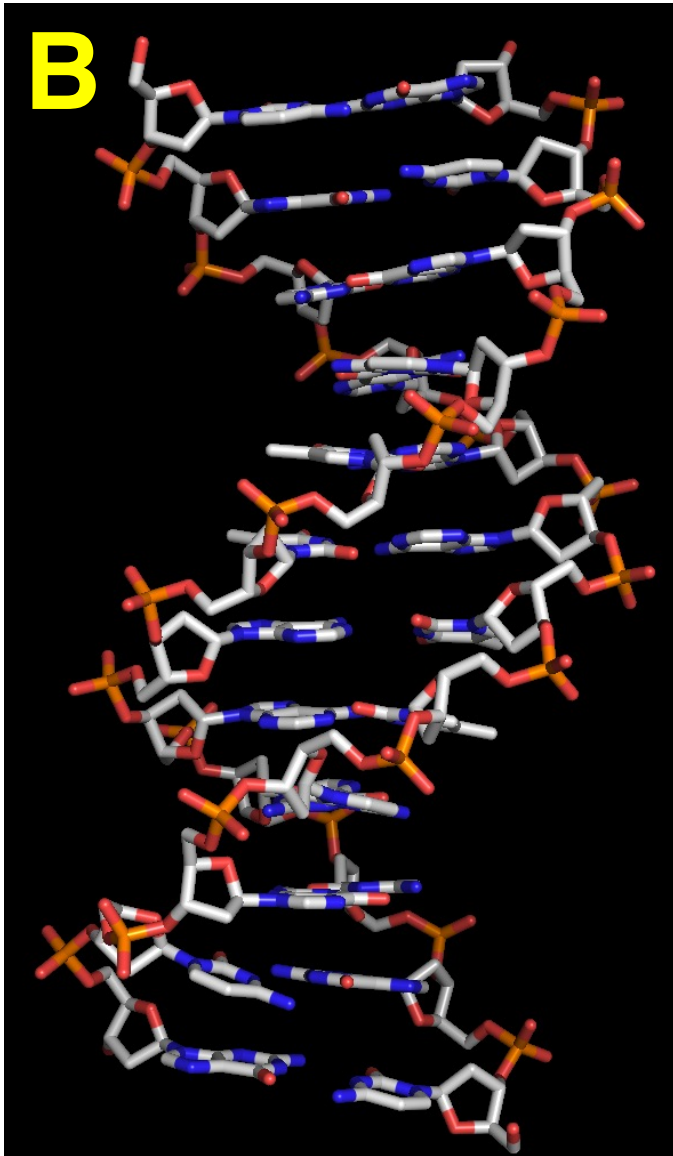
A: A is more compact than B (shorter helical pitch)



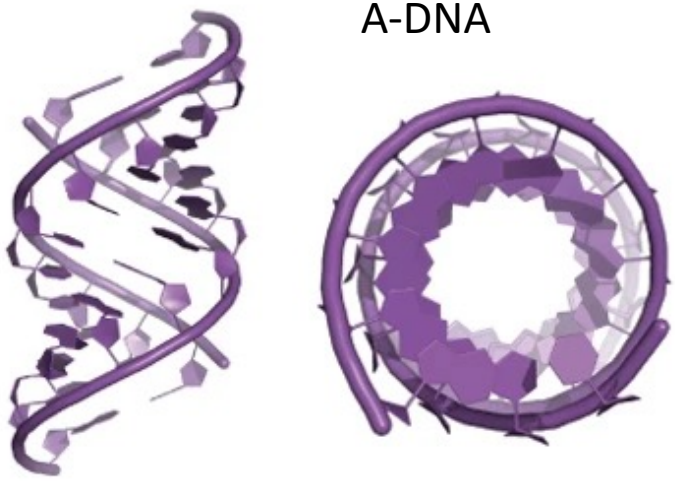
B: Base pairs are more tilted in A vs. B.

C: Sugar pucker is C3´endo in A, C2´endo in B

D: Major groove is deeper in B than A

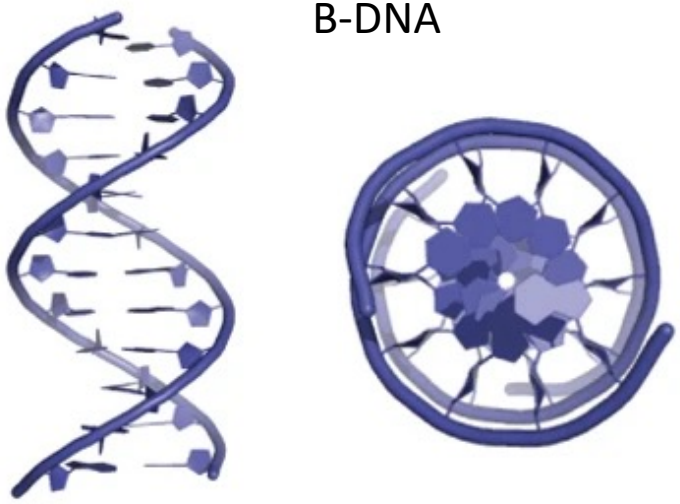


A-DNA



BP rise ~2.5 Å Diameter ~23 Å

B-DNA



BP rise ~3.4 Å Diameter ~18 Å

	B-DNA	A-DNA
Sugar pucker	C2'-endo	C3'-endo
Rise/residue	3.4 Å	2.6 Å
Residues/turn	10.5	11
Helical twist	34°	33°
Diameter	20 Å	26 Å
Tilt	6°	20°
Propellor twist	12°	15°

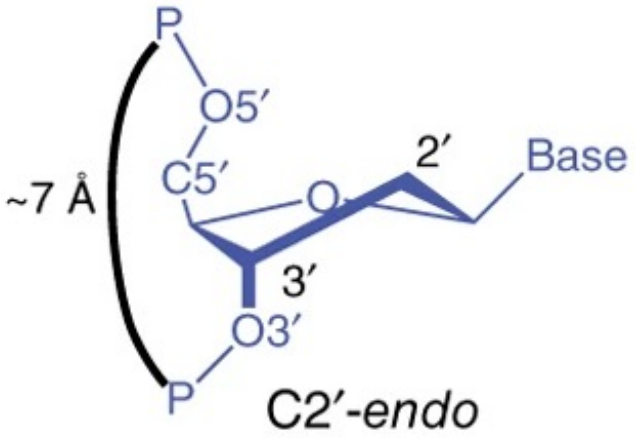
Major differences :

- **A DNA is shorter than B DNA: 1 helix turn is 28.6 A vs 34 A for B DNA. This is due to the 3´ endo sugar pucker in A**

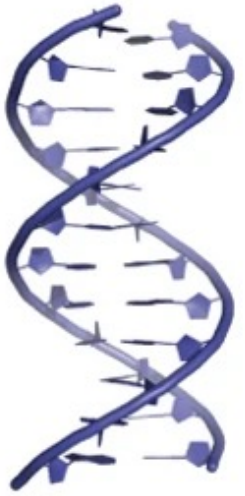
- **Bases of A-DNA are shifted away from the helical axis. This results in a deep major groove and in a shallow minor groove. There is a 6 A hole in helical projection.**

Exact values need not to be remembered…

Impact of Sugar puckering (C2´or C3´ endo) on double helix geometry

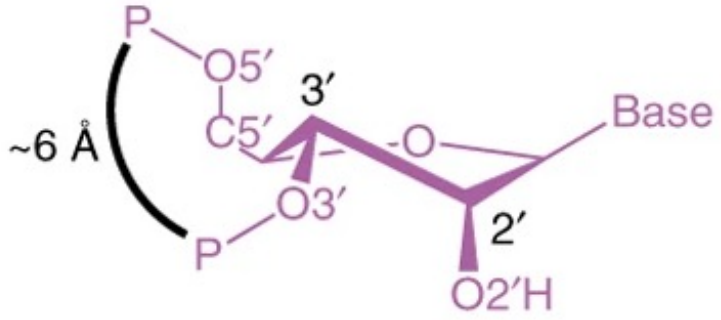


C2′-endo

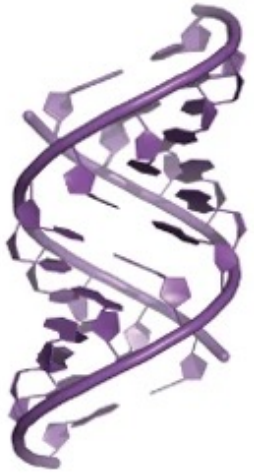


B-Form (DNA)

BP rise ~3.4 Å



C3′-endo

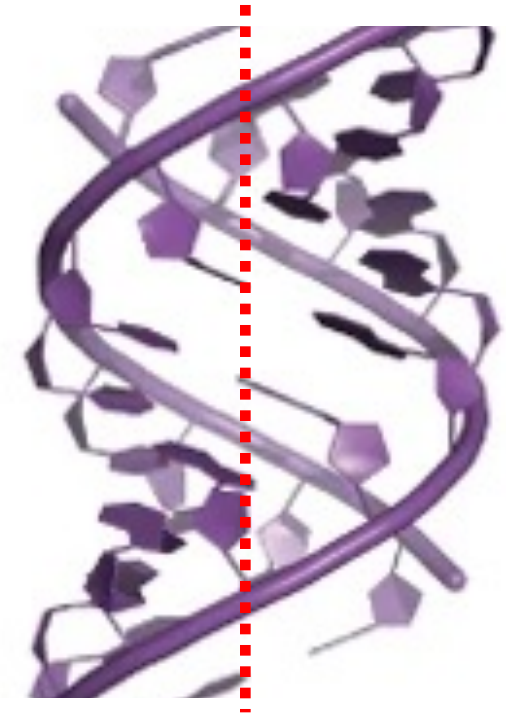


A-Form (DNA or dsRNA)

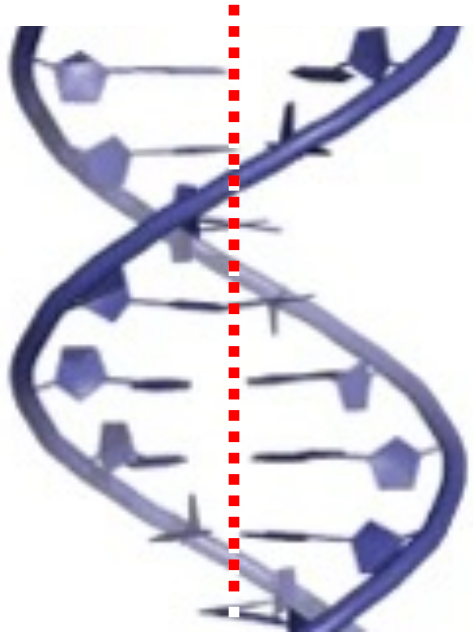
BP rise ~2.5 Å

Base tilting: A vs. B DNA

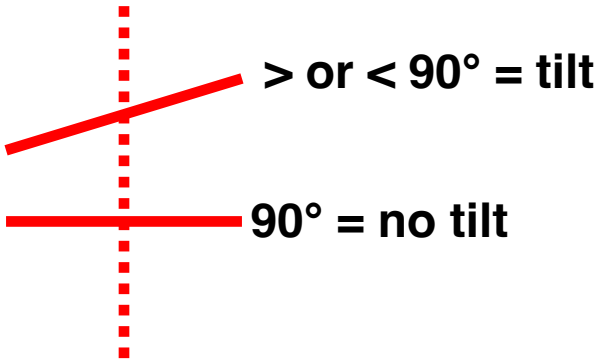
A-DNA/dsRNA



B-DNA

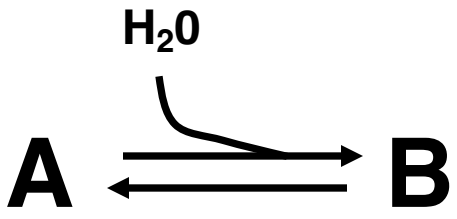


Consider Angles between the base pairs and helical axis (vertical dashed lines):



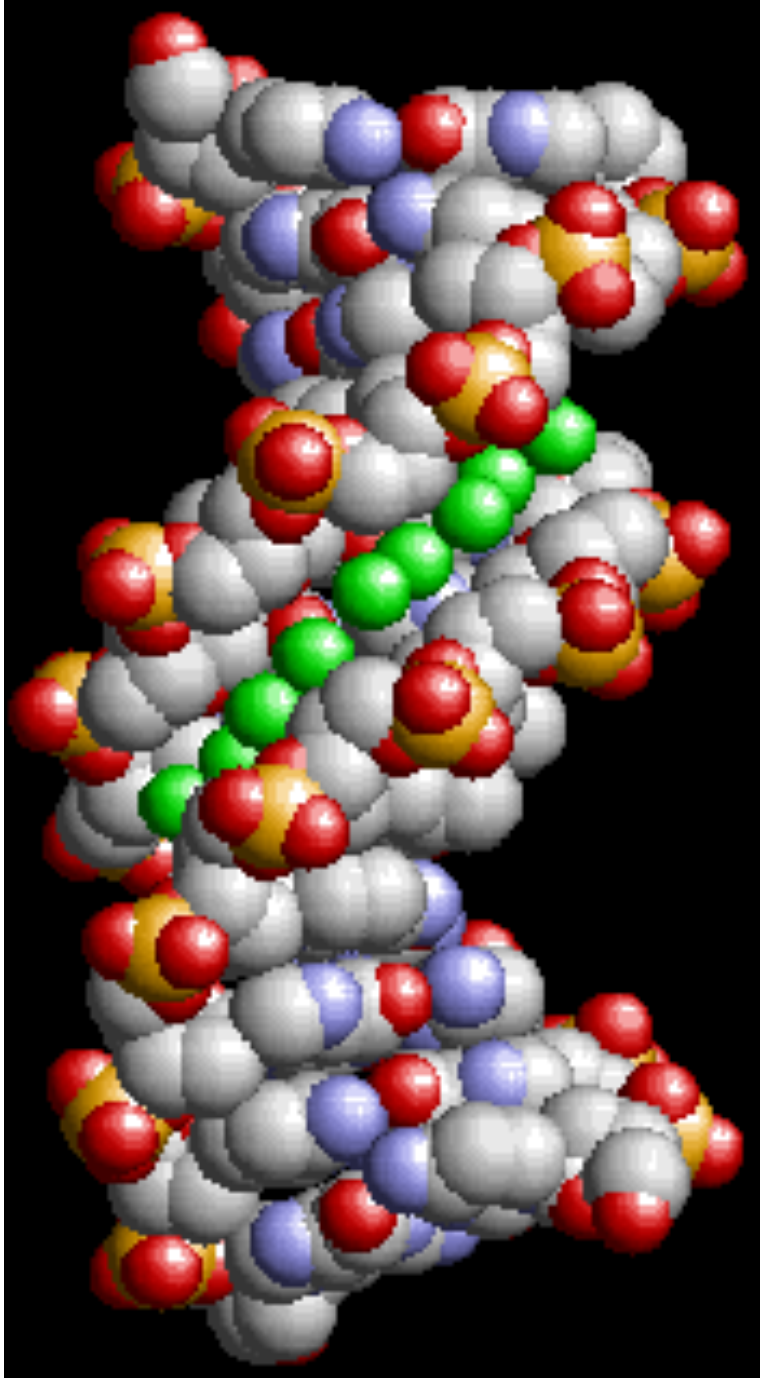
PyMol: A-form_dsRNA.pse

Water is essential in the transition A <--> B DNA



A water spine (green dots) has been proposed to exist in the minor groove of B-DNA that would stabilize the B-form

This concept is controversial and will not be detailed further

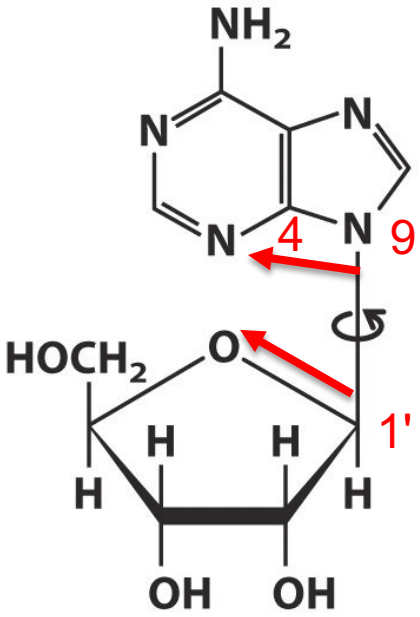


Glycosidic bond Anti /Syn conformations

Anti and Syn conformations are defined based on the torsion angle of the glycosidic bond

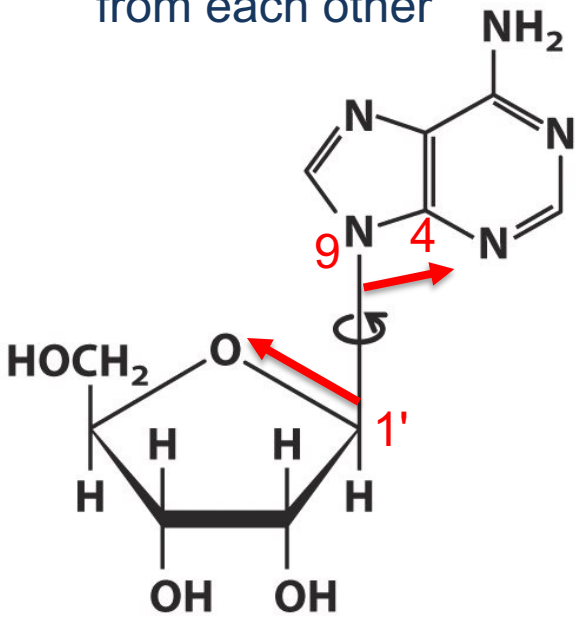
The sequence of atoms chosen to define the torsion angle to define anti/syn conformation is: O4'-C1'-N9-C4 for purines - O4'-C1'-N1-C2 for pyrimidines.

Syn A/G:
C1´-O and N9-C4 are pointing in same direction



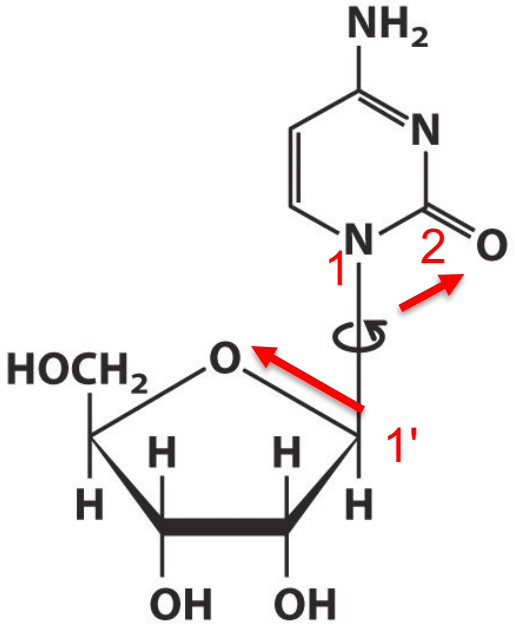
syn-Adenosine

Anti A/G:
C1´-O and N9-C4 are pointing away from each other



anti-Adenosine

Anti C/T:
C1´-O and N1-C2 are pointing away from each other

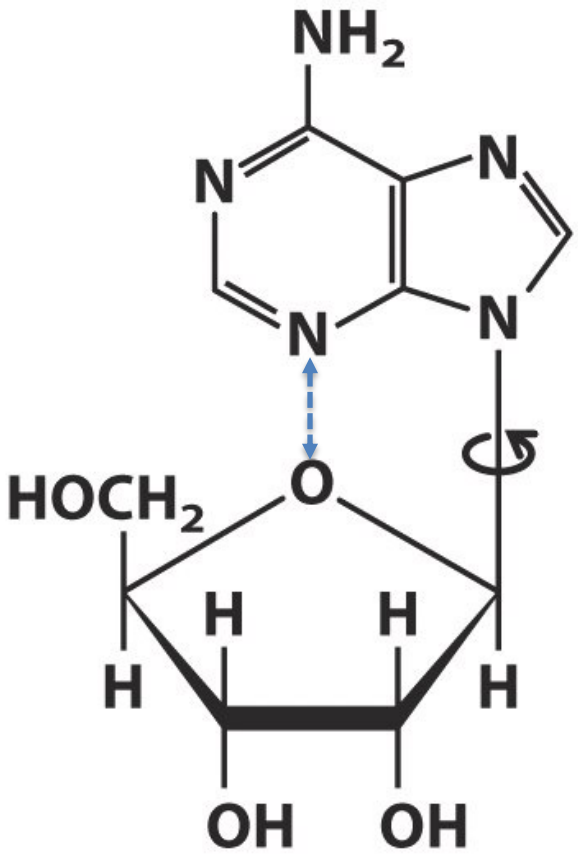


anti-Cytidine

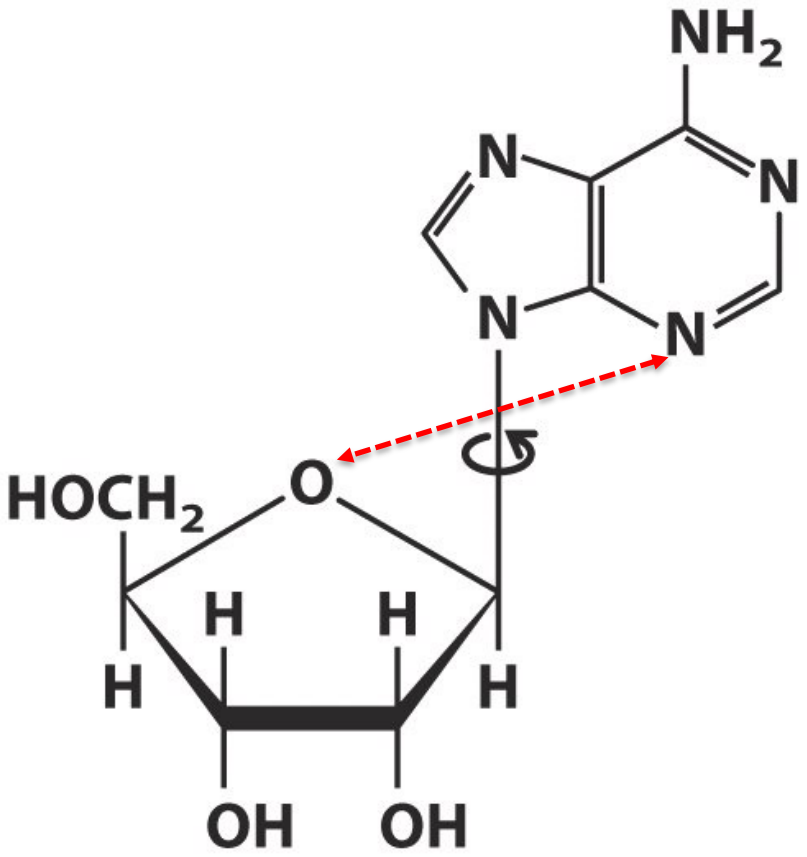
How to recognize Anti /Syn Glycosidic bonds

