Structural Bioinformatics

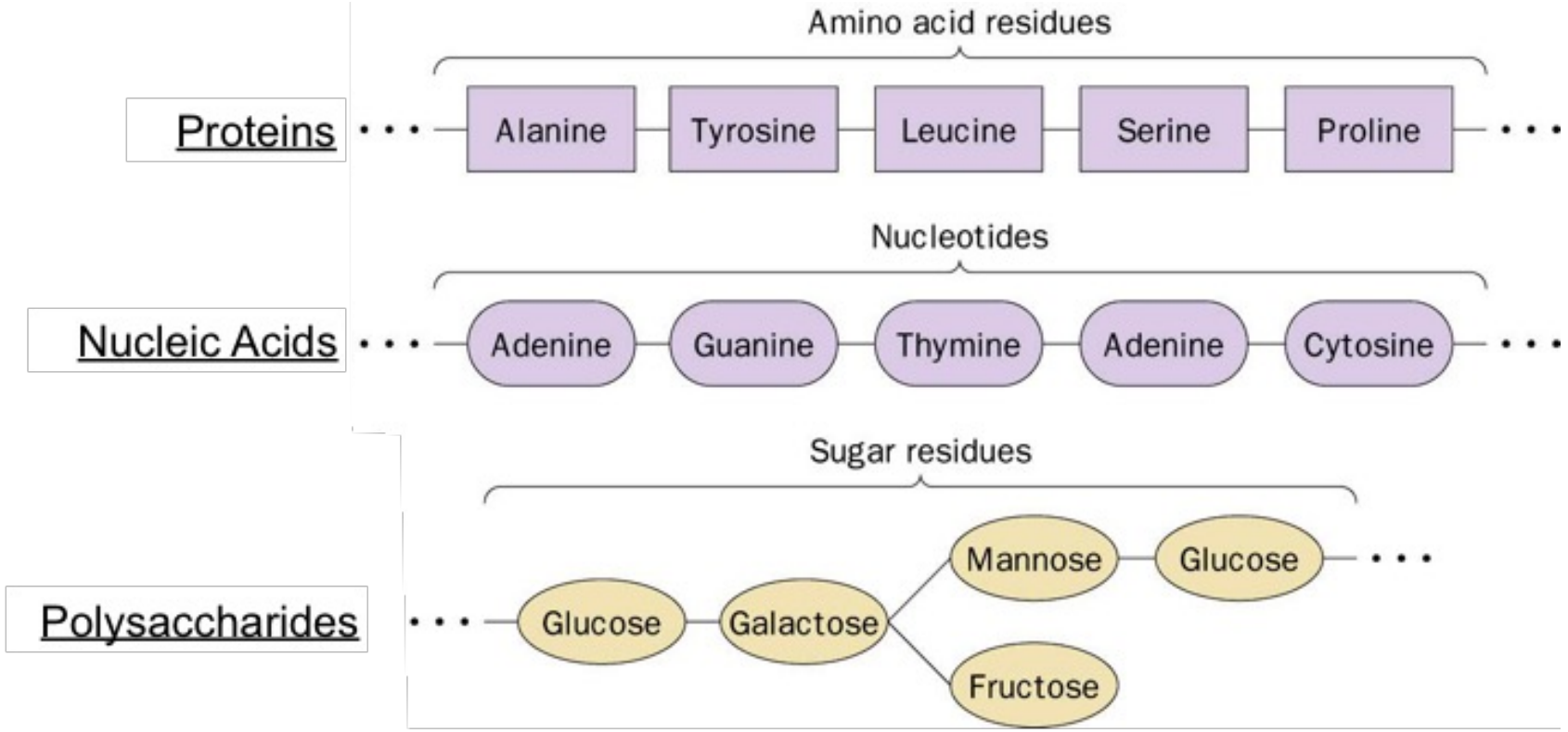
GENOME 541

Spring 2022

Lecture 4: Nucleic Acids

Frank DiMaio (dimaio@uw.edu)

The major biopolymers

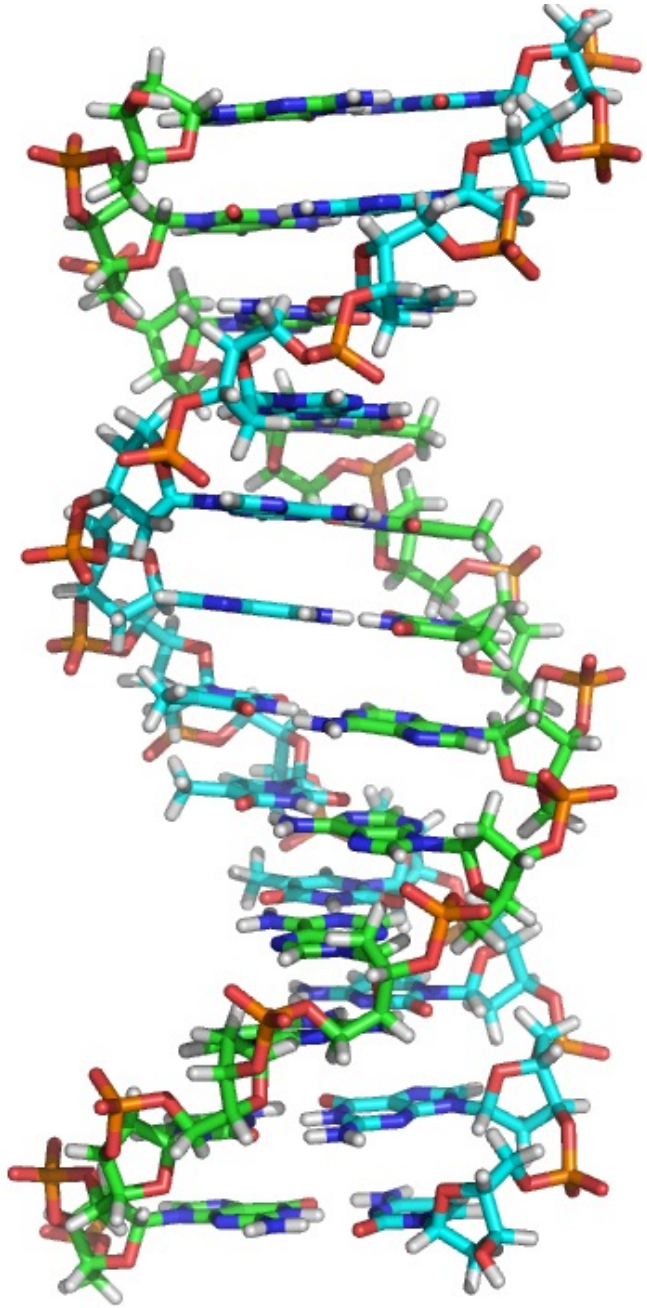


Building blocks:

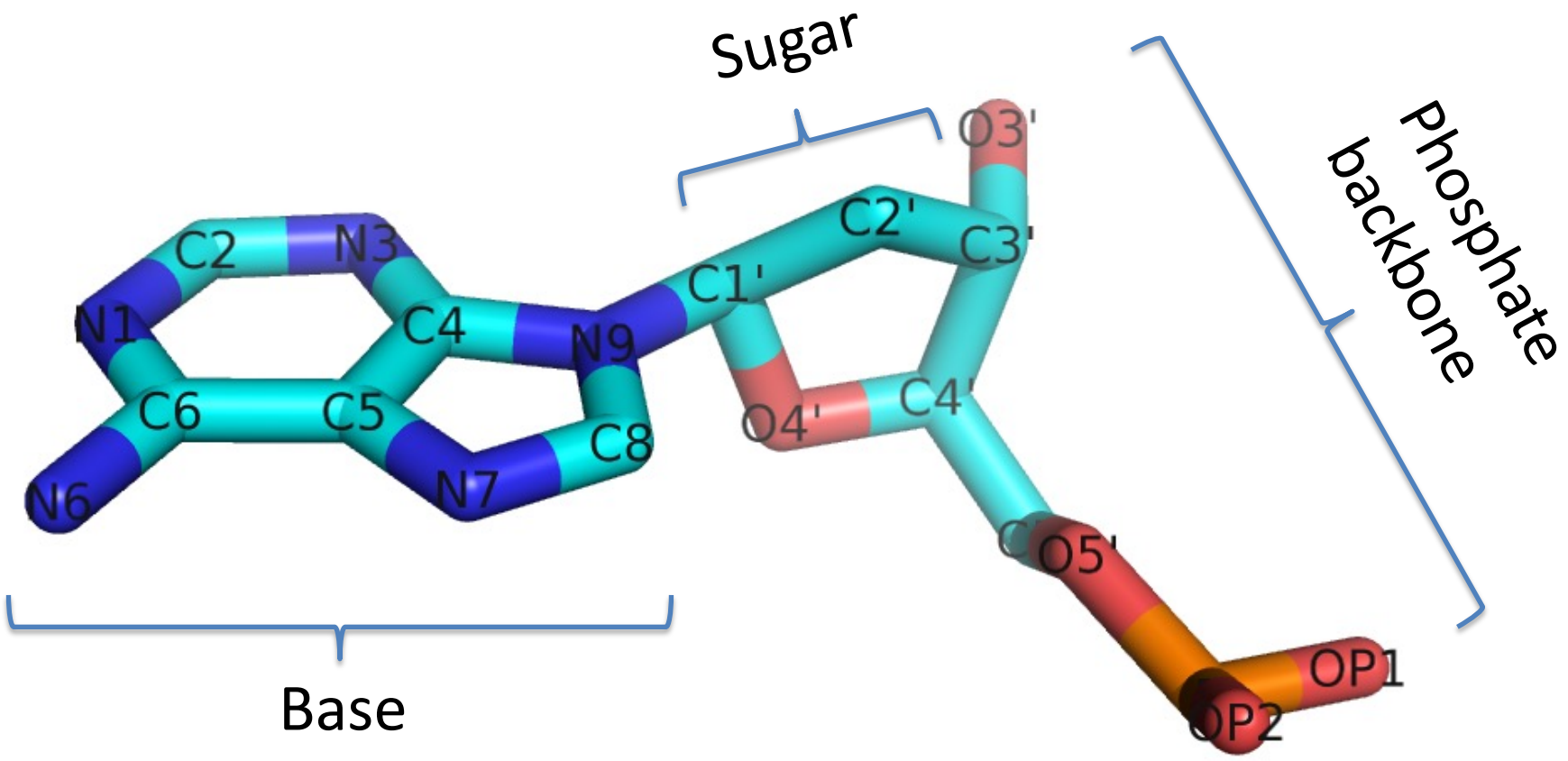
- **Proteins:** Amino Acids
- **Nucleic Acids:** Nucleotides
- **Polysaccharides:** Sugars

DNA structure

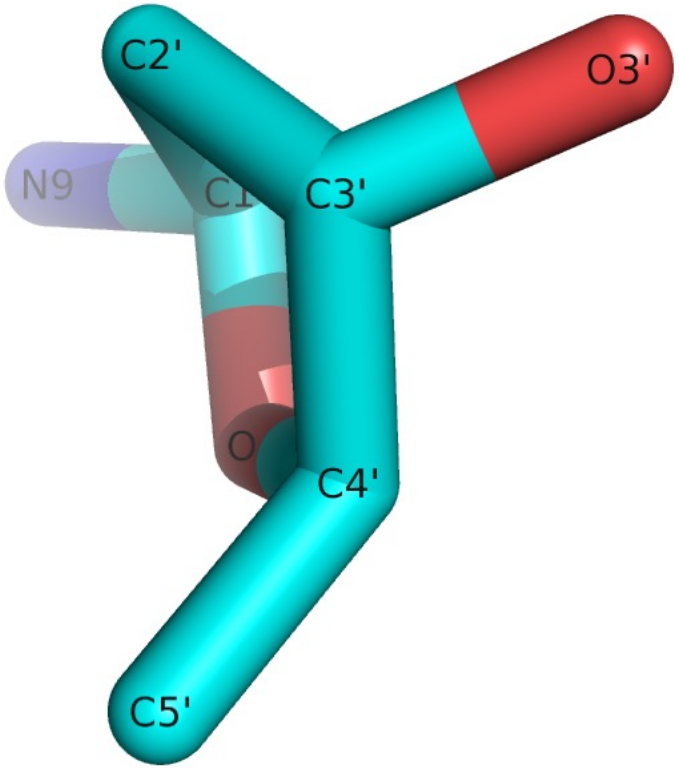
- B-form DNA
  - right-handed anti-parallel double helix
  - ~10 base pairs per turn
  - 3.4 Å rise per base pair
  - C2’ endo sugar pucker
  - A:T and G:C base pairs
  - wide major groove, narrow minor groove



