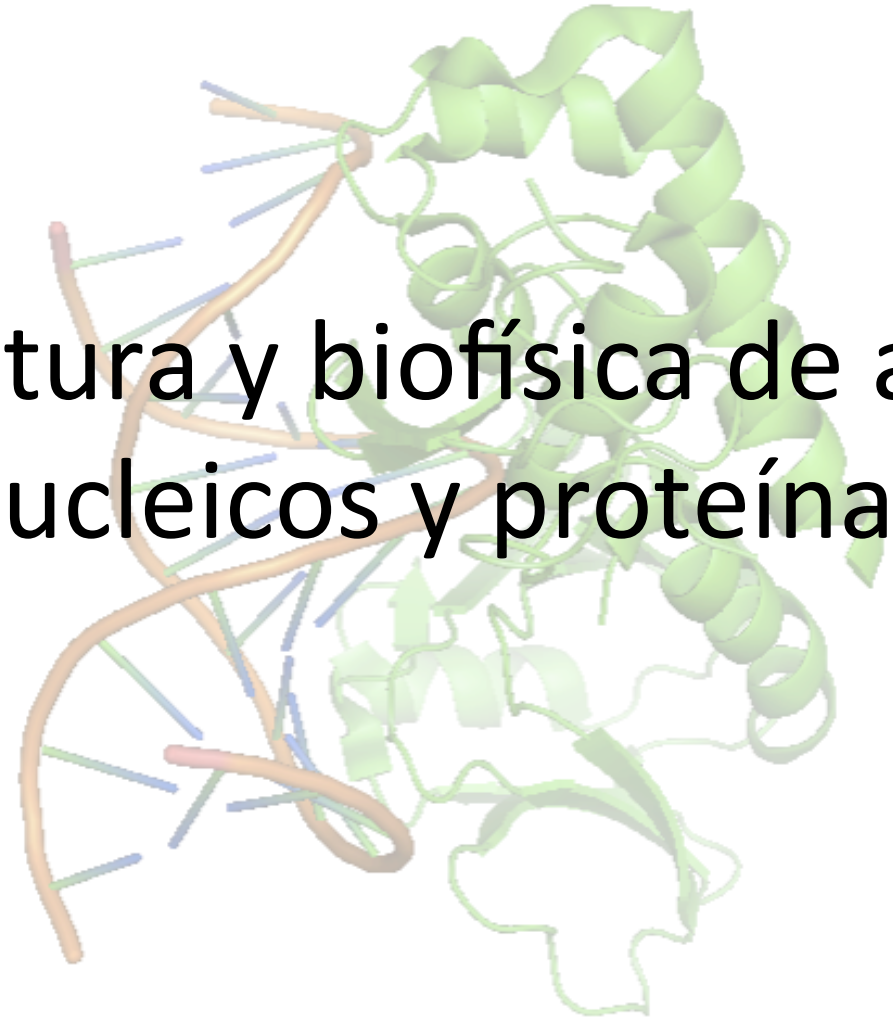


Estructura y biofísica de ácidos nucleicos y proteínas

David Dufour Rausell
ddufour@pcb.ub.cat

Summary

Proteins

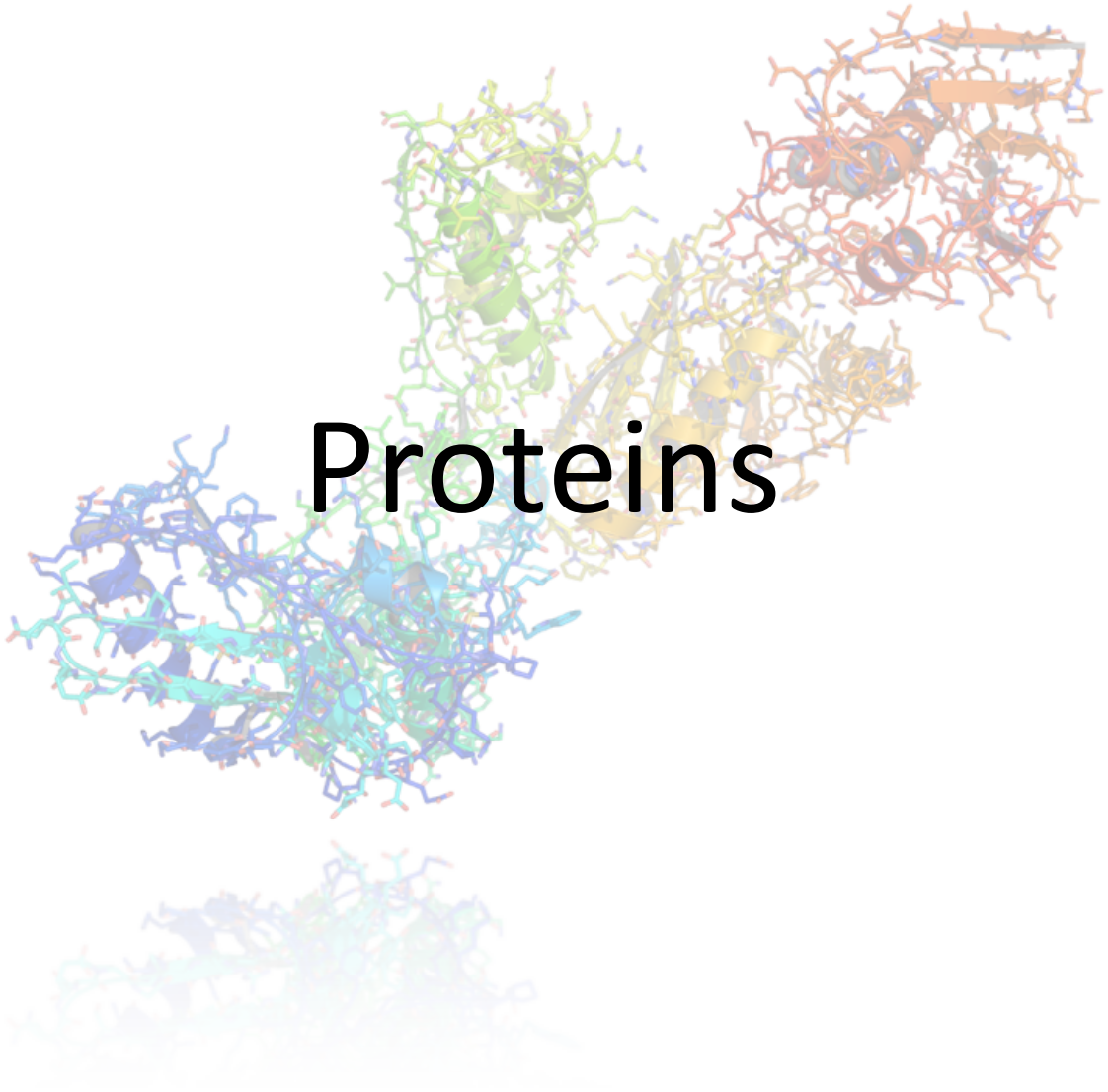
- Primary to quaternary structure
- Peptide Bond
- Amino acids
- Torsion angles
- Ramachandran plot
- Secondary structures: α-helix, β-sheets
- Fold space
- SCOP, CATH, PDB

Summary

Nucleic Acids (DNA, RNA)

- Bases, sugar, phosphate
- Phosphodiester bond
- Numbering system
- Sugar Puckering
- Orientation around glycosidic angle
- C4'-C5' torsion angle
- RNA rotamers
- Base pairs
- Triples
- Helical parameters
- Helical grooves
- Secondary structures in RNA
- Forces that drive RNA folding
- NDB

Proteins



Nomenclature

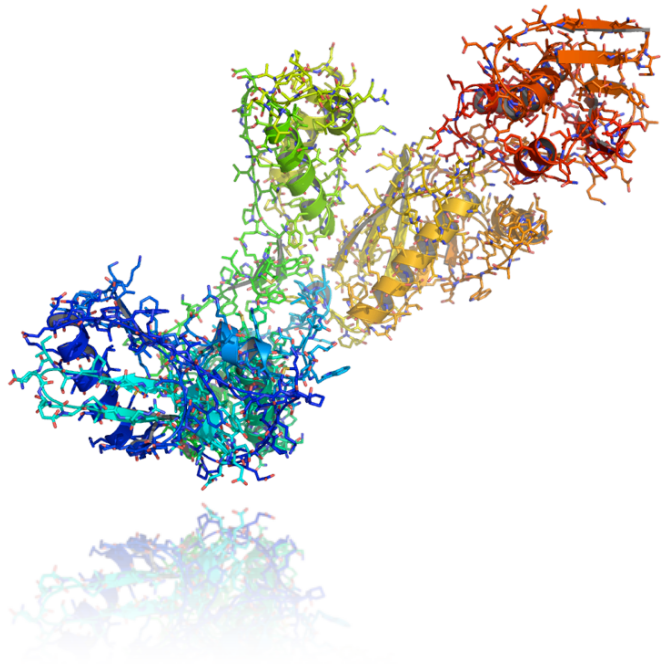
Fold: Three dimensional conformation of a protein sequence (usually at domain level).

Domain: Structurally globular part of a protein, which may independently fold.

Secondary Structure: Regular sub-domain structures composed by alpha-helices, beta-sheets and coils (or loops).

Backbone: Protein structure skeleton composed by the carbon, nitrogen and oxygen atoms.

Side-Chain: Specific atoms identifying each of the 20 residues types.

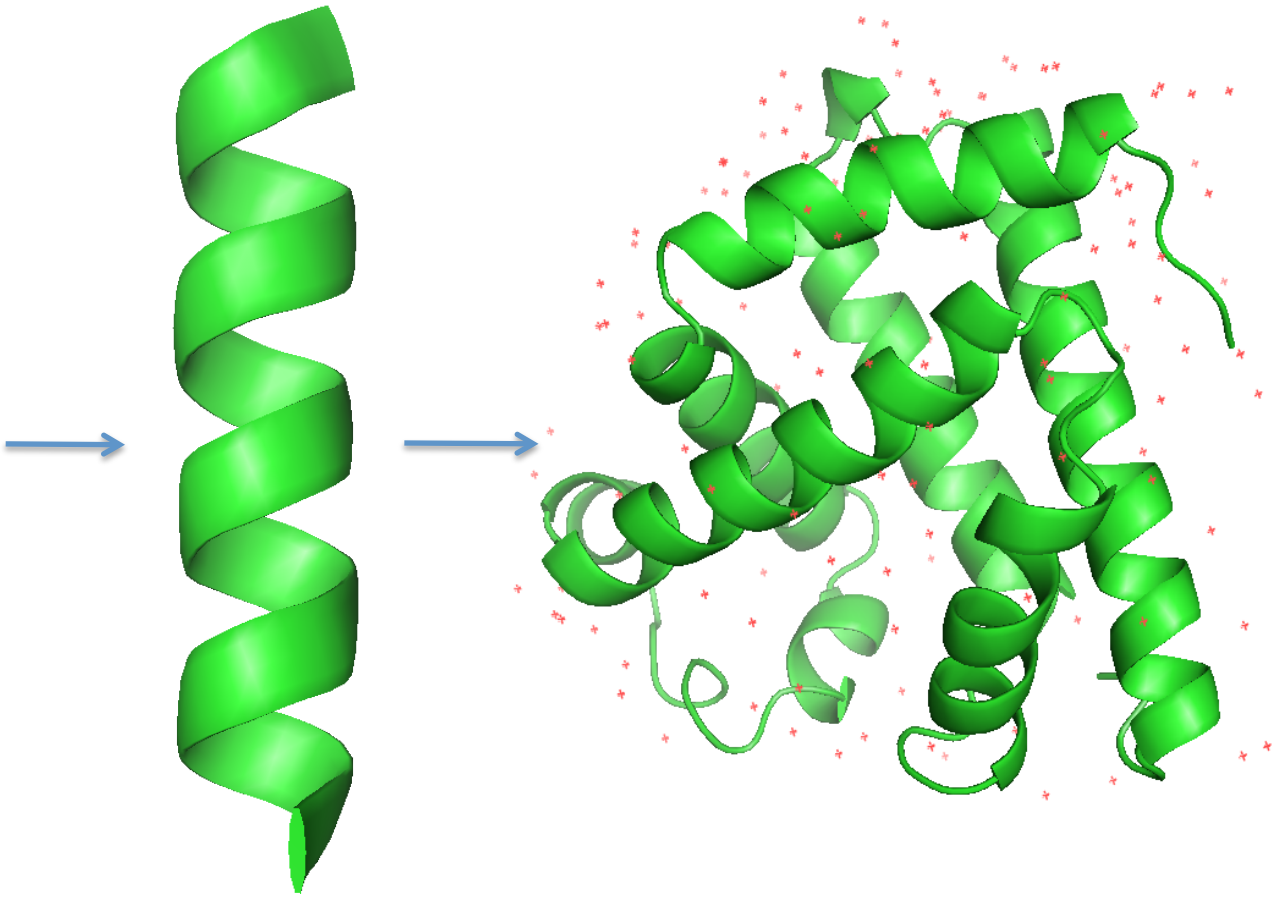


Primary structure

Secondary structure

Tertiary structure

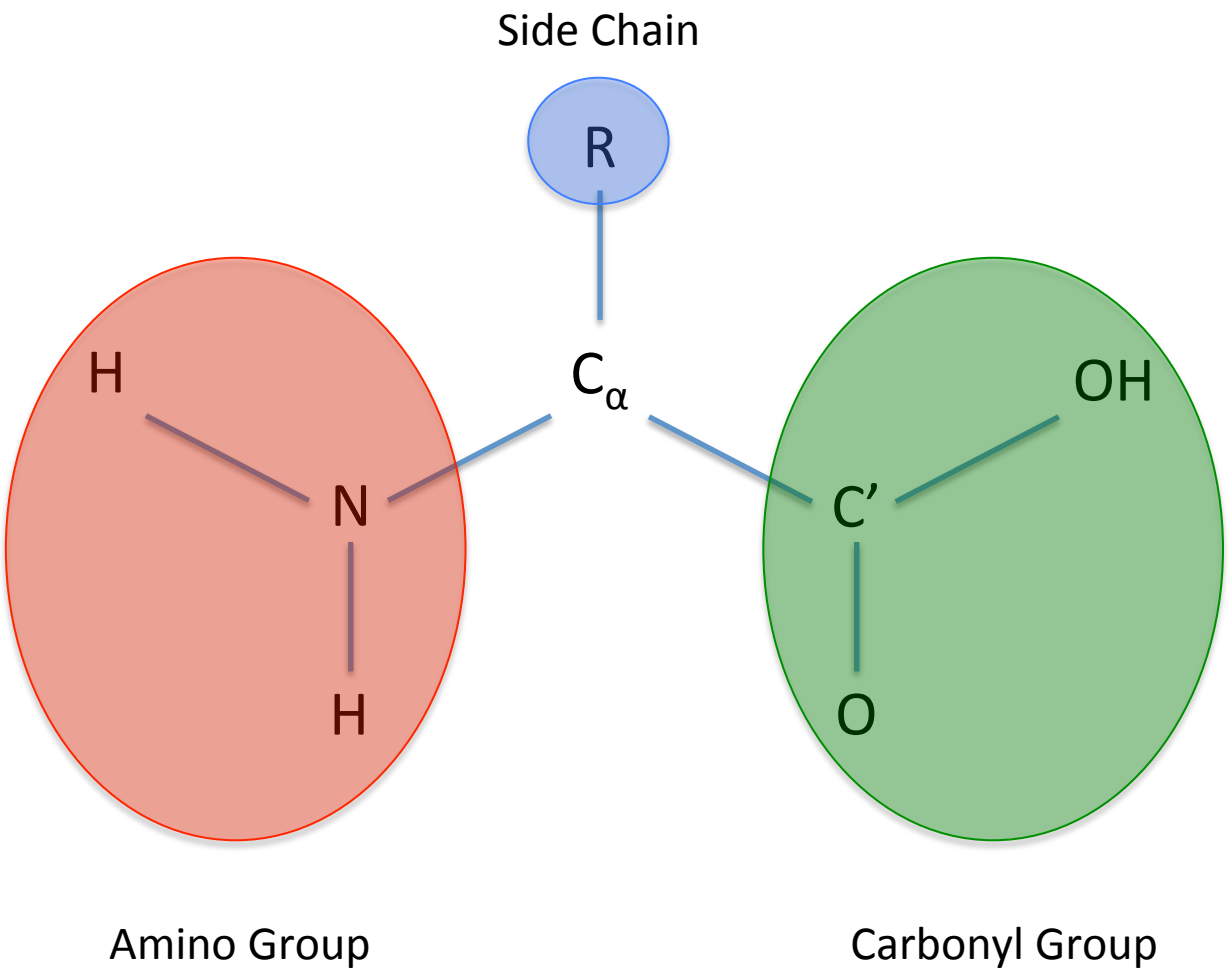
MVLSEGEWQLVLHVWAKVEA
DVAGHGQDILIRLFKSHPETLEK
FDRFKHLKTEAEMKASEDLKKH
GVTVLTALGAILKKKGHHEAELK
PLAQSHATKHKIPIKYLEFISEAII
HVLHSRHPGNFGADAQGAMN
KALELFRKDIAAKYKELGYQG



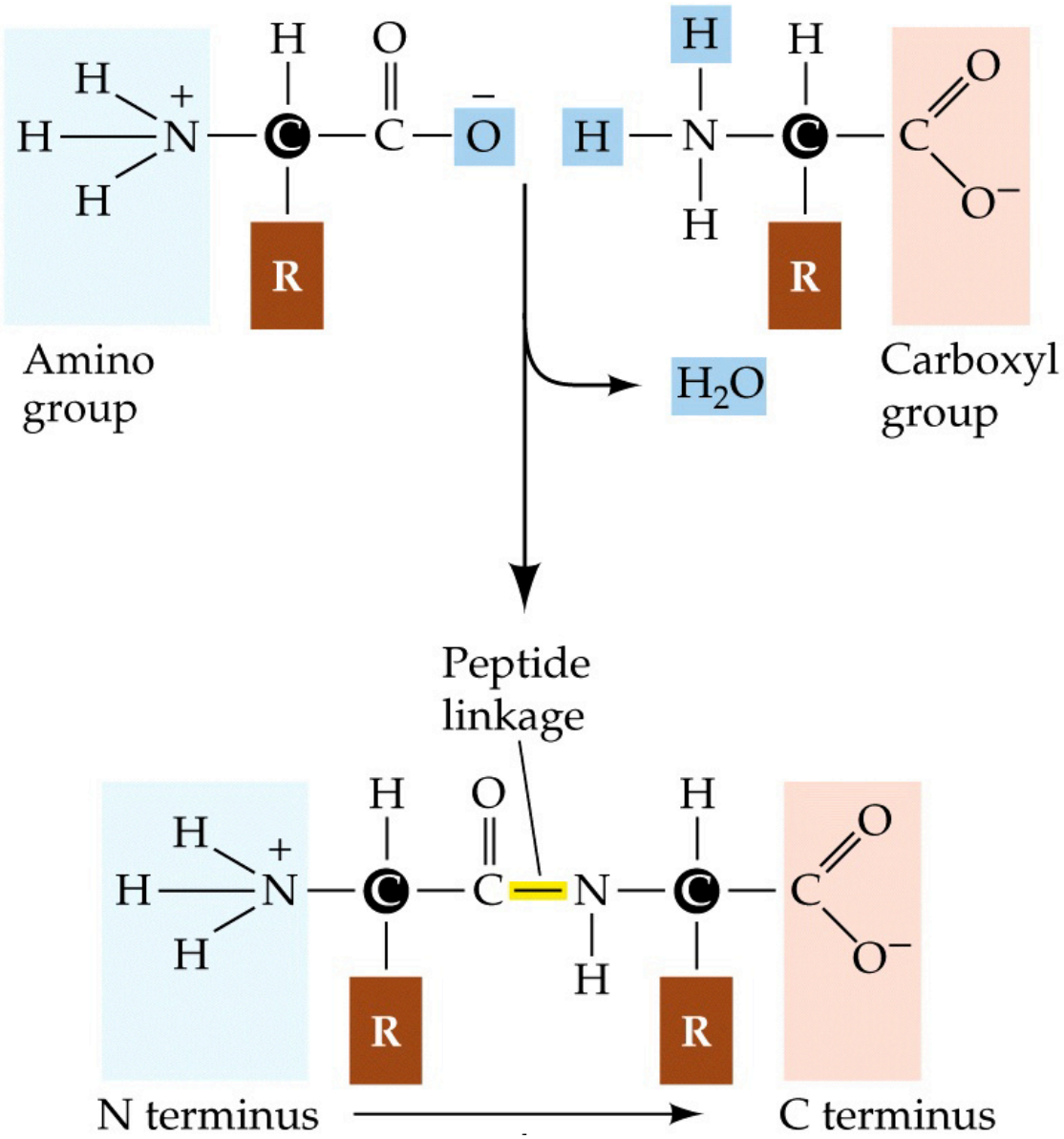
(Myoglobin)