Syn A/G:
N_3 of base is **close/ on top** of **cyclic O of ribose**

Anti A/G:
N_3 of base is **away** from the **cyclic O of ribose**



***syn*-Adenosine**

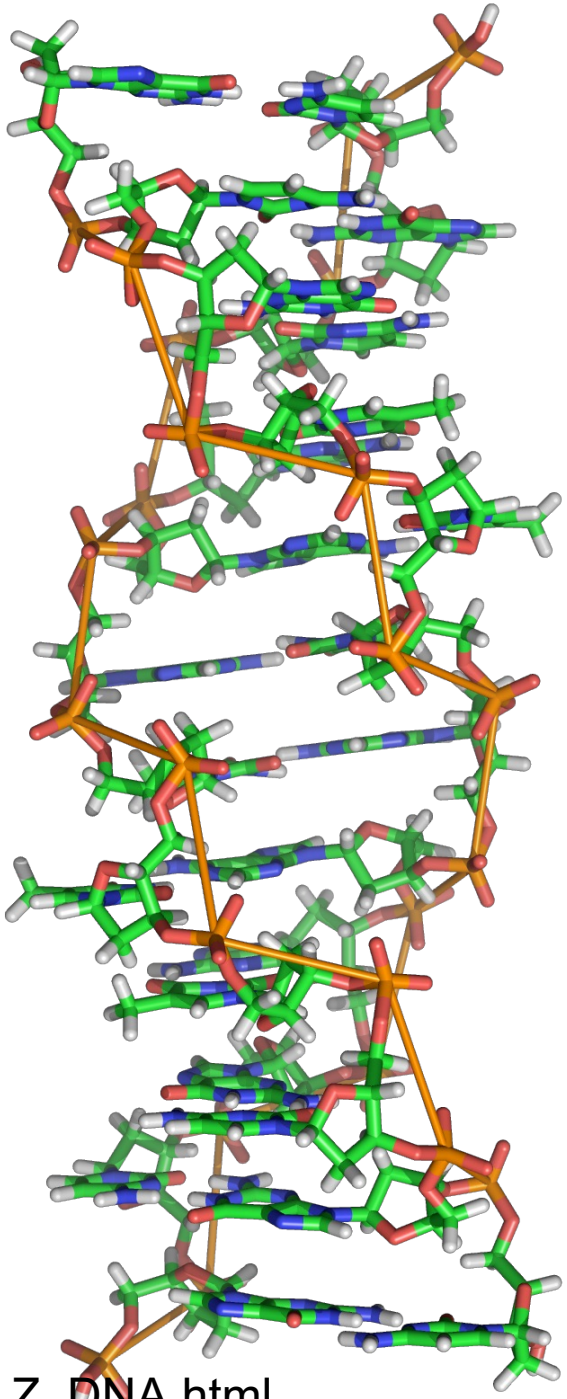


***anti*-Adenosine**

Z-DNA

- Occurs in DNA sequences with stretches of consecutive G-C base pairs

- Requires high salt concentrations for forming stable structures in vitro



- **Left handed** Helix

- jagged backbone

- G nucleotides switch their glycosidic bond conformations:
  anti -> *Syn*

- C nucleotides:
  No change

PDB ID:
1DCG

PyMol: ZDNA.pse

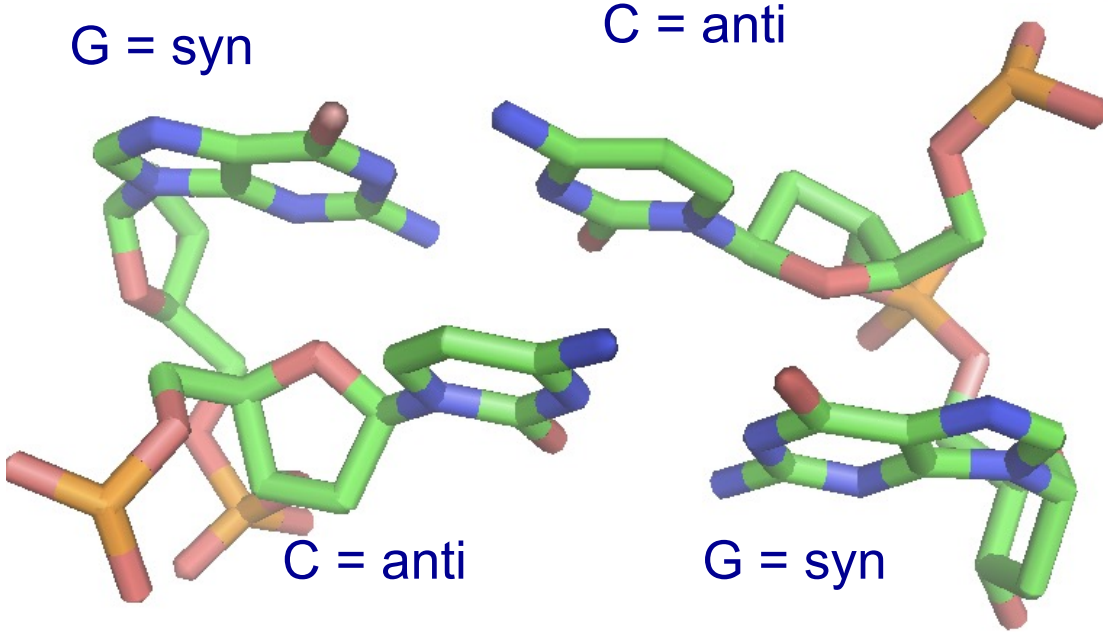
www.mun.ca/biology/scarr/A_B_Z_DNA.html

Consecutive Nucleotides flipping in Z-DNA

Focus on two consecutive base pairs

G = syn

C = anti



C = anti

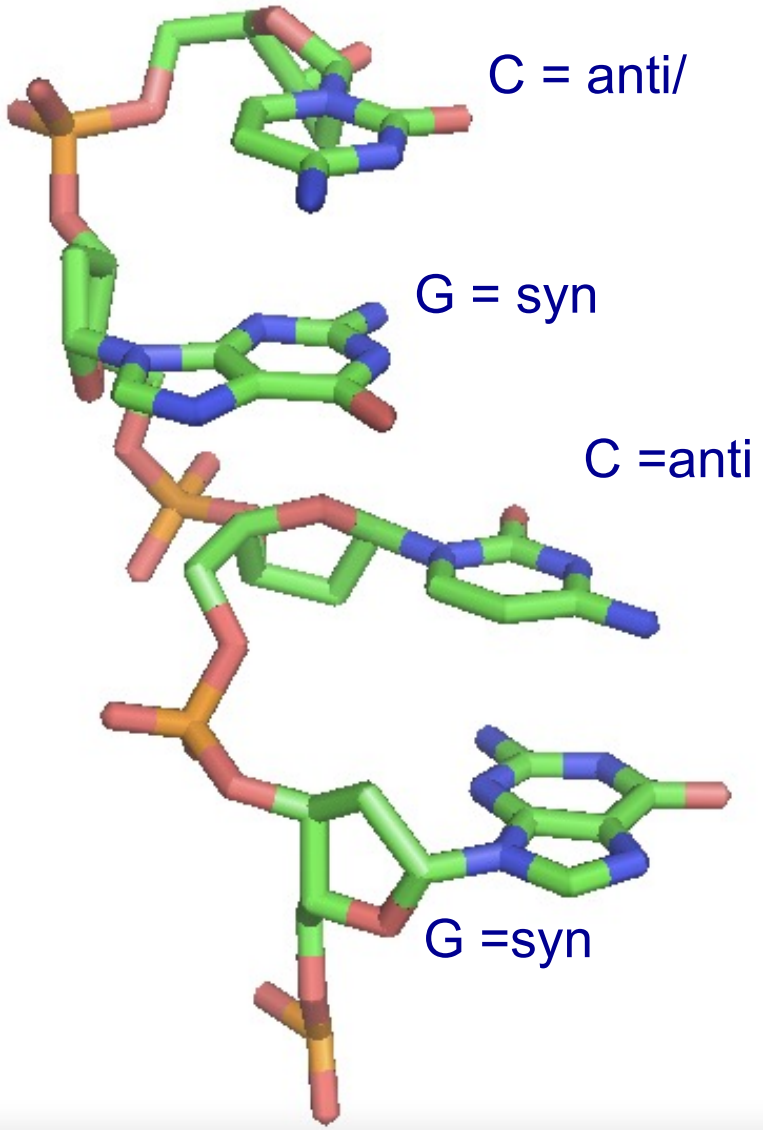
G = syn

Focus on one strand

C = anti/

G = syn

C =anti



G =syn

Can you recognize Glycosidic Bonds Conformations?

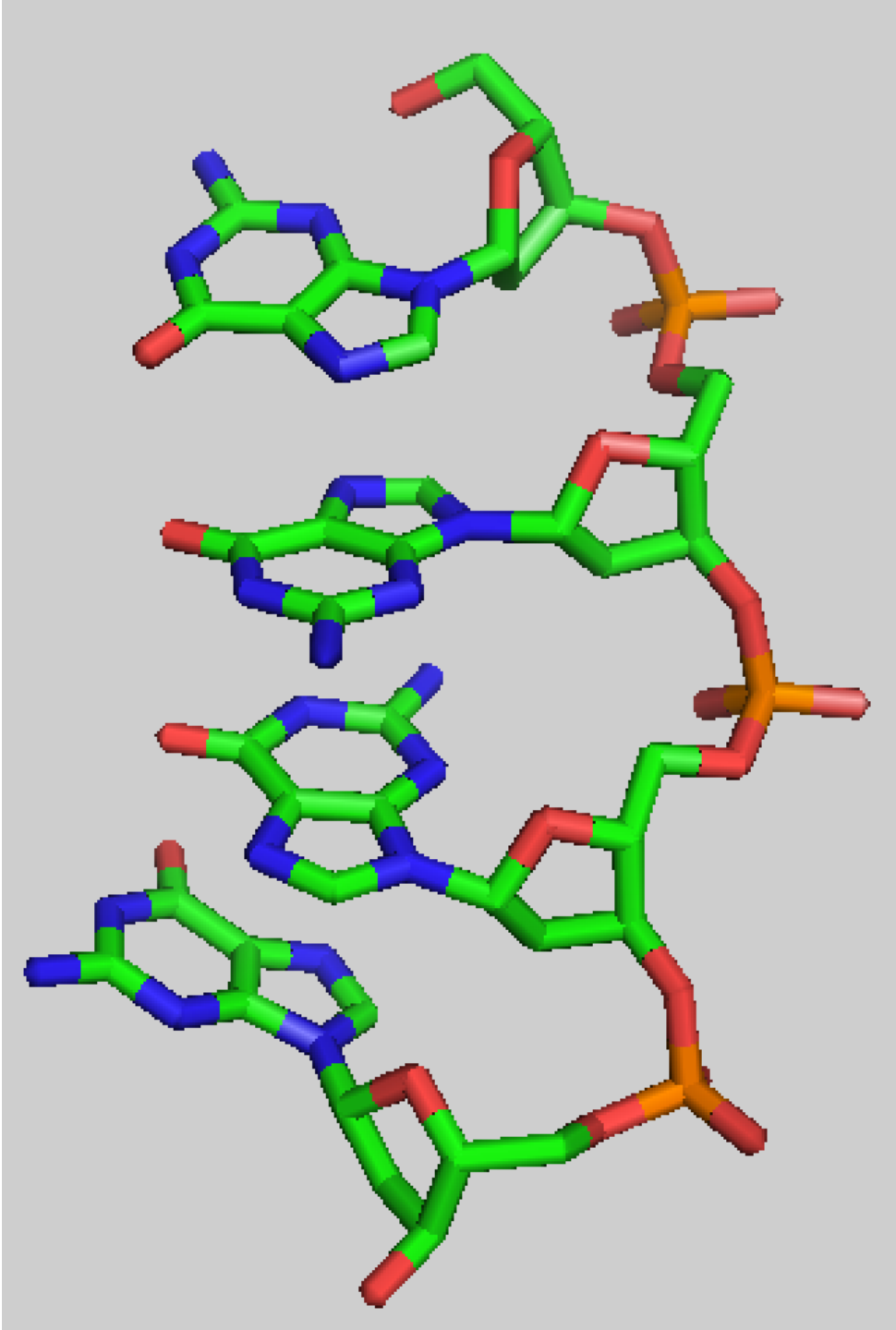
(Top to bottom)

A: Anti, Syn, Anti, Syn

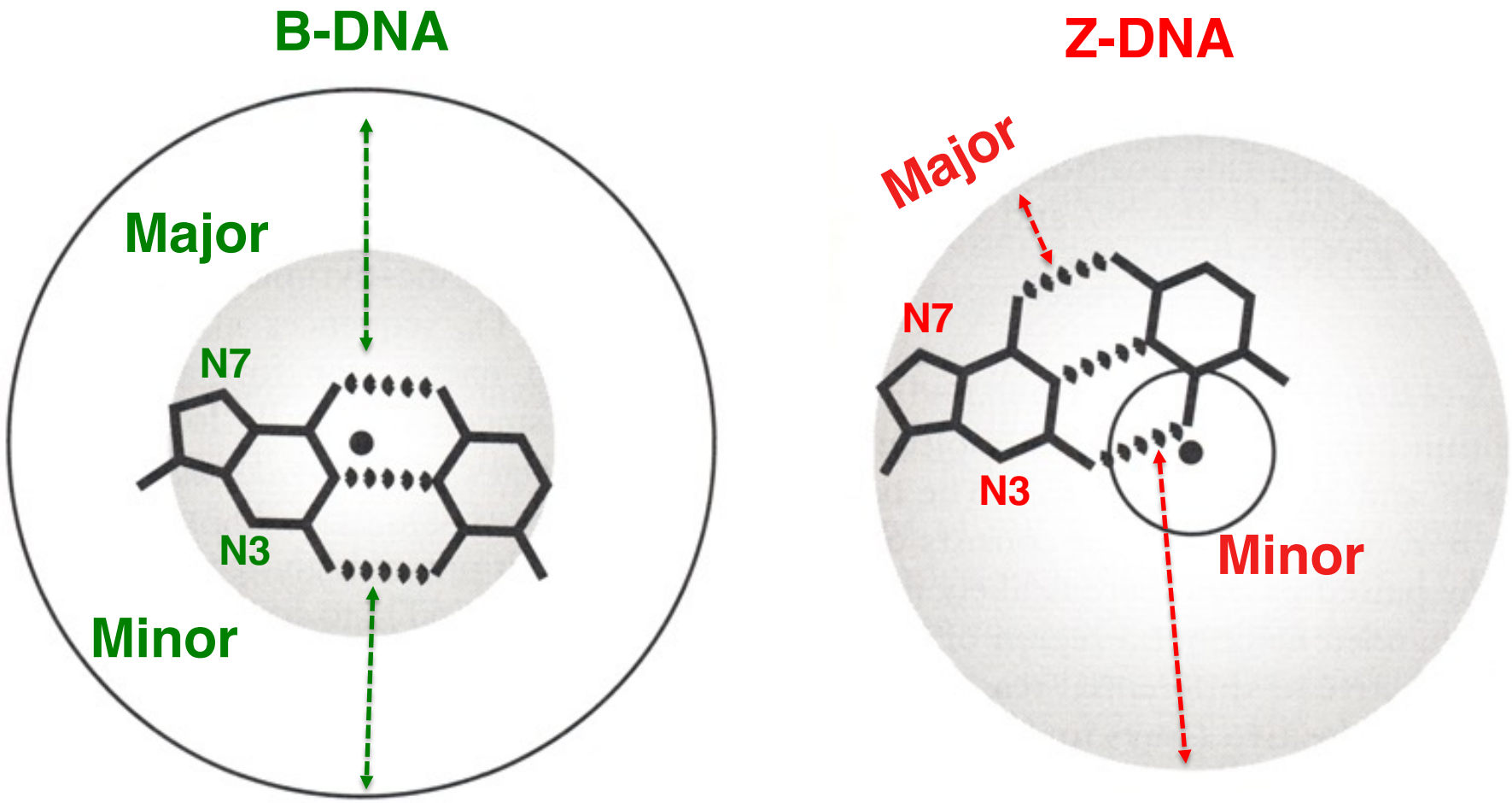
B: Syn, Syn, Anti, Anti

C: Syn, Anti, Syn, Anti

D: Anti, Anti, Syn, Syn



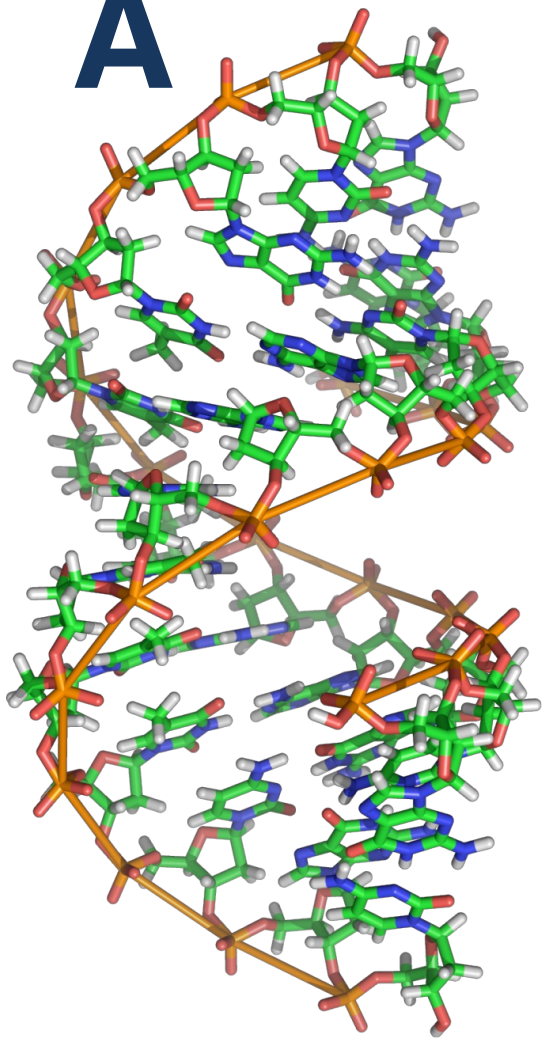
The major groove is very shallow in Z-DNA:

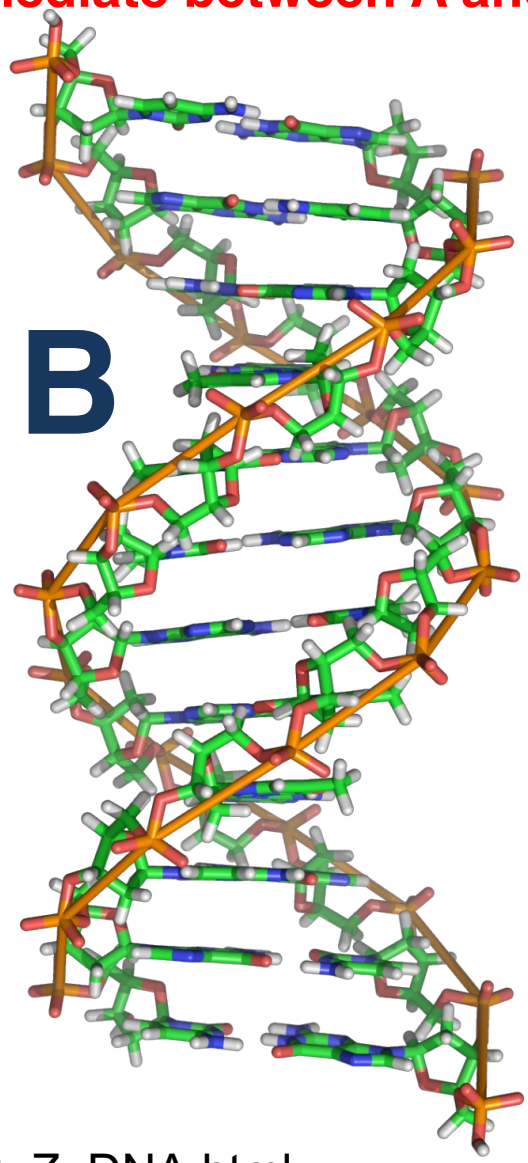


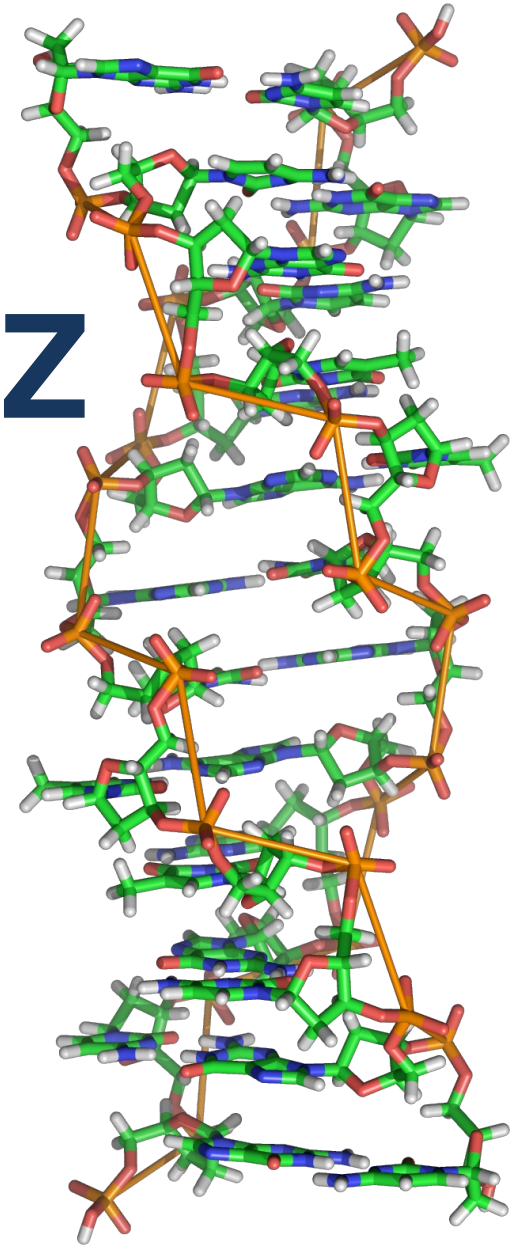
PyMol: ZDNA.pse

Summary of ABZ structures/ conformations

!! There are Helical Conformations other than A/B/Z
e.g.: conformations intermediate between A and B