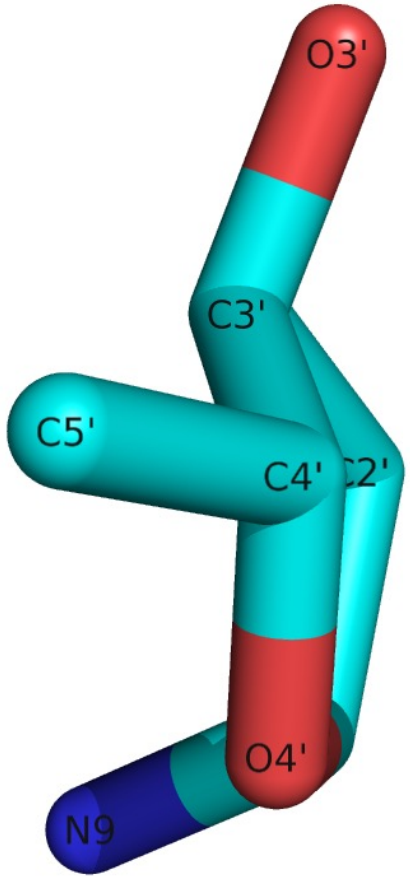
DNA structure: a single nucleotide



DNA structure: sugar pucker



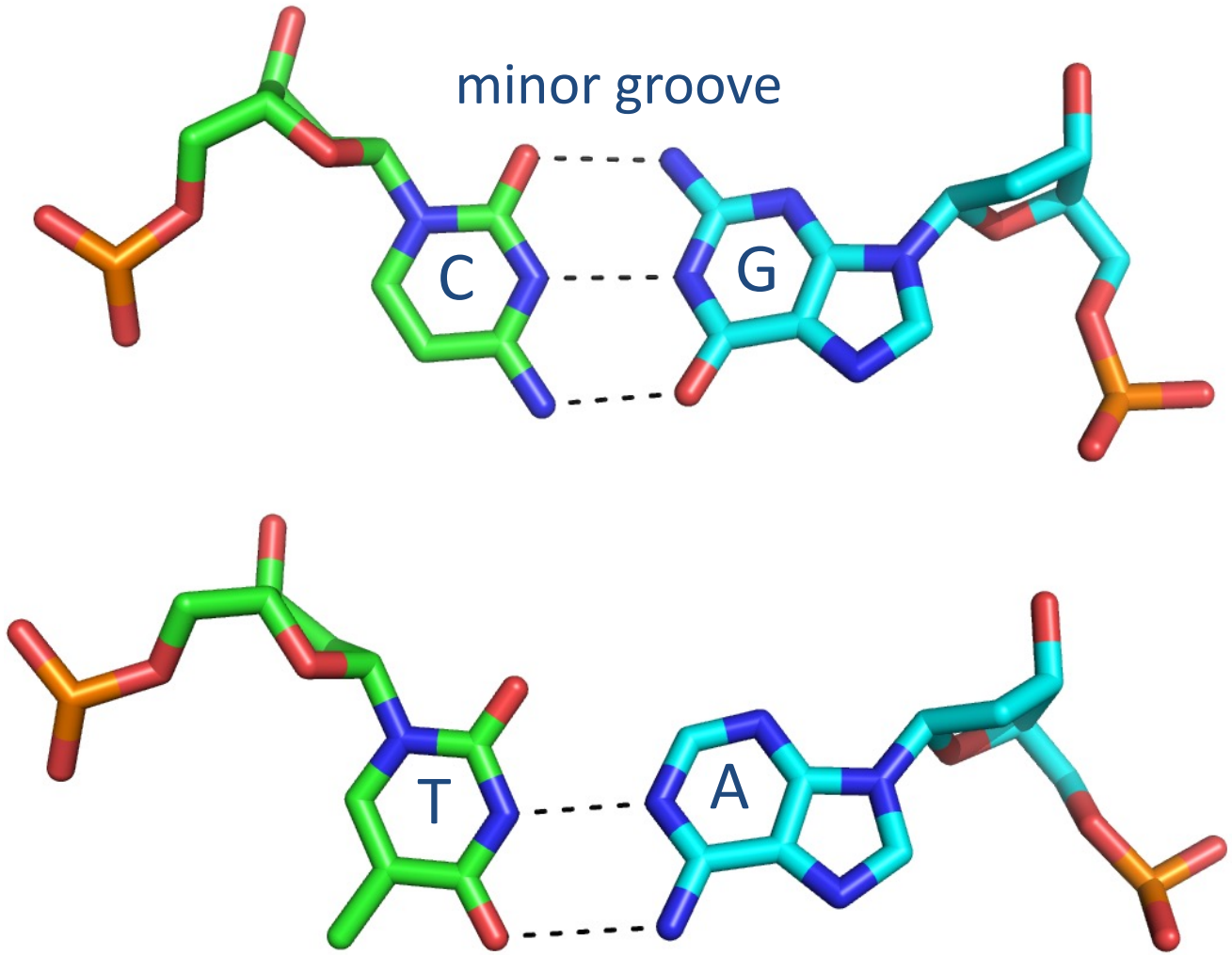
B-form: C2’-endo pucker



A-form: C3’-endo pucker

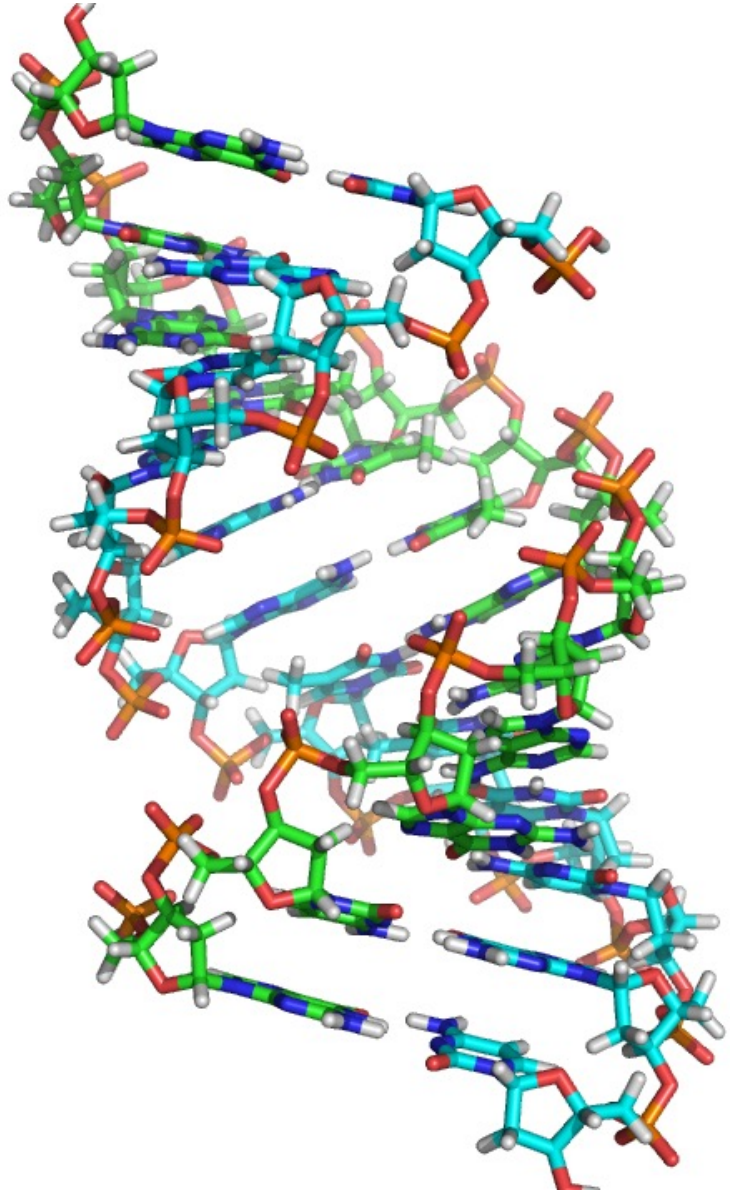
The ribose ring is not planar

DNA structure: Watson-Crick base-pairing



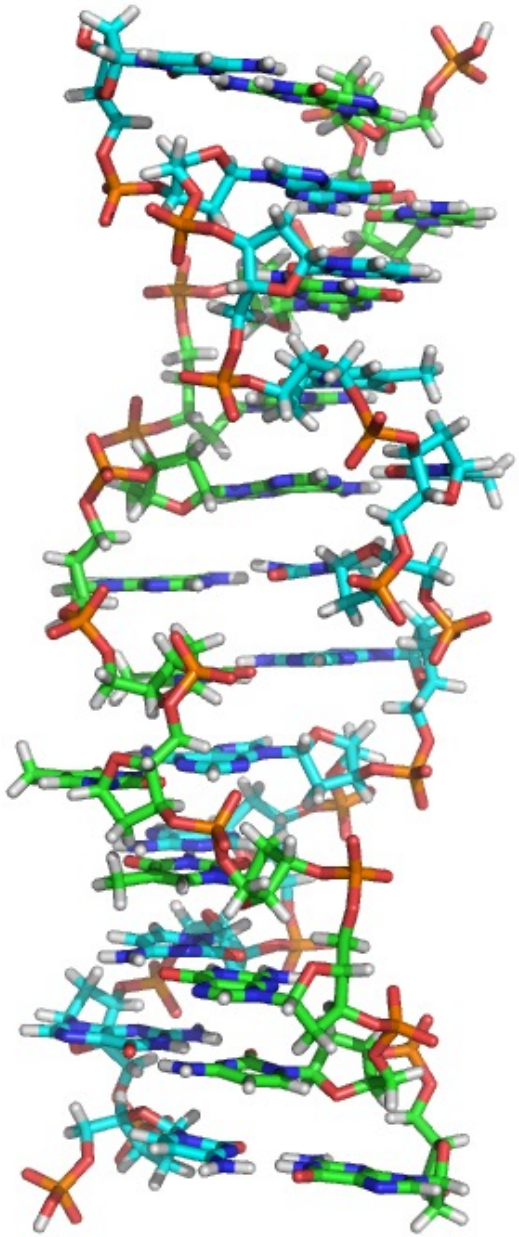
DNA structure

- A-form DNA
  - right-handed anti-parallel double helix
  - ~11 base pairs per turn
  - 2.56 Å rise per base pair
  - C3’ endo sugar pucker
  - A:T and G:C base pairs
  - narrow and deep major groove, wide and shallow minor groove



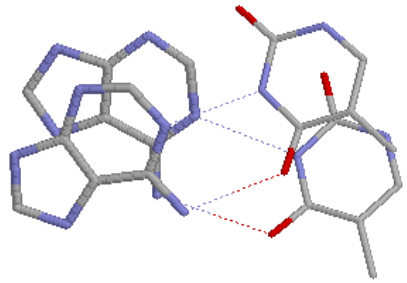
DNA structure

- Z-form DNA
  - left-handed anti-parallel double helix
  - alternating C:G and G:C base pairs
  - found under high salt conditions
  - rare in nature



Factors Stabilizing the DNA Duplex

1. **“Hydrophobic interactions,” base stacking**
   - vertical base stacking interactions make duplex formation enthalpically favored

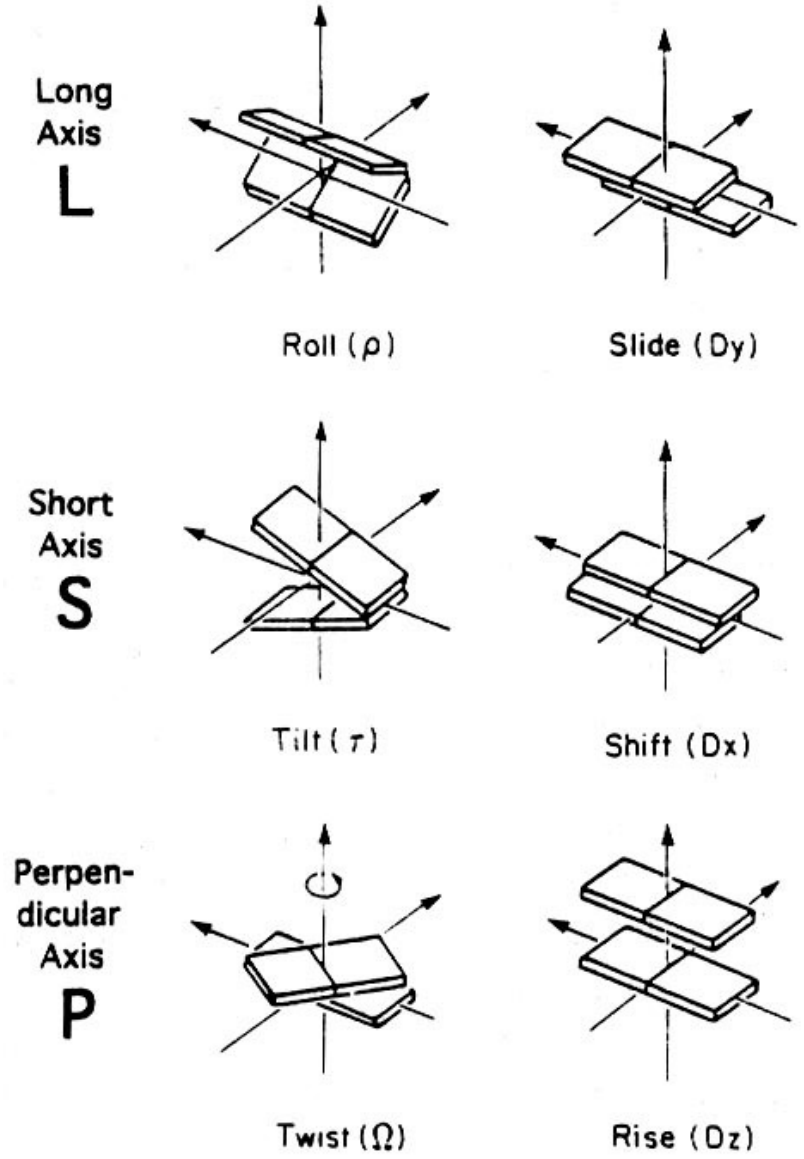


2. **Ionic interactions**
   - duplex becomes more stable as ionic strength increases
   - presence of positive counterions partially neutralizes negative charges of backbone phosphates