Quaternary structure

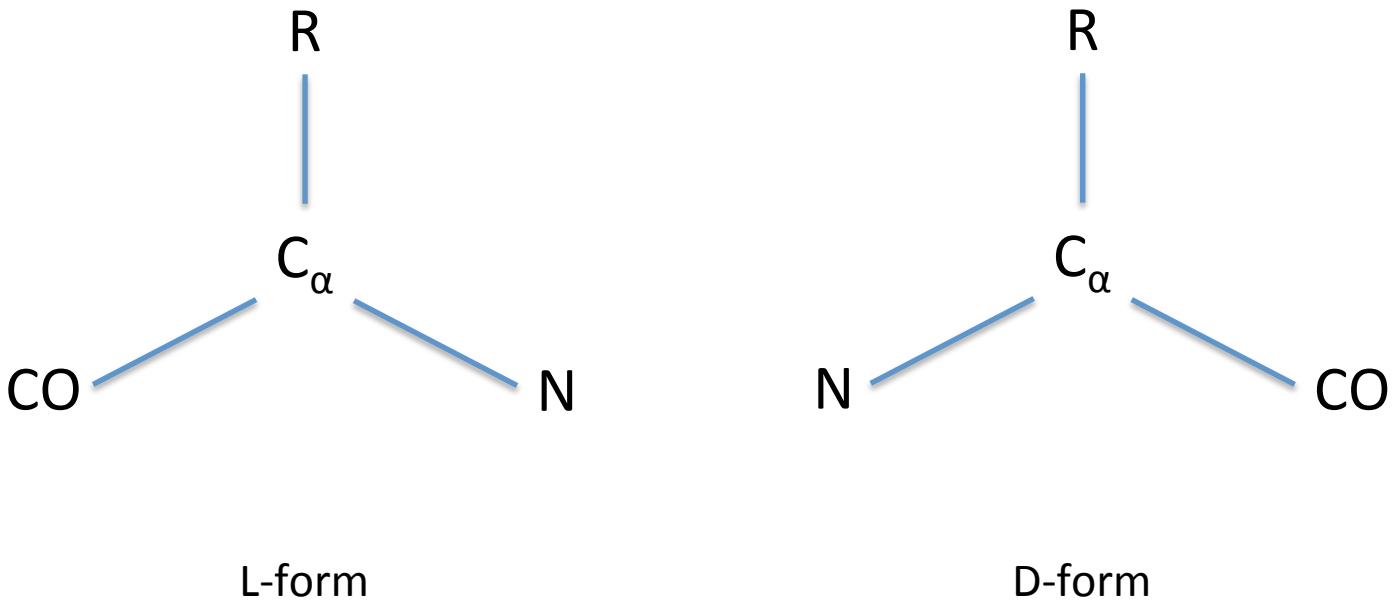
Proteins are polypeptide chains



Peptide Bond



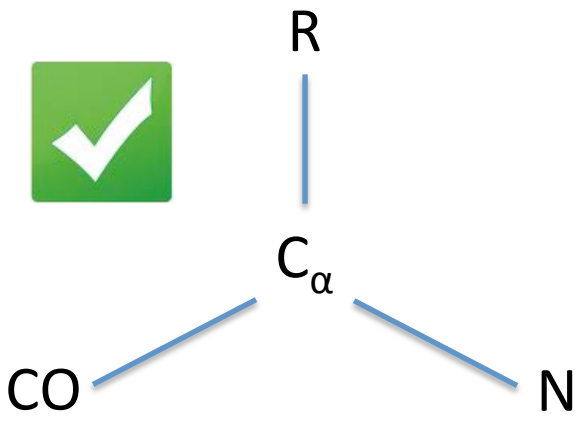
L- and D- forms



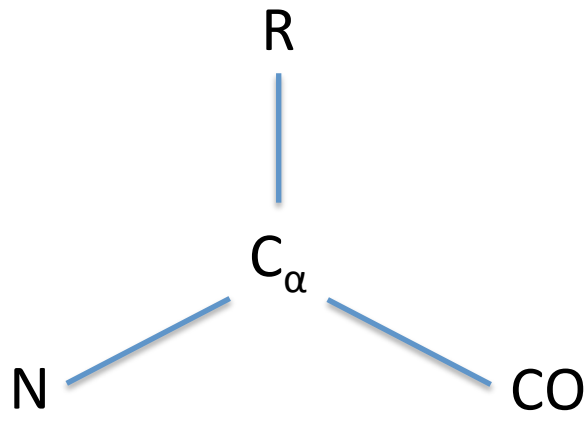
L-form

D-form

L- and D- forms

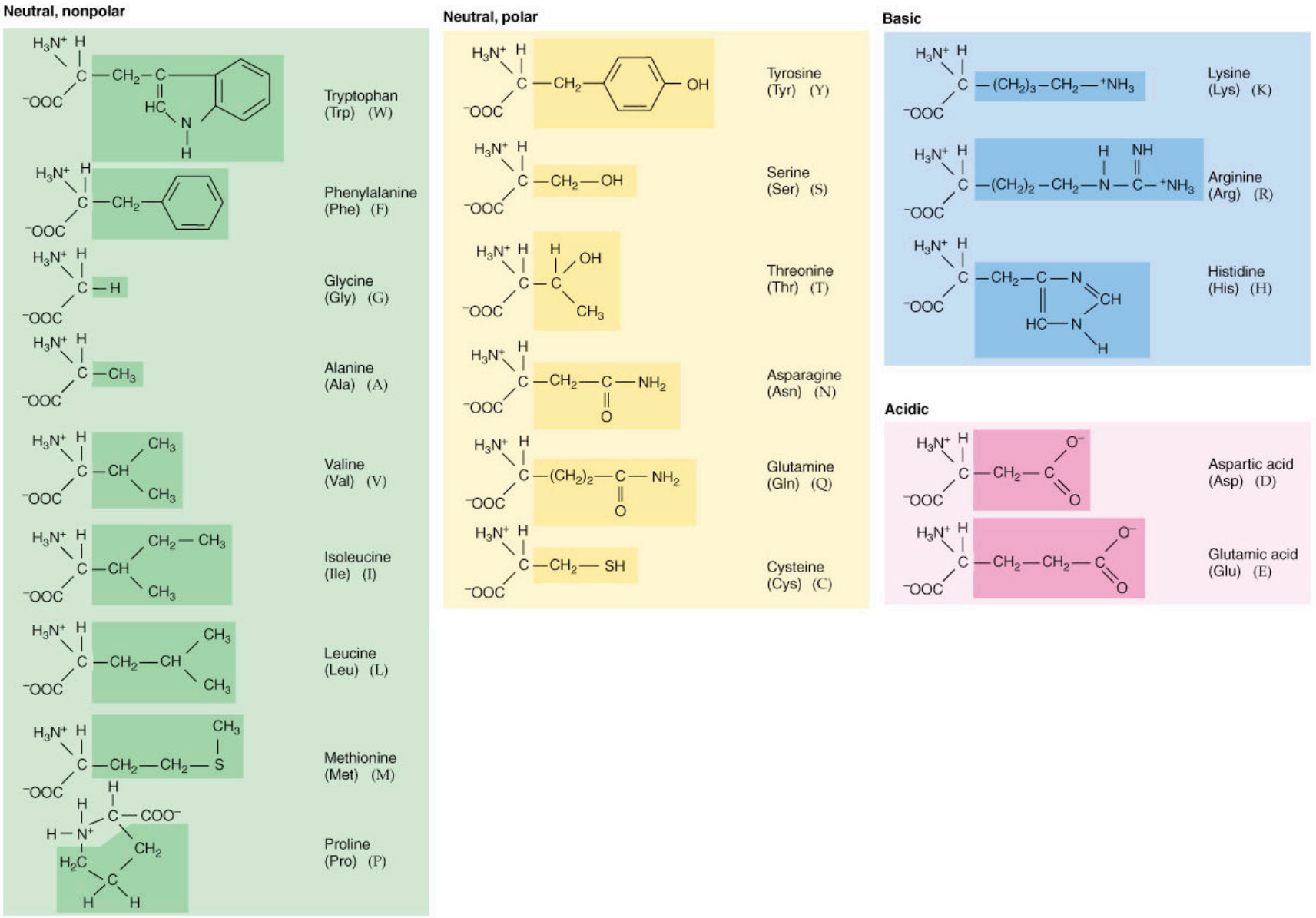


L-form



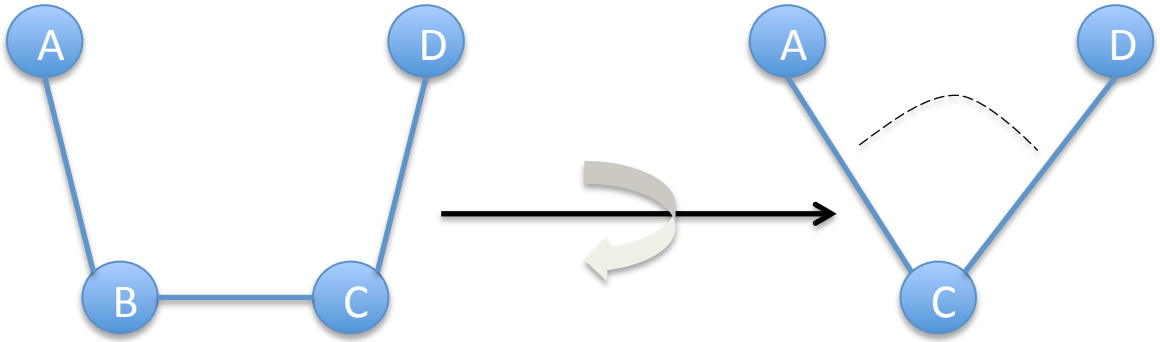
D-form

The 20 amino acids

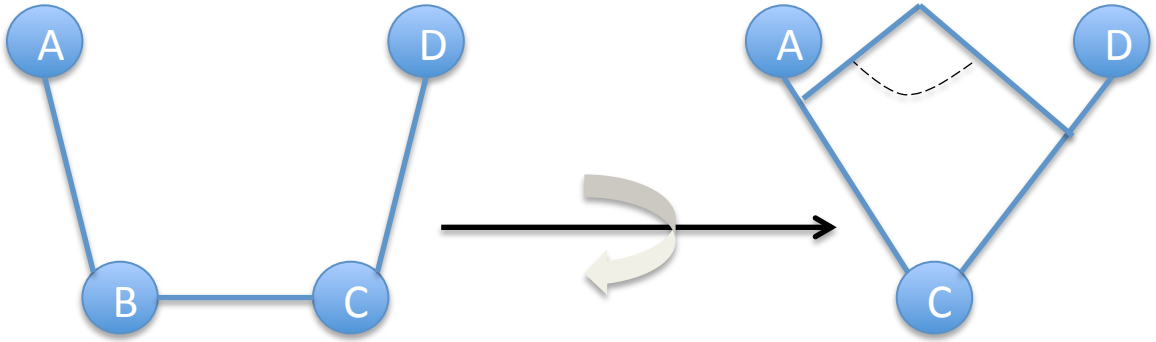


Torsion and dihedral angles

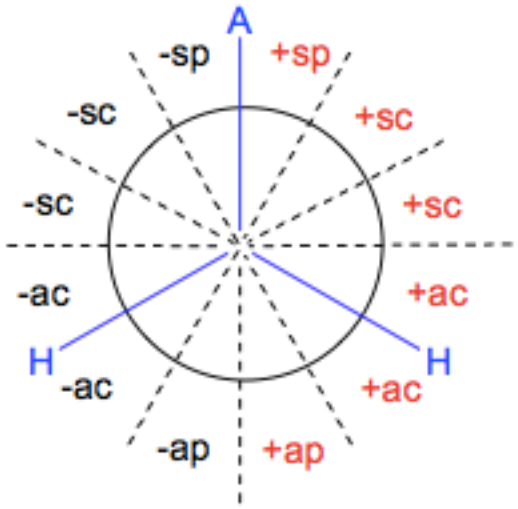
Torsion angle



Dihedral angle

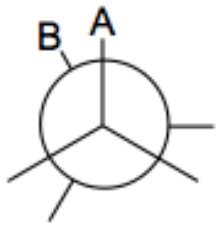


Torsion angle classification

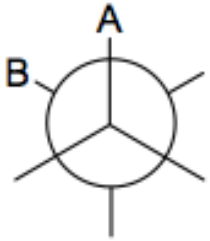


Klyne-Prelog system

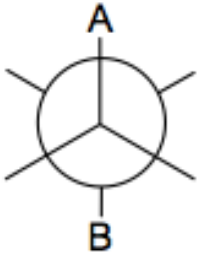
s: syn
p: periplanar
a: anti
c: clinal



Cis

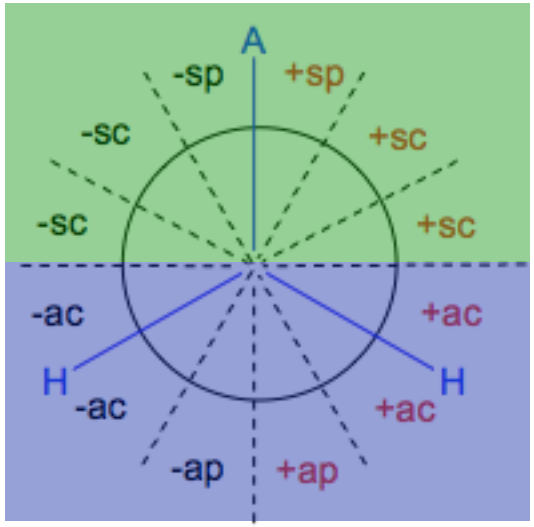


Gauche



Trans

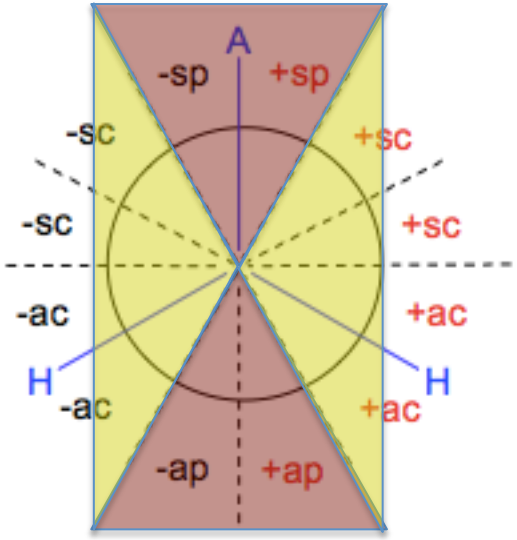
Torsion angle classification



s: syn
p: periplanar
a: anti
c: clinal

Klyne-Prelog system

Torsion angle classification



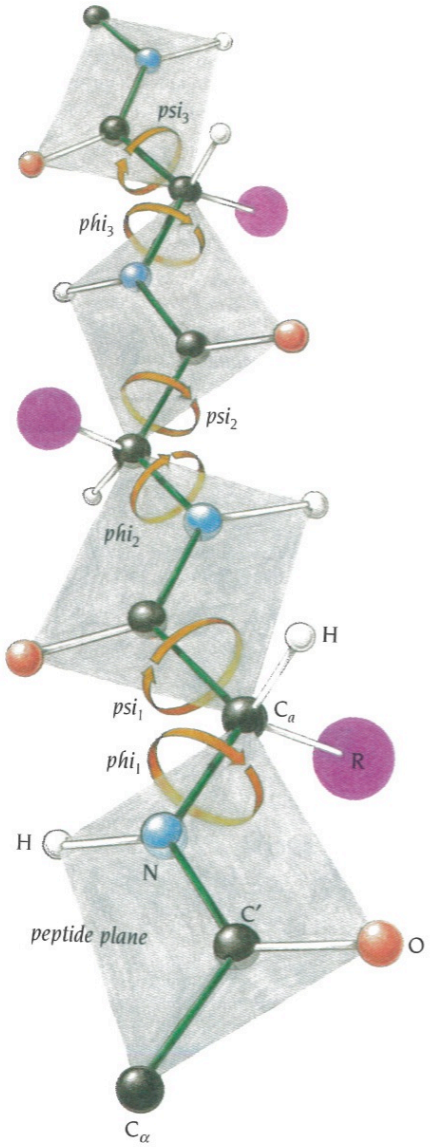
Klyne-Prelog system

s: syn
p: periplanar
a: anti
c: clinal

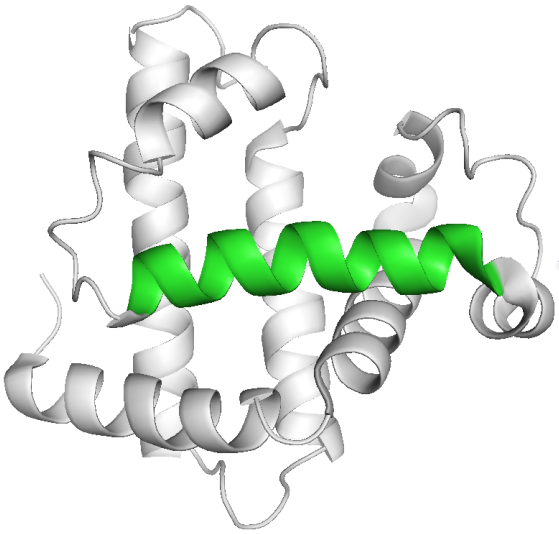
Peptide units = suits

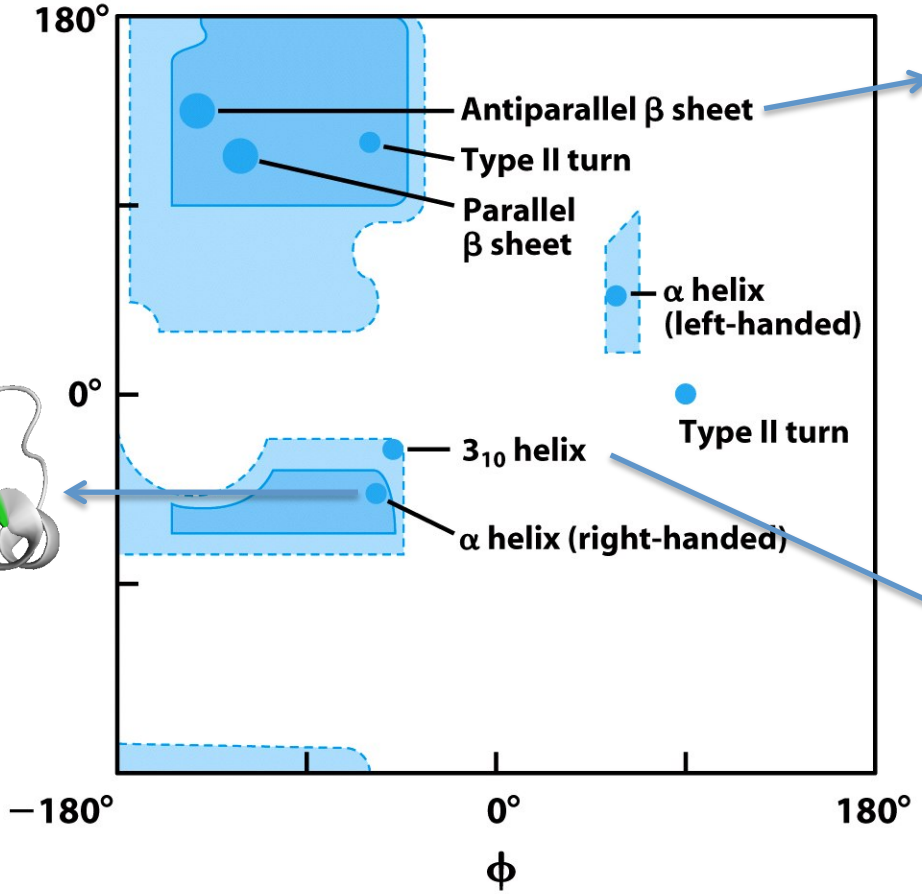
Phi (φ) = C'-C_α-N-C'$^{+1}$,
rotation around C_α-N