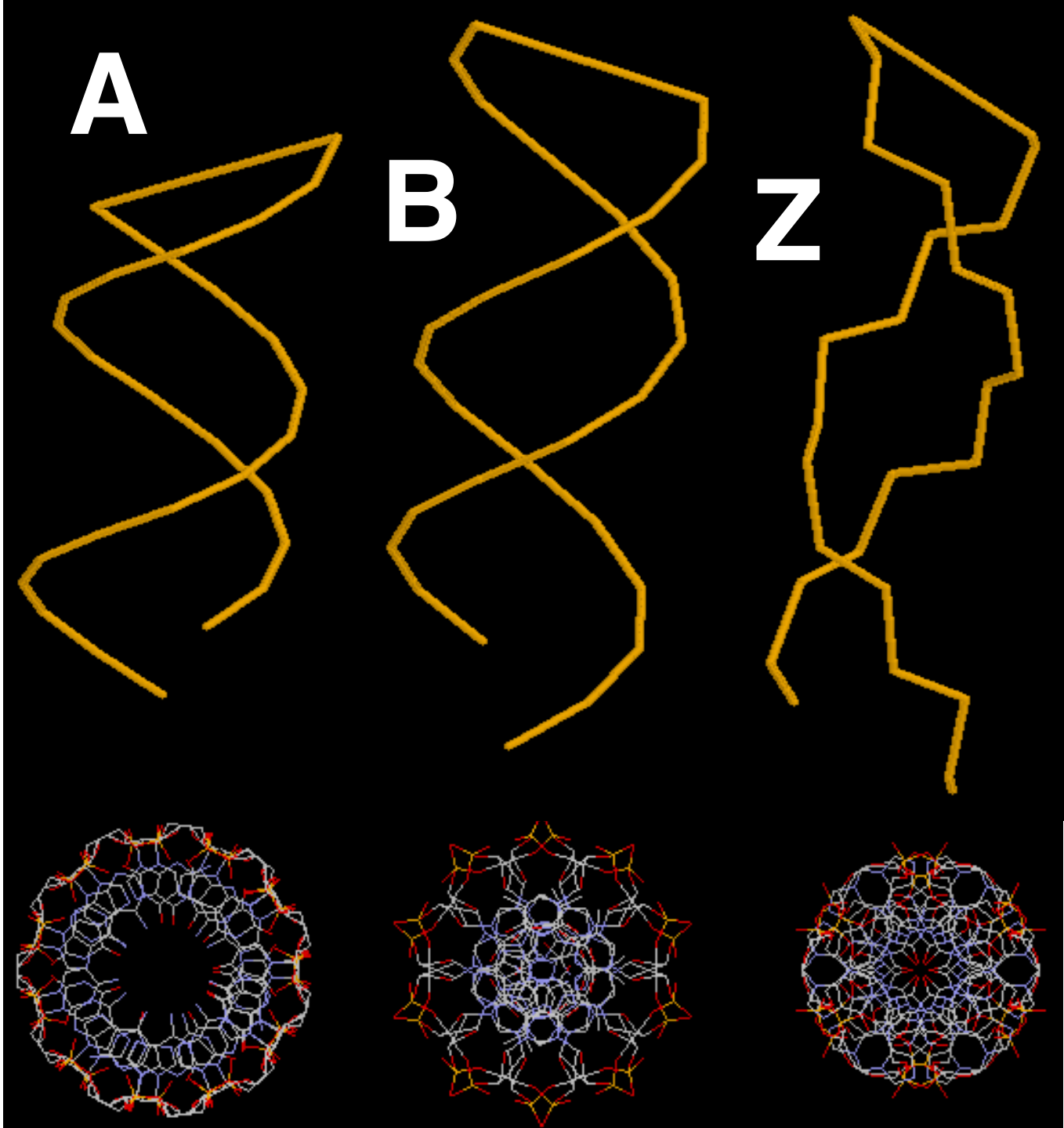


www.mun.ca/biology/scarr/A_B_Z_DNA.html

Summary of ABZ structures/ conformations

Backbone Profiles

Helical Projections



What natural nucleic acid form does this cartoon represent?

A: A-form

B: B-form

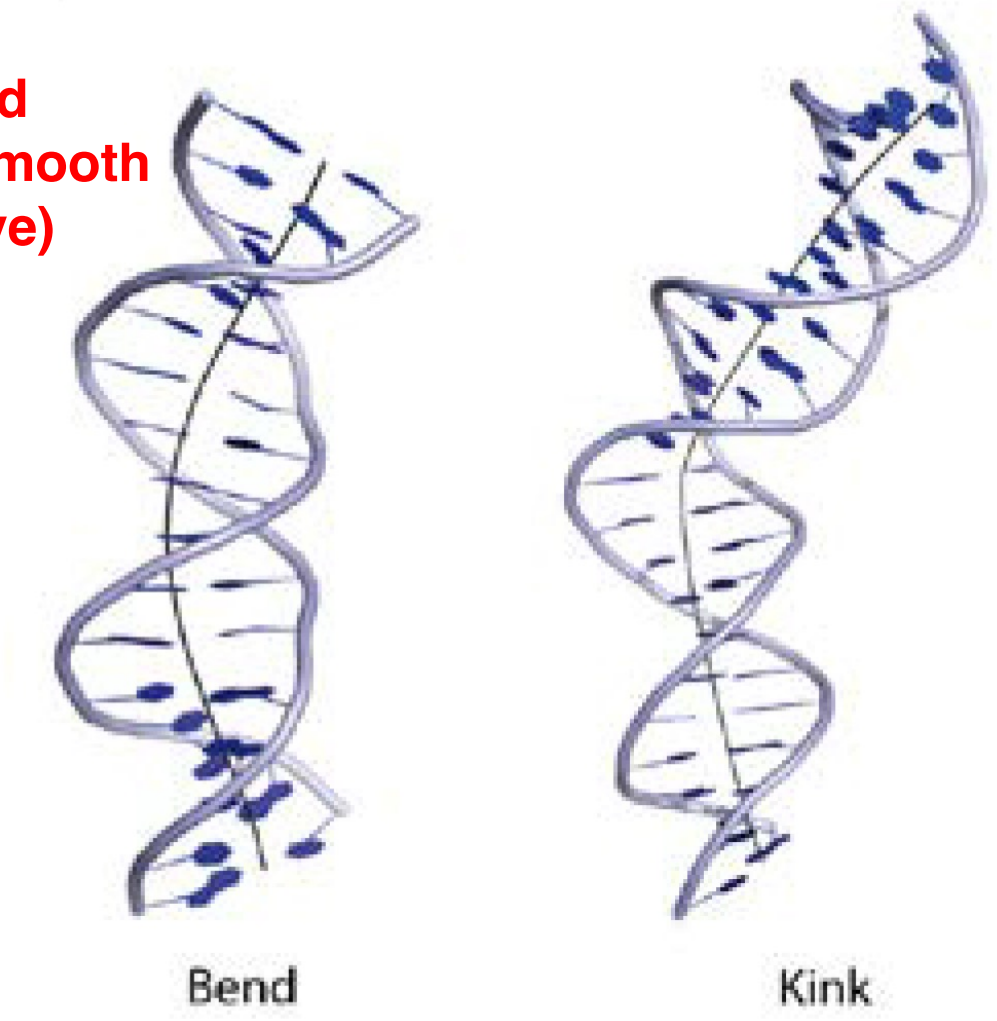
C: Z-form

D: None of the above

California Science Center, Expo Park

Double-stranded helices can be distorted and form bent/kinked structures:

Bend (= smooth curve)



Harteis & Schneider
Int. J. Mol. Sci. 2014

Kink = sharp change in helical axis

This can be due to:

- **Intrinsic DNA sequences/ features that increase the roll parameter, resulting in bend or kinks**
- **Binding of a protein that forces/favors a bent/kinked conformation**

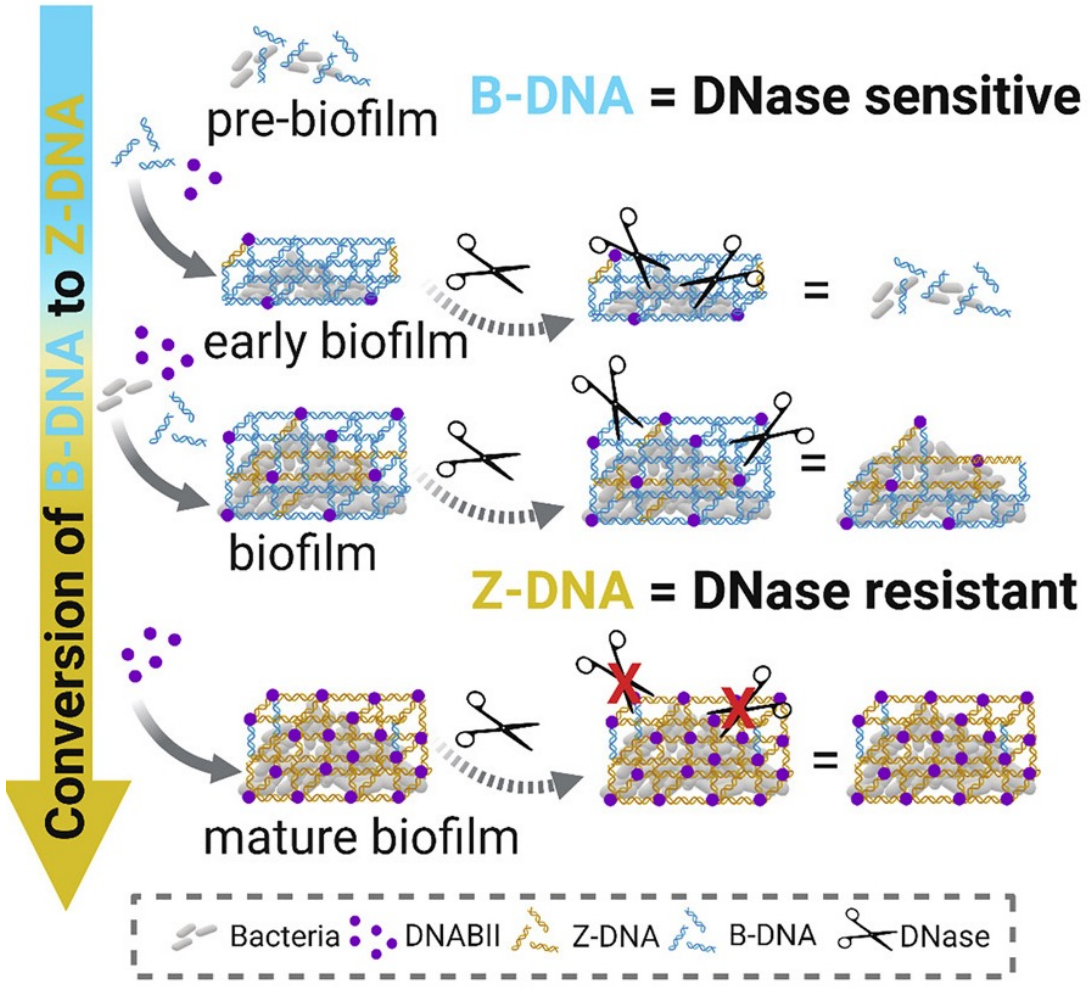
Why study nucleic acid structure?

- Structure and Sequence-specific Recognition by DNA binding proteins // see next

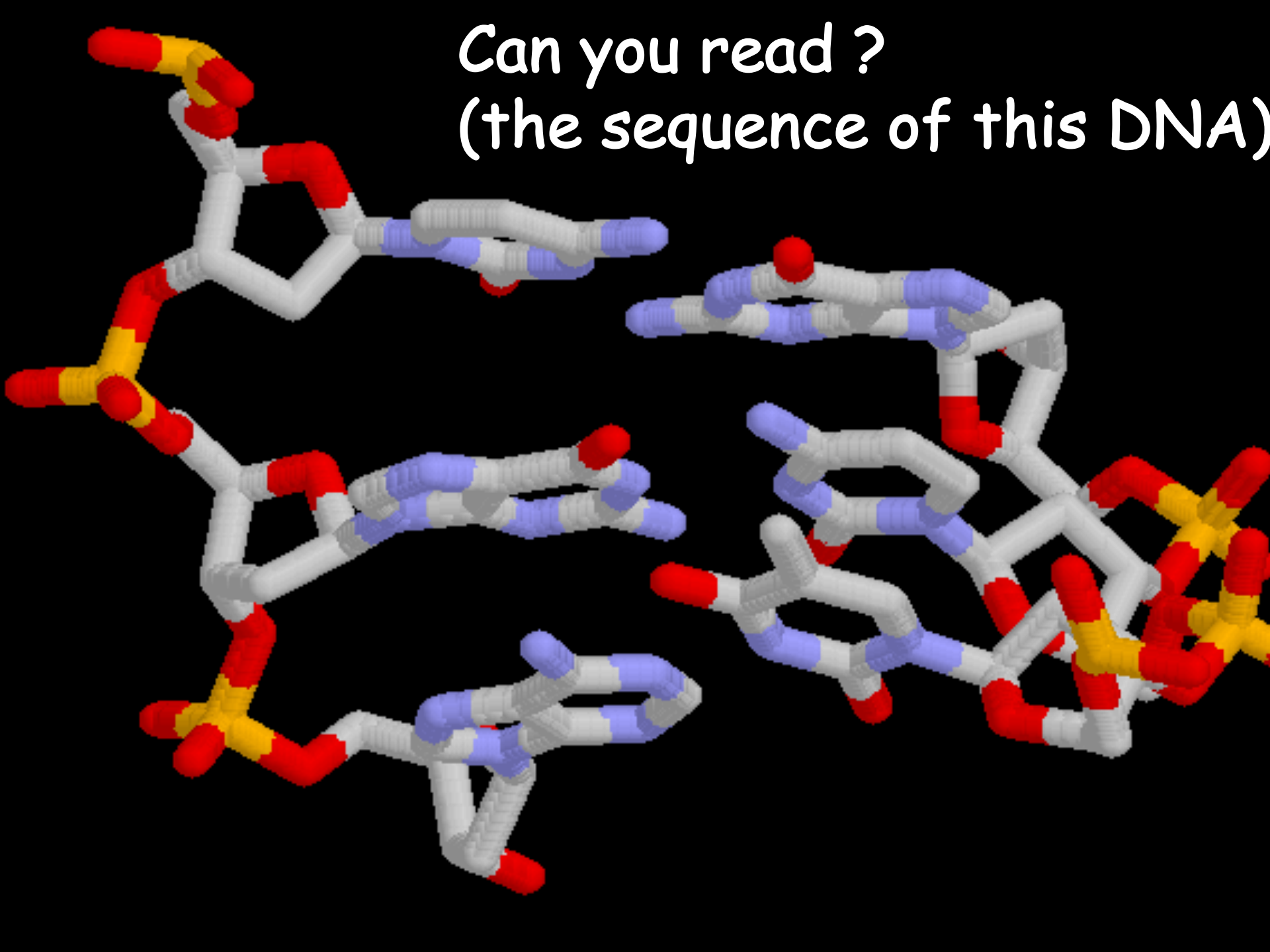
- Some non B-DNA structures are biologically relevant:
  - dehydrated living forms
  - dsRNA is A form (see PDB: 2KYD)
  - DNA/RNA duplex (replication, transcription) is A form

- Z-DNA might be associated with sequences that regulate DNA transcription

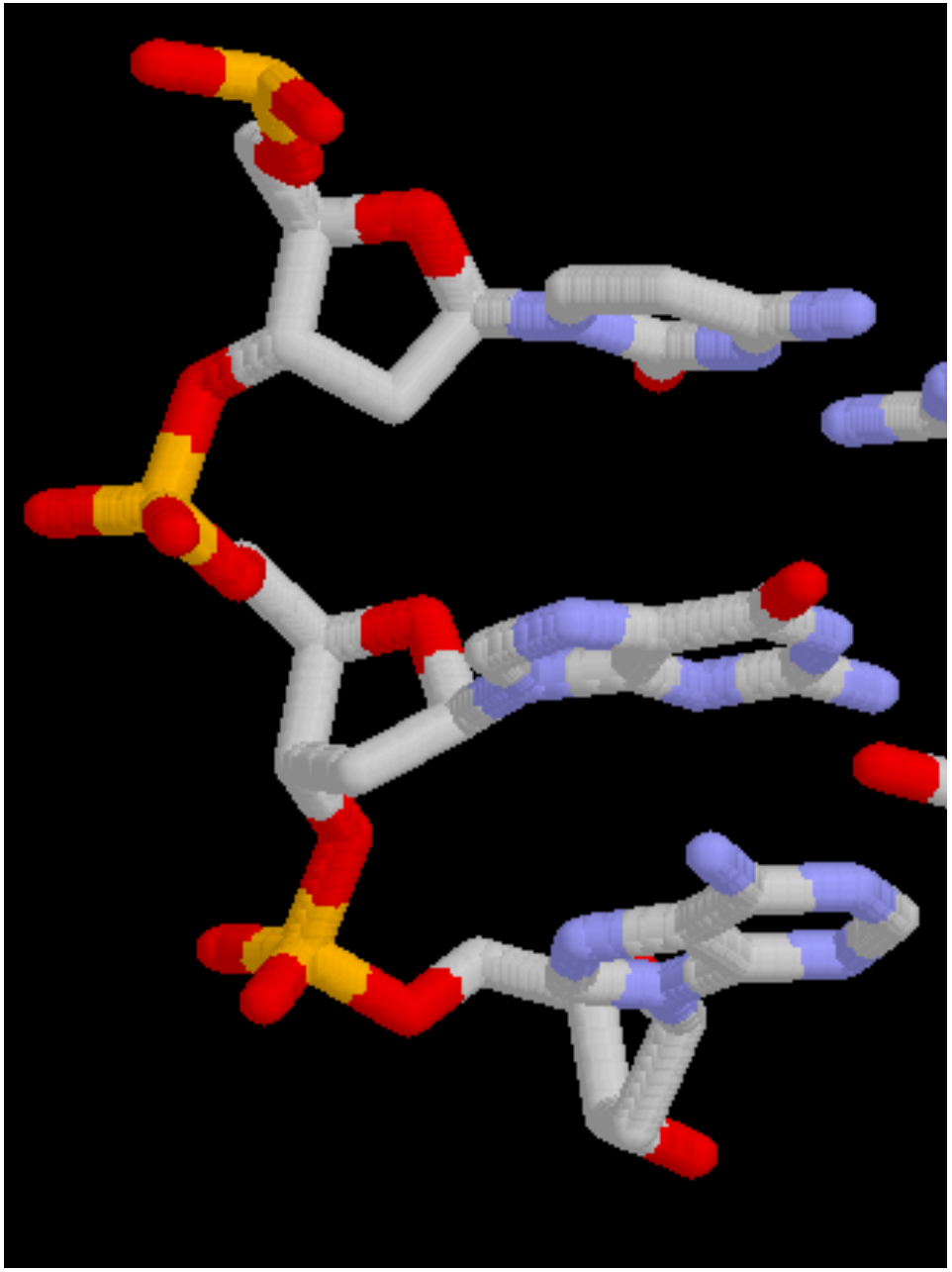
- Z-DNA is found in bacterial biofilms (Buzzo et al. Cell 2021)



Can you read ?
(the sequence of this DNA)



Can you read the sequence of this DNA Strand?