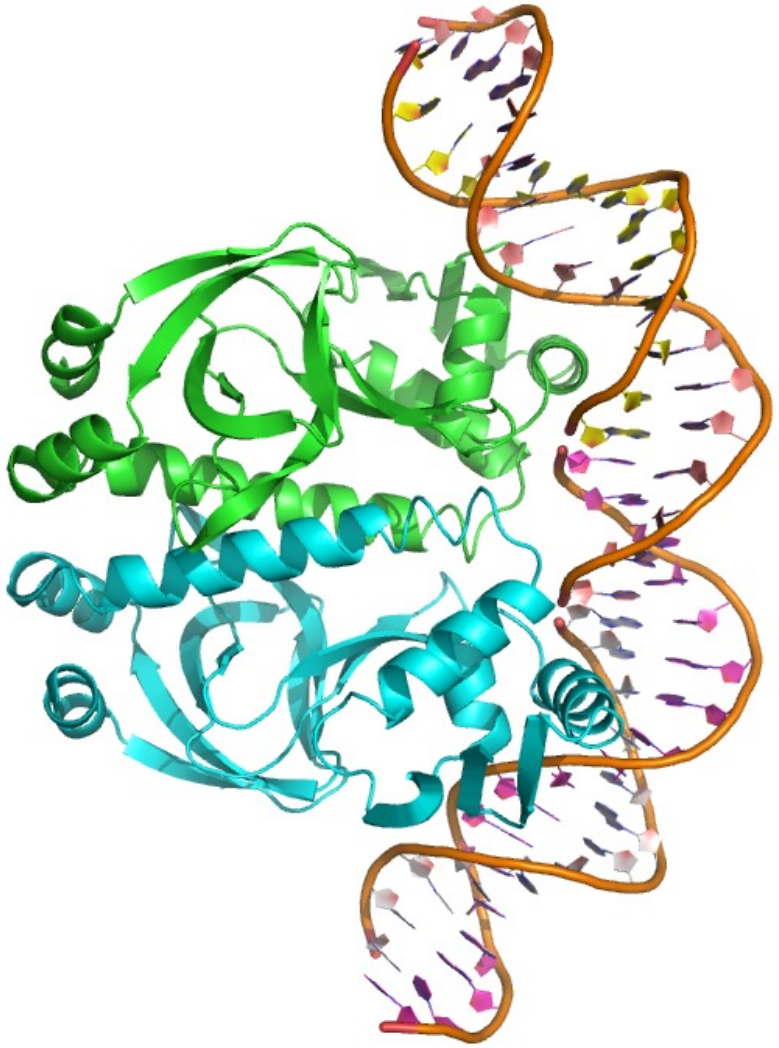
3. **Hydrogen bonding between base pairs**

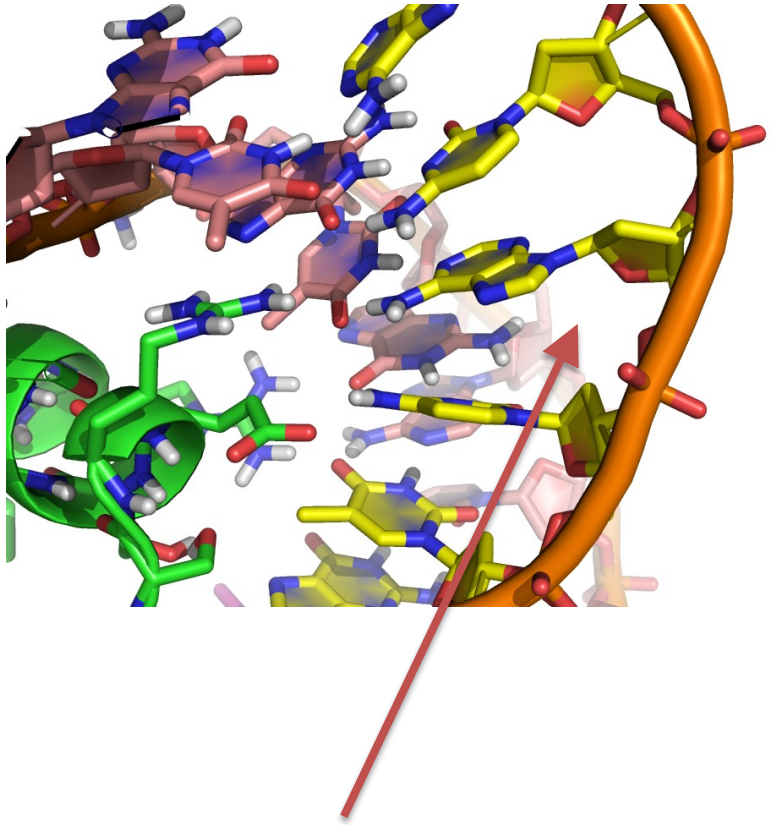
DNA bending

- B-form DNA bends in three major modes:
  - major kinking (CAP)
  - writhe (TBP)
  - smooth continuous bending (Mat a1/alpha2 homeodomain)
- Different base steps have different intrinsic bending propensities
  - pyrimidine-purine base steps can form sharp kinks (e.g. T-A steps)



DNA bending: kinking in CAP:DNA

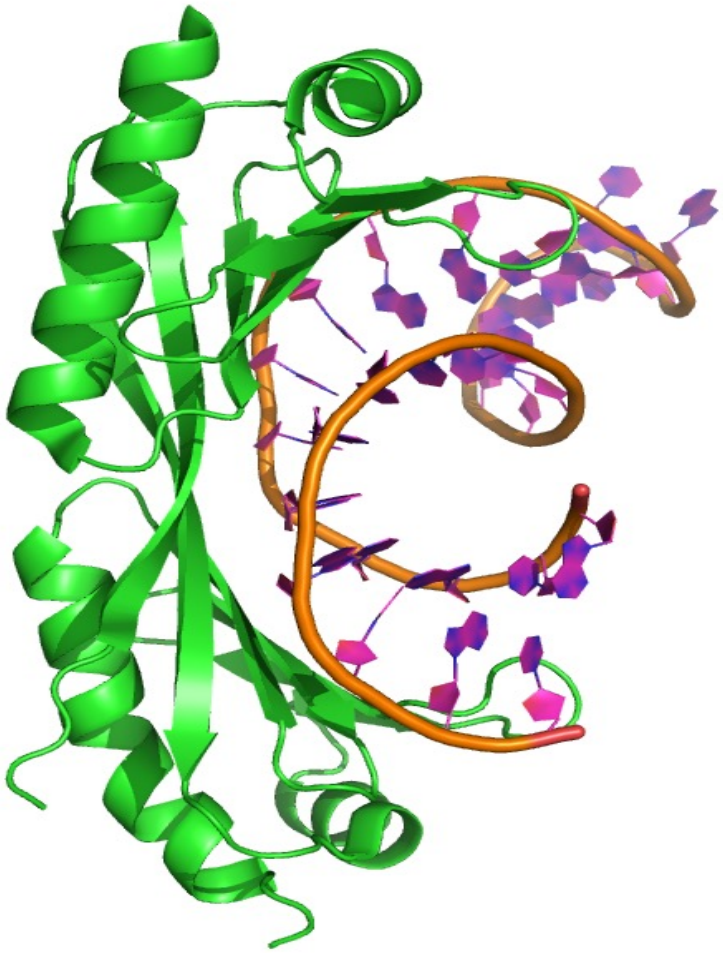




Kink at C-A step (pyrimidine-purine)

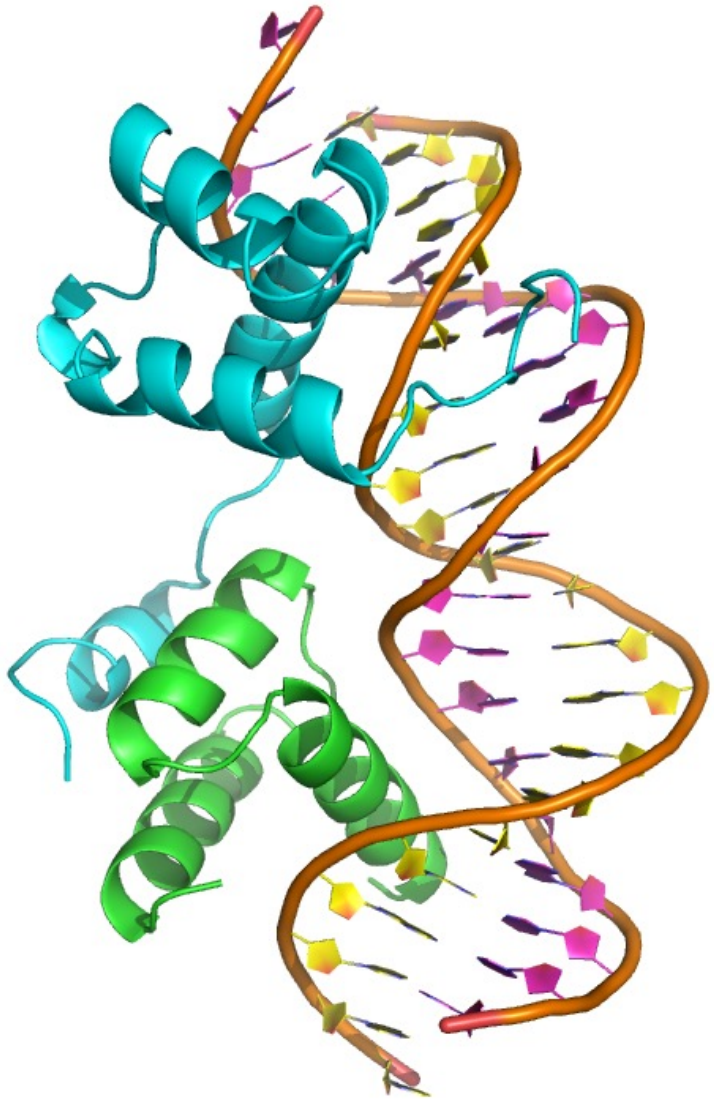
1CGP

DNA bending: writhing in TBP:DNA



1YTB

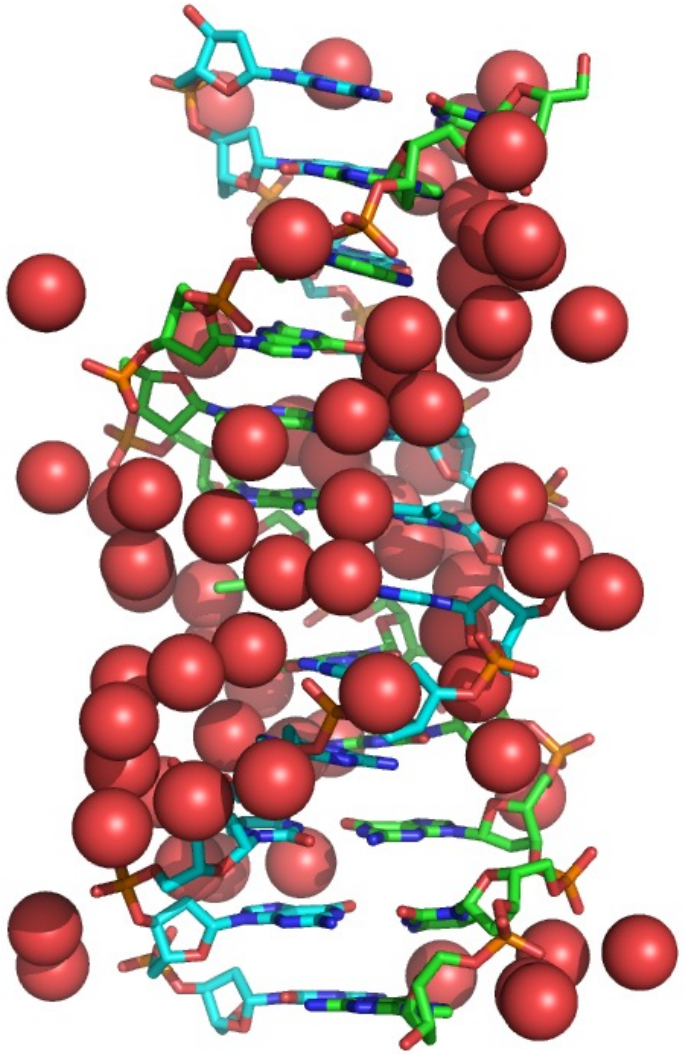
DNA bending: smooth bending in MAT a1-alpha2:DNA