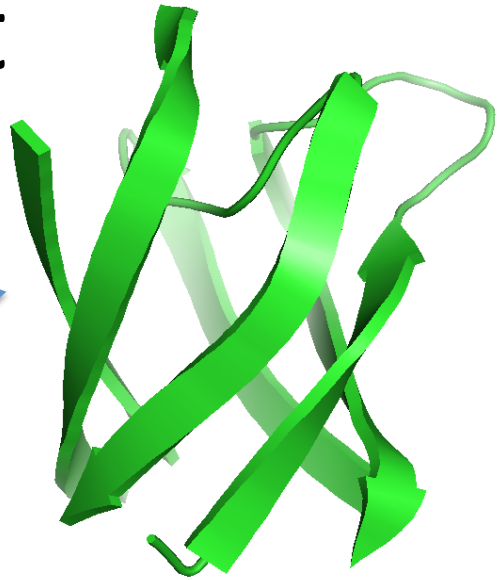
Psi (Ψ) = N-C'-C_α-N^{+1},
rotation around C'-C_α



Ramachandran plot







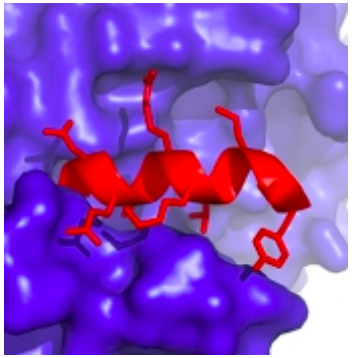
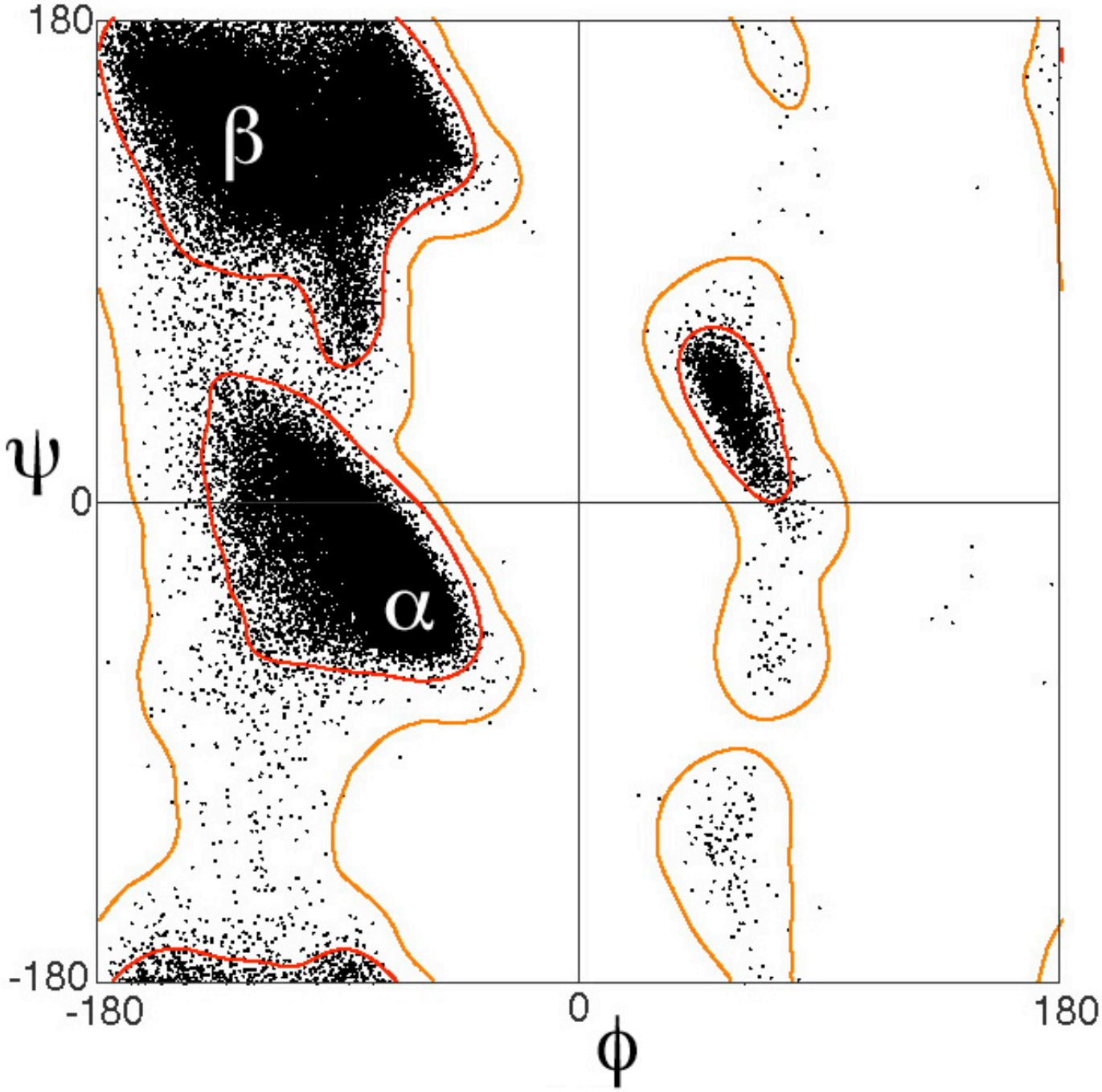
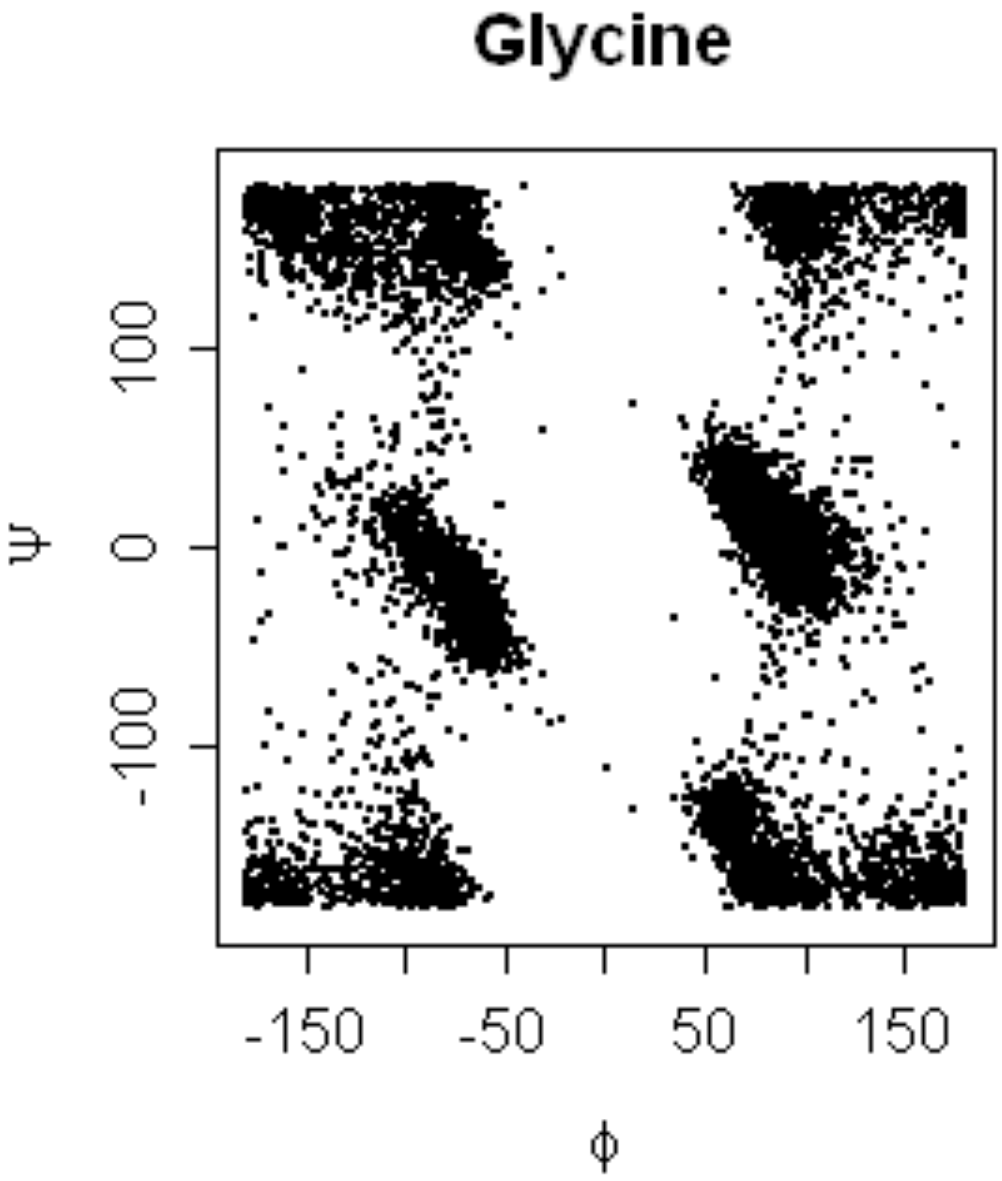
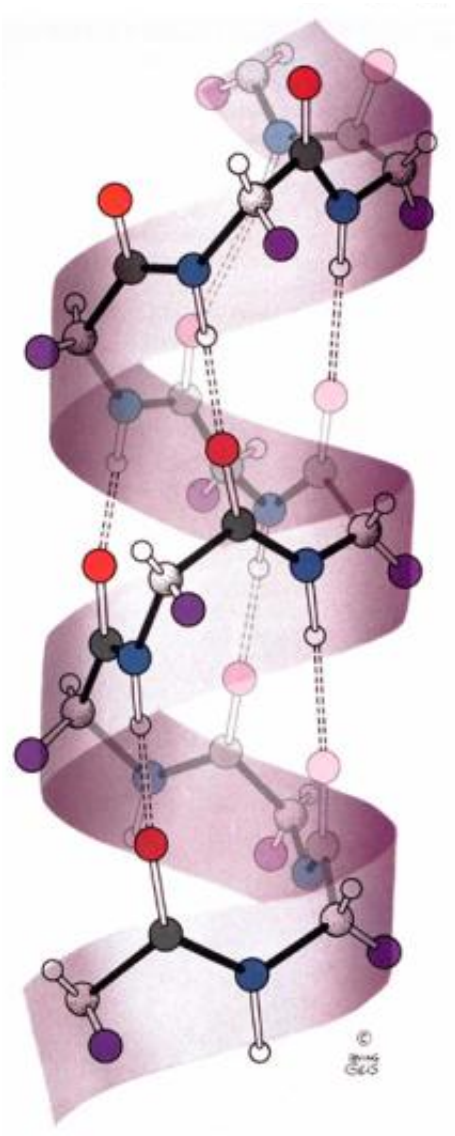