Top to bottom:

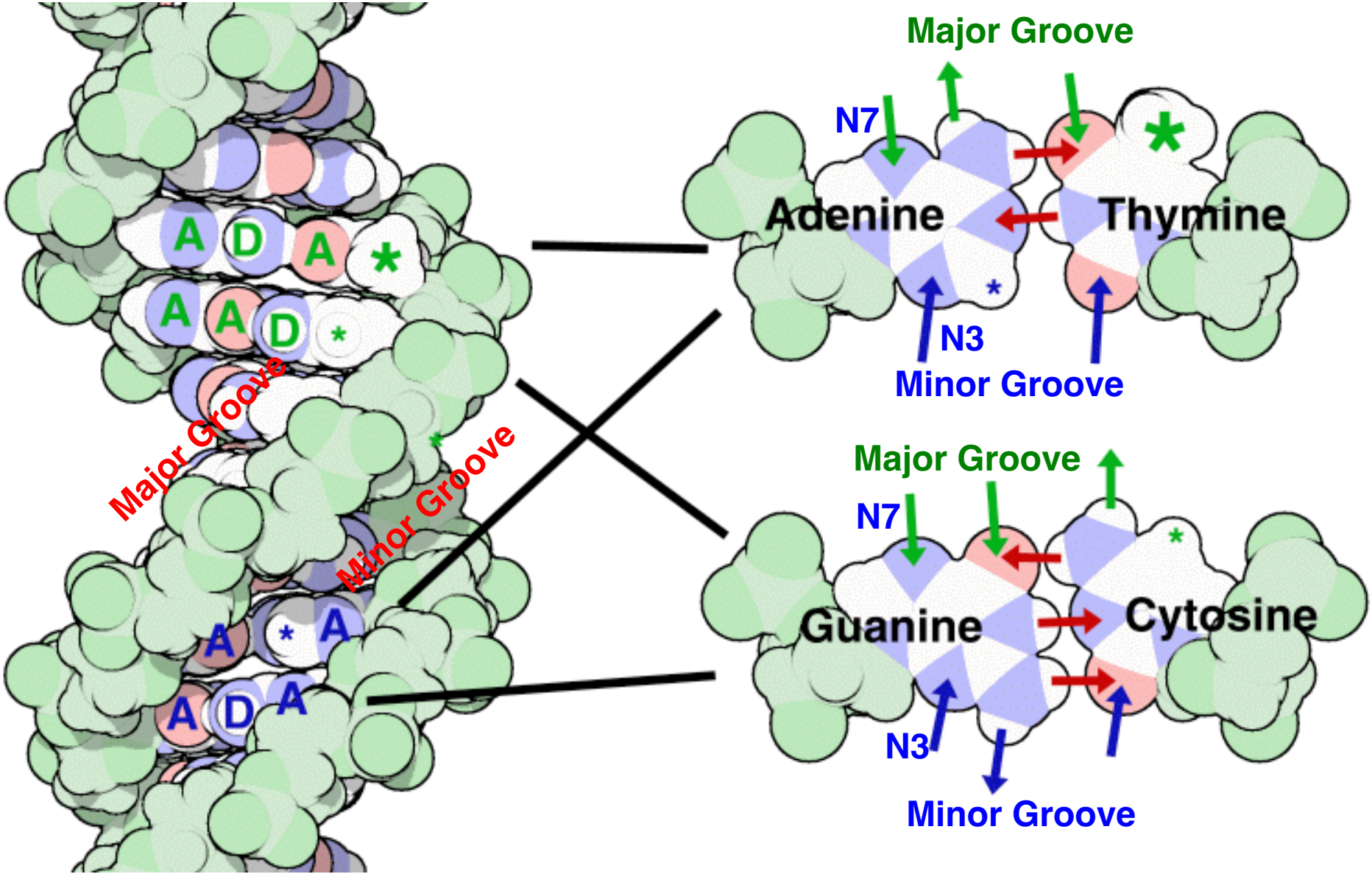
A: 5´TGA3´

B: 3´CGA5´

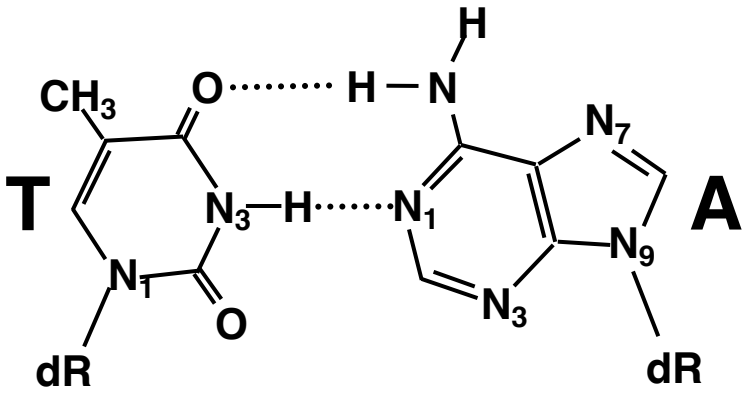
C: 5´CGA3´

D: 5´CAG3´

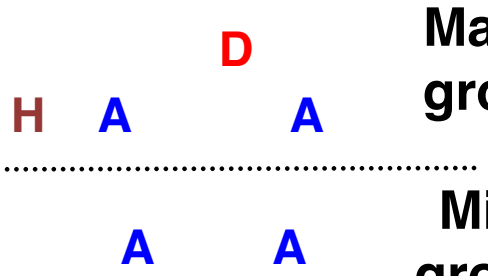
Sequence-Specific Recognition of DNA by proteins: What do proteins "see" ?

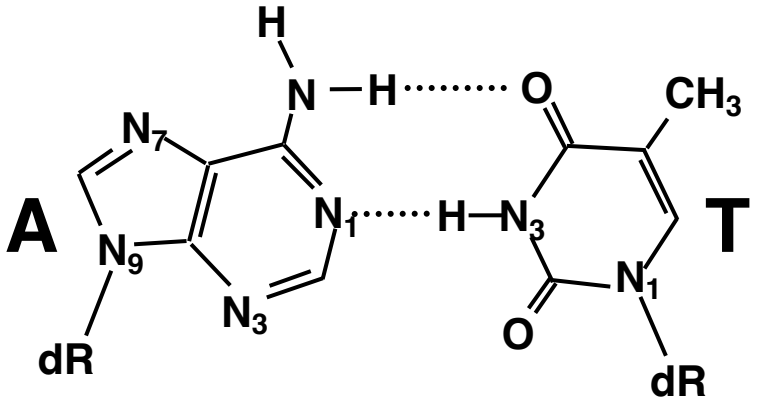


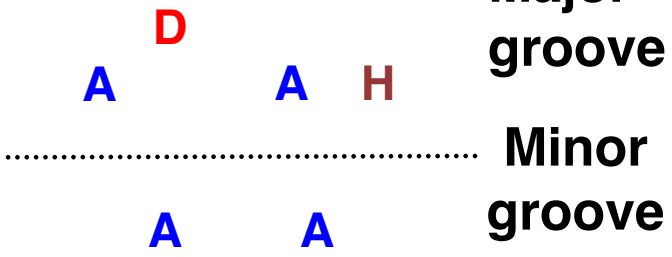
Recognition of Specific sequences by DNA-binding proteins



Distribution of H-bonds Donors (D) Acceptors (A) and Hydrophobic groups (H) available for recognition



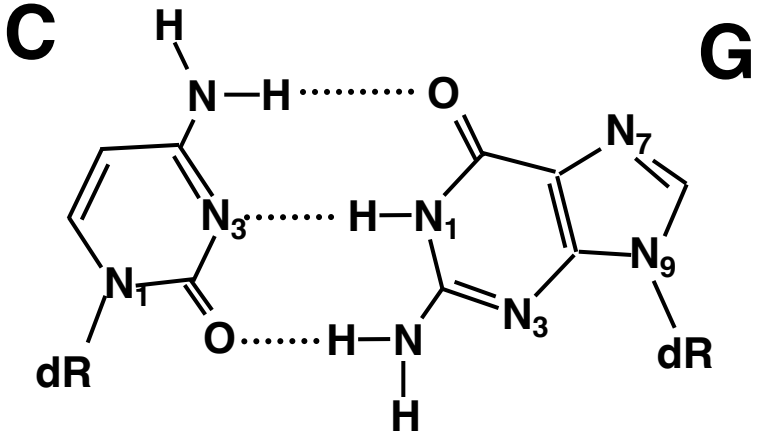




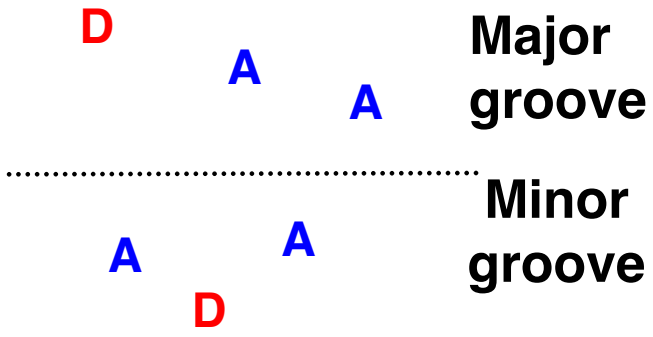
Conclusion: *Sequence-Specific* DNA binding proteins can differentiate A-T base pairs from T-A base pairs if they bind from the major groove side, but not from the minor groove side

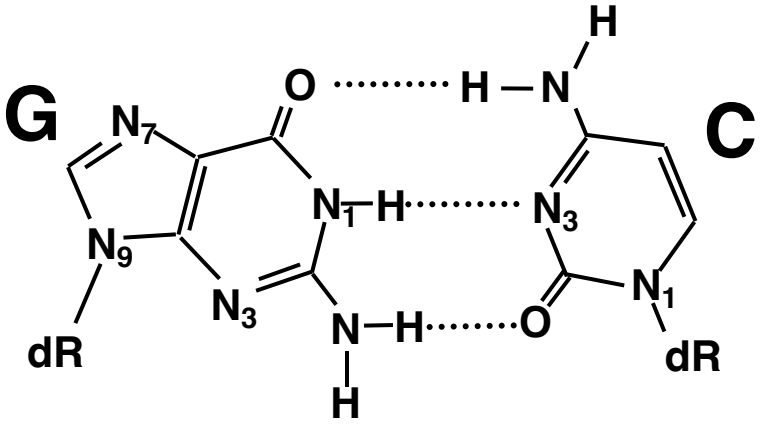
Concept is important to remember to understand transcription regulation

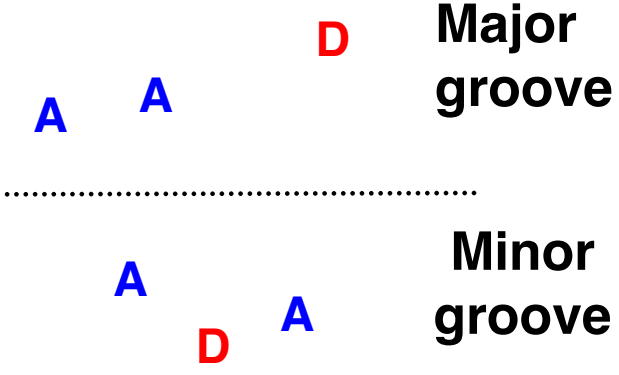
Recognition of Specific sequences by DNA-binding proteins



Patterns of H-bonds Donors (D), Acceptors (A), and Hydrophobic groups (H) available for recognition

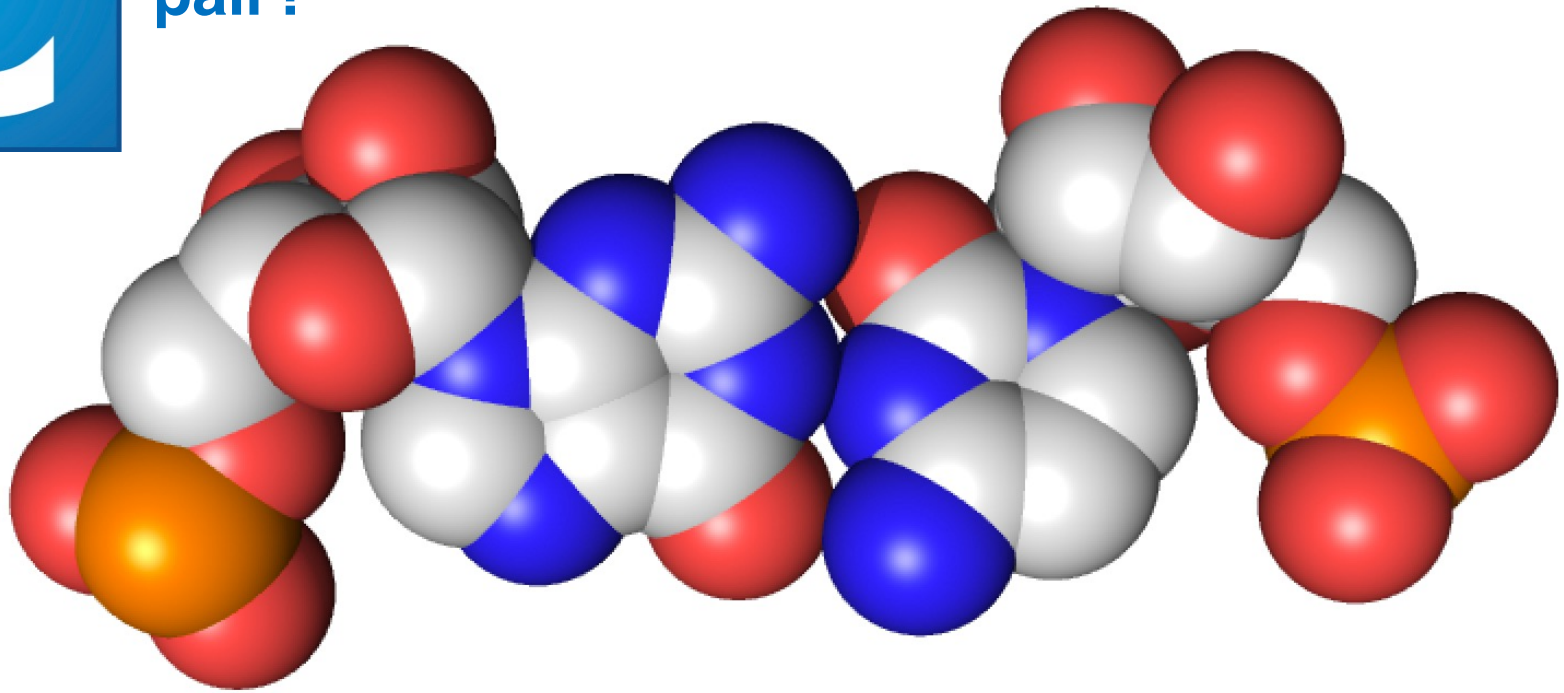






Conclusion: *Sequence-Specific* DNA binding proteins can differentiate G-C base pairs from C-G base pairs if they bind from the major groove side, but not from the minor groove side

What is the H-bond donor (D)/Acceptor (A) pattern in the MAJOR GROOVE for this base pair?



A: A D A D

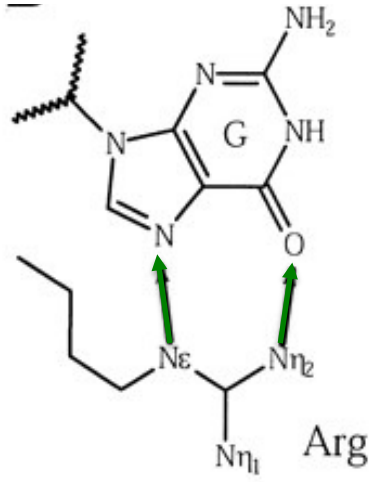
B: A A D

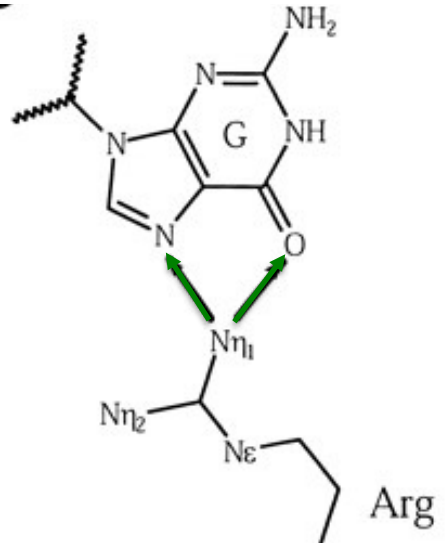
C: D A A

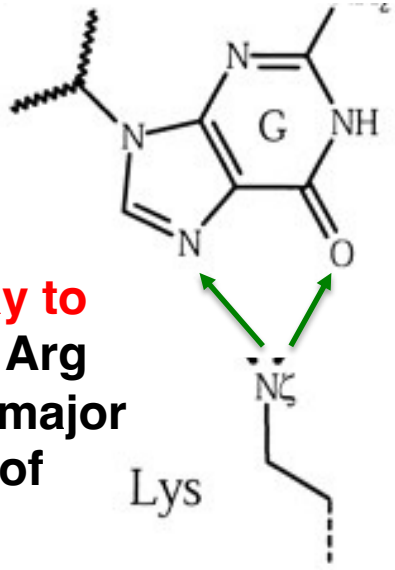
D: D A A D

Examples of interactions of amino acids with the major groove side of bases or base pairs

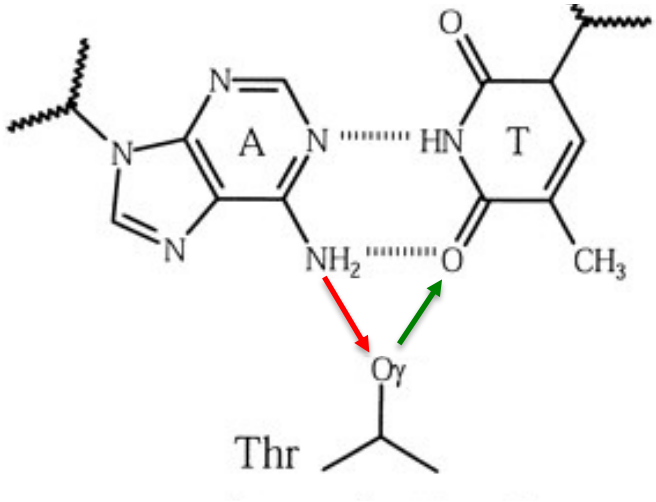
• **These will contribute to sequence recognition and generally be found in proteins that bind DNA in a sequence specific manner**







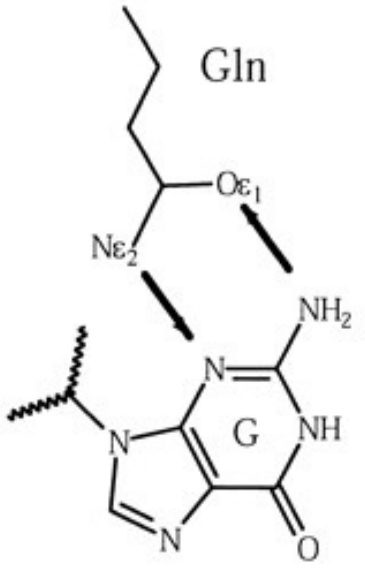
• **There is more than one way to "skin the cat": for instance, Arg can recognize a G from the major groove side using a variety of geometries (see top interactions)**

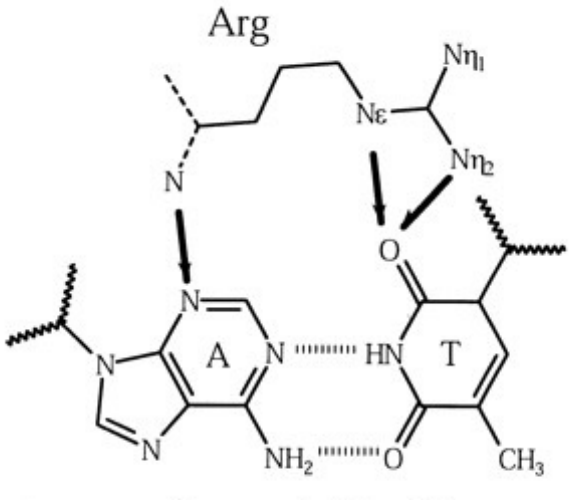


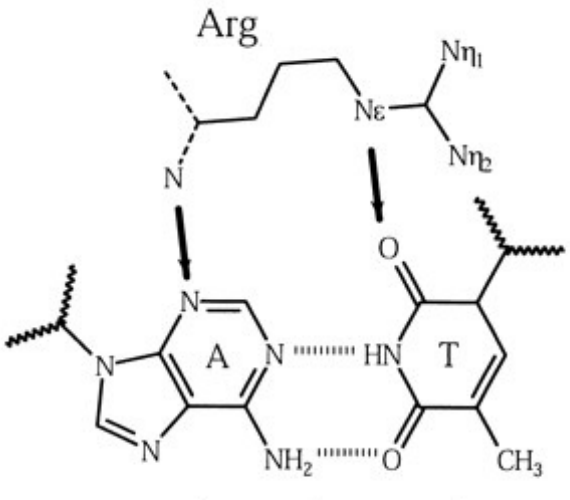
Examples of interactions of amino acids with the minor groove side of bases or base pairs

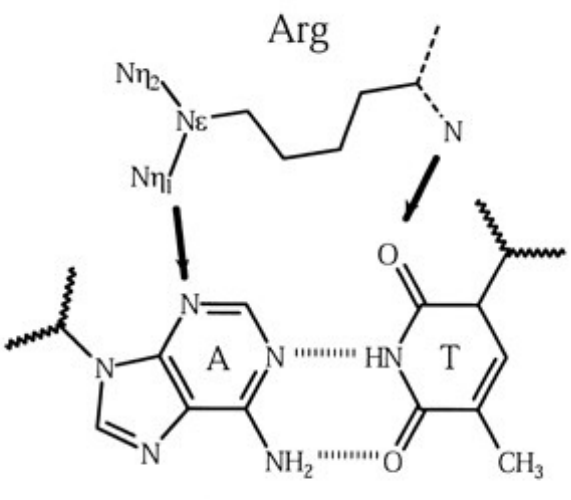
• In general, these interactions do NOT contribute to sequence recognition

• However these interactions can be made by both sequence-specific and non-sequence specific DNA binding proteins -- any sequence-specific DNA binding proteins also has some level of non sequence-specific binding sites that contribute to general DNA binding). These interactions will just not contribute to sequence specificity



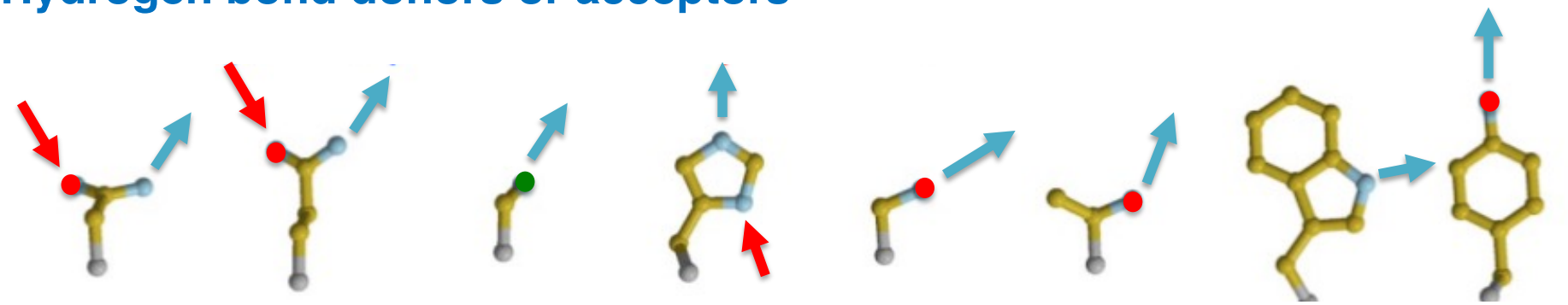






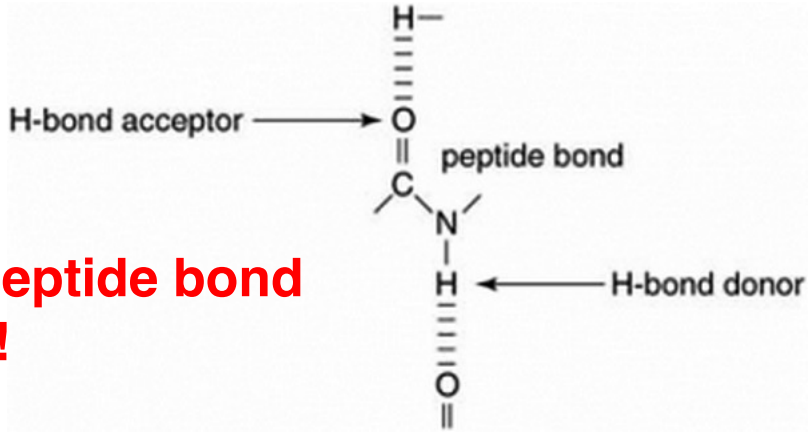
PyMol: sequence_specific_recognition.pse

Amino acids side groups and atoms color coded by chemical make up: Hydrogen bond donors or acceptors



Asn Gln Cys His Ser Thr Trp Tyr

Also: Arg, Lys, Asp, Glu – see 2 slides from now



Also: peptide bond atoms!