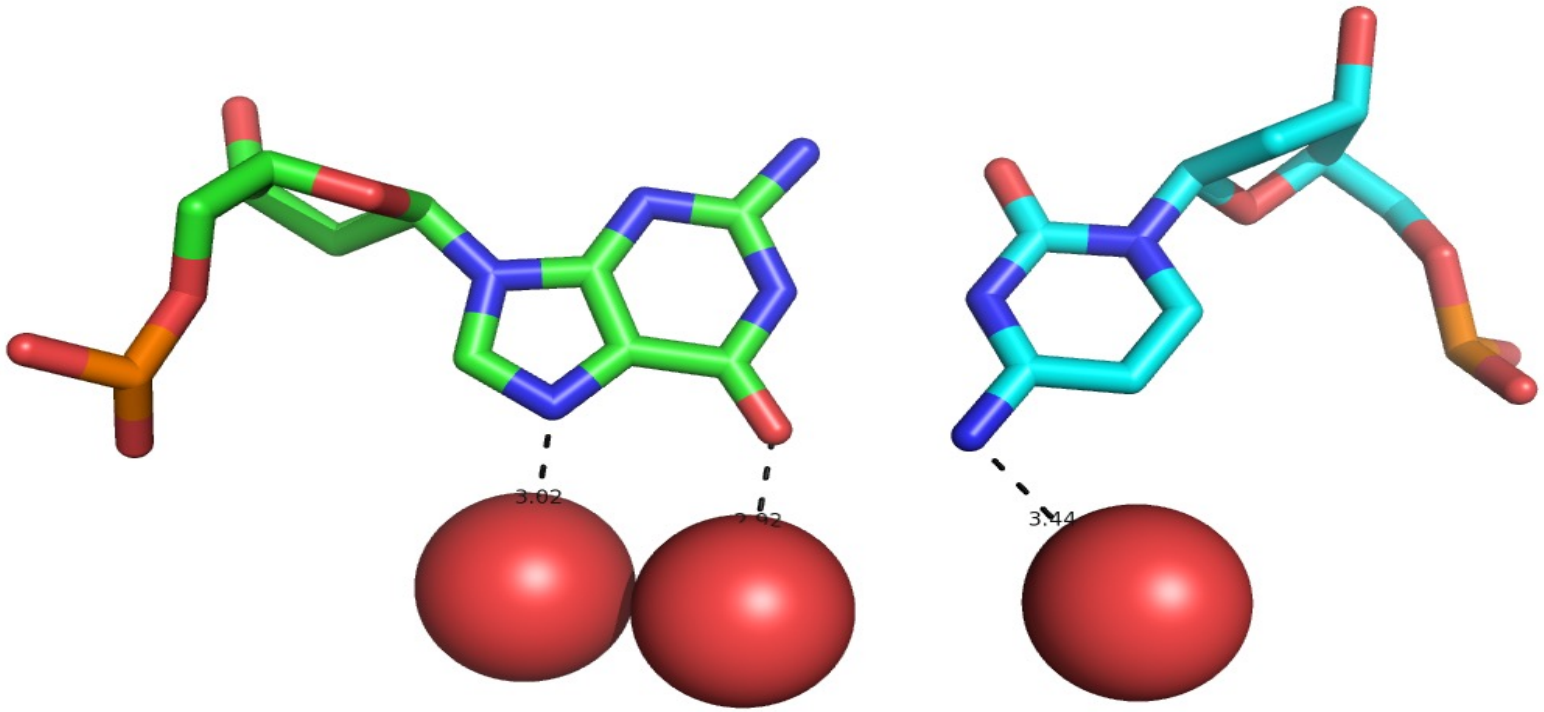
1YRN

DNA hydration

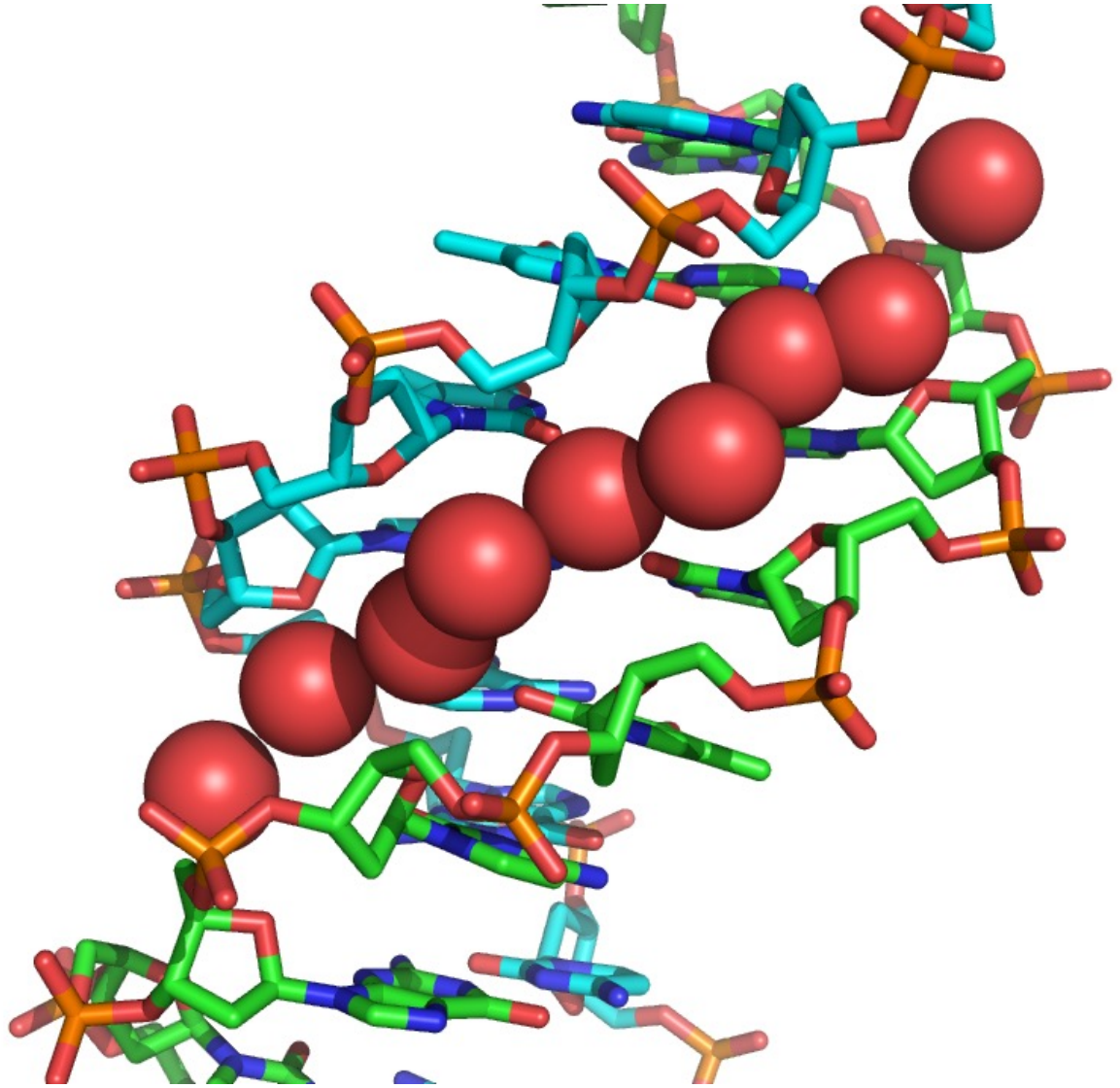
- DNA is highly hydrated under physiological conditions
- Specific ordered water locations have been identified through analysis of high-resolution DNA crystal structures
  - major groove base waters
  - minor groove spine of hydration



DNA hydration: major groove waters



DNA hydration: minor groove waters

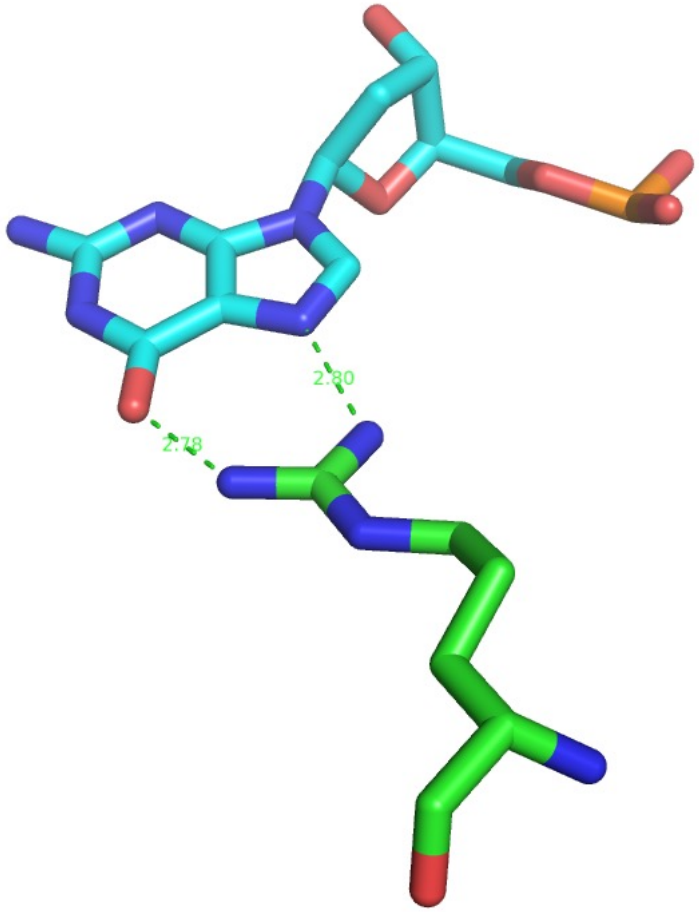


1BNA

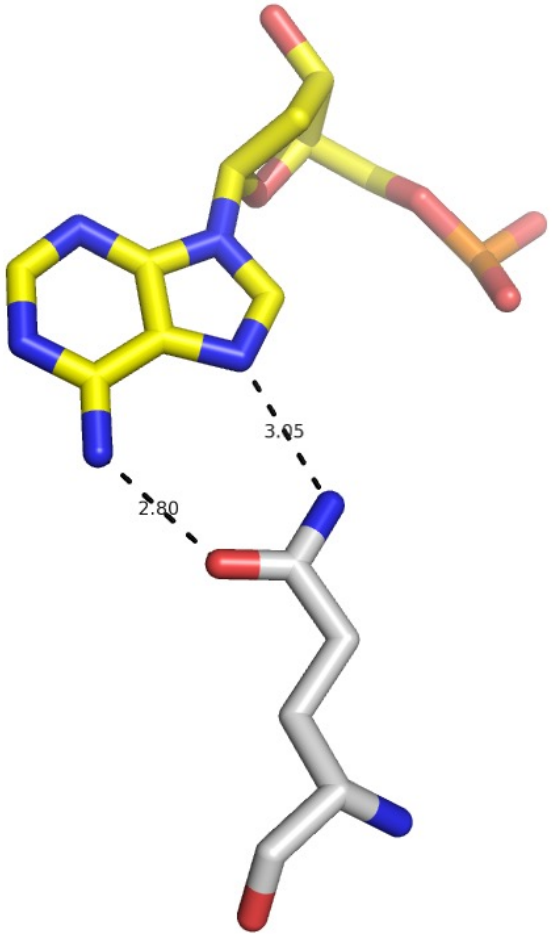
DNA recognition

- Direct readout
  - protein recognizes specific pattern of hydrogen bond donors/acceptors, packing sites
  - major groove usually targeted due to uniqueness of hbond pattern
- Indirect readout
  - protein recognizes DNA shape
  - sequence-specific DNA bending
  - phosphate backbone contacts often important