Figure 4-9a Principles of Biochemistry, 4/e

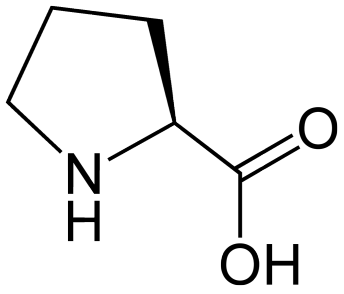

180
β
ψ
0
α
-180
-180
0
180
ϕ


Glycine
ψ
100
0
-100
-150
-50
50
150
ϕ

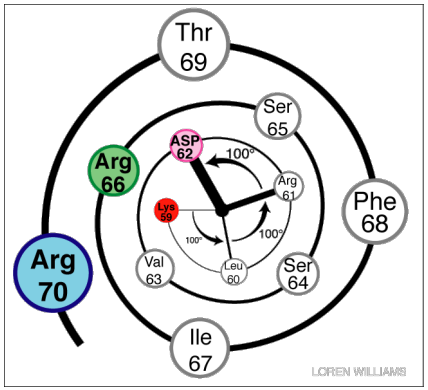
α-helix



- Backbone angles:
  - ψ = -50º
  - ϕ = -60º
- Dipole moment
- Proline bending

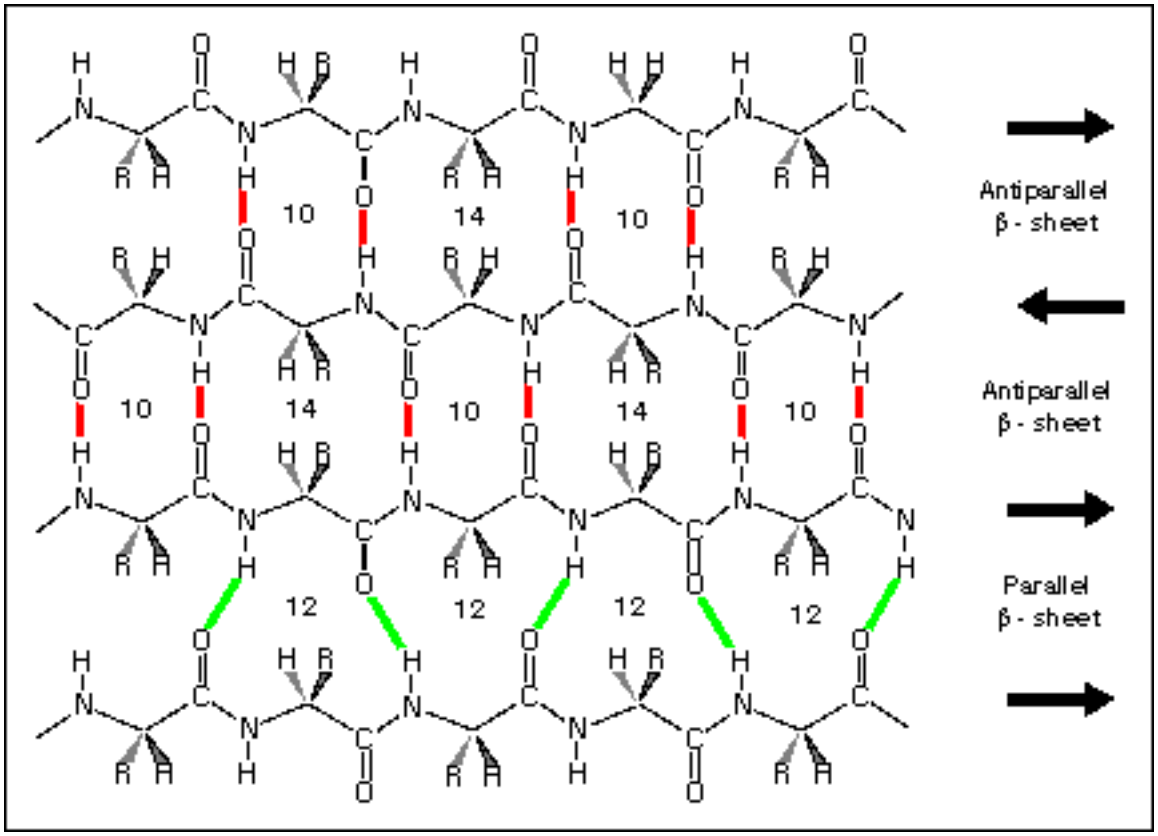


Proline

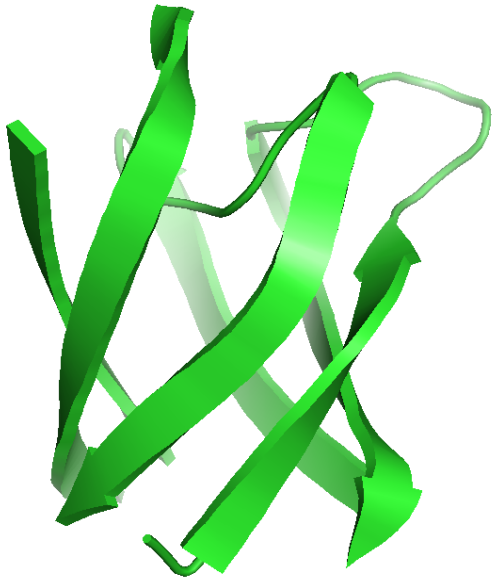


Helical wheel projection

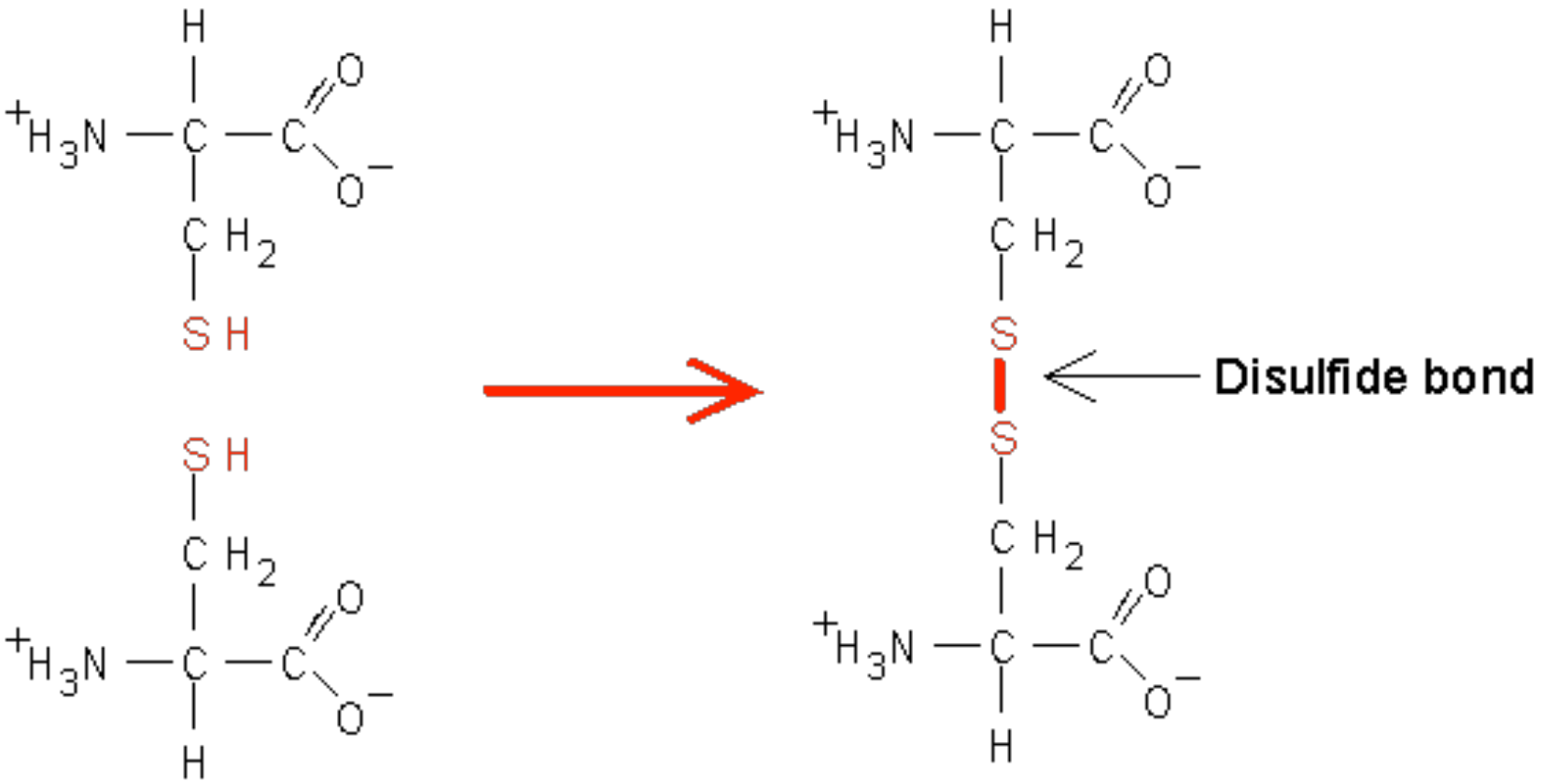
β-sheet



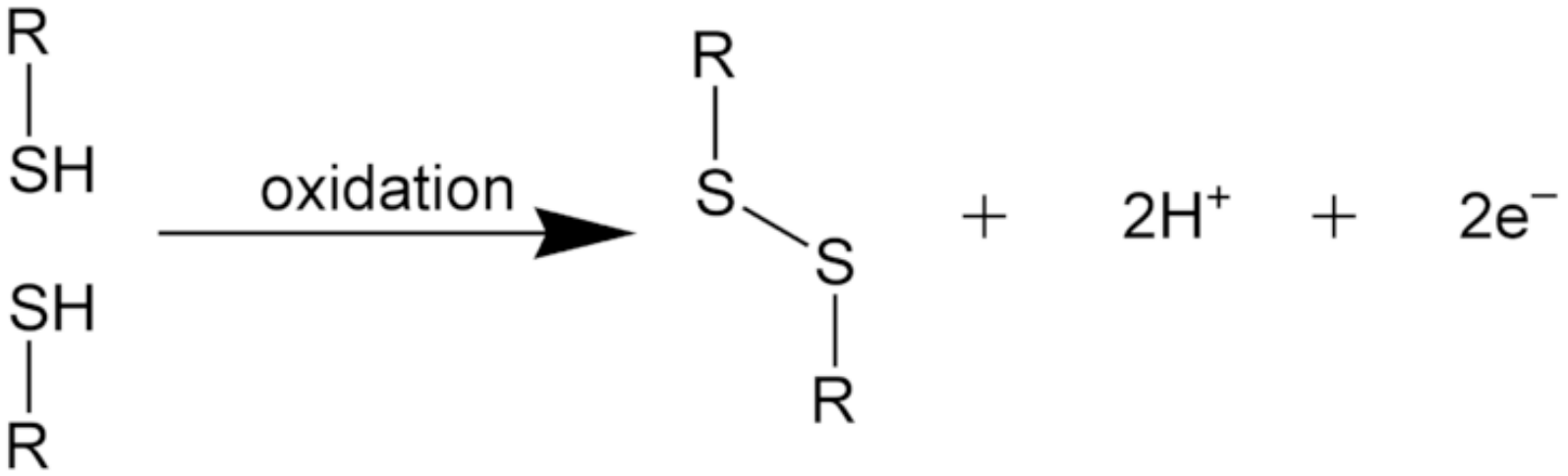
- First quadrant in Ramachandran's plot
- Parallel or antiparallel



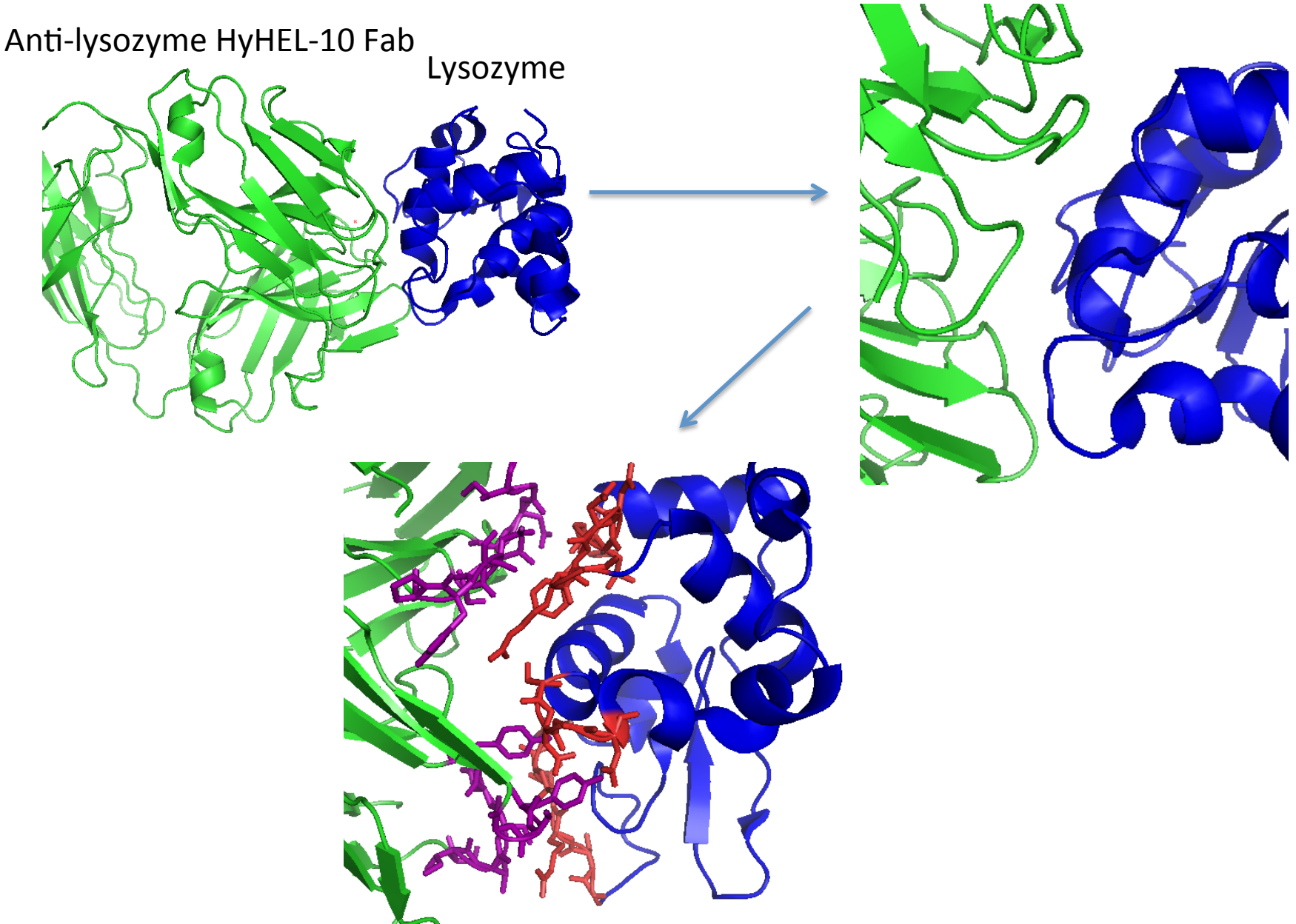
Cysteine disulfide bond



Cysteine disulfide bond

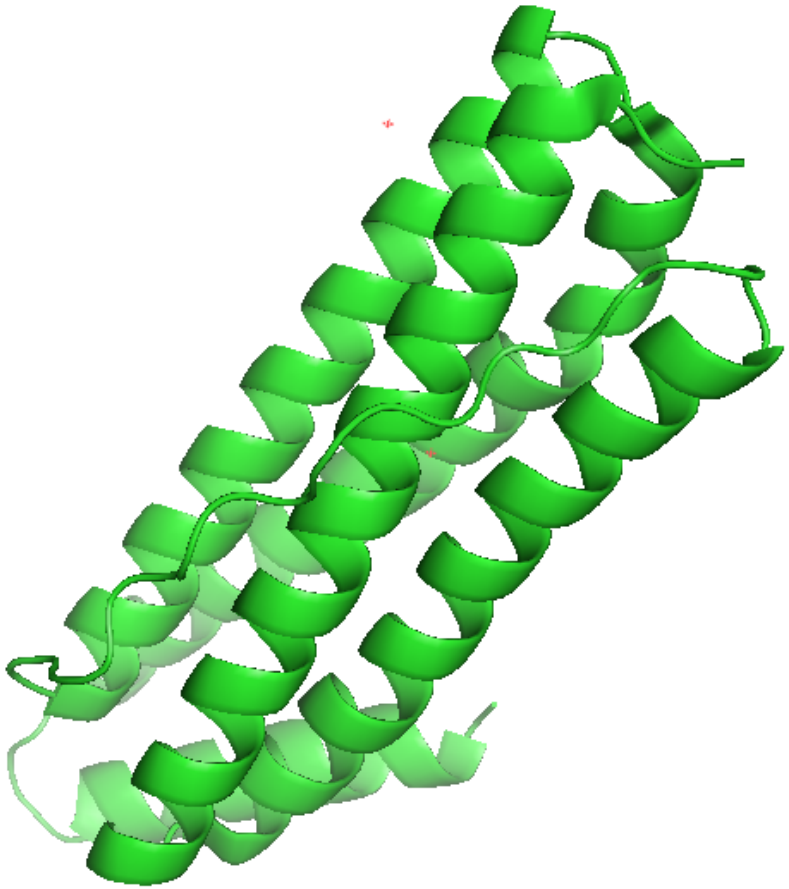


Loops and function



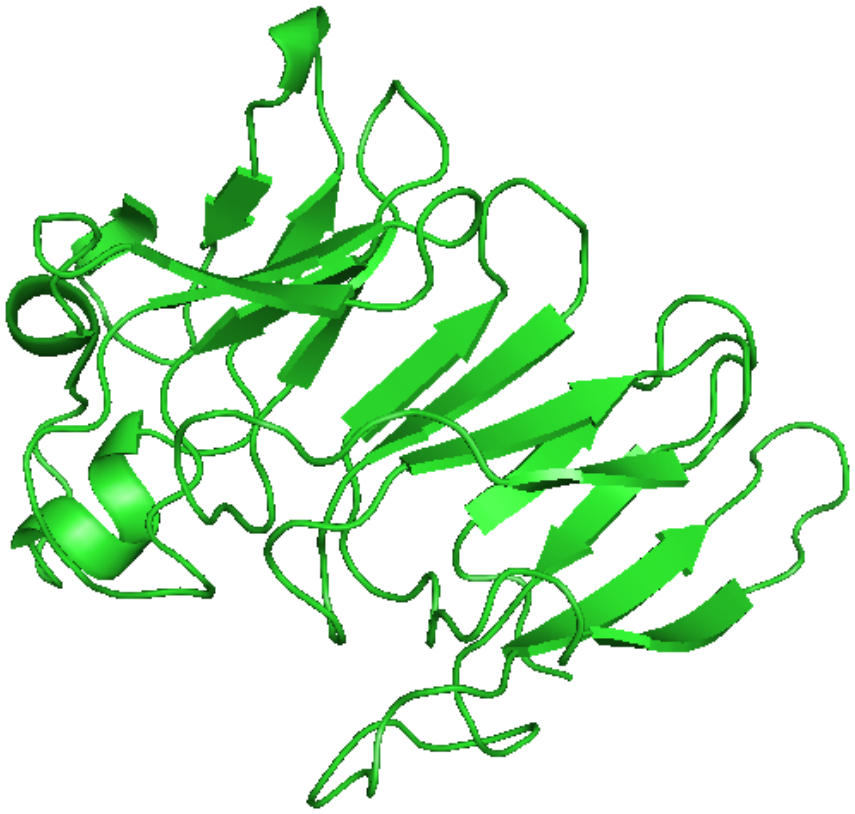
Tertiary structures - Fold space

Mainly alpha:



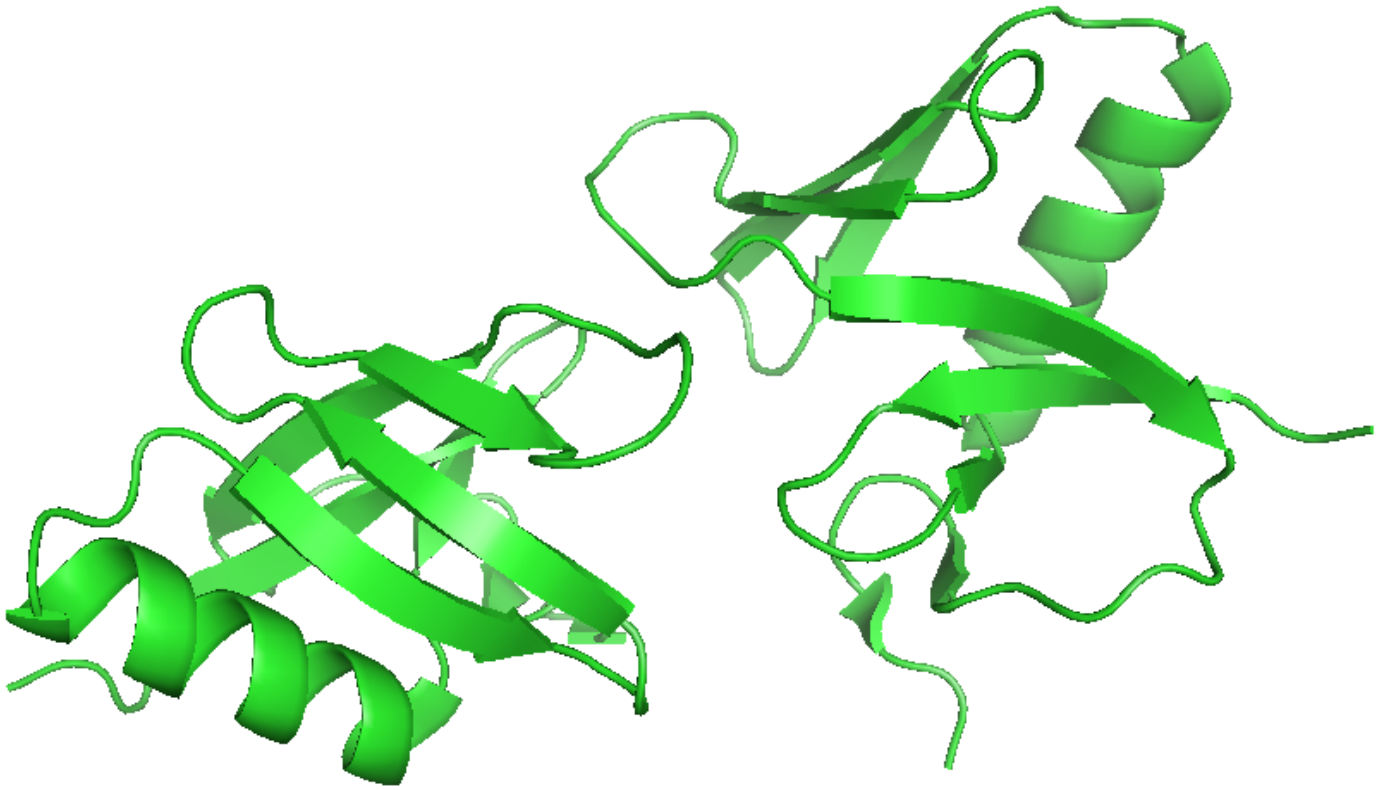
Tertiary structures - Fold space

Mainly beta:



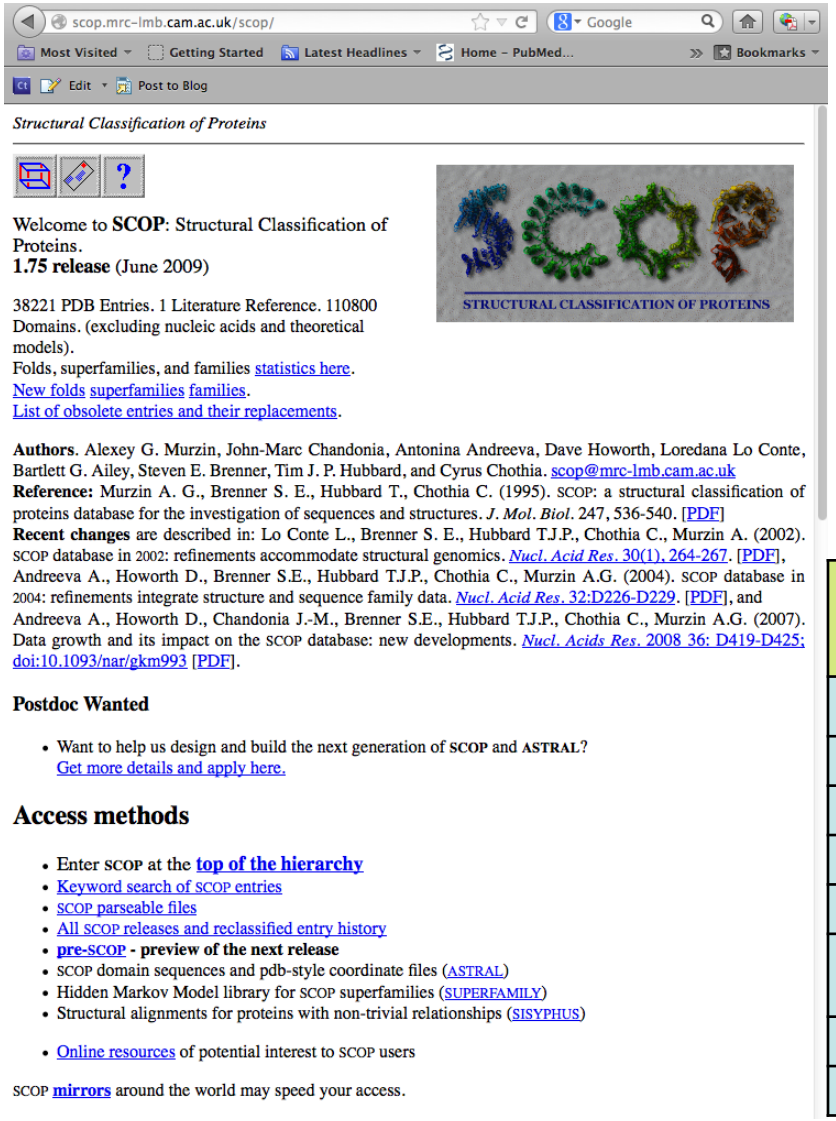
Tertiary structures - Fold space

Alpha beta:



SCOP$_{1.75}$ database

http://scop.mrc-lmb.cam.ac.uk/scop/



Murzin A. G.,el at. (1995). J. Mol. Biol. ***247****, 536-540.*

- ✓ **Largely recognized as "standard of gold"**
- ✓ **Manually classification**
- ✓ **Clear classification of structures in:**

CLASS
FOLD
SUPER-FAMILY
FAMILY

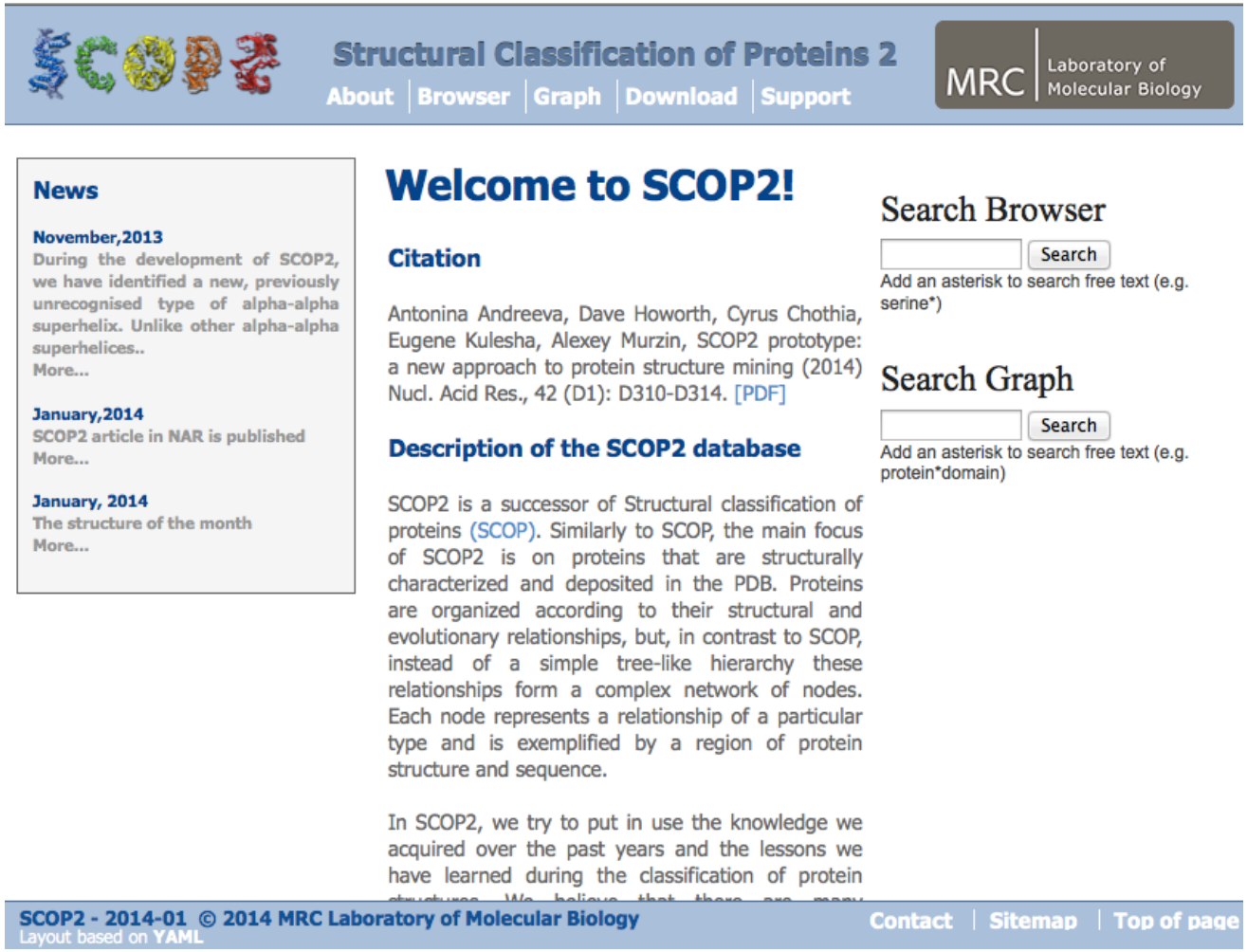
- ✓ **Some large number of tools already available**

Manually classification
Not 100% up-to-date
Domain boundaries definition

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	284	507	871
All beta proteins	174	354	742
Alpha and beta proteins (a/b)	147	244	803
Alpha and beta proteins (a+b)	376	552	1055
Multi-domain proteins	66	66	89
Membrane and cell surface proteins	58	110	123
Small proteins	90	129	219
Total	1195	1962	3902

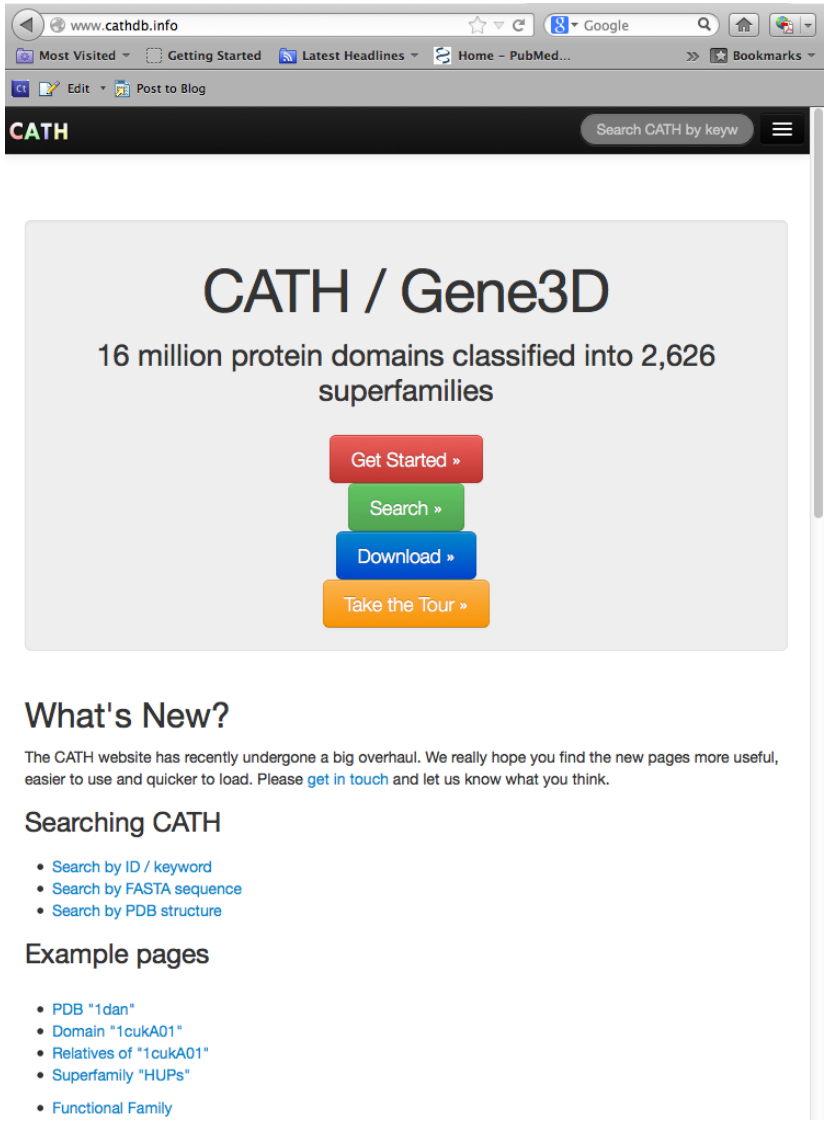
SCOP2 database

http://scop2.mrc-lmb.cam.ac.uk/



CATH$_{3.5}$ database

http://www.cathdb.info



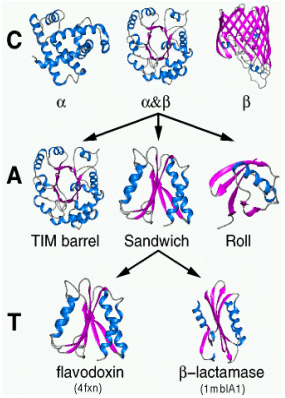
Uses FSSP for superimposition

- ✓ Recognized as "standard of gold"
- ✓ Semi-automatic classification
- ✓ Clear classification of structures in:

CLASS
ARCHITECTURE
TOPOLOGY
HOMOLOGOUS SUPERFAMILIES

- ✓ Some large number of tools already available
- ✓ Easy to navigate

Semi-automatic classification
Domain boundaries definition



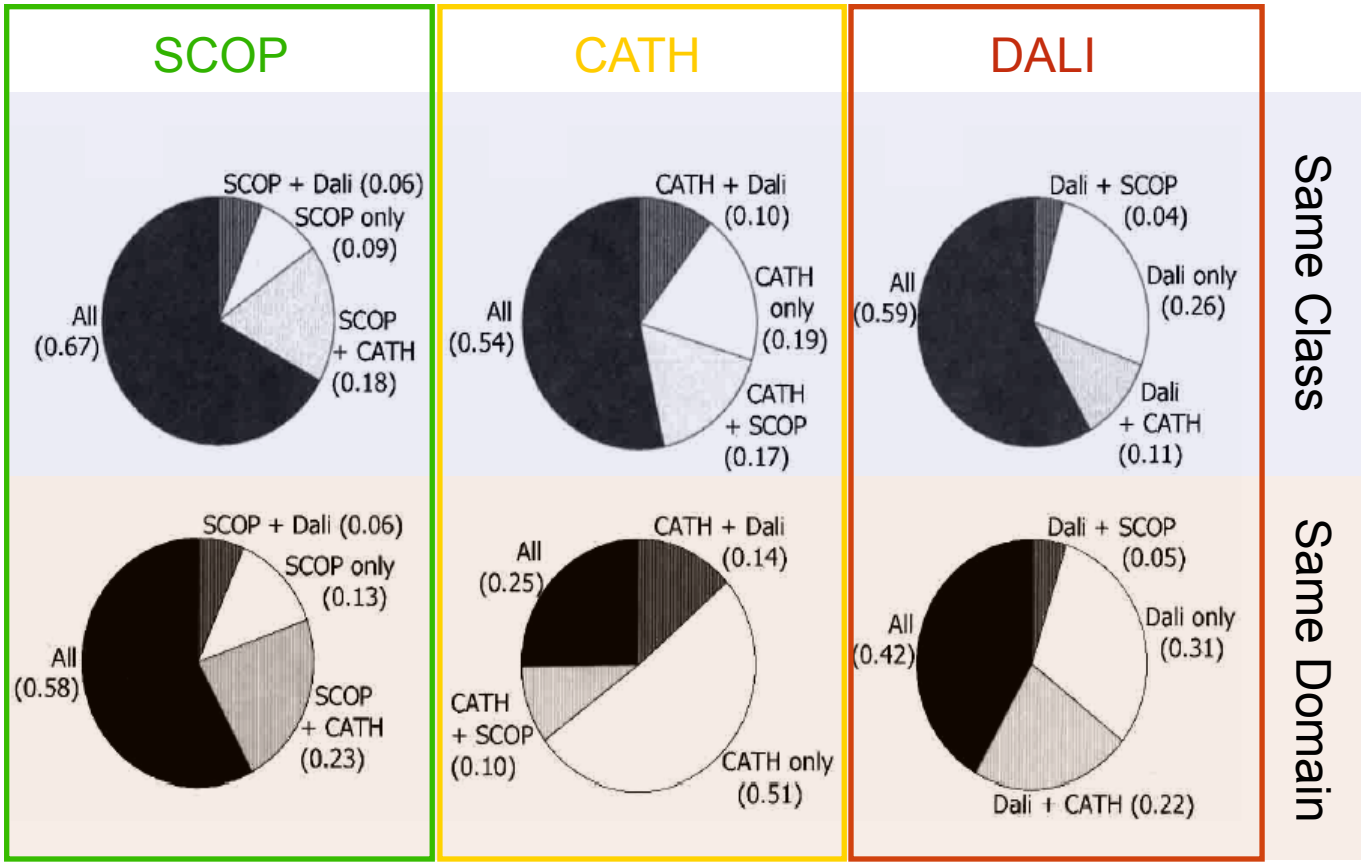
173,536	CATH Domains
2,626	CATH Superfamilies
51,334	PDBs

*Orengo, C.A., et al. (1997) Structure. **5**. 1093-1108.*

Classification of the structural space

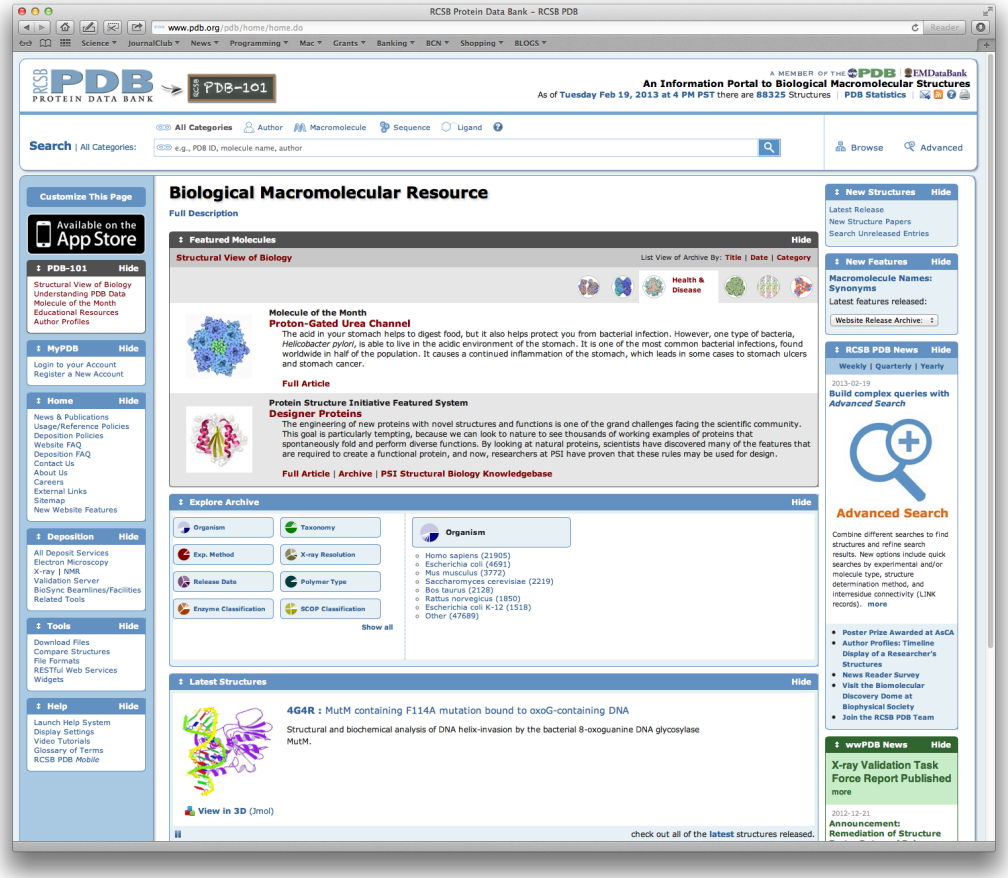
Not an easy task!