- **Will establish hydrogen bonds with bases donors or acceptors**

- **depending on the side of the base (major or minor groove), these will contribute to sequence specific recognition… or not.**

For DNA binding proteins:

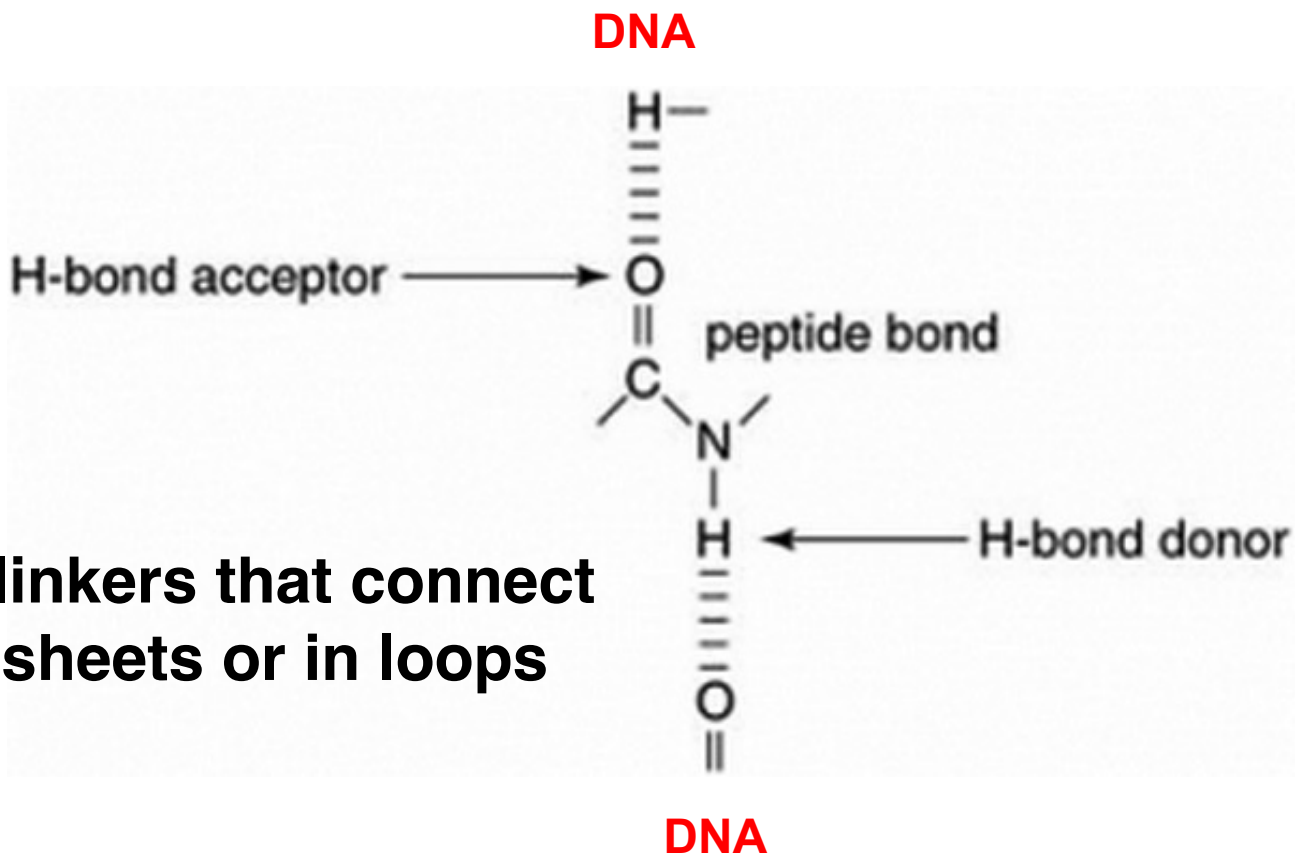
In what domains of proteins are the main chain (or peptide bond) N-H and C=O groups more likely to be involved in hydrogen bonding with the bound DNA?

A: α-helices

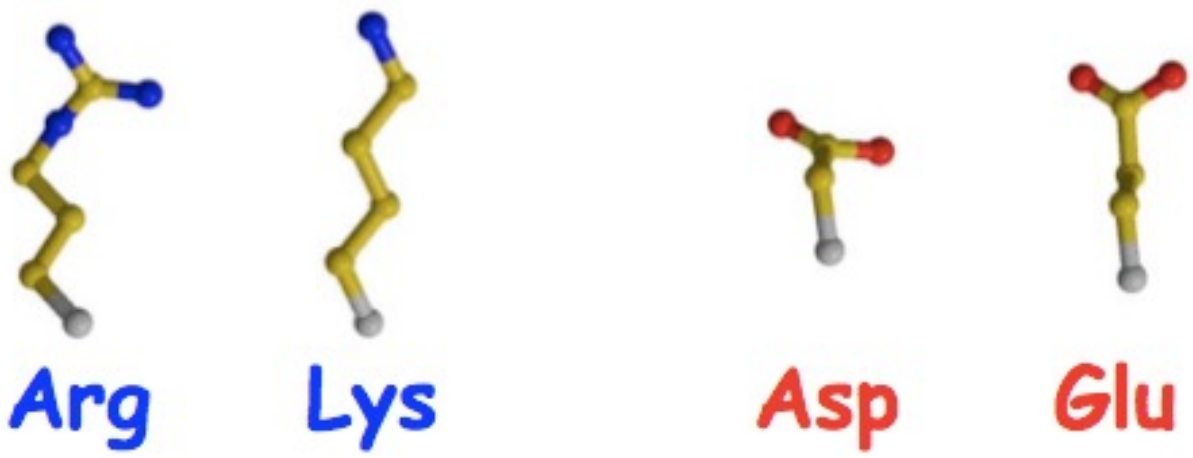
B: β–sheets

C: In unstructured linkers that connect α-helices and/or β-sheets or in loops

D: All the above



Amino acids side groups and atoms color coded by chemical make up: positively and negatively charged side chains

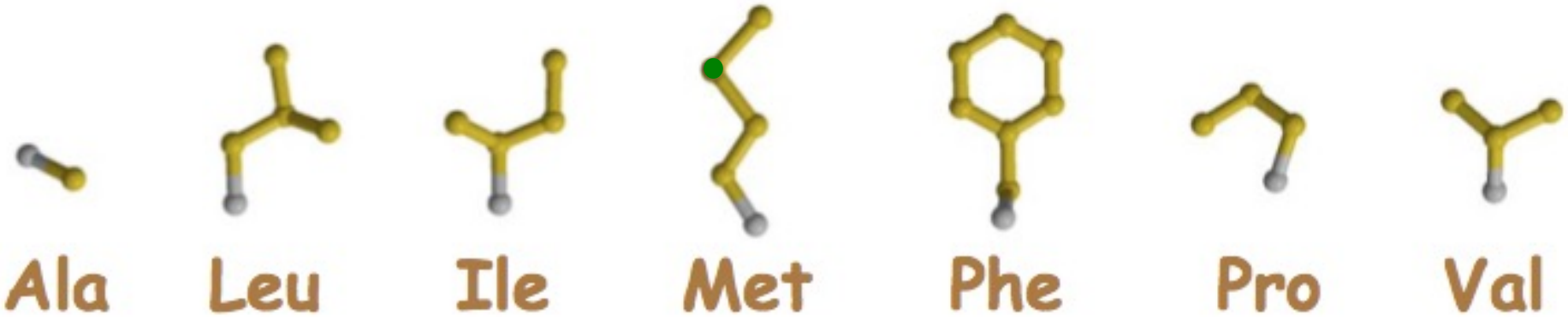


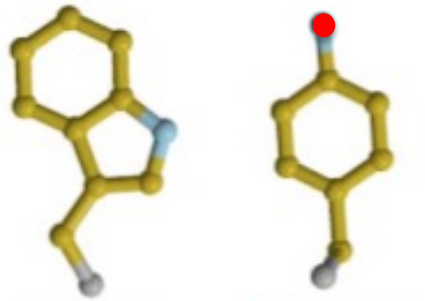
Positively Charged:
Will interact electrostatically with negatively charged phosphate oxygens

Negatively Charged:
May interact indirectly with the backbone through binding of positively charged ions (K^+, Mg^{++})

Note that these side chains can also serve as H-bonds donors/acceptors

Amino acids side groups and atoms color coded by chemical make up: Hydrophobic side chains



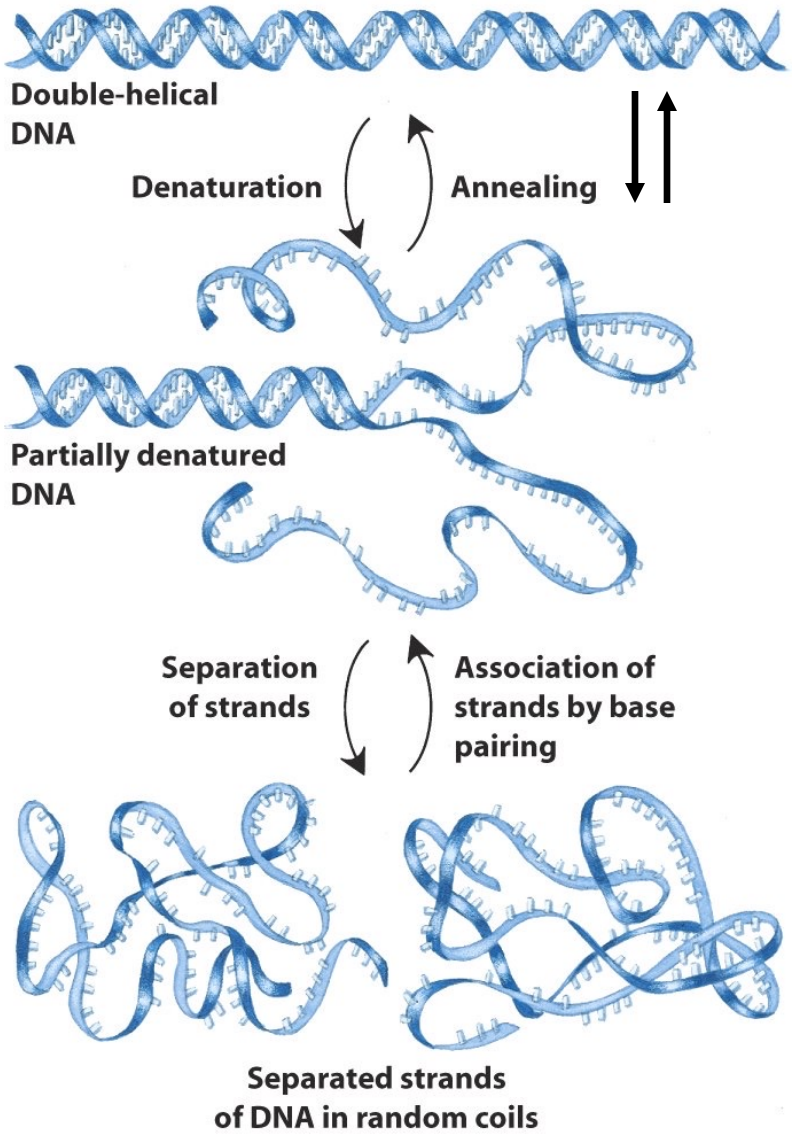


These will make aromatic π stacking interactions with bases

in general, these interactions will not contribute to sequence-specific recognition

Physical forces governing nucleic acid structure

DS DNA (Helix)



2 SS DNAs ("random coils")

What influences the equilibrium ?

(important because DNA is "opened" during replication and transcription)

In favor of double-stranded DNA

- Base stacking interactions (major)
  - Pi stacking / dipole-dipole coupling
  - Release of otherwise ordered water molecules
- Hydrogen bonds between strands (minor)

In favor of single-stranded DNA

- Electrostatic repulsion between strands

- Entropic considerations:
  - Increased entropy for ssDNA vs dsDNA (increased conformational and translational components of entropic considerations)

Why is the contribution of H-bonds to DNA duplex stability considered minor?

A: Because the ΔG of H-bonds is negligible compared to that of hydrophobic interactions

B: Because scientists are not able to calculate their contribution precisely so they neglect their contributions

C: Because H-bonds are much weaker than covalent bonds

D: because H-bonds between 2 bases are replaced by H-bonds between bases and water during denaturation so the net energy change is close to zero

Gibbs free energy

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

Entropy

$$S = -k_B \sum_{n=1}^{W} P_n \ln P_n$$

k_B: Boltzmann constant
W: number of possible configurations
P_n: probability of state n

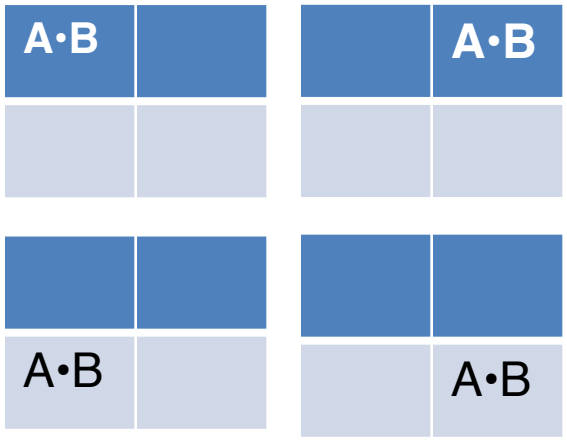
• Conformational Entropy

dsDNA:
Relatively rigid with limited flexibility
⇒ low conformational entropy

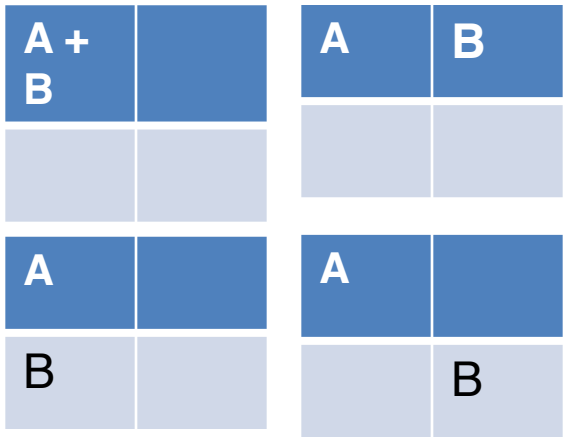
ssDNA:
More flexible with rotational movement of the strand
⇒ higher conformational entropy

• Translational Entropy

Linked to the number of possible states of the thermodynamical system



dsDNA:
strands A and B are linked
= 4 possible states

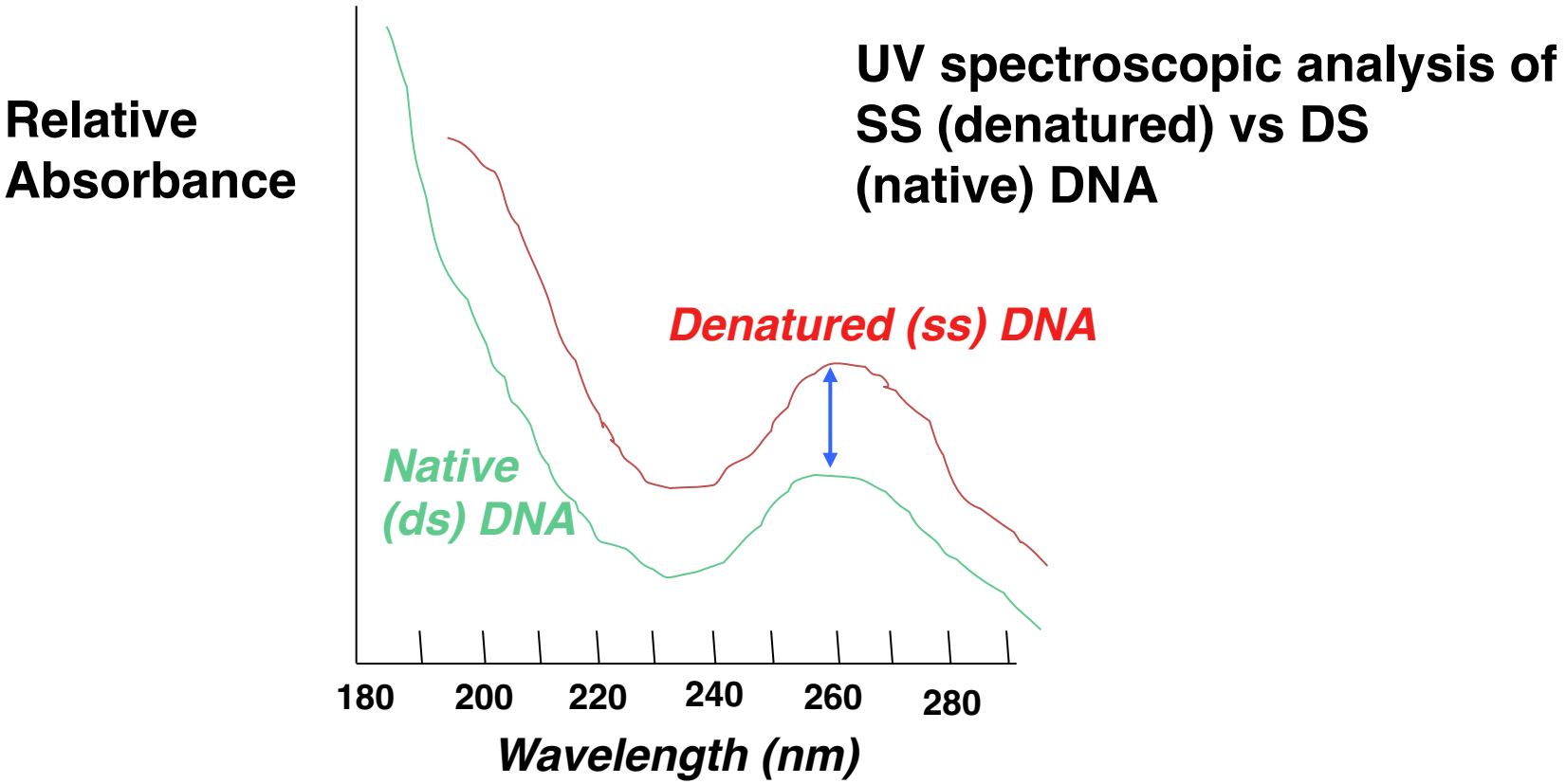


ssDNA:
strands A and B are free from each other
= 16 possible States

(repeat 3x with A in other positions)