DNA recognition: direct readout

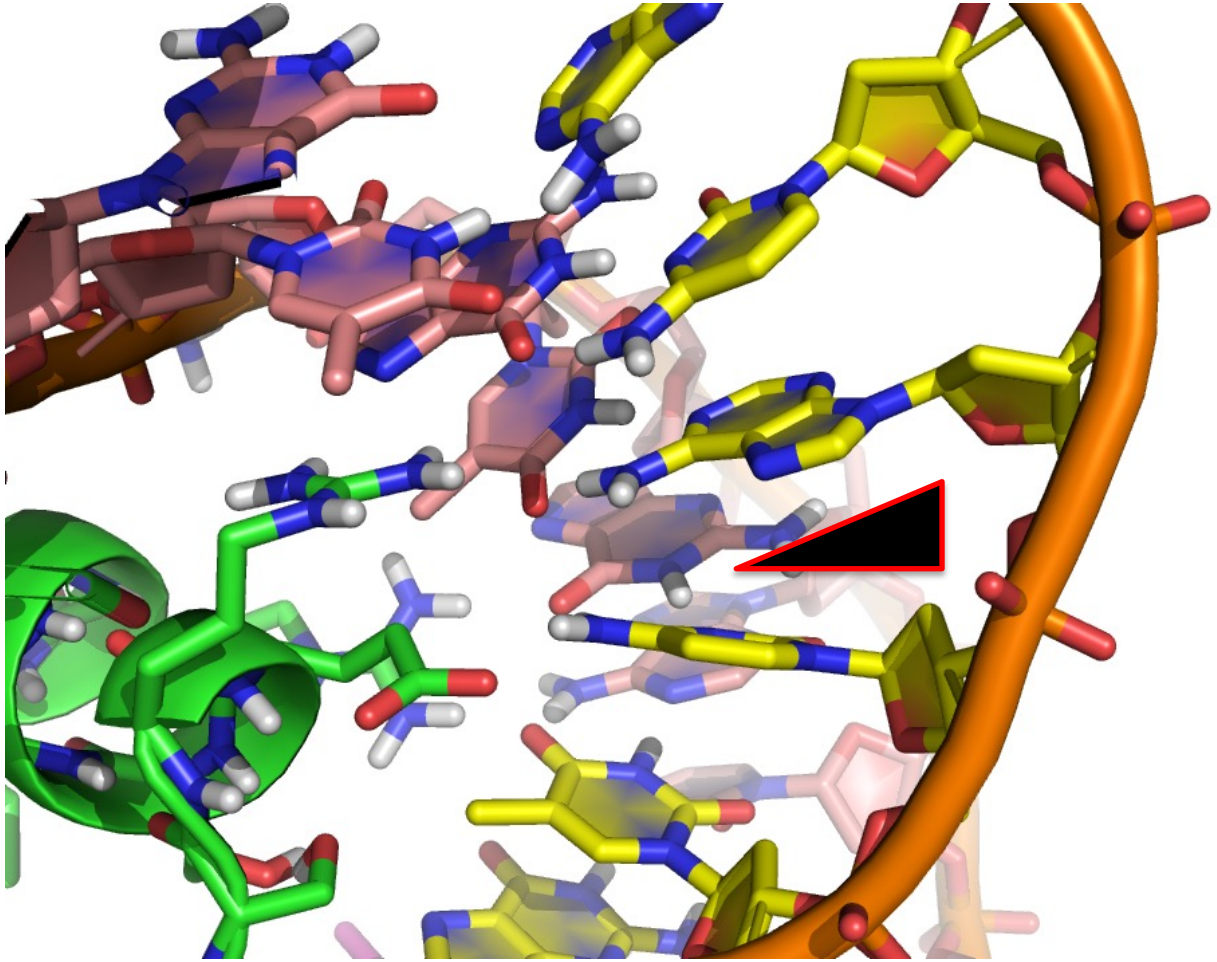


Arg-Gua hbonds



Asn-Ade hbonds

DNA recognition: indirect readout



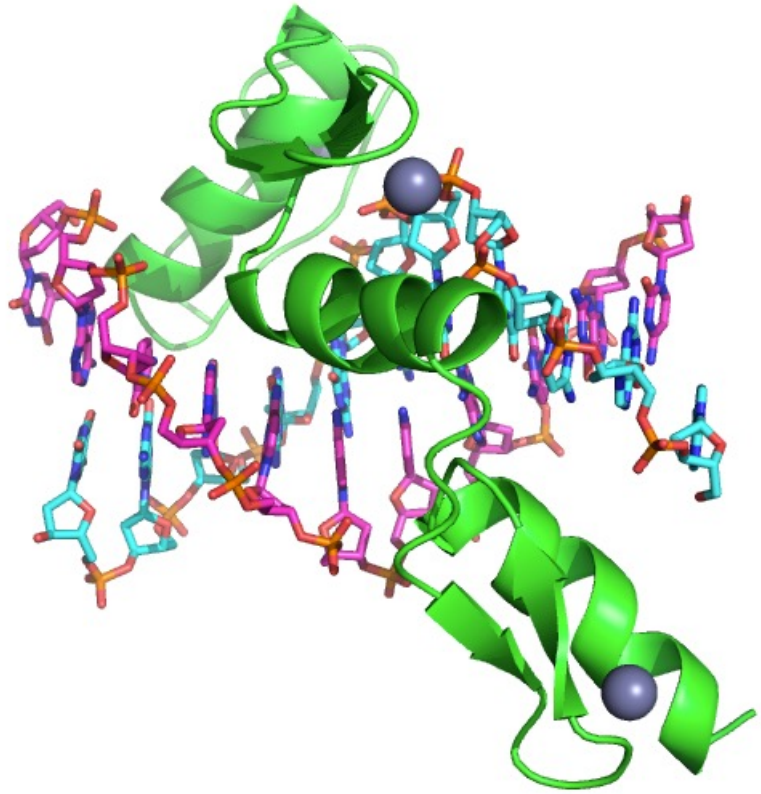
Kink at pyrimidine-purine base step

1CGP

DNA recognition: major families

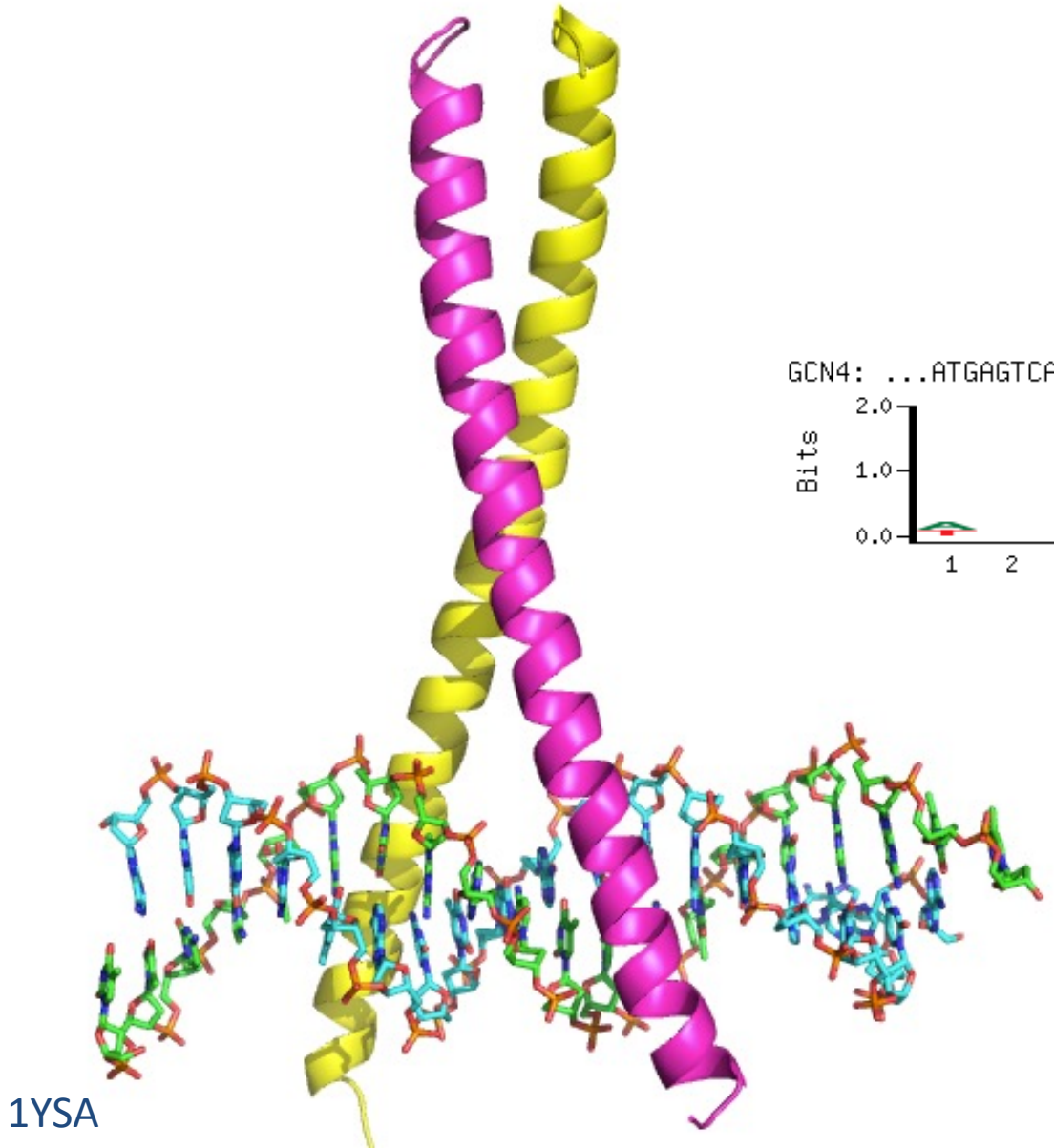
- Helix-turn-helix (1cgp)
  - Homeodomain (1b72)
- Zinc finger (1aay)
- bZIP (1ysa)
- bHLH (1mdy)

C2H2 zinc finger: Zif268



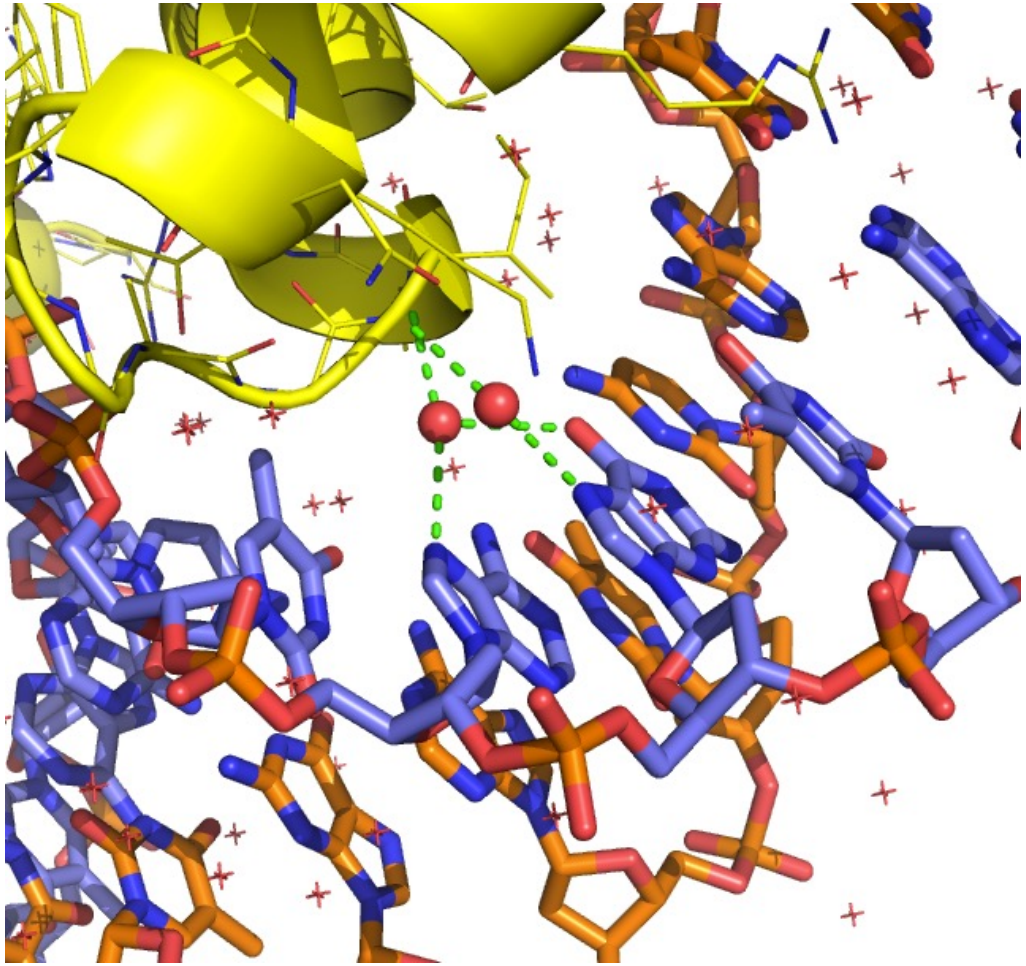
1AAY

bZIP: GCN4



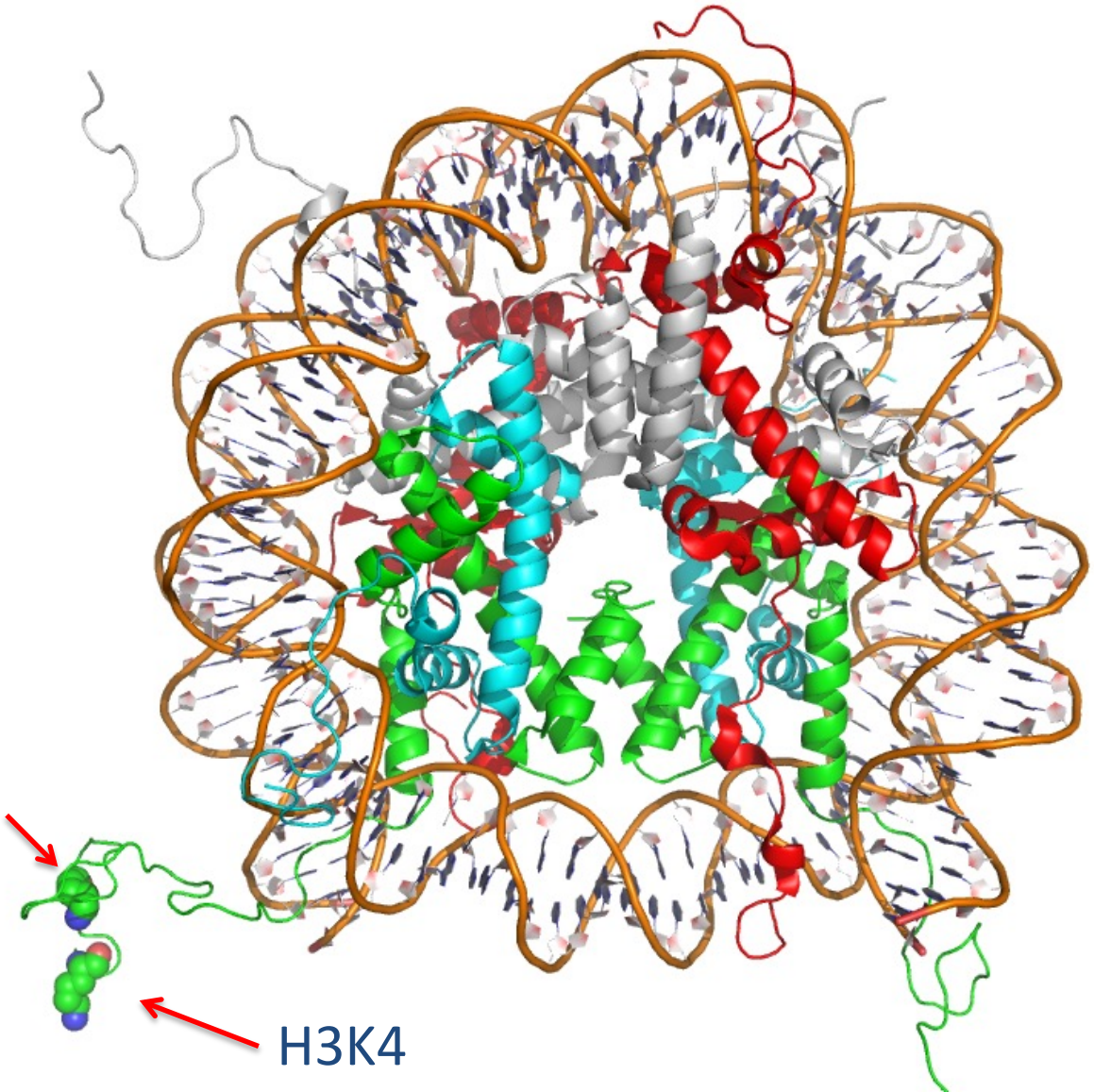
1YSA

Water-mediated interactions: Trp repressor



1TRO

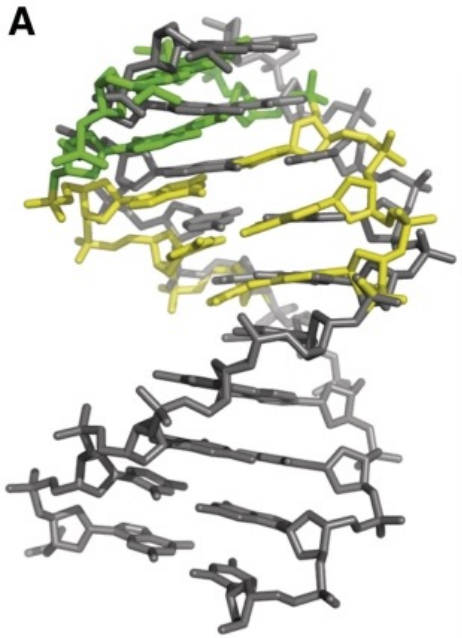
DNA is wrapped around nucleosomes

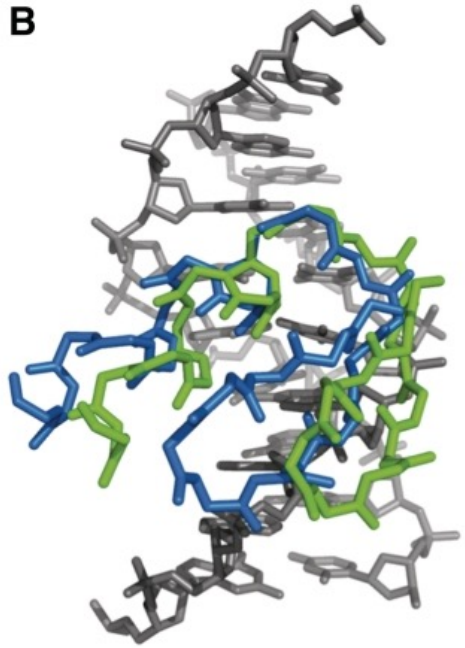


DNA bending leads to sequence preferences for nucleosome positioning

Protein-DNA interfaces require new sampling moves

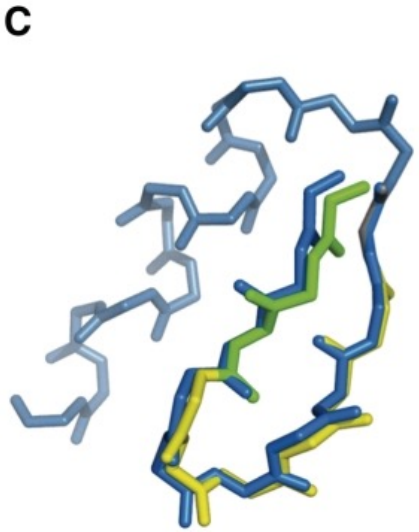
Double-helical DNA fragment insertions preserve base-pairing outside the region of fragment insertion