Domain definition AND domain classification

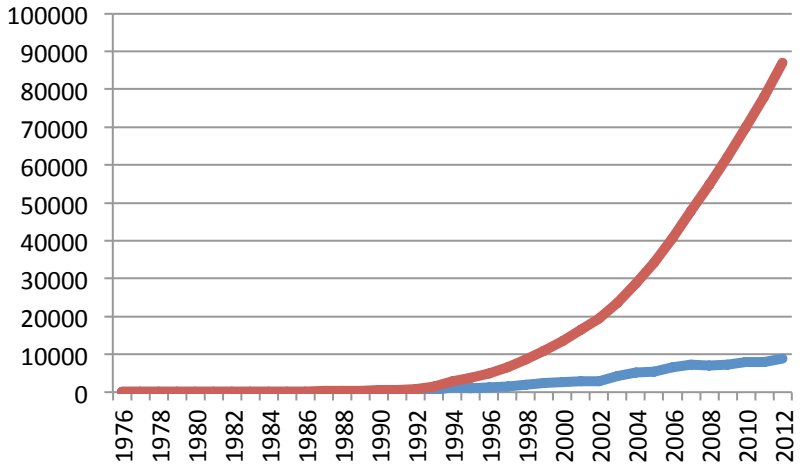


*Day, et al. (2003) Protein Sciences, **12** pp2150*

PDB

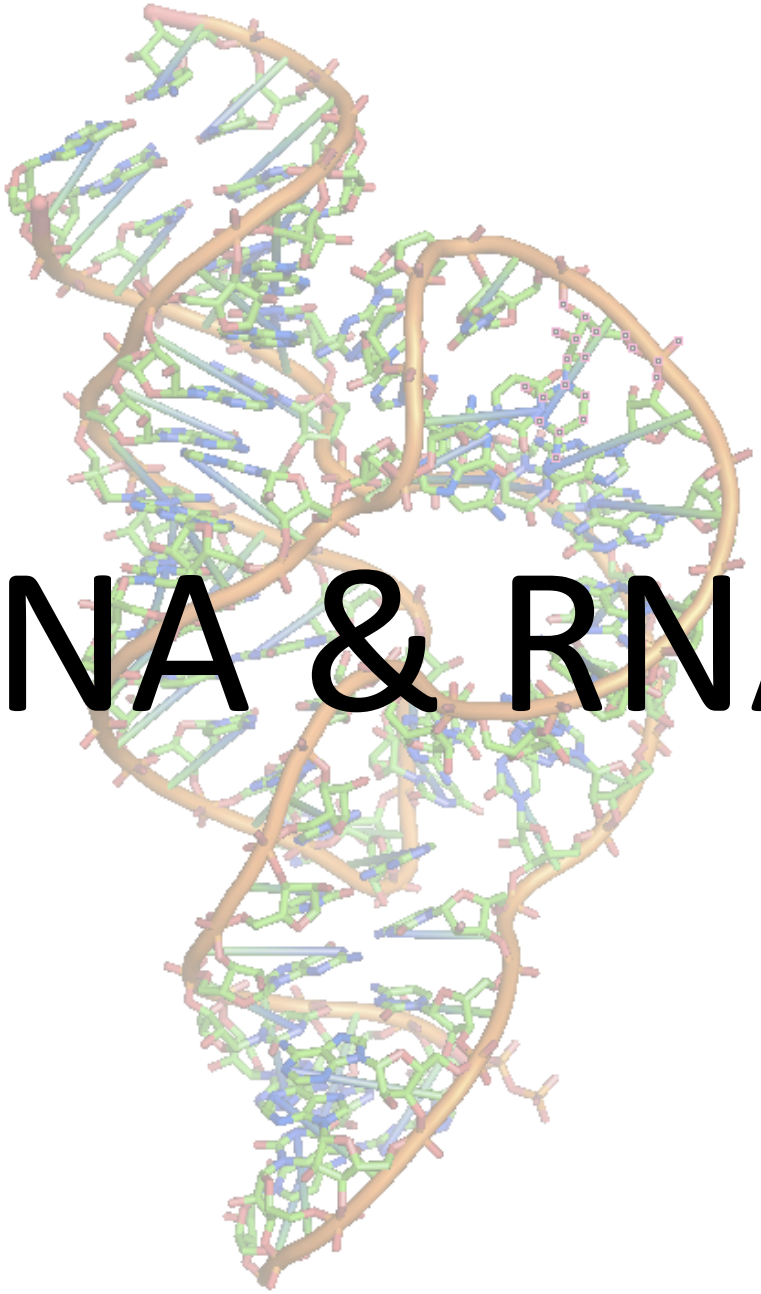


Yearly and total PDB structures per year



Yearly

Total

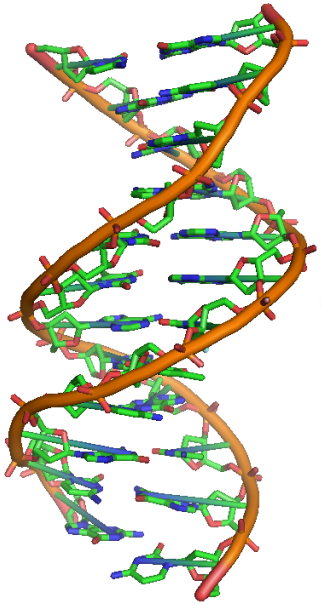


DNA & RNA

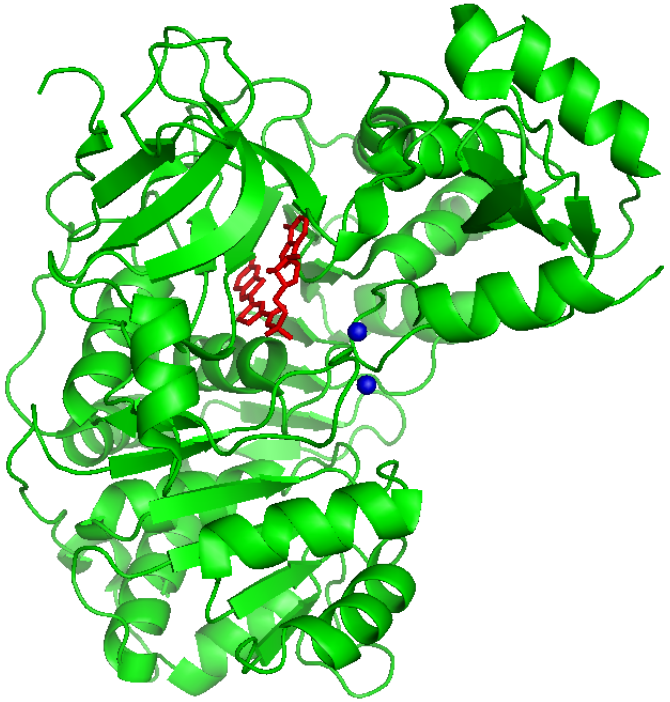
RNA function

- 1) Informative
- 2) Structural
- 3) Catalytic
- 4) Regulatory

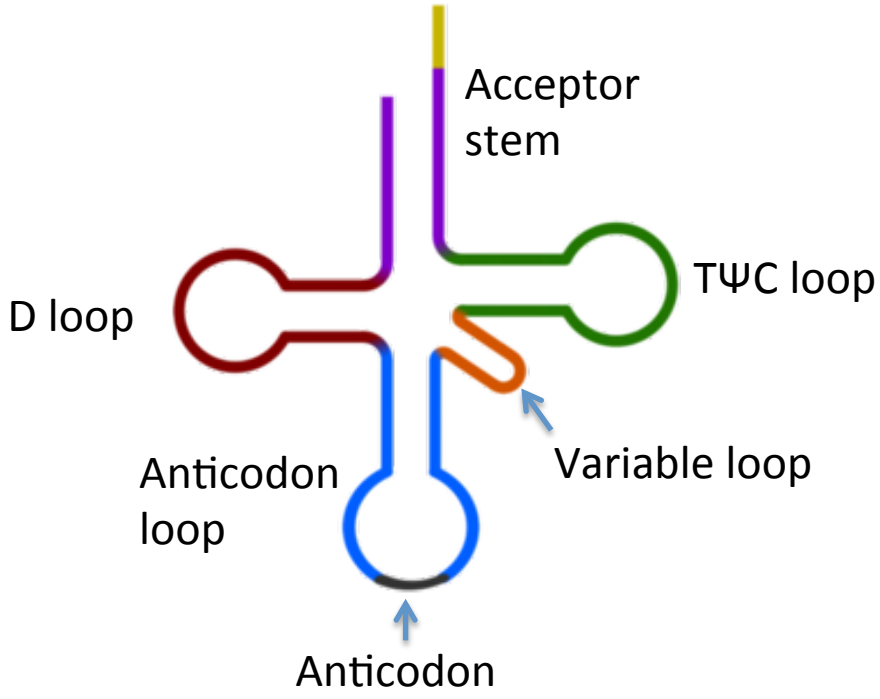
- 1) Messenger (mRNA)

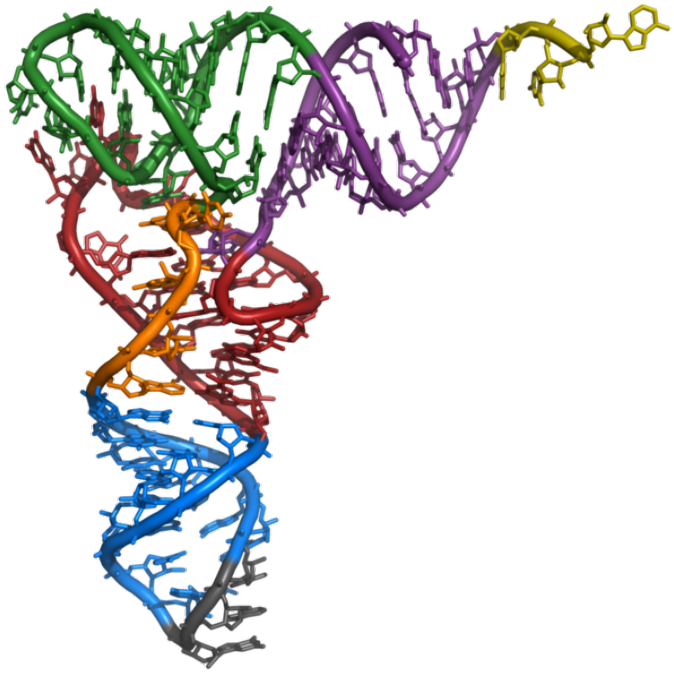


CAGACCGAAGUGAUGAAG
CGAUUGGUUAAUCUGGGC
AAAGCGUCUGAAGGUUGU
GGUUUCGAC

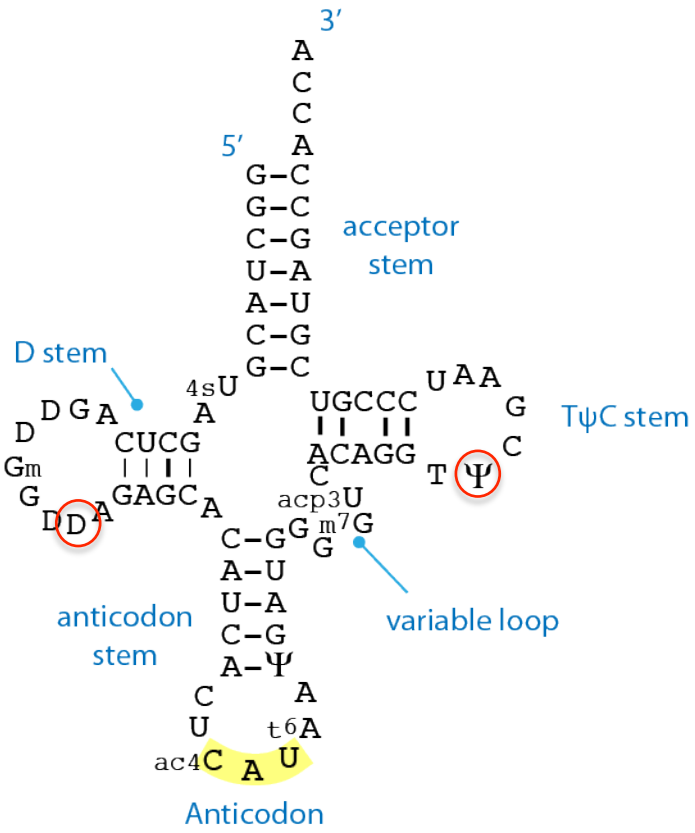


- 2) Transfer (tRNA)

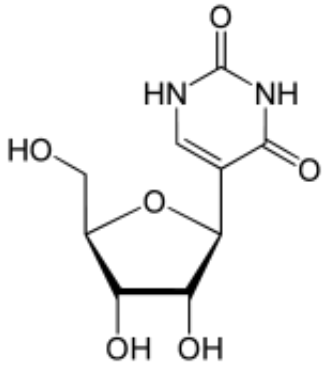




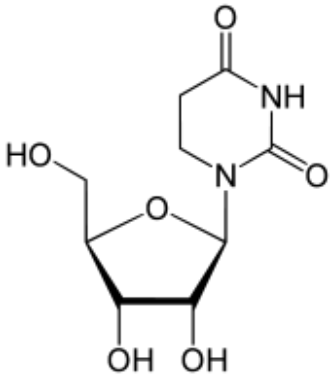
- tRNA modifications



Ψ = Pseudouridine

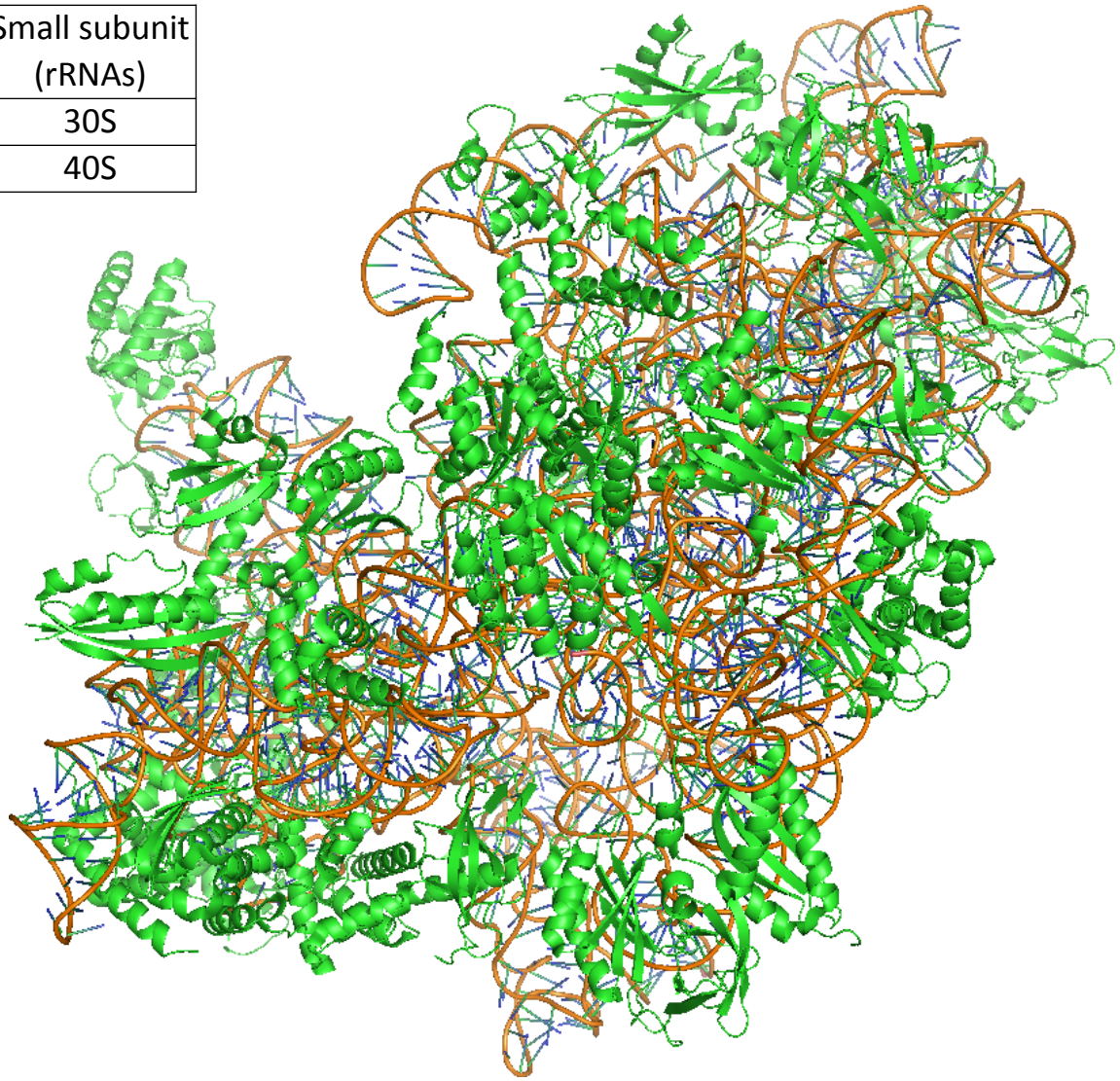


D = Dihydrouridine



- 3) Ribosomal (structural) (rRNA)

Type	Size	Large subunit (rRNAs)	Small subunit (rRNAs)
prokaryotic	70S	50S	30S
eukaryotic	80S	60S	40S

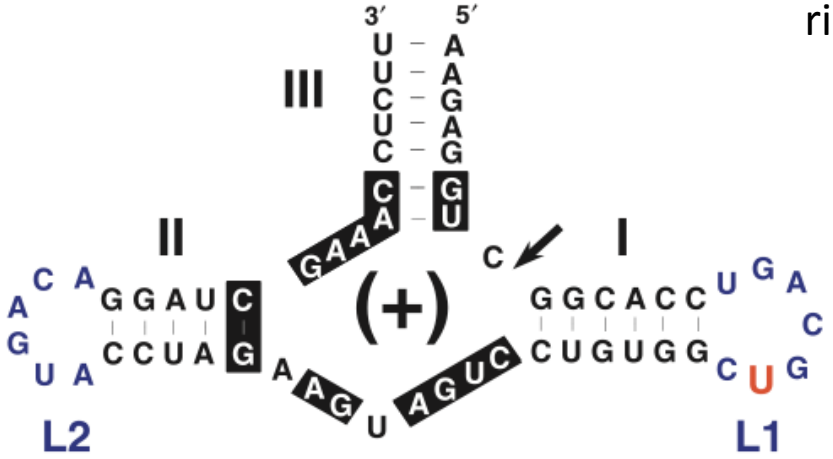


- 4) Genomic

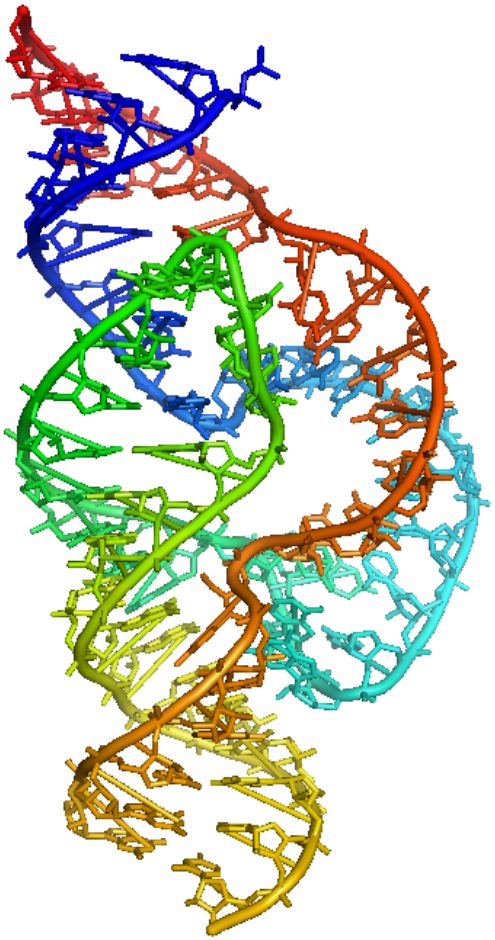
- RNA virus
  - Single stranded RNA virus (ssRNA)
    - Positive strand RNA virus
    - Negative strand RNA virus
  - Double stranded RNA virus (dsRNA)
- RNA satellites
  - SatRNAs of *Cucumber Mosaic Virus*
  - SatRNAs of *Turnip Crinkle Virus*
  - …

- Retroviruses

- 5) Catalytical (ribozymes)