Higher Translational Entropy

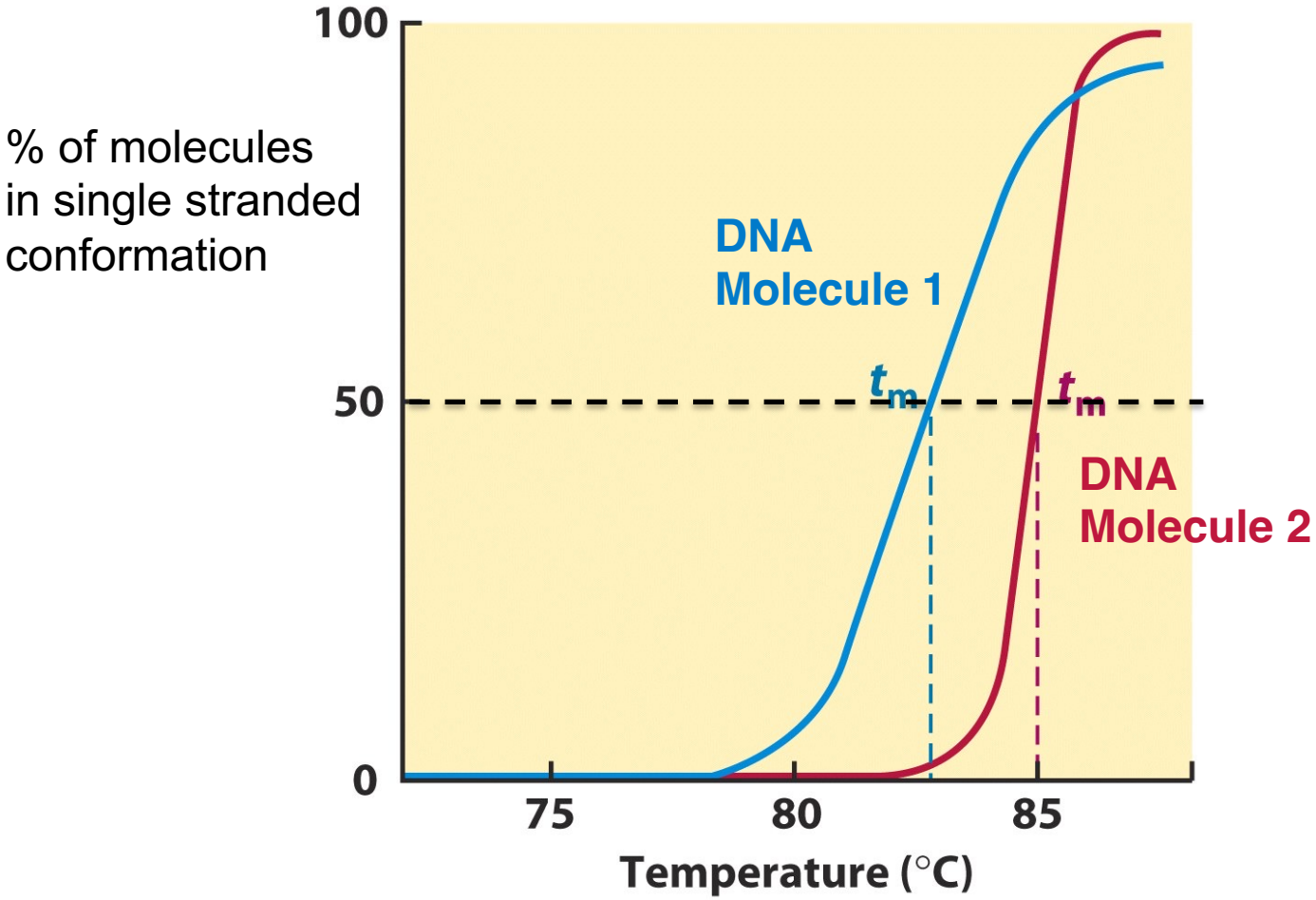
DNA denaturation can be measured using UV absorbance measurements



Hyperchromic Effect: UV absorbance of SS DNA > that of native DNA

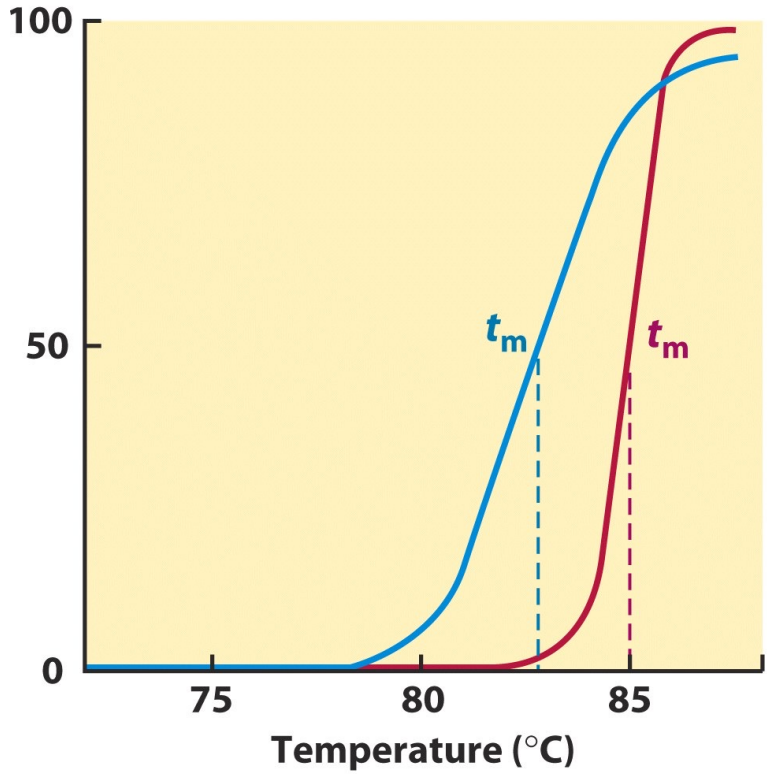
Why? In DNA found in a single stranded conformation, electrons in the bases orbital systems are more excitable by UV because bases are no longer stacked regularly -> these electrons are less involved in pi-stacking interactions

Measuring DNA denaturation as a function of temperature



"melting curves" for two different DNA molecules (red and blue) show different "melting points" = 2 different T_m values

What do these DNA melting curves remind you of?



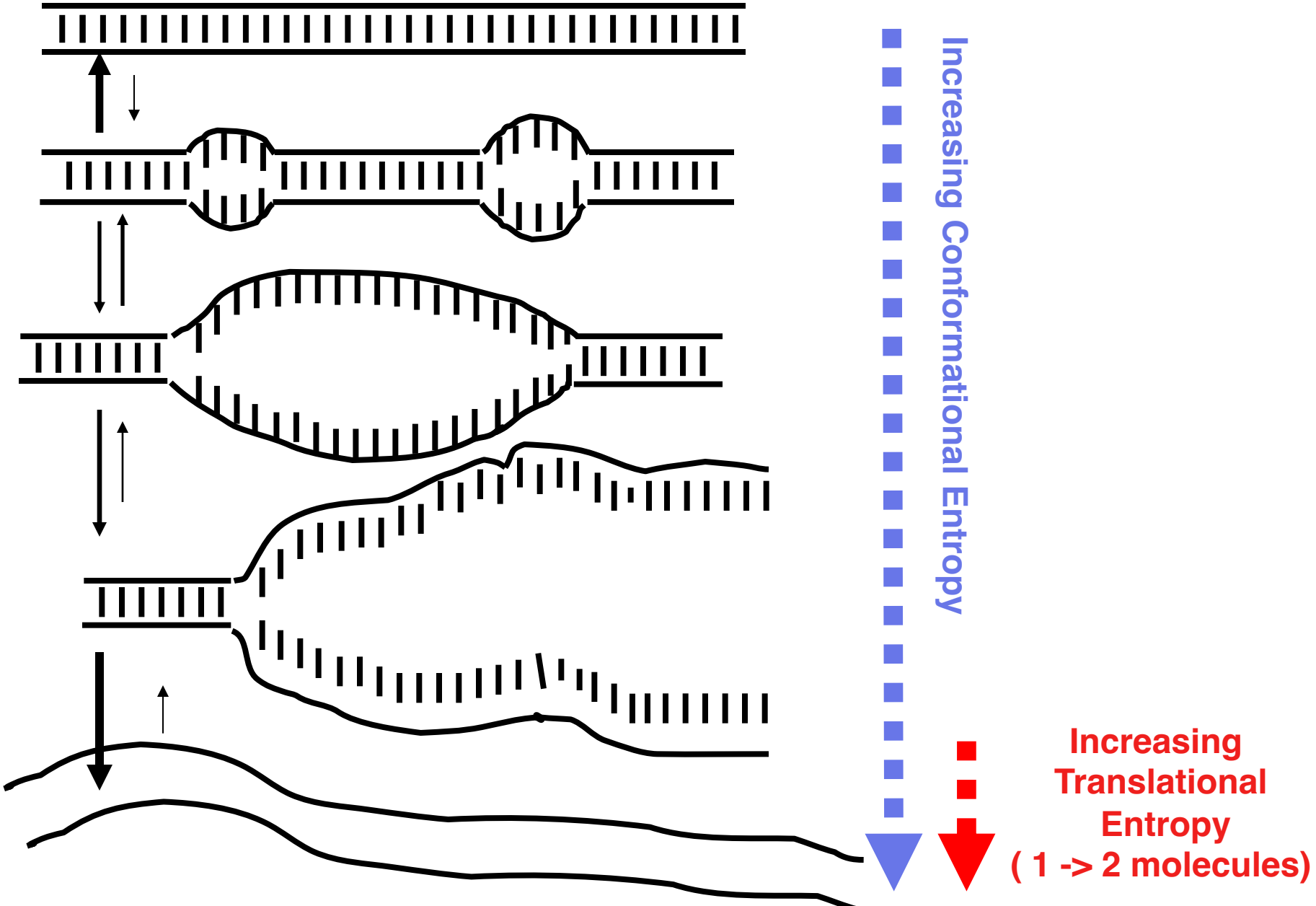
A: Poisson distribution

B: oxygen binding/dissociation by Hemoglobin = sigmoid curve

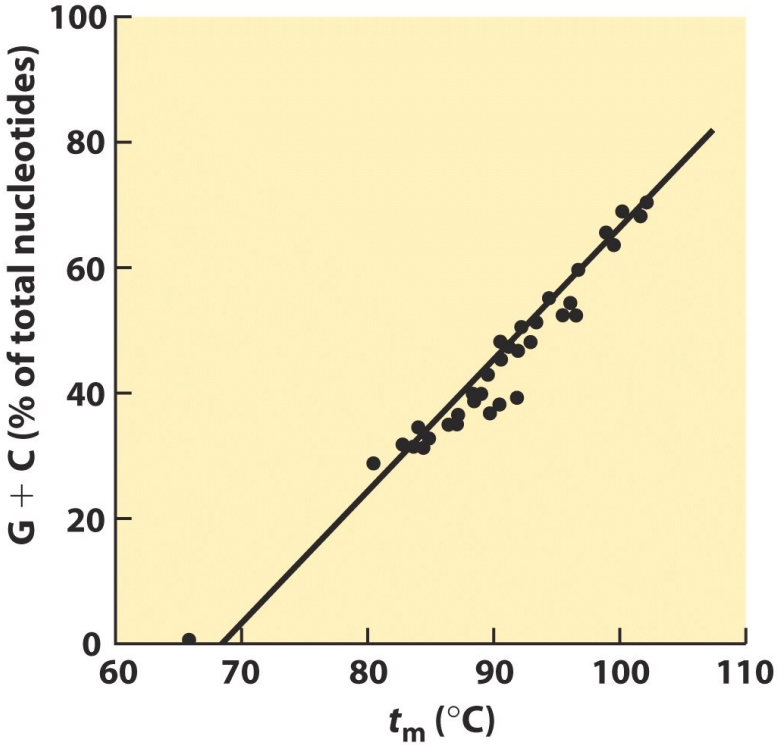
C: Case counts in the early days of a pandemic = exponential curve

D: Gaussian curve = expected grading for a "curved" class

DNA melting is a cooperative process:
this explains the sigmoid shape of DNA denaturation curves



The Tm of a DNA molecule is a linear function of its G-C content but NOT because of a higher number of H-bonds: 3(GC) vs 2 (AT)



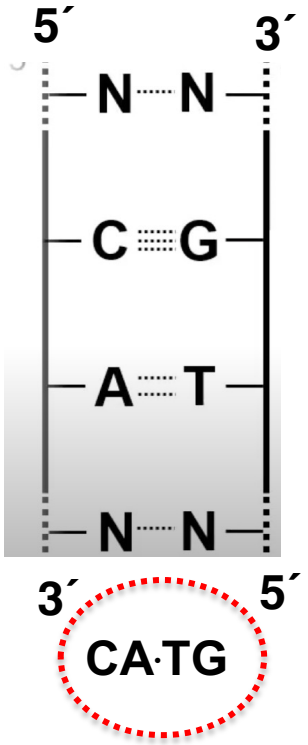
Effect of G-C content on stability is Due to higher stacking energy provided by G-C base pairs in general compared to AT base pairs – especially cumulative effect of consecutive G-C base pairs

Measuring the ΔG of various consecutive base pairs:

	ΔG of stacking (kcal/mol)
GC·GC	-2.17
AC·GT	-1.81
GG·CC	-1.44
GA·TC	-1.43
AT·AT	-1.34
AA·TT	-1.11
AG·CT	-1.06
CG·CG	-0.91
CA·TG	-0.55
TA·TA	-0.19

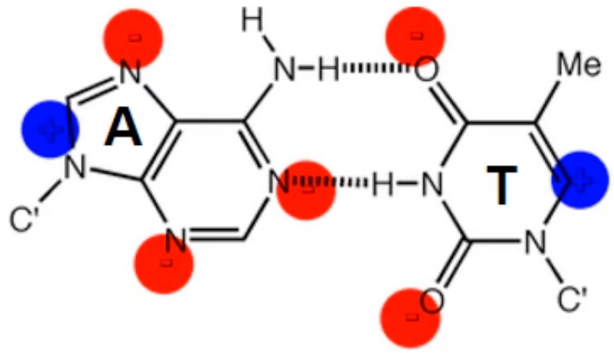
Protozanova et al. J.Mol.Biol. 2004

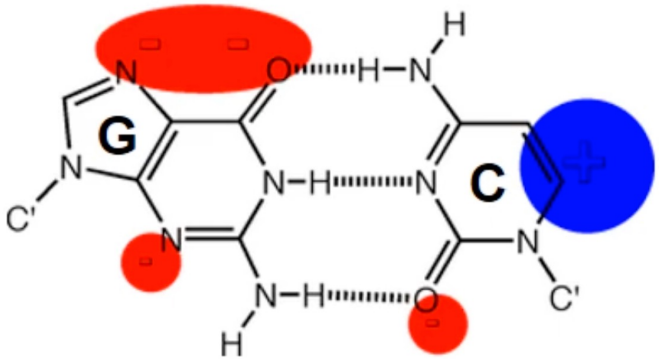
How to read this table: Example of a CA•TG Stack:



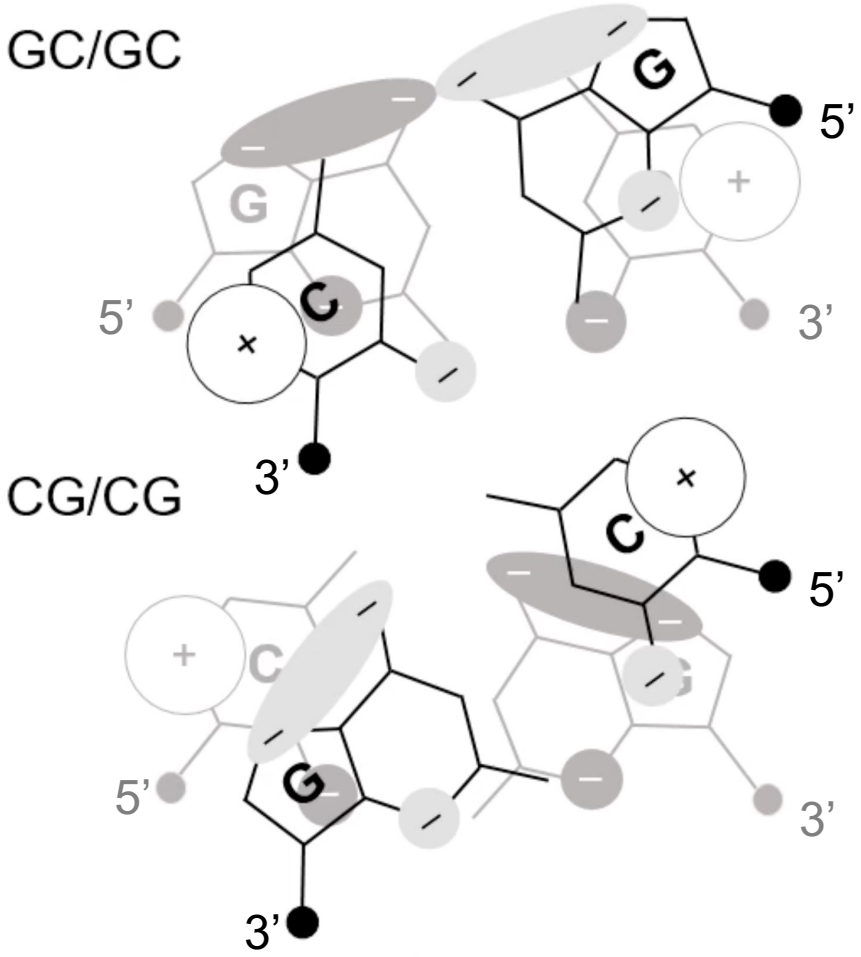
https://www.youtube.com/watch?v=PM3IlvFoLAg&ab_channel=DerekMcLachlin

The identity and orientation of stacked bases affects its favorability





CA Hunter (1996) *BioEssays* 18:157-62



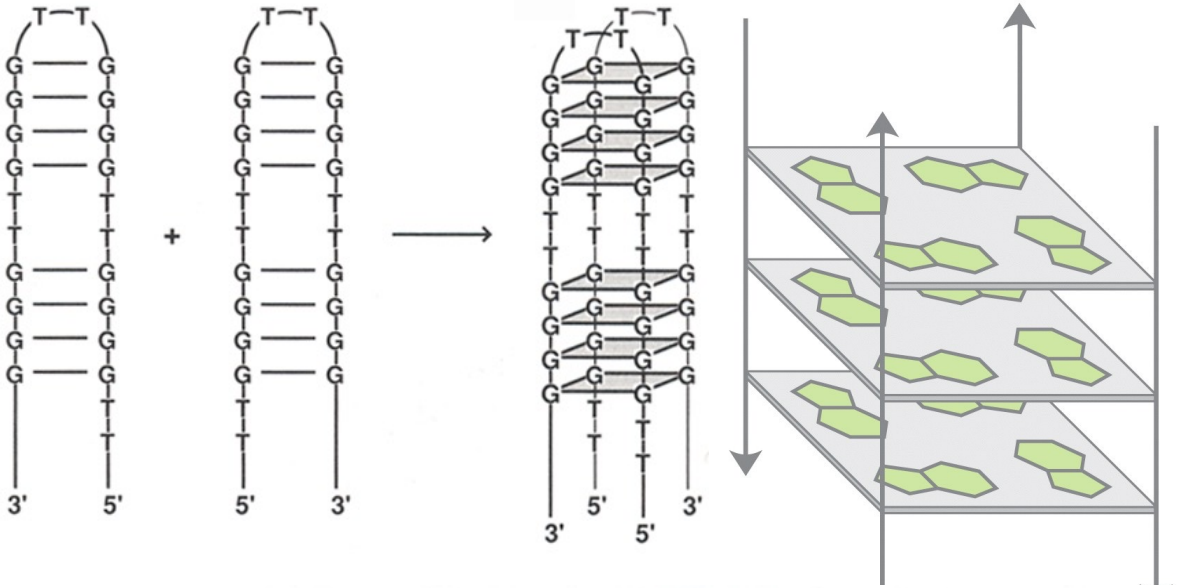
	ΔG of stacking (kcal/mol)
GC·GC	-2.17
CG·CG	-0.91

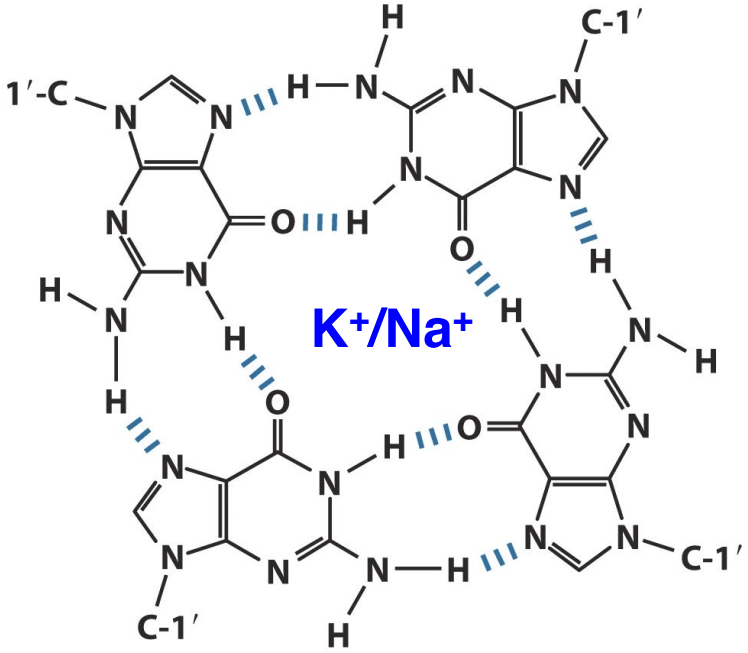
https://www.youtube.com/watch?v=PM3lIvFoLAg&ab_channel=DerekMcLachlin

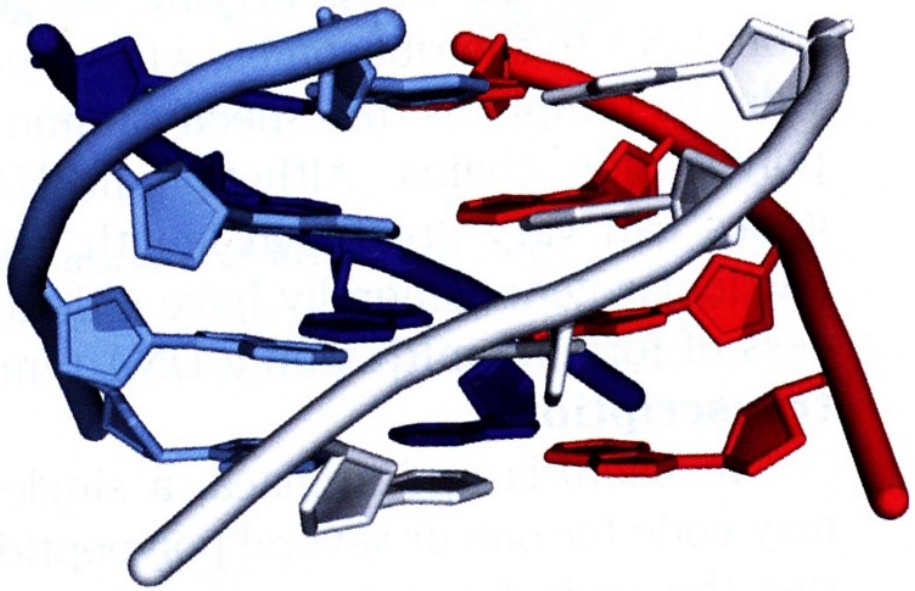
Example of intrinsic DNA Tertiary Structure

G-quadruplex structures in telomeric DNA: case of (T2G4) repeats in telomeric DNA

4 G-quartets stack onto each other







G-quartet = 4 G bases interact with each other

PDB ID:156D

G-quadruplexes protect the ends of your chromosomes!