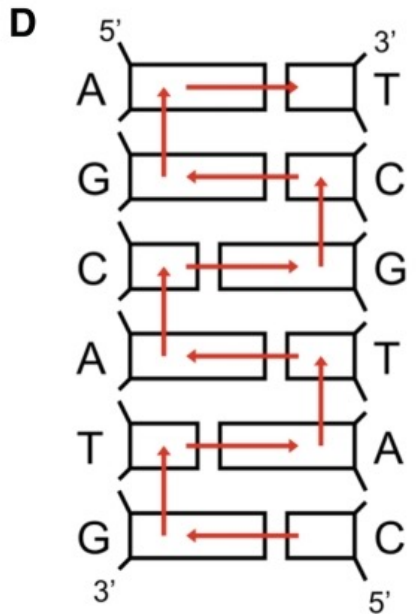




Interface moves sample the protein-DNA rigid body orientation using homologous structures as templates

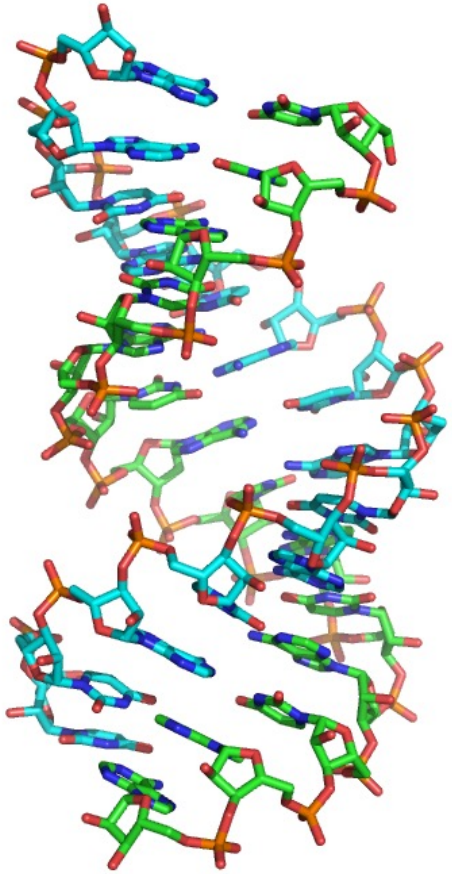
Protein fragment insertions sample backbone conformation without perturbing DNA or binding mode



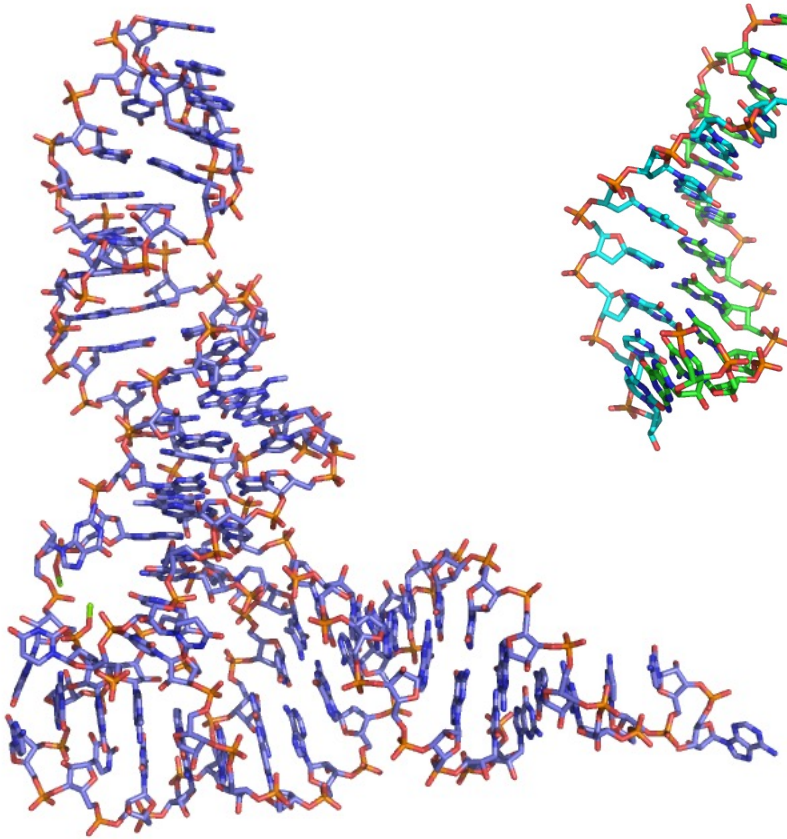


Kinematic structure for DNA allows torsion-space (internal coordinate) sampling while maintaining the DNA duplex

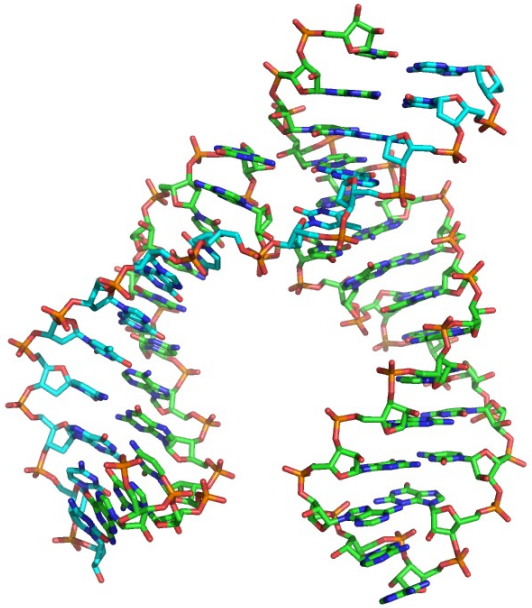
RNA structures are highly diverse



RNA duplex



transfer RNA



hammerhead ribozyme

6TNA,1HMH

Examples of RNA structural motifs

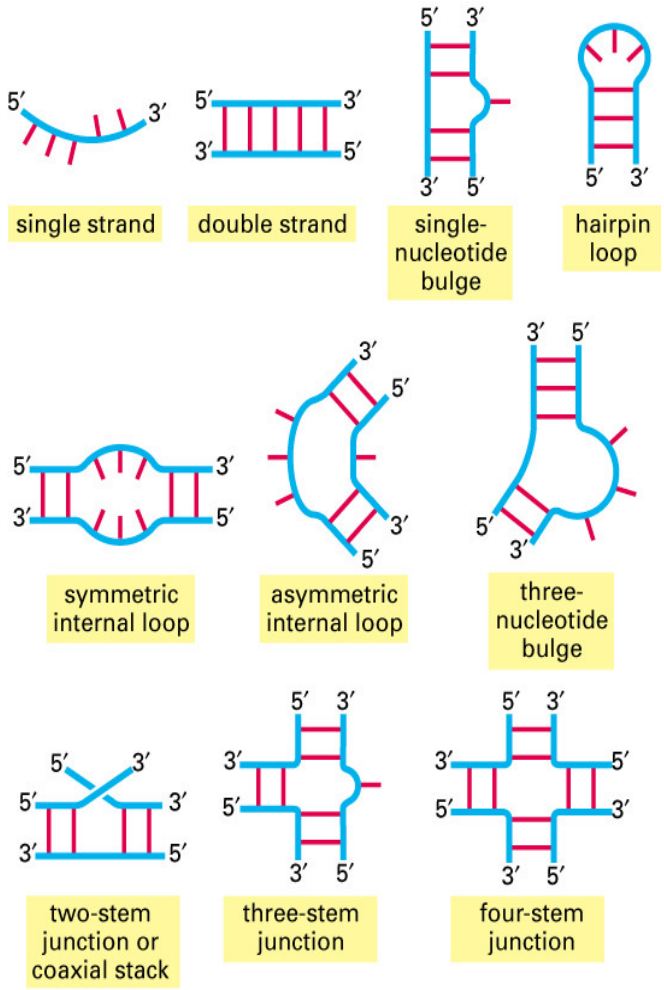


Figure 6–94. Molecular Biology of the Cell, 4th Edition.

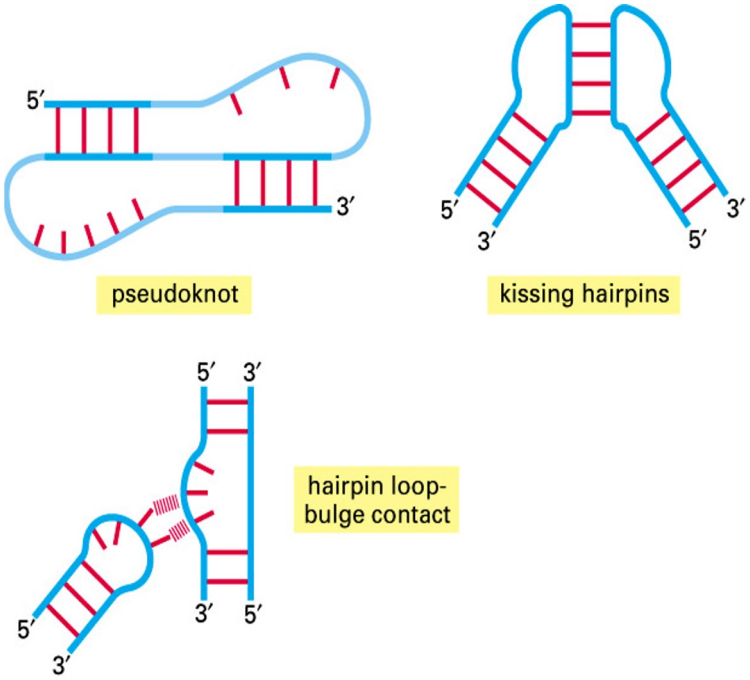
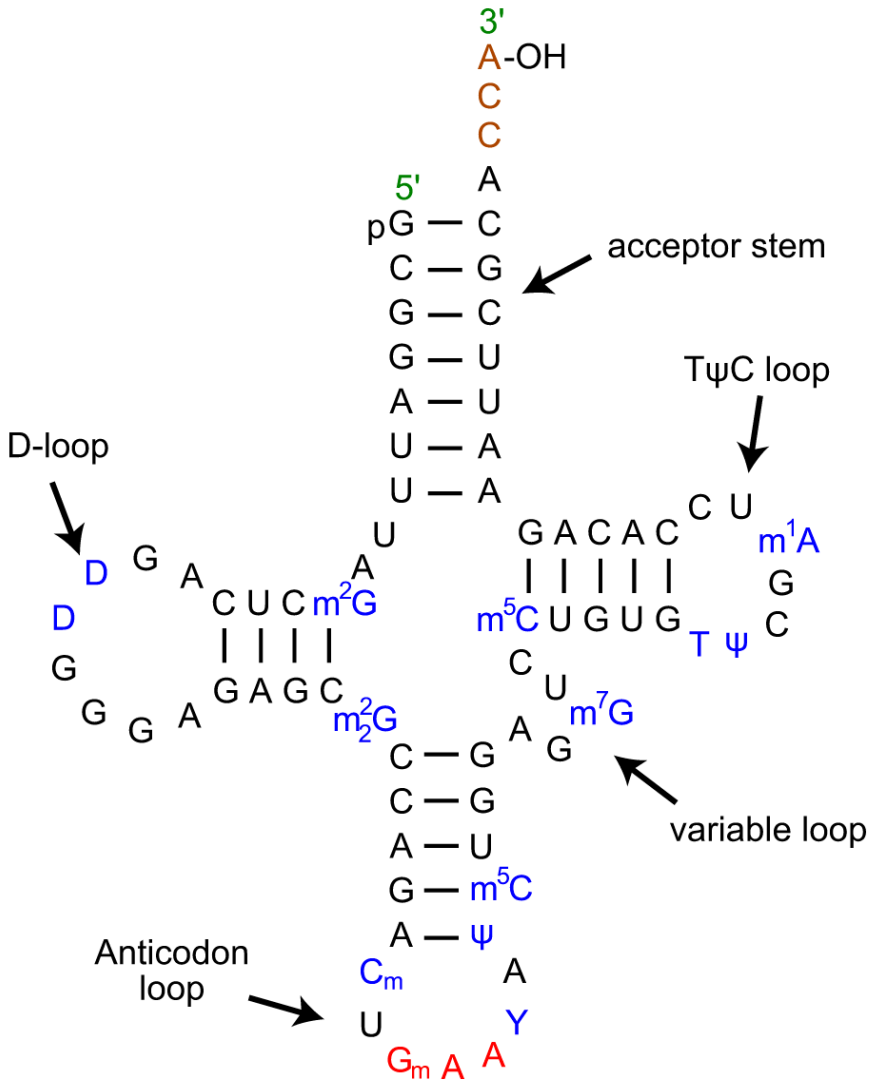
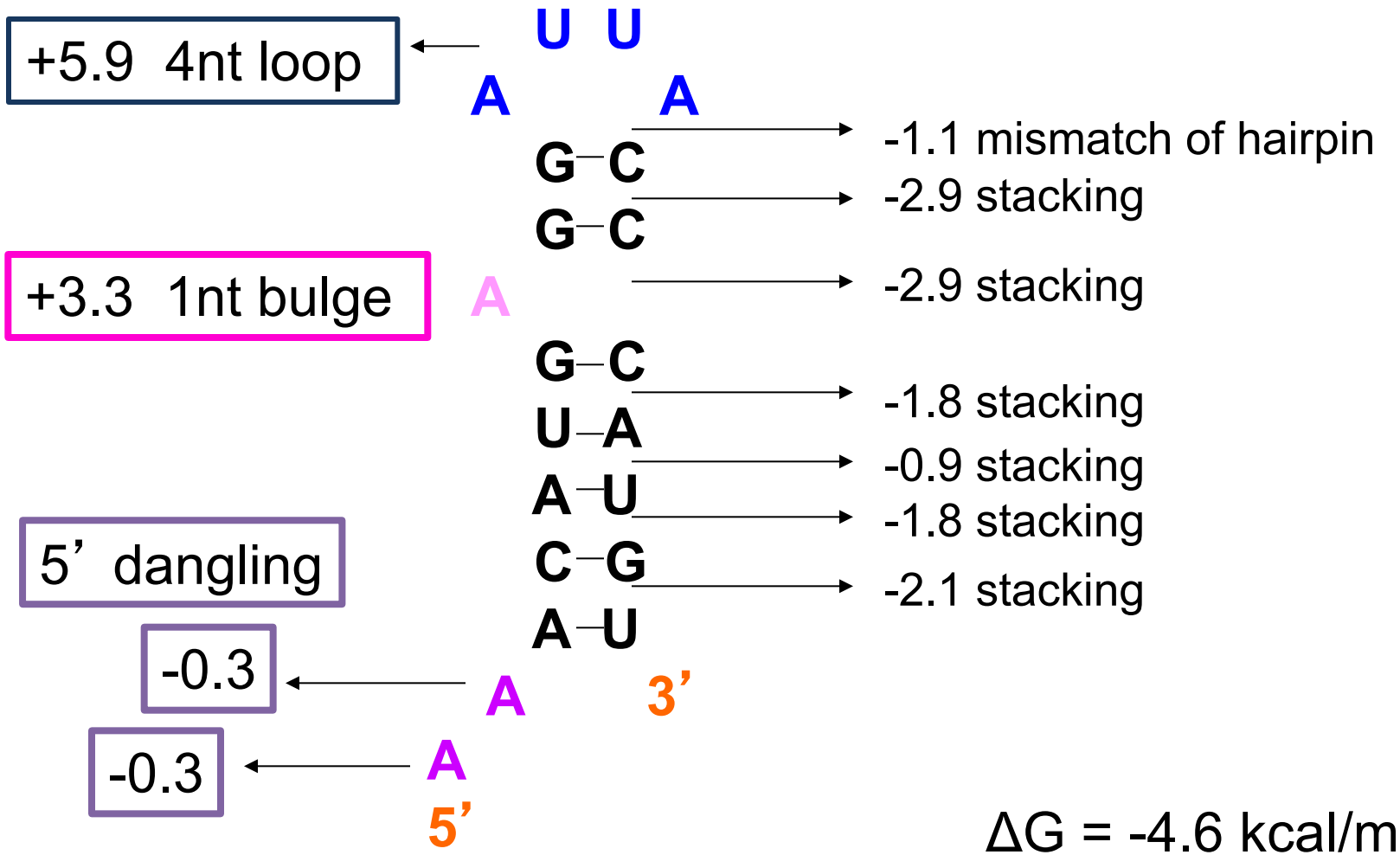