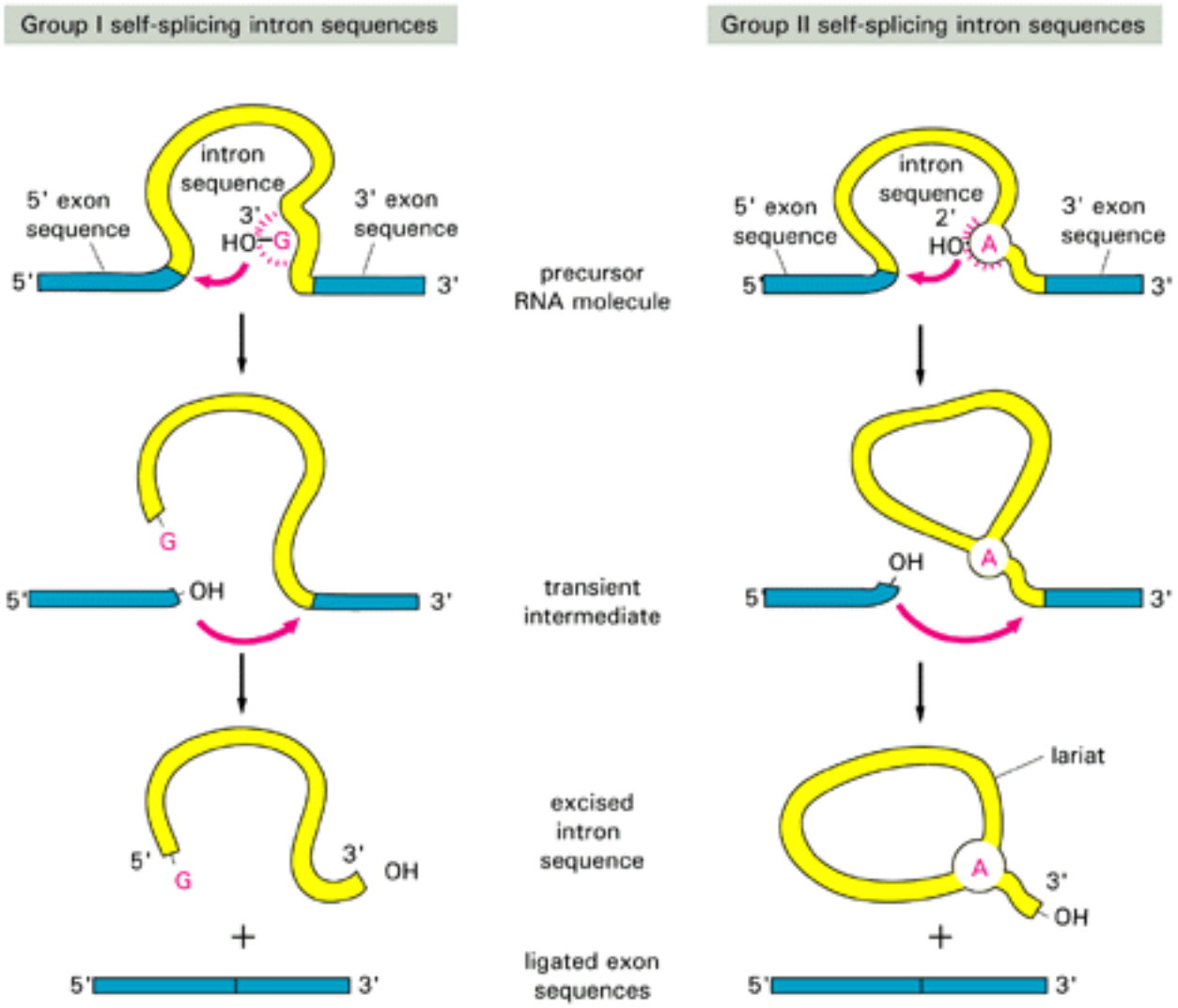


Hammerhead ribozyme

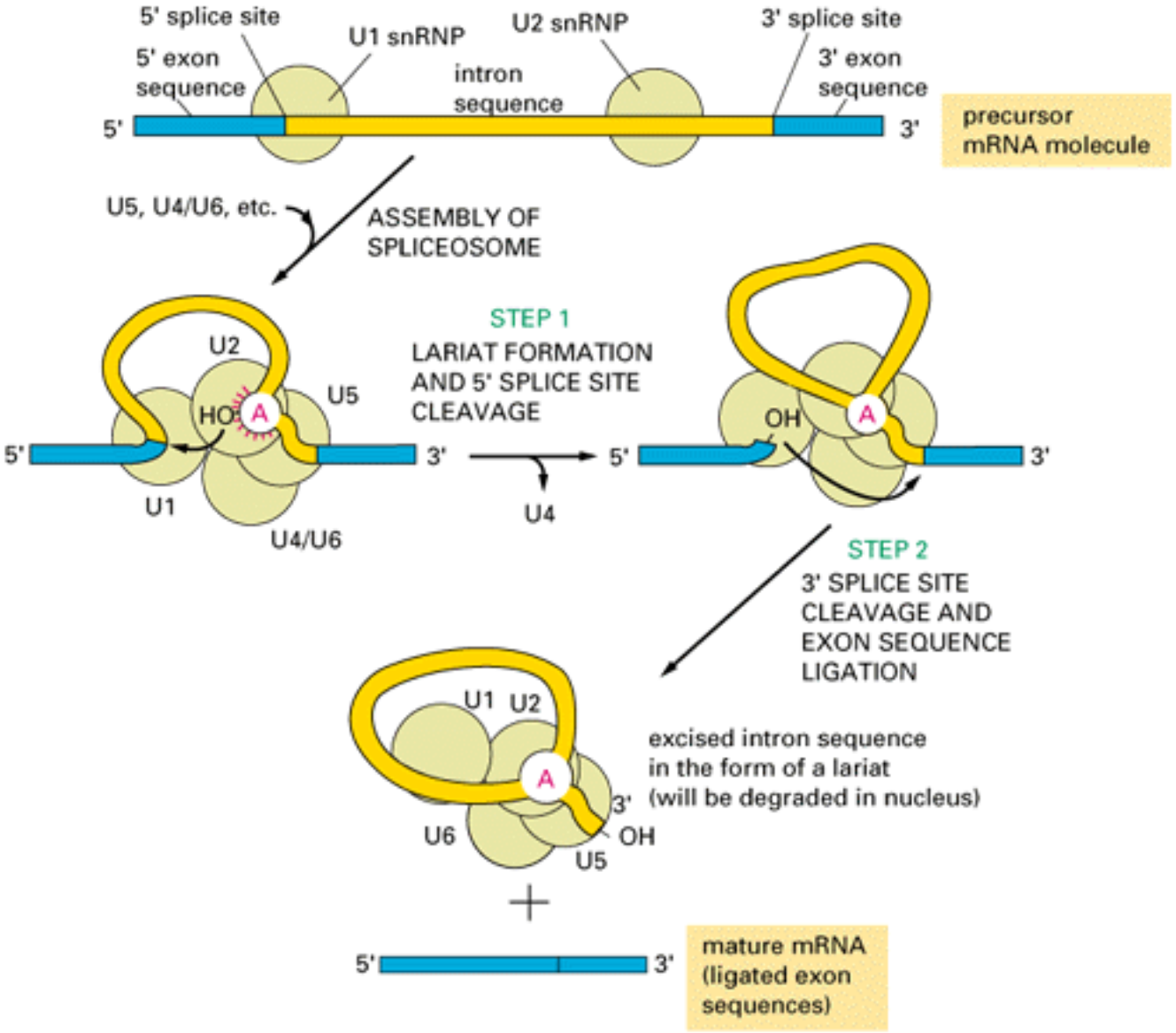


- 5) Catalytical (ribozymes)
- Examples:
  - Peptidyl transferase 23S rRNA
  - Group I and II introns
  - RNAse P
  - Hammerhead ribozyme
  - Hairpin ribozyme
  - …

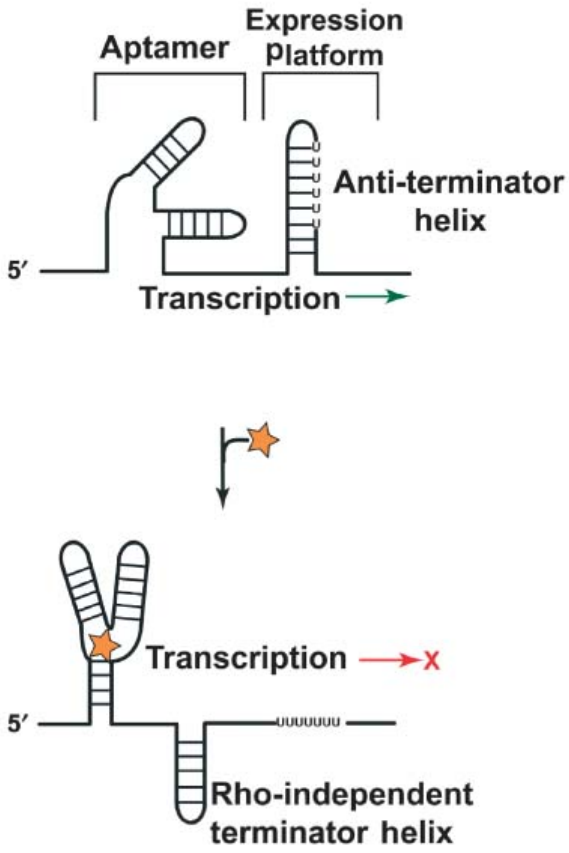
Introns


Group I self-splicing intron sequences
Group II self-splicing intron sequences
intron sequence
5' exon sequence
3' exon sequence
3'
HO
G
5'
3'
precursor RNA molecule
G
5'
OH
3'
transient intermediate
5'
G
3'
OH
excised intron sequence
+
5'
3'
ligated exon sequences
intron sequence
2'
A
lariat

Type III introns



- 6) Molecular sensor (riboswitches)



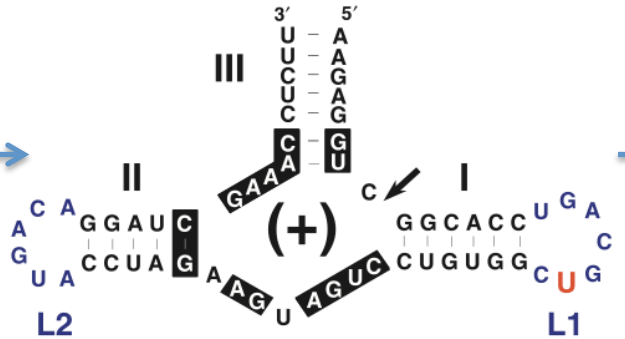
- Small RNAs (sRNA)
  - Small nuclear RNAs (snRNA)
  - Small nucleolar RNAs (snoRNA)
  - Micro RNAs (miRNA)
  - Silencing RNAs (siRNA)
  - Piwi-interacting RNAs (piRNA)
- Long non-coding RNAs (lncRNA)
- others…
- To be discovered…
- RNA therapeutics

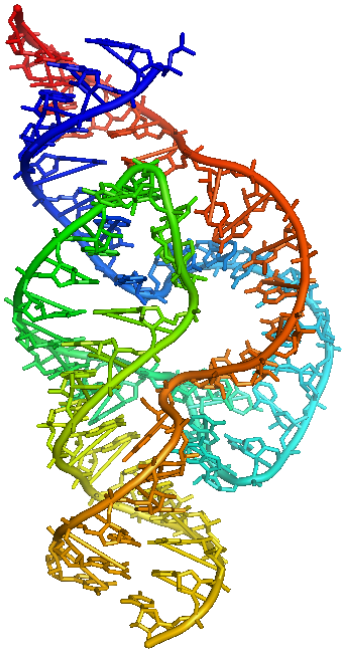
Primary structure

Secondary structure

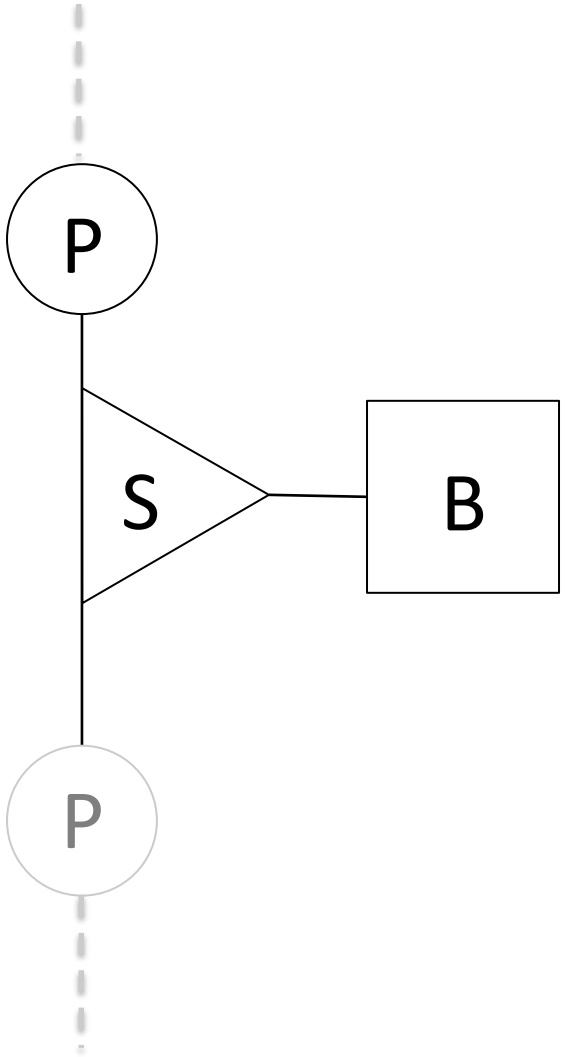
Tertiary structure (3D)

CAGACCGAAGUGAUGAAGC
GAUUGGUUAAUCUGGGCA
AAGCGUCUGAAGGUUGUG
GUUUCGAC





Nucleotide structure



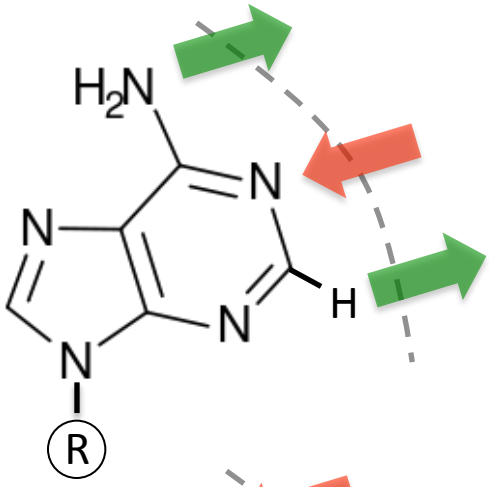
P = Phosphate
S = Sugar (ribose, deoxyribose)
B = nucleotidic Base

Bases

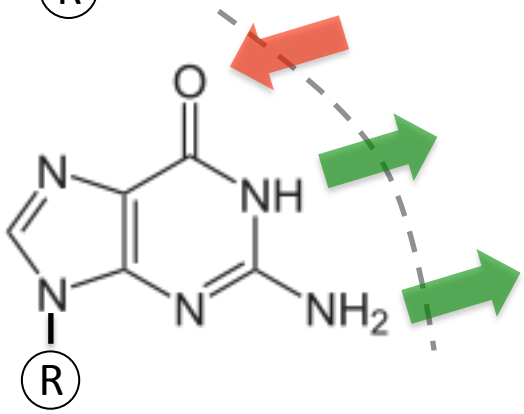
(aka nucleotidic bases, aromatic bases)

Purines

Adenine
(A)

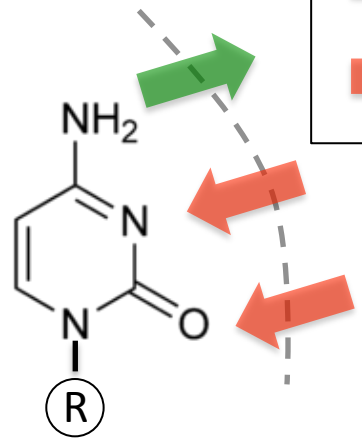


Guanine
(G)

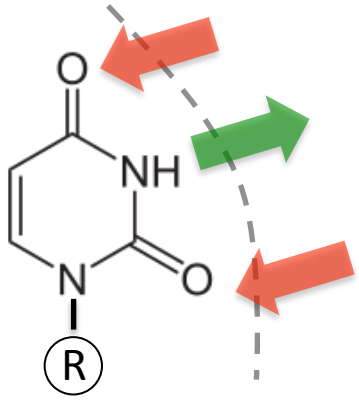


Pyrimidines

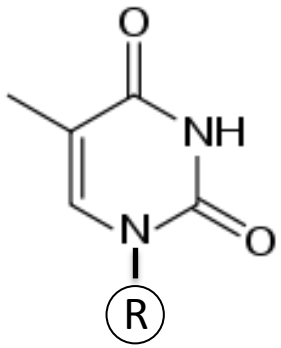
Cytosine
(C)

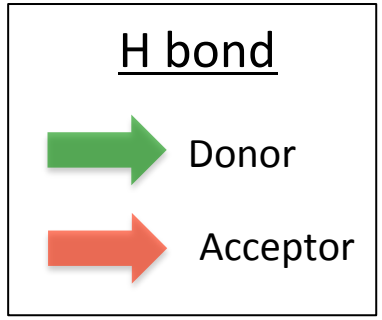


Uracil
(U)



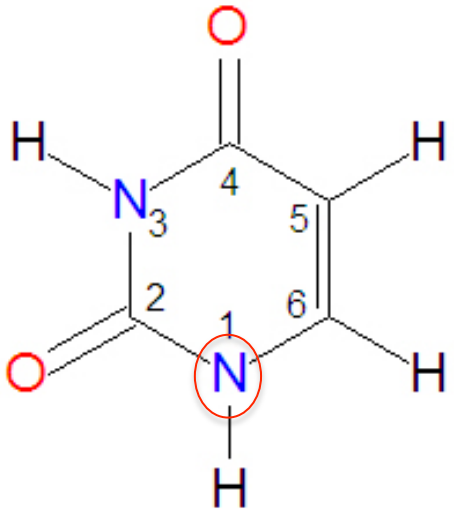
Thymidine
(T)



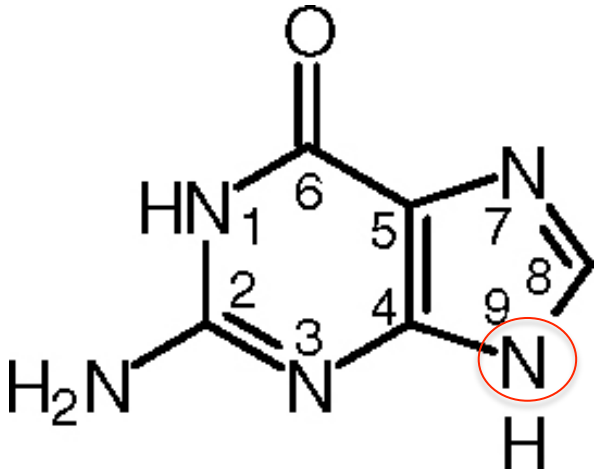


Numbering system - bases

Uracil

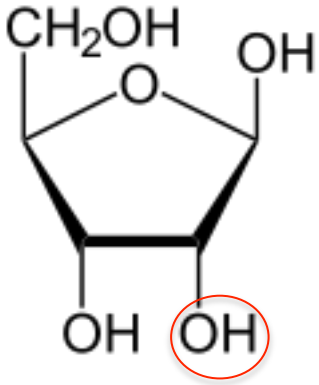


Guanine

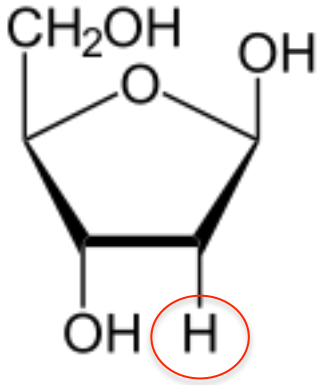


Sugars

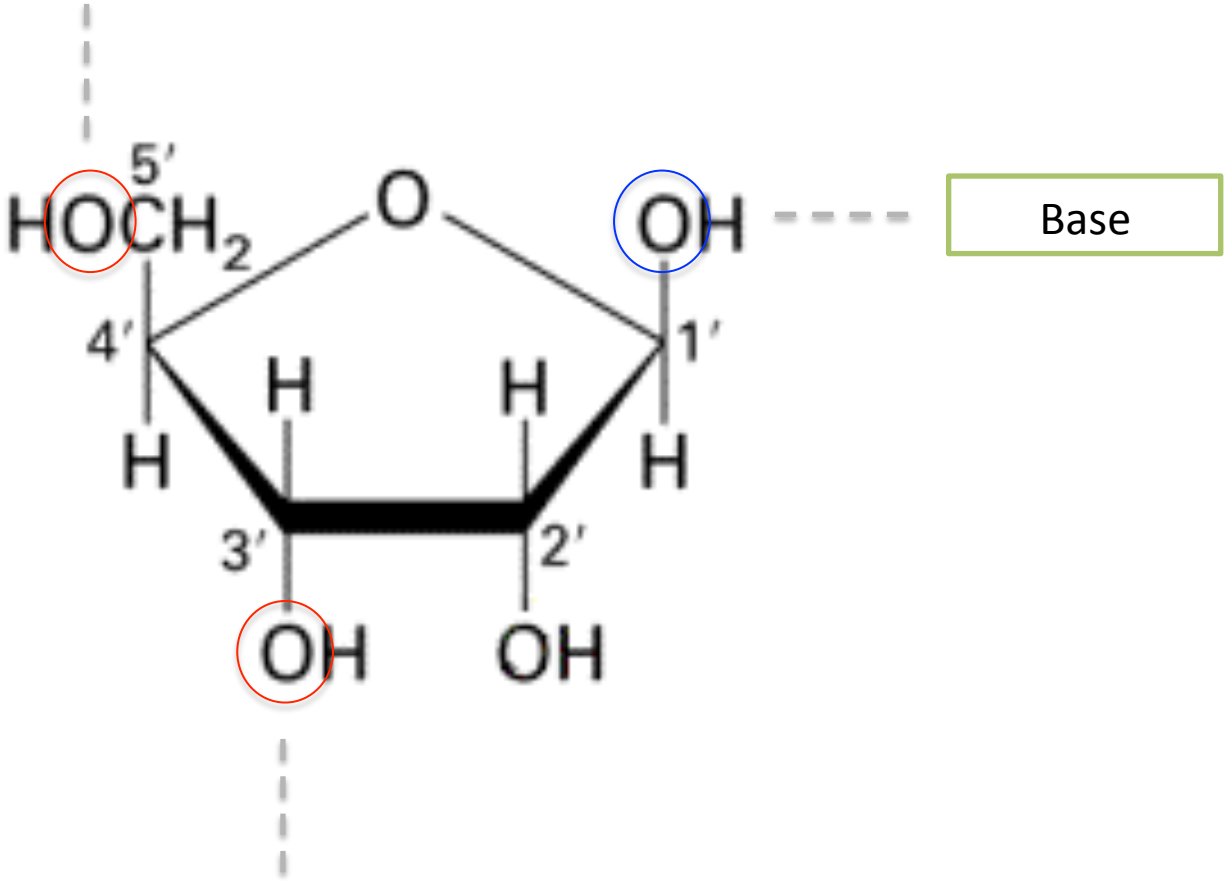
β-D-ribose



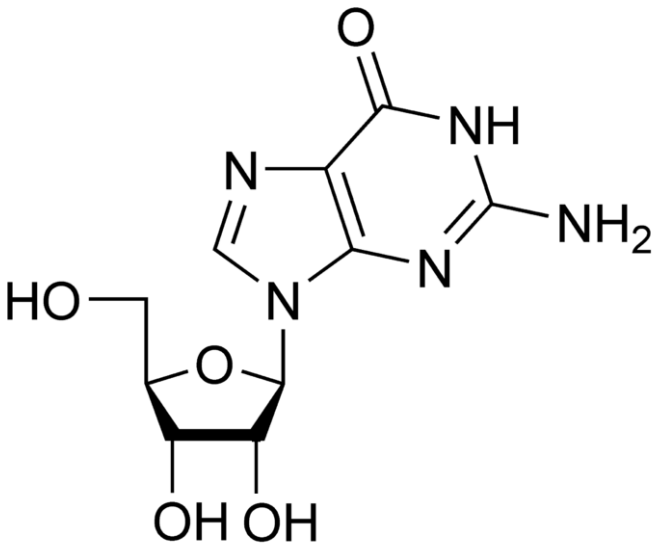
β-D-2-Deoxyribose



Numbering system -ribose

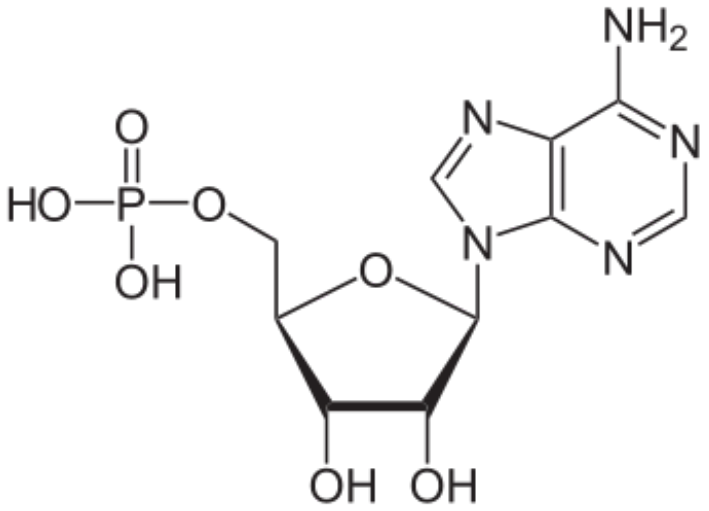


Nucleoside



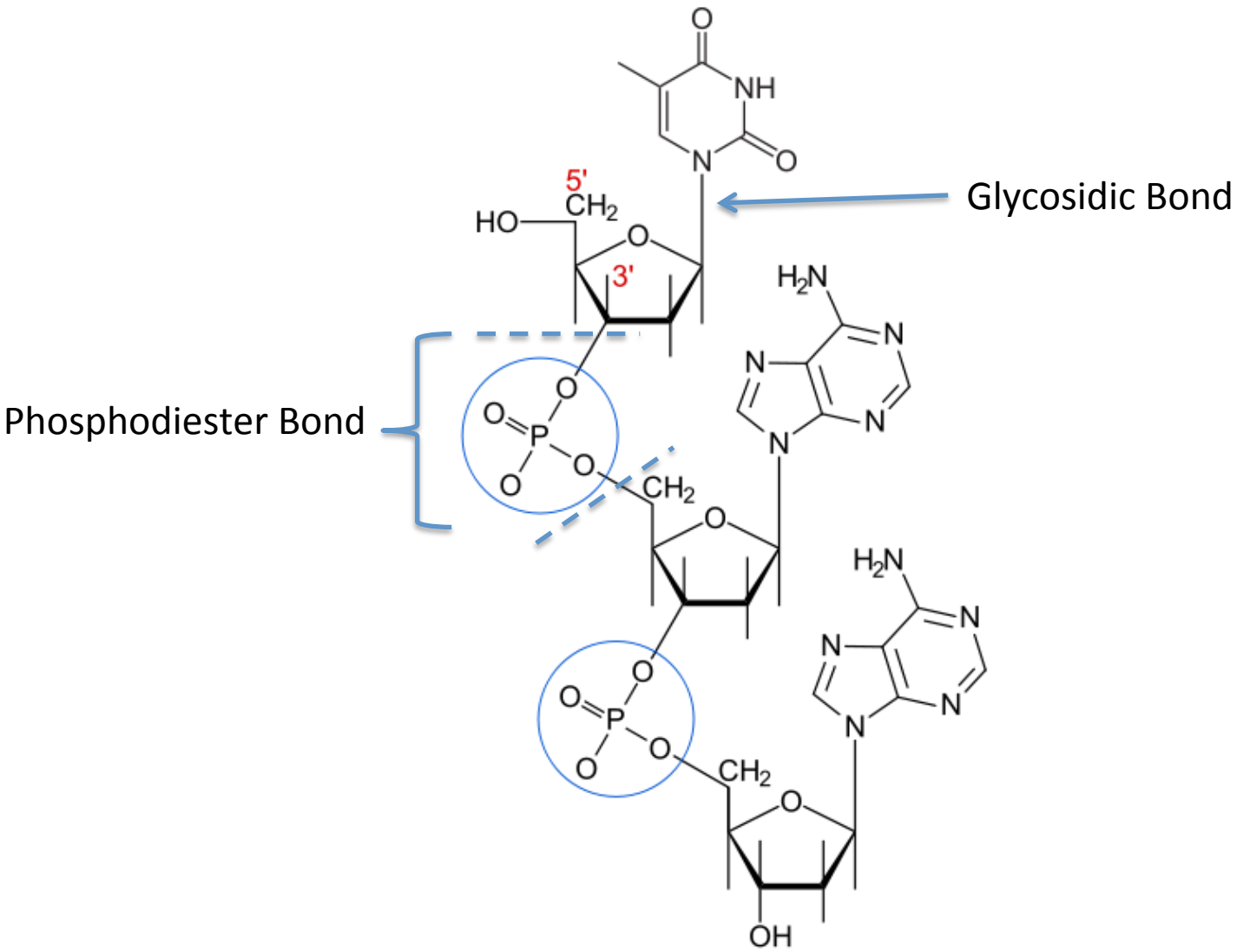
Sugar + base

Nucleotide monophosphate

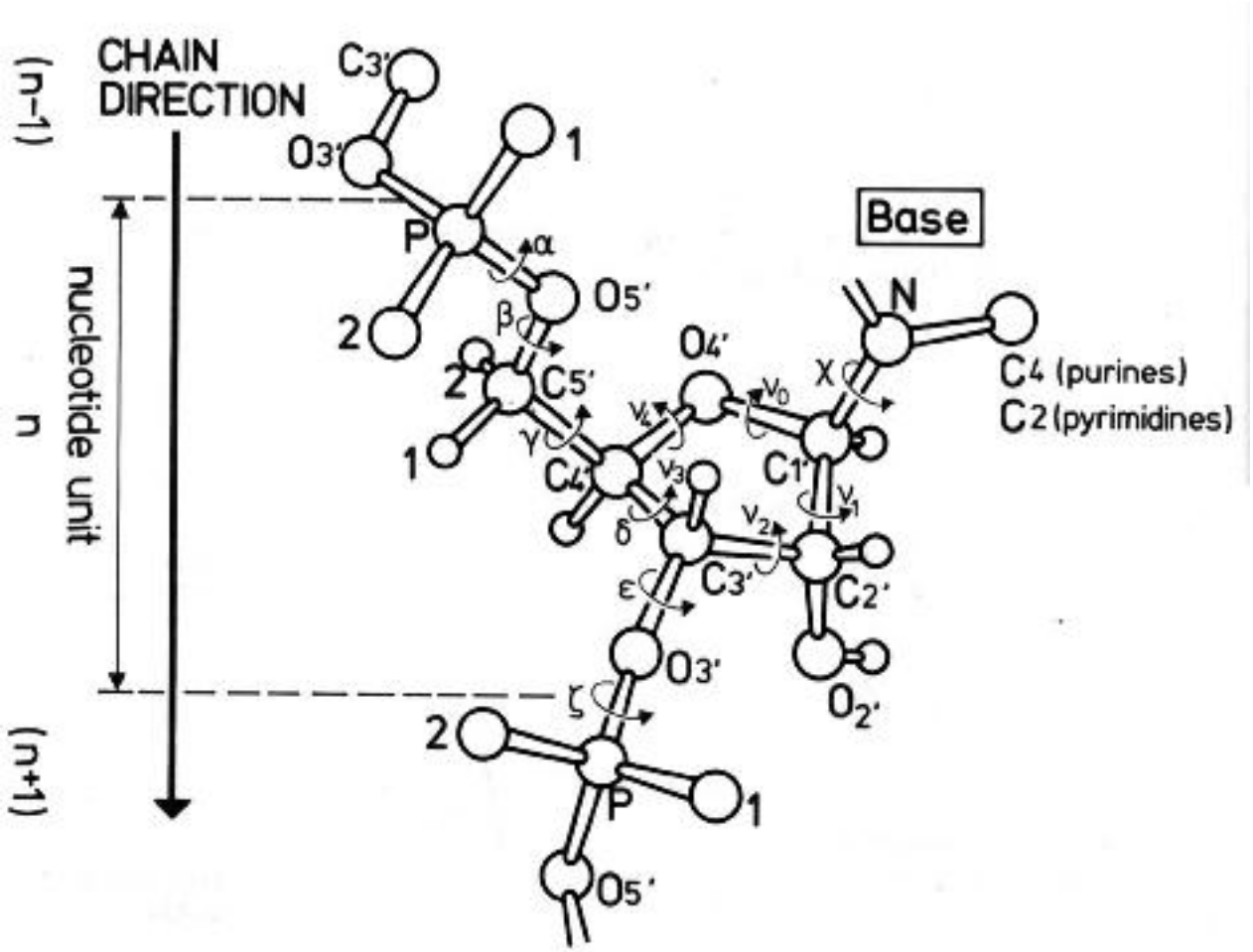


Phosphate + sugar + base

Phosphodiester Bond



Nucleic acid torsion angles



Ramachandran plot

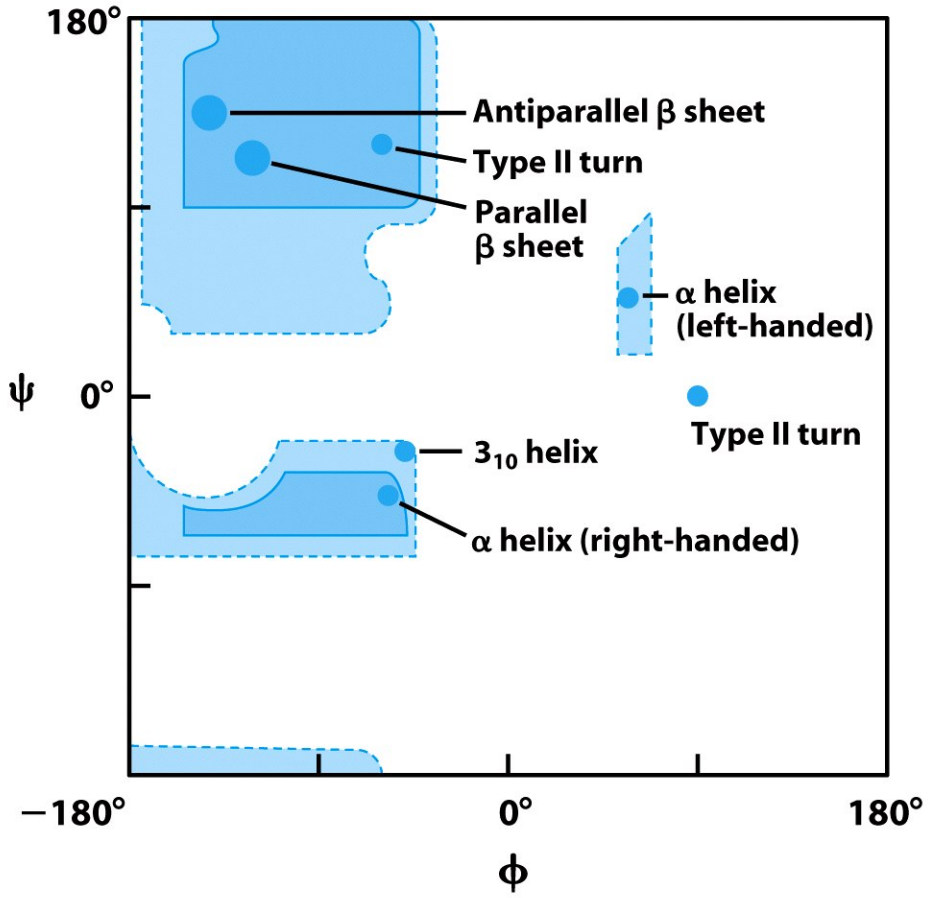
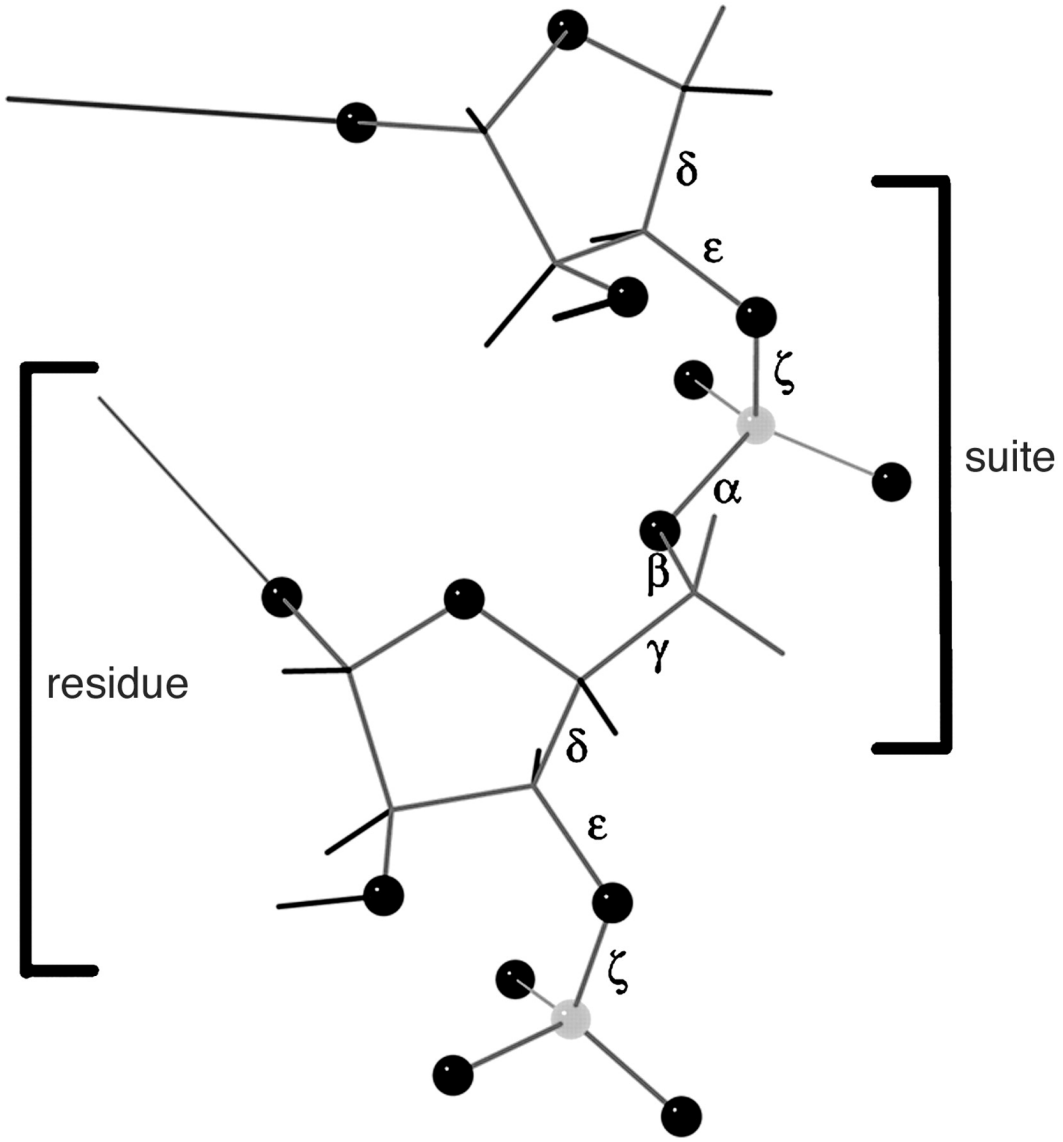