PyMol: Quadruplex.pse

RNA structures

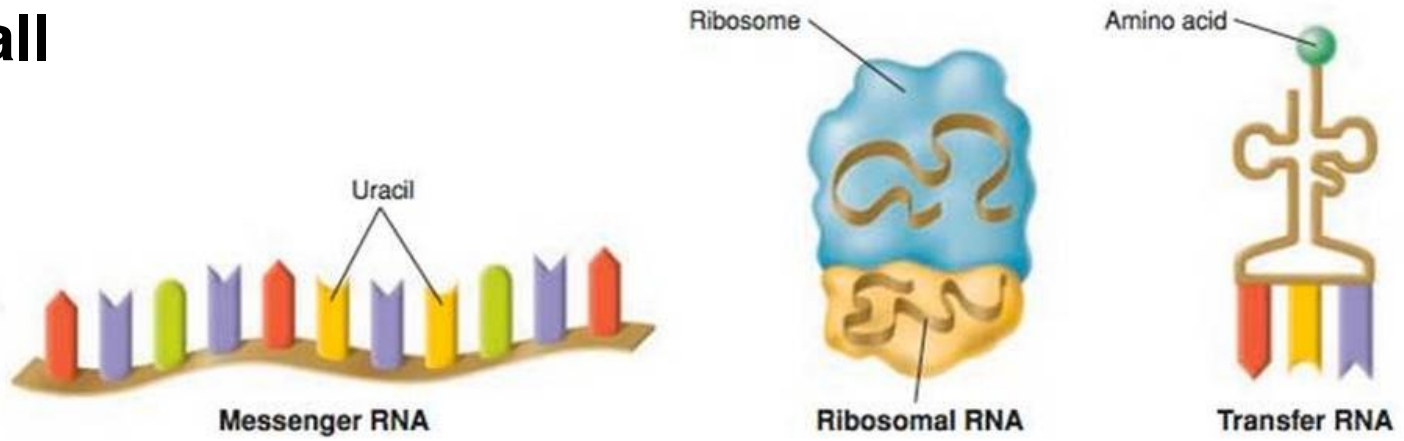
Is this a good representation of RNA structures?

A: Yes for mRNA and tRNA, not for rRNA

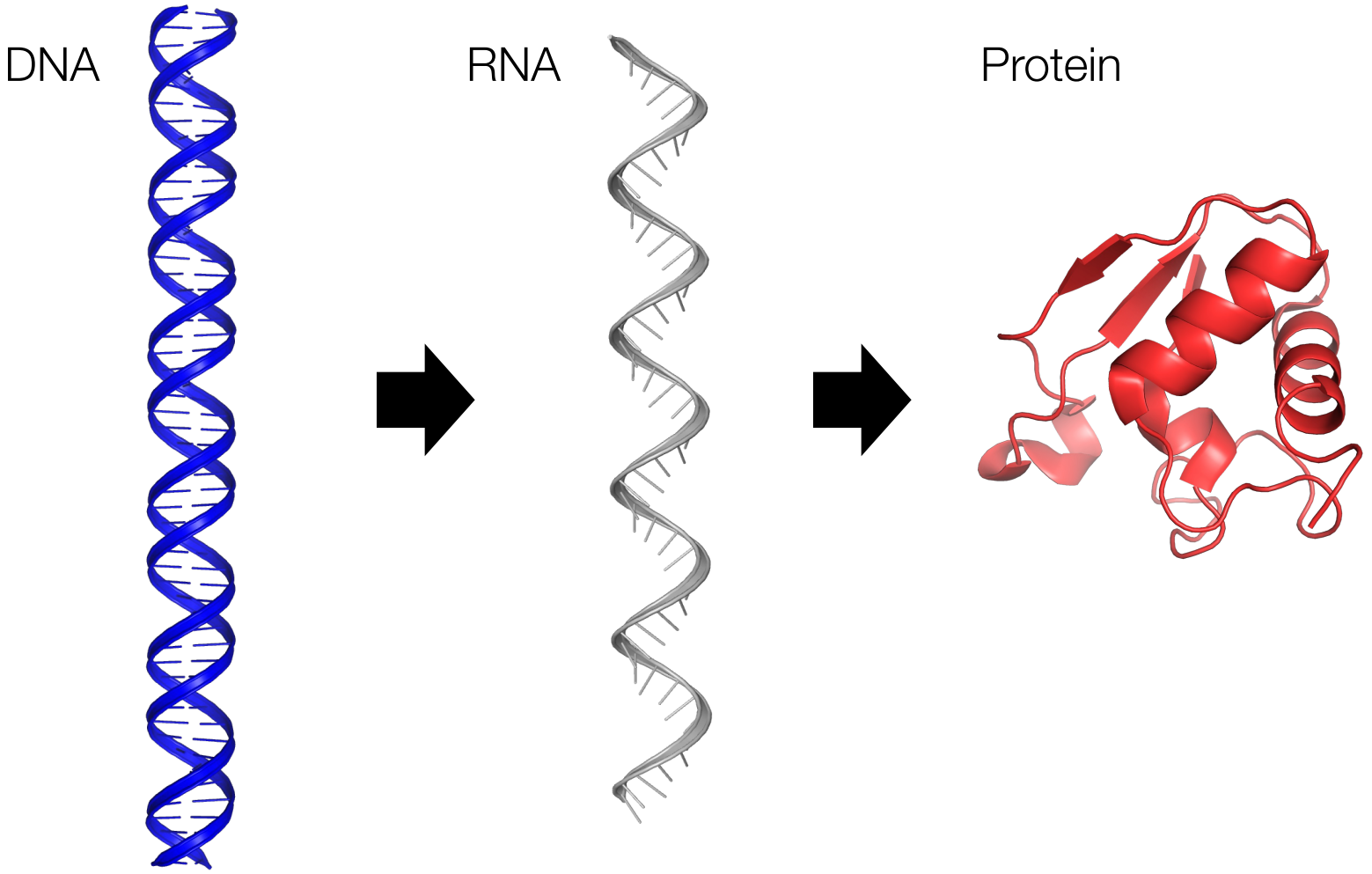
B: No mainly because mRNA is shown completely unstructured

C: Yes for tRNA only

D: No they are all oversimplified



RNA plays a central role in biology
RNA molecules typically exist as single strands

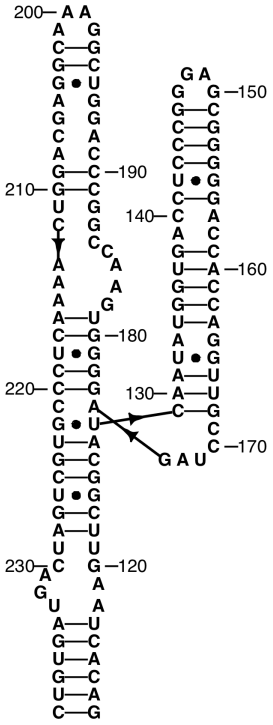


RNA can fold into secondary and tertiary structures
Structures can form *within* a single RNA strand

Sequence

GACACUAAGUUCGGCAU
CAAUAUGGUGACCUCCC
GGGAGCGGGGGACCACC
AGGUUGCCUAGAGGGGU
GAACCGGCCCAGGUCGG
AAACGGAGCAGGUCAAA
ACUCCCGUGCUGAUCAG
UAGUGU

Secondary structure



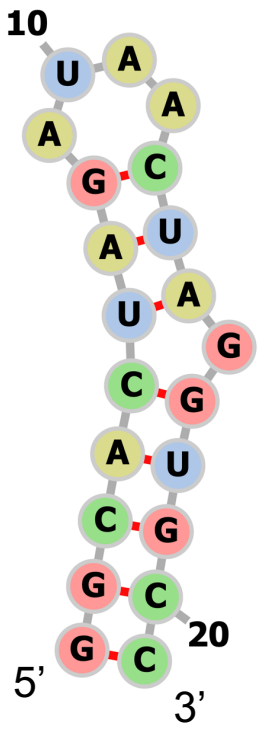
Tertiary structure

RNA-protein complexes

Signal Recognition Particle
Oubridge et al., 2002

RNA secondary structure dot bracket notation

- Unpaired nucleotides are represented as dots
- Base pairs are represented as parentheses (Watson-Crick pairs)




```
GGCACUAGAUAACUAGGUGCC
((((((((.....)))).)))))
```

What secondary structure does this RNA sequence adopt?

AUUAGCCGUAAUCGGC

A: ((((((...))))))

B: (((....)))).....

C: ....(((....)))

D: ((((....((((....

Predicting RNA secondary structure

Nearest neighbor parameters

Table 4: RNA Thermodynamic Parameters for INN-HB Nearest-Neighbor Model, 1 M NaCl, pH 7[a]

parameters	ΔG°_{37} (kcal/mol)
5′AA3′ 3′UU5′	−0.93
5′AU3′ 3′UA5′	−1.10
5′UA3′ 3′AU5′	−1.33
5′CU3′ 3′GA5′	−2.08
5′CA3′ 3′GU5′	−2.11
5′GU3′ 3′CA5′	−2.24
5′GA3′ 3′CU5′	−2.35
5′CG3′ 3′GC5′	−2.36
5′GG3′ 3′CC5′	−3.26
5′GC3′ 3′CG5′	−3.42

Xia et al. *Biochemistry* **37**, 14719-14735 (1998)

Experimentally measured thermodynamic parameters, measured via optical melting experiments

Predicting RNA secondary structure

Nearest neighbor parameters

Table 4: RNA Thermodynamic Parameters for INN-HB Nearest-Neighbor Model, 1 M NaCl, pH 7[a]

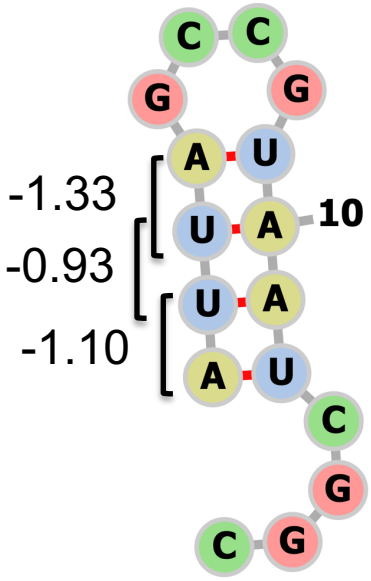
parameters	ΔG°_{37} (kcal/mol)
5′AA3′ 3′UU5′	−0.93
5′AU3′ 3′UA5′	−1.10
5′UA3′ 3′AU5′	−1.33
5′CU3′ 3′GA5′	−2.08
5′CA3′ 3′GU5′	−2.11
5′GU3′ 3′CA5′	−2.24
5′GA3′ 3′CU5′	−2.35
5′CG3′ 3′GC5′	−2.36
5′GG3′ 3′CC5′	−3.26
5′GC3′ 3′CG5′	−3.42

Xia et al. *Biochemistry* **37**, 14719-14735 (1998)

**There are also some parameters for terminal base pairs, hairpins, etc. that we are ignoring here

AUUAGCCGUAAUCGGC

((((....))))....



$\Delta G = -1.10 + -0.93 + -1.33$

$\Delta G = -3.36$ kcal/mol

Predicting RNA secondary structure

Nearest neighbor parameters

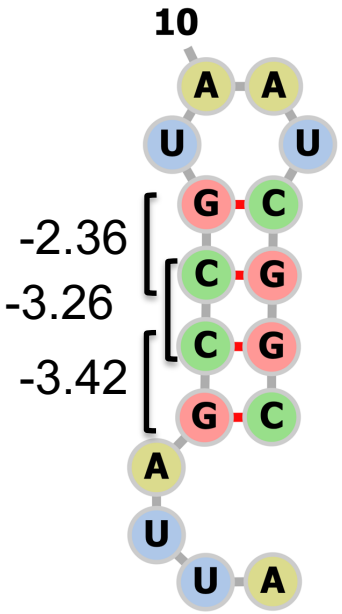
Table 4: RNA Thermodynamic Parameters for INN-HB Nearest-Neighbor Model, 1 M NaCl, pH 7[a]

parameters	ΔG°_{37} (kcal/mol)
5′AA3′ 3′UU5′	−0.93
5′AU3′ 3′UA5′	−1.10
5′UA3′ 3′AU5′	−1.33
5′CU3′ 3′GA5′	−2.08
5′CA3′ 3′GU5′	−2.11
5′GU3′ 3′CA5′	−2.24
5′GA3′ 3′CU5′	−2.35
5′CG3′ 3′GC5′	−2.36
5′GG3′ 3′CC5′	−3.26
5′GC3′ 3′CG5′	−3.42

Xia et al. *Biochemistry* **37**, 14719-14735 (1998)

AUUAGCCGUAAUCGGC

. . . . ((((. . .))))



$\Delta G = -3.42 + -3.26 + -2.36$

$\Delta G = -9.04$ kcal/mol

RNAfold WebServer: http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi
RNAeval WebServer: http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAeval.cgi

What secondary structure does this RNA sequence adopt?

AUUAGCCGUAAUCGGC

A: ((((((...))))))

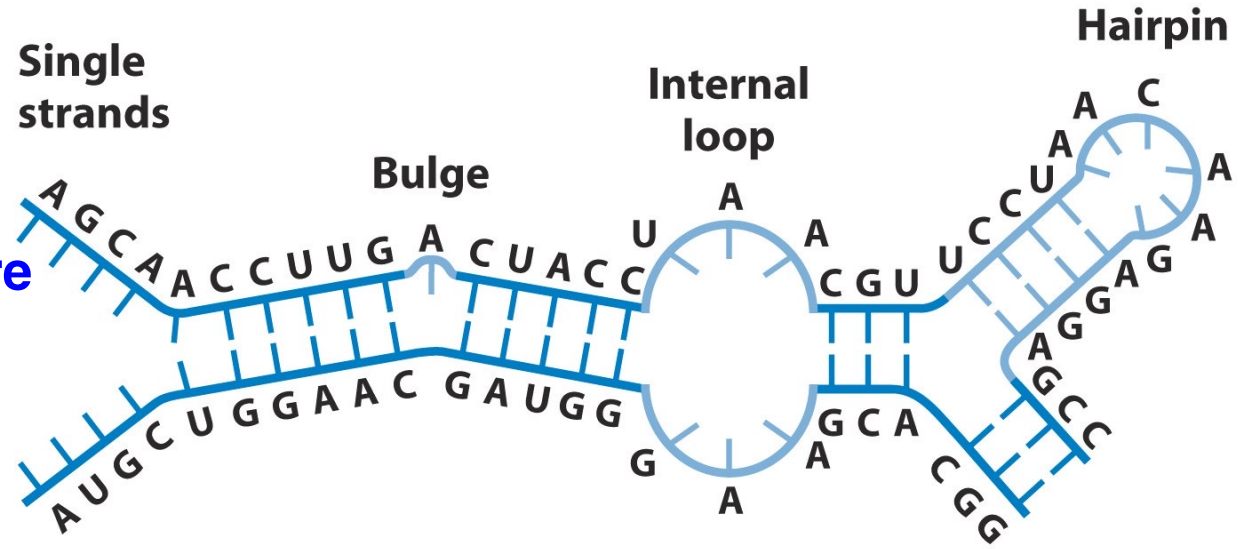
B: (((...))).... ΔG = -3.36 kcal/mol

C: ....(((...))) ΔG = -9.04 kcal/mol

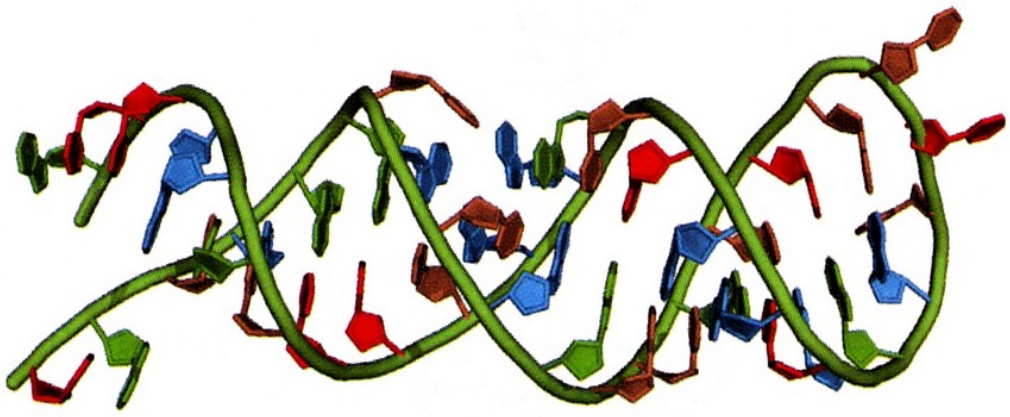
D: (((....(((....

Secondary and Tertiary Structure of RNA

Basic elements of Secondary Structure in RNAs:



A hairpin segment in pseudo 3D:



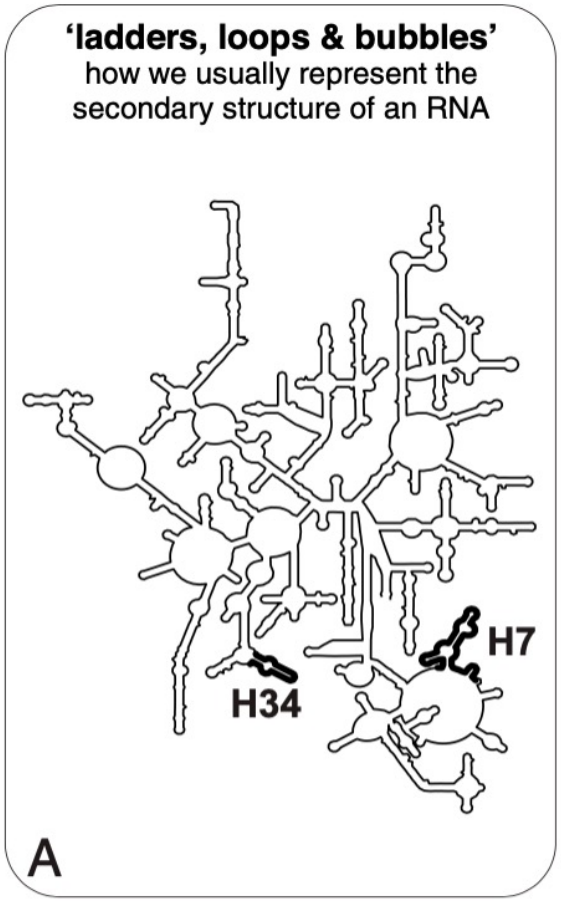
Single strandedness nature of RNA makes it able to "fold" on itself and base-pair with complementary segments within the same molecule

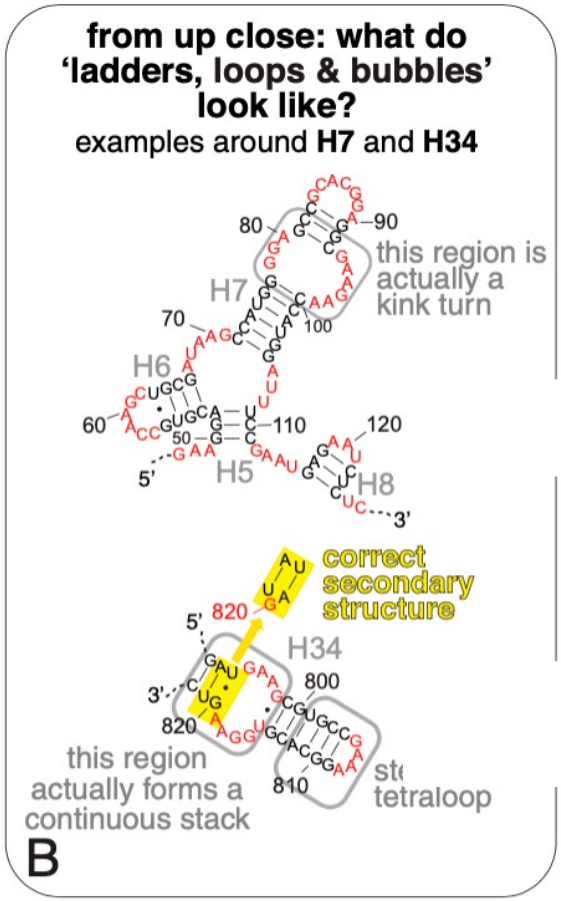
Some General Rules about RNA 3D structures/folding

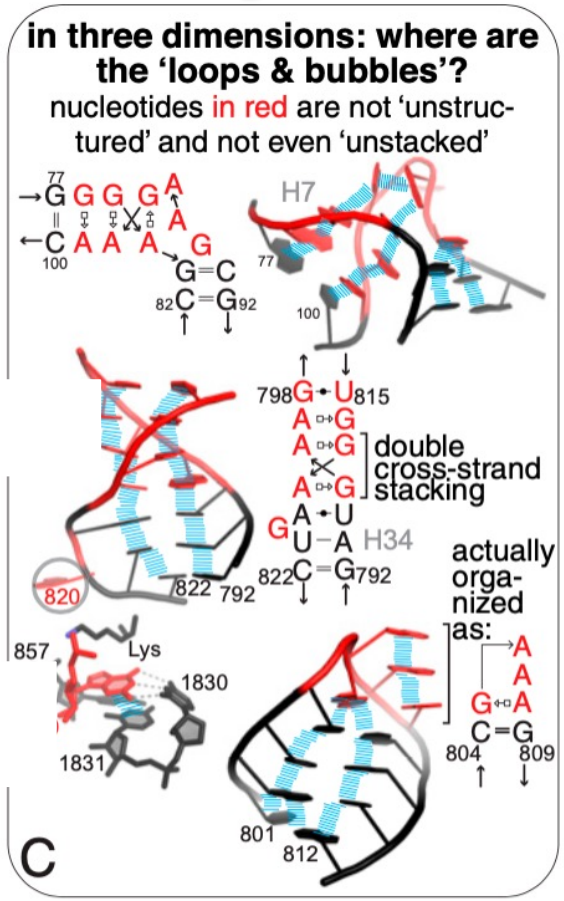
RNAs are (in general) compact molecules - not limp spaghettis...

- **Watson–Crick Pairing is important but a bit overrated**
- **Non-Watson–Crick (aka non-canonical) pairing occurs frequently and is very important!**
- **Base Stacking is a major driver for RNA structure**

From:
Vicens and Kieft
PNAS 2022





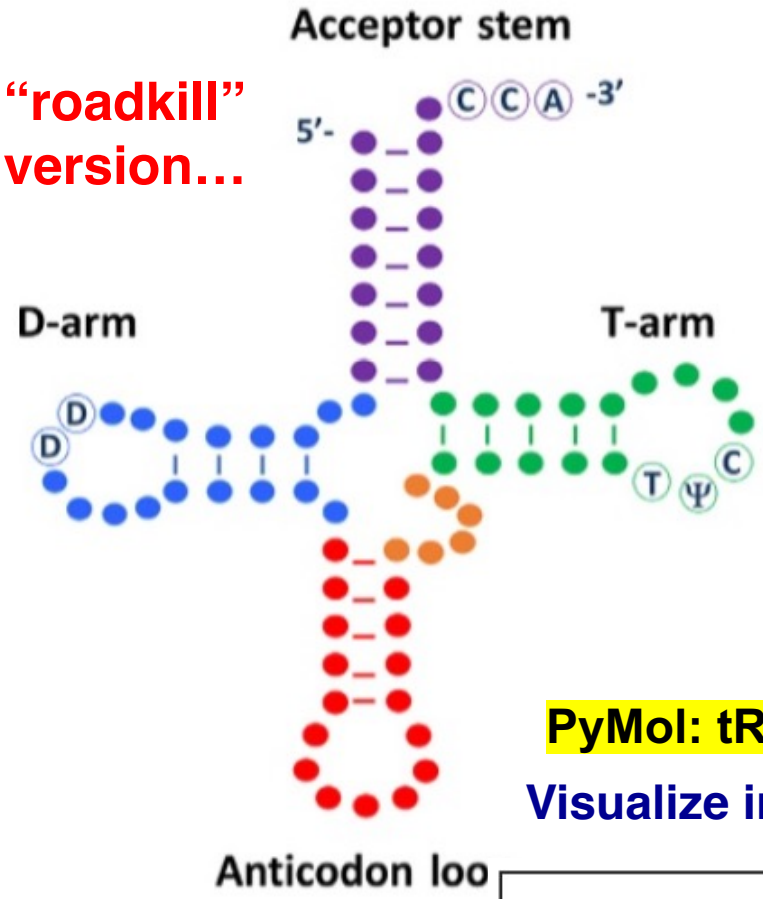


PyMol: GAAA stem loop.pse

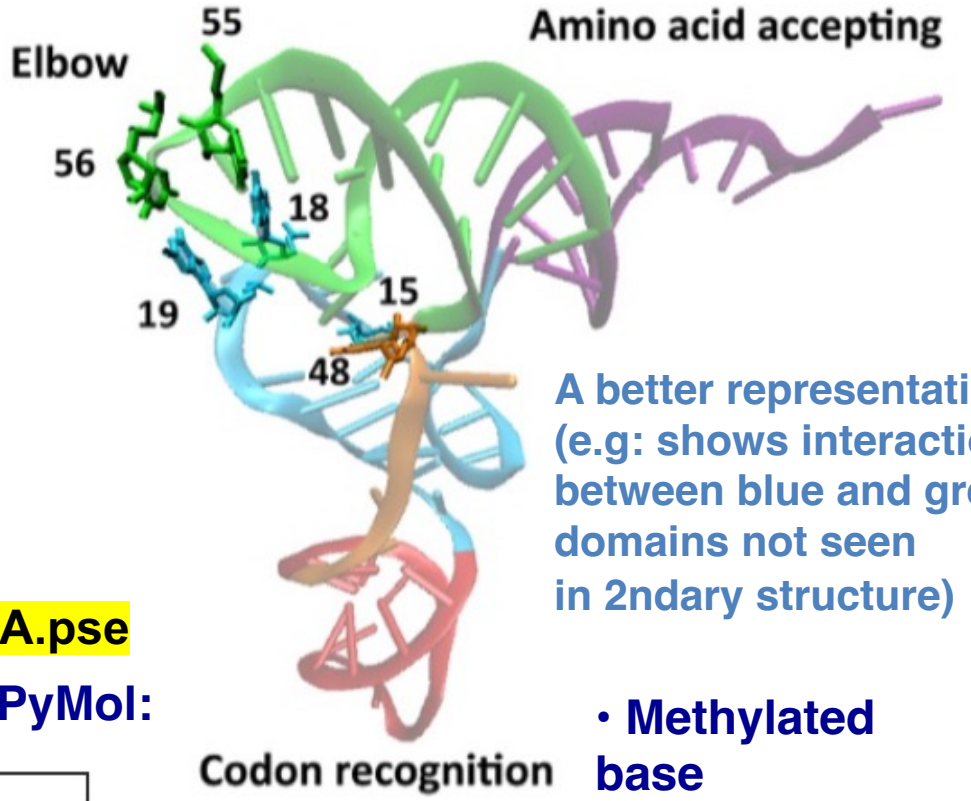
Secondary vs 3D Structures: Example of transfer RNA (tRNA)

tRNA cloverleaf secondary structure

tRNA L-shaped tertiary structure



"roadkill" version…



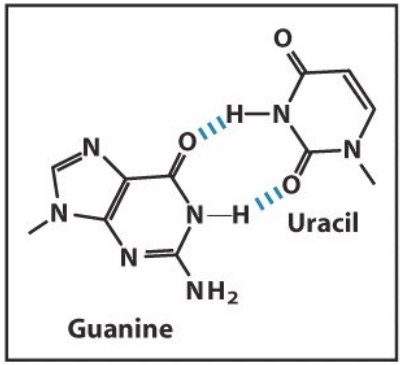
A better representation.. (e.g: shows interactions between blue and green domains not seen in 2ndary structure)

PyMol: tRNA.pse

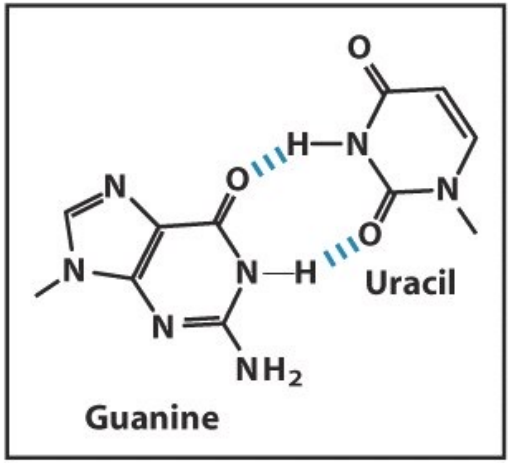
Visualize in PyMol:

- **Methylated base**
- **Non canonical base pairs: e.g. A:A**
- **Cations promote 3D folding**
- **Structure is not static!**

- **Several G:U base pairs: "A" form tolerates the presence of G:U base pairs**



Why does the A-form of dsRNA tolerate the presence of G-U base pairs?



A: Because the major groove is deeper which allows it to fit

B: Because G:U base pairs are isosteric to Watson-Crick base pairs in A-type double helices, but not in B-type.

C: Because divalent cations frequently bind RNA, which stabilize weak base pairs like G:U

D: Because the wider diameter of A-type double helices allows for more flexibility in the geometry of the base pairs that can be included

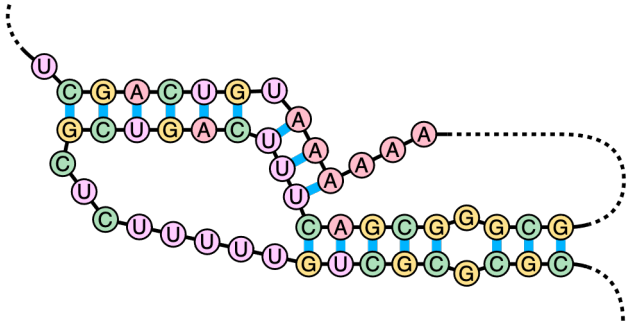
Some structural motifs that help RNAs fold into 3D structures

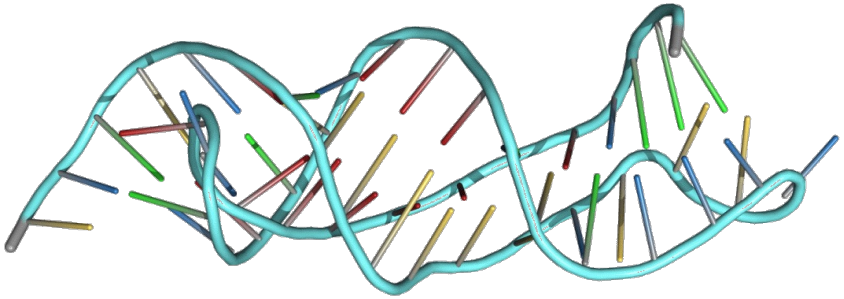
Pseudoknots and kissing loops

Pseudoknots

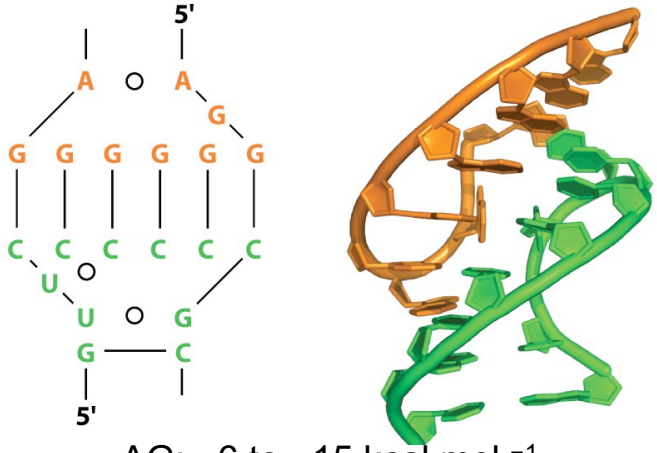
Half of one stem is between the two halves of another stem

```
((((((..[[[...))))))....]]]
```

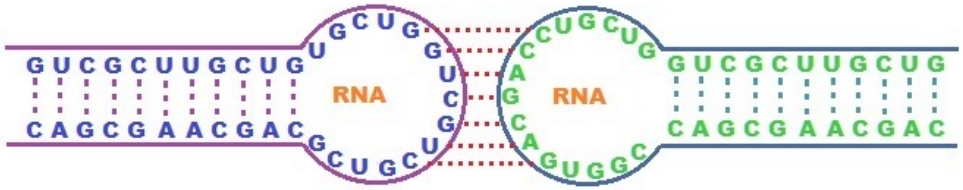




Kissing loops



ΔG: −6 to −15 kcal mol.$^{-1}$

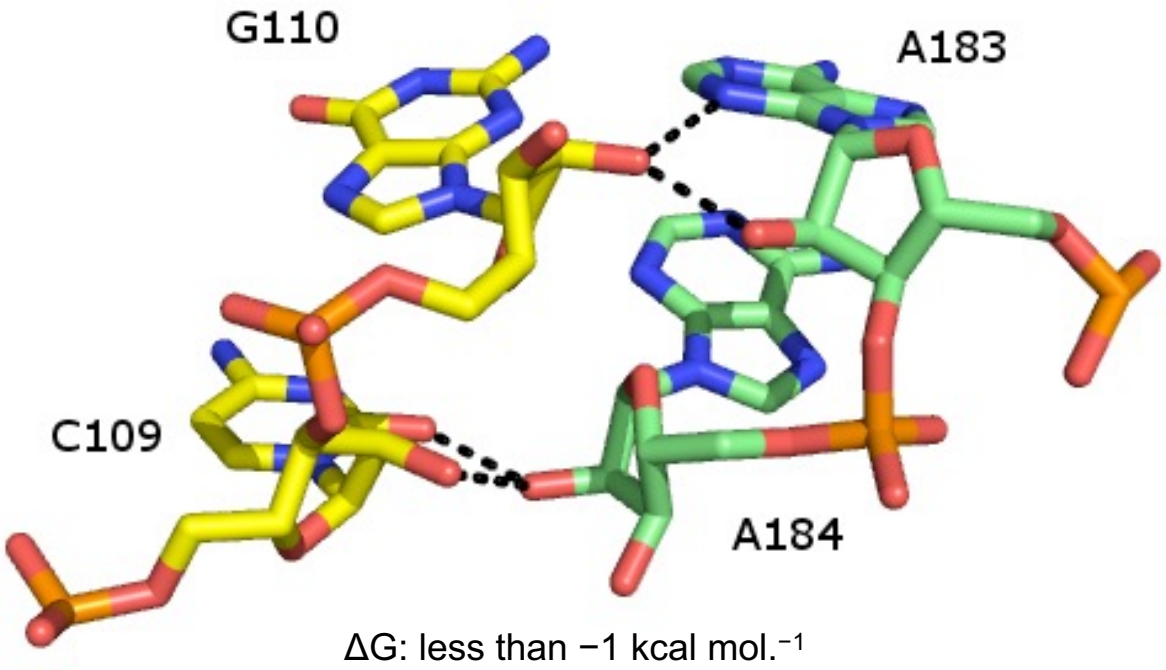


- Watson-Crick base pairs between two stem loops

https://en.wikipedia.org/wiki/Pseudoknot

Ribose zippers

Hydrogen bonding between the 2'OH groups on two different strands

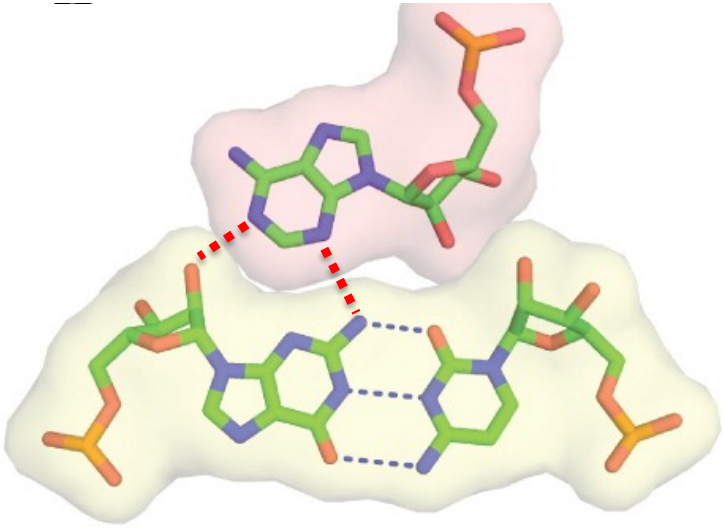


ΔG: less than −1 kcal mol.$^{-1}$

Brings two strands/backbone close together

A-minor interactions

Adenosine is inserted into the minor groove of an RNA double helix.
Found in ~25% of adenosines of small ribosomal RNA



Woodson SA. 2010.
Annu. Rev. Biophys. 39:61–77

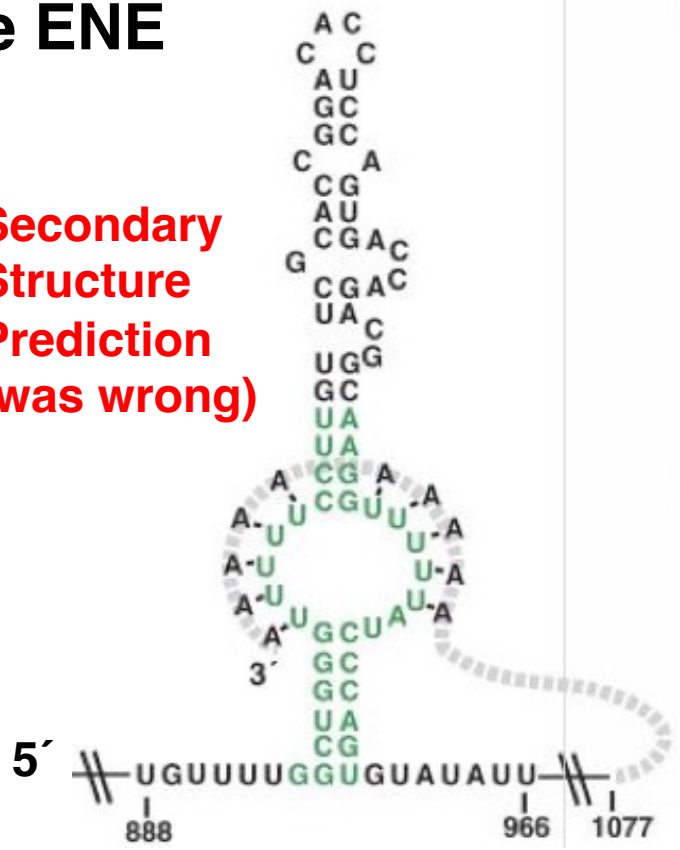
ΔG: less than −1 kcal mol.$^{-1}$

Triple Helical Structures: Example of the ENE Structural element in a viral RNA

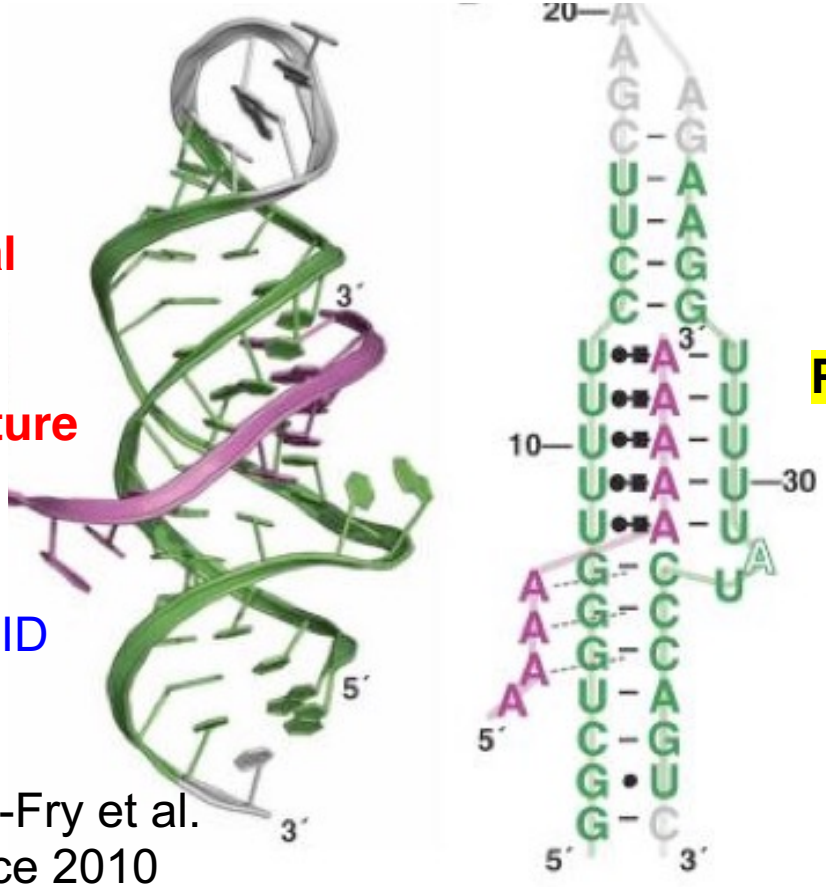
Kaposi's sarcoma-associated herpesvirus (KSHV) produces the PAN RNA, a highly abundant, nuclear noncoding RNA

A triple helical structure at the 3´-end of the PAN RNA is important to protect this RNA against degradation by cellular ribonucleases.

Secondary Structure Prediction (was wrong)



Actual Triple Helix structure



PDB ID 3P22

Mitton-Fry et al. Science 2010

PyMol: TripleHelix_ENE_A9.pse

U-A-U base pairs