Figure 6–95. Molecular Biology of the Cell, 4th Edition.

Secondary structure of yeast Phe tRNA



Free energy computation predicts RNA secondary structure (mfold)



ΔG = -4.6 kcal/mol

Mfold algorithm

(Zuker & Stiegler, NAR 1981 9(1):133)

W(i,j) – min free energy formed from subsequence [i…j]

V(i,j) – min free energy from all substructures where I and j pair

$$V(i,j) = \min \begin{cases} E(FH(i,j)) & (1) \\ \min_{i<k<m<j} E(FL(i,j;k,m)) + V(k,m) & (2) \\ \min_{i+1<k<j-2} W(i+1,k) + W(k+1,j-1) & (3) \end{cases}$$

hairpin

stacking / bulge / interior loop

closed bifurcation (multiple loops)

$$W(i,j) = \min \begin{cases} W(i+1,j) & (4)a \\ W(i,j-1) & (4)b \\ V(i,j) & (1-3) \\ \min_{i<k<j-1}(W(i,k) + W(k+1,j) & (5) \end{cases}$$

i or j unpaired

i and j paired

open bifurcation

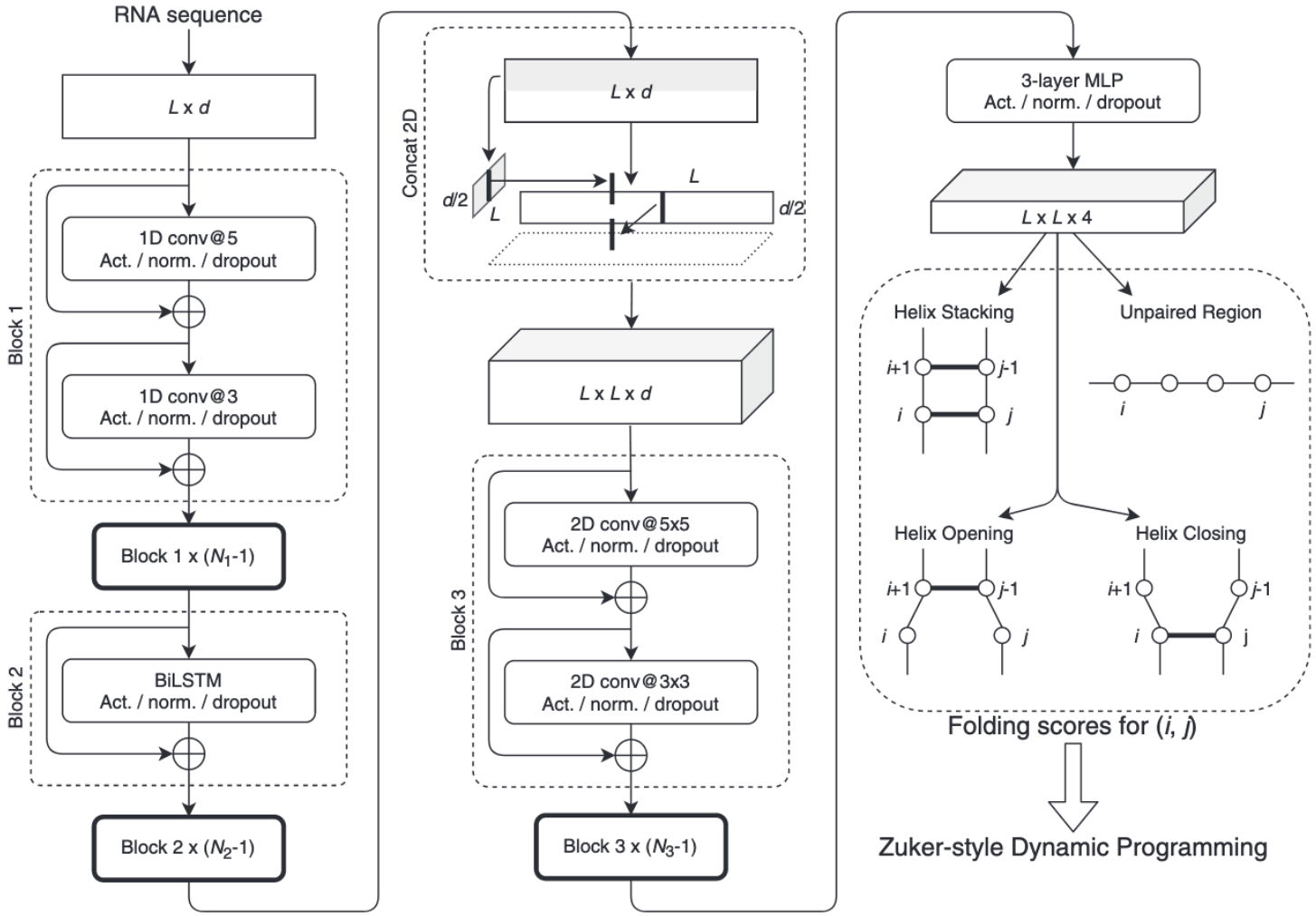
nature COMMUNICATIONS

ARTICLE

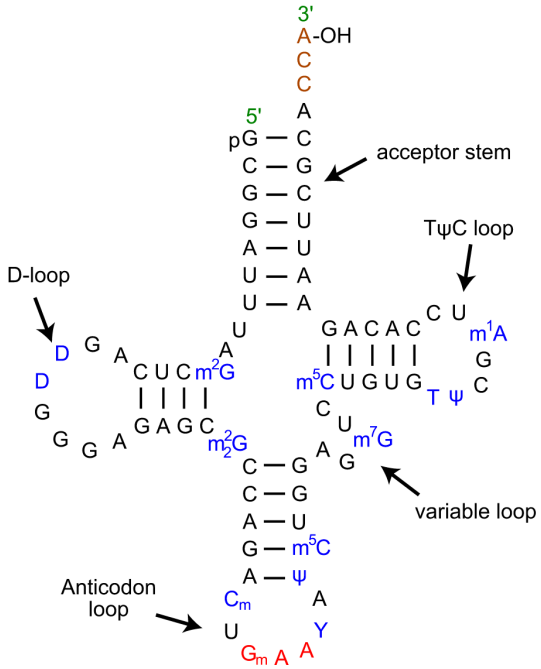
https://doi.org/10.1038/s41467-021-21194-4 OPEN

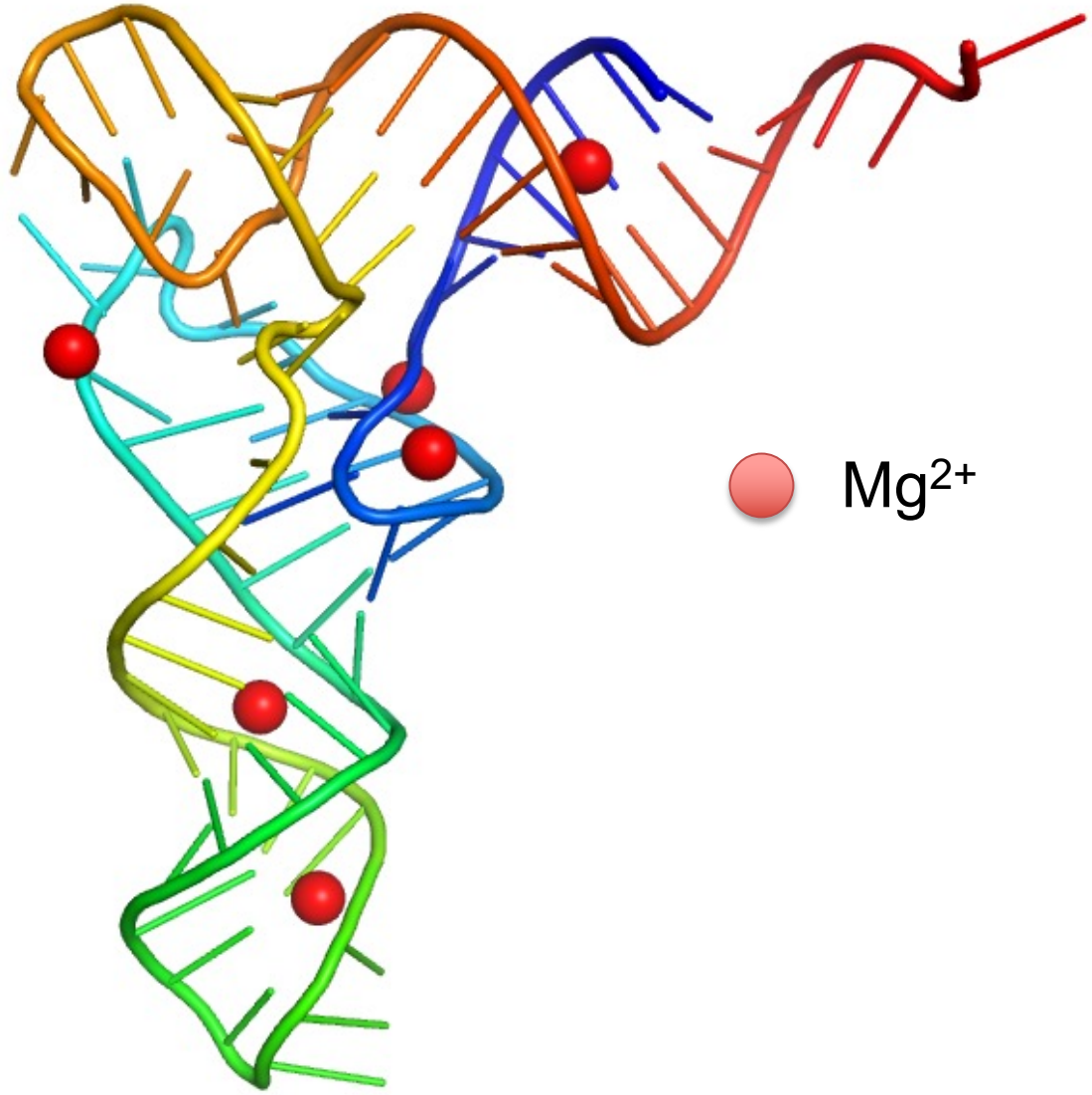
RNA secondary structure prediction using deep learning with thermodynamic integration

Kengo Sato [1✉], Manato Akiyama[1] & Yasubumi Sakakibara[1]

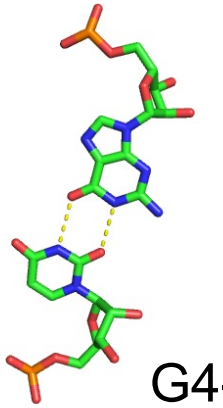


3D structure of yeast Phe tRNA fold

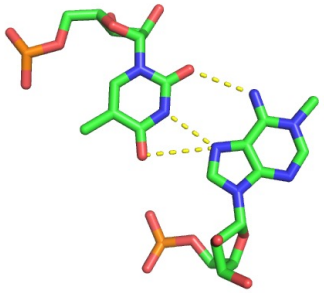




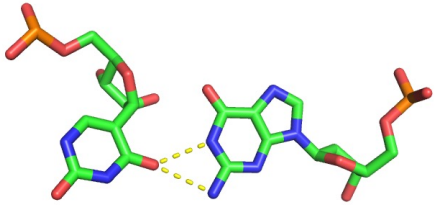
Non-WC base pairs and base triples in yeast tRNA Phe



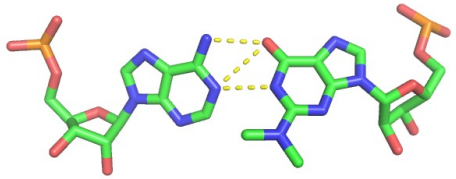
G4-U69



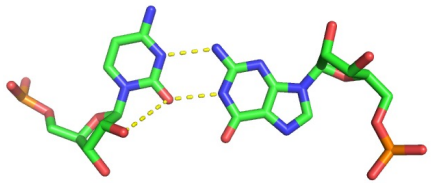
m^1A58-T54



G18-Ψ55

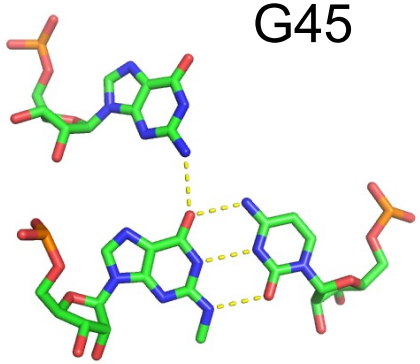


m^2G26-A44



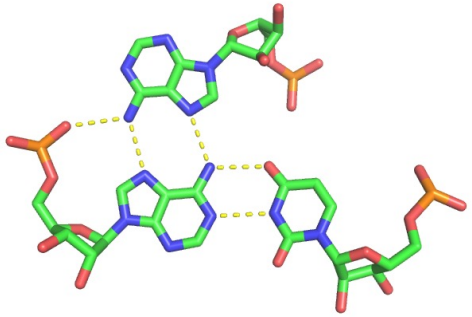
G15-C48

G45



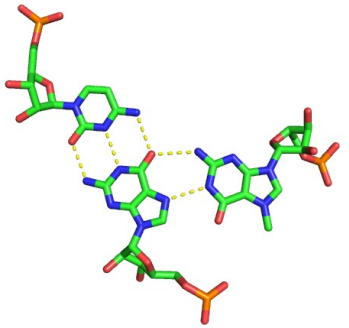
m^2G10-C25

A9



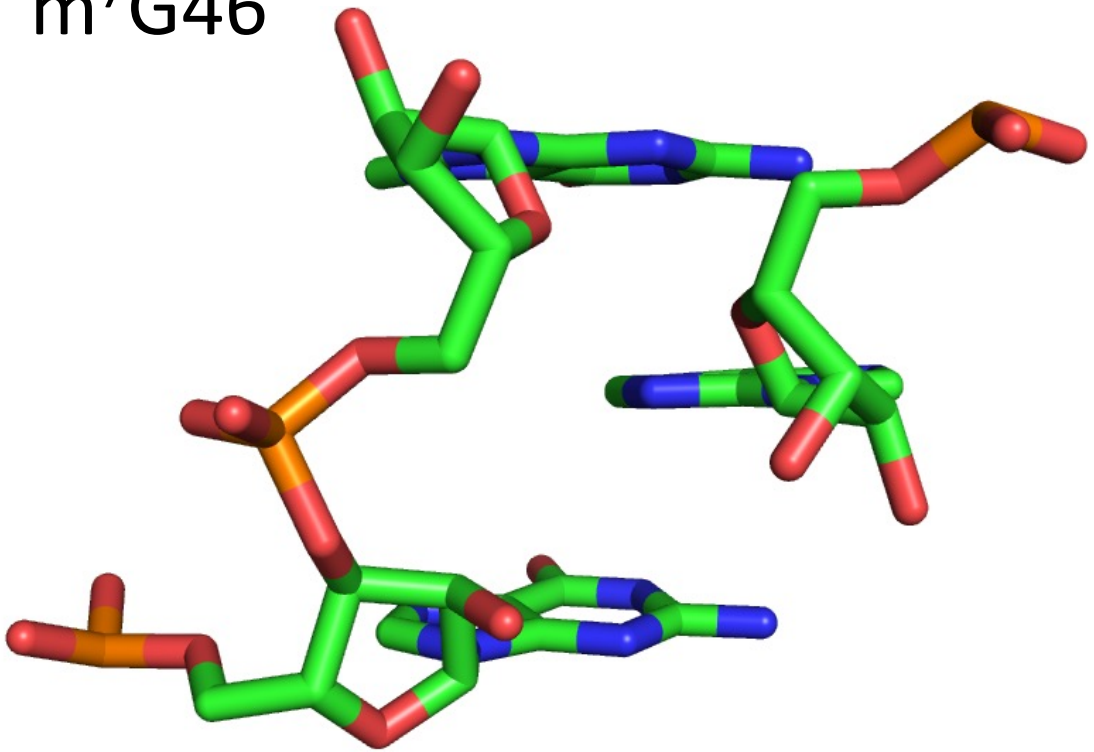
A23-U12

m^7G46

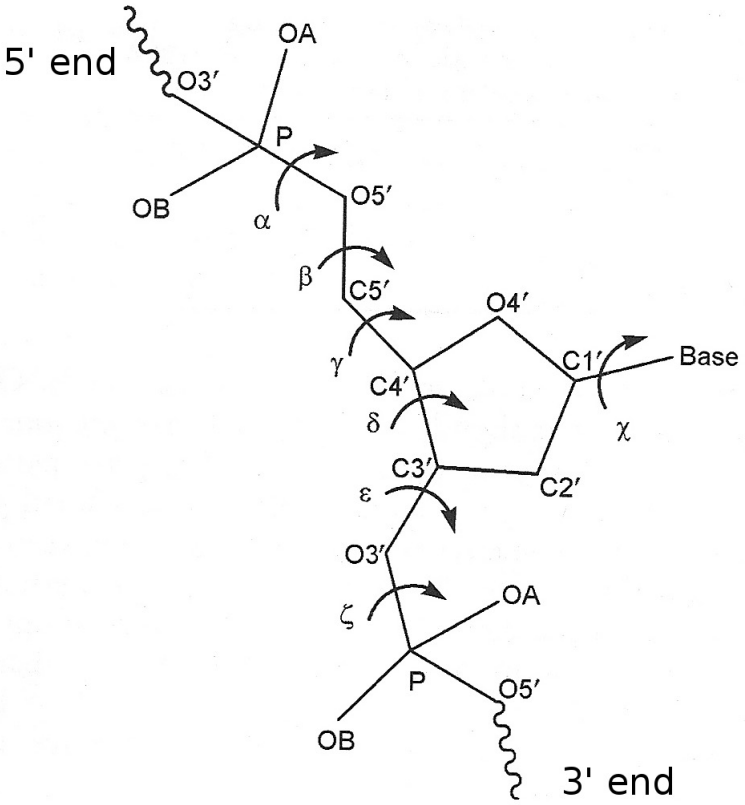


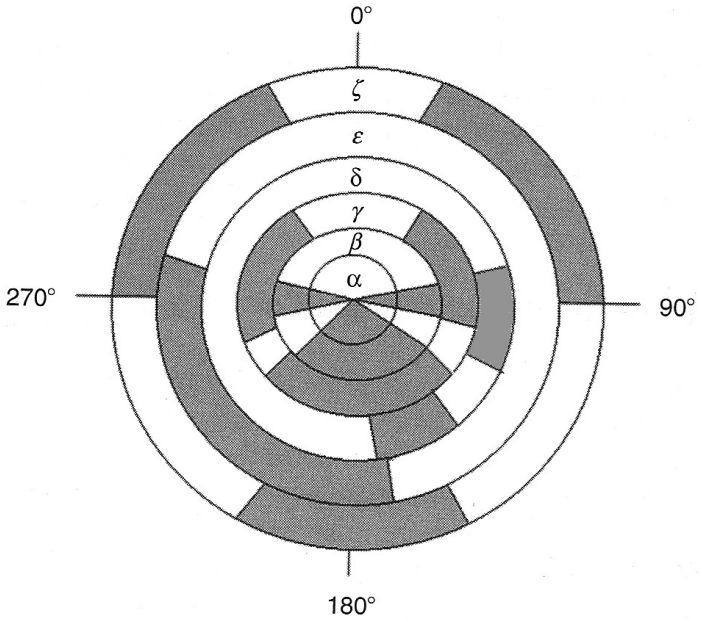
G22-C13

A9 intercalates between adjacent G45 and m^7G46 in yeast tRNA Phe

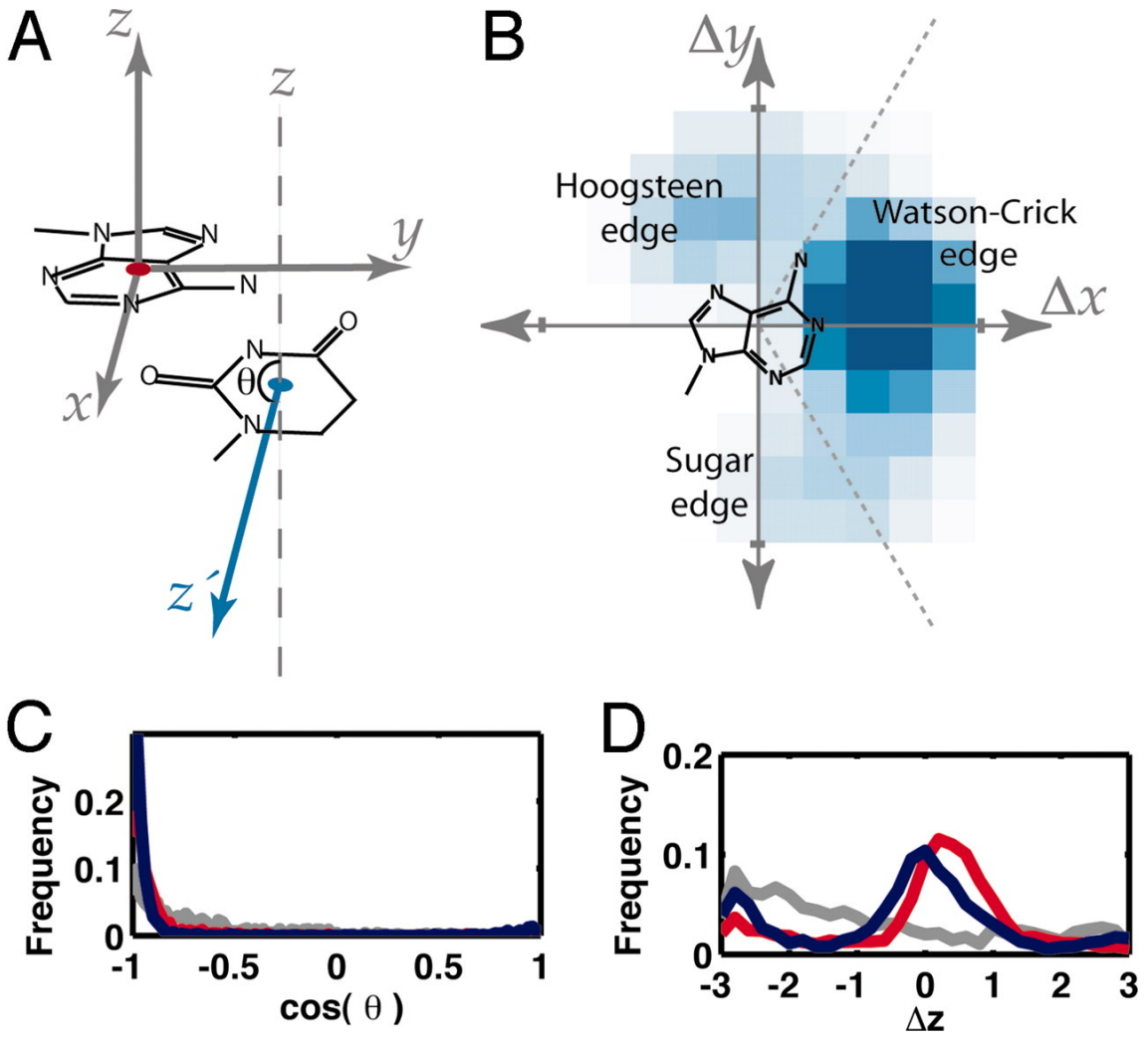


Six backbone dihedral angles (α–ζ) per nucleotide





Prediction of RNA tertiary structure



PNAS, 2009

De Novo RNA Tertiary Structure Prediction at Atomic Resolution Using Geometric Potentials from Deep Learning

Robin Pearce[a], Gilbert S. Omenn[a,c], Yang Zhang[a,b,*]

[a]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109 USA; [b]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA; [c]Departments of Internal Medicine and Human Genetics and School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA.

*To whom correspondence should be addressed. E-mail: zhng@umich.edu

ABSTRACT

Experimental characterization of RNA structure remains difficult, especially for non-coding RNAs that are critical to many cellular activities. We developed DeepFoldRNA to predict RNA structures from sequence alone by coupling deep self-attention neural networks with gradient-based folding simulations. The method was tested on two independent benchmark datasets from Rfam families and RNA-Puzzle experiments, where DeepFoldRNA constructed models with an average RMSD=2.69 Å and TM-score=0.743, which outperformed state-of-the-art methods and the best models submitted from the RNA-Puzzles community by a large margin. On average, DeepFoldRNA required ~1 minute to fold medium-sized RNAs, which was ~350-4000 times faster than the leading Monte Carlo simulation approaches. These results demonstrate the major advantage of advanced deep learning techniques to learn more accurate information from evolutionary profiles than knowledge-based potentials derived from simple statistics of the PDB library. The high speed and accuracy of the developed method should enable large-scale atomic-level RNA structure modeling applications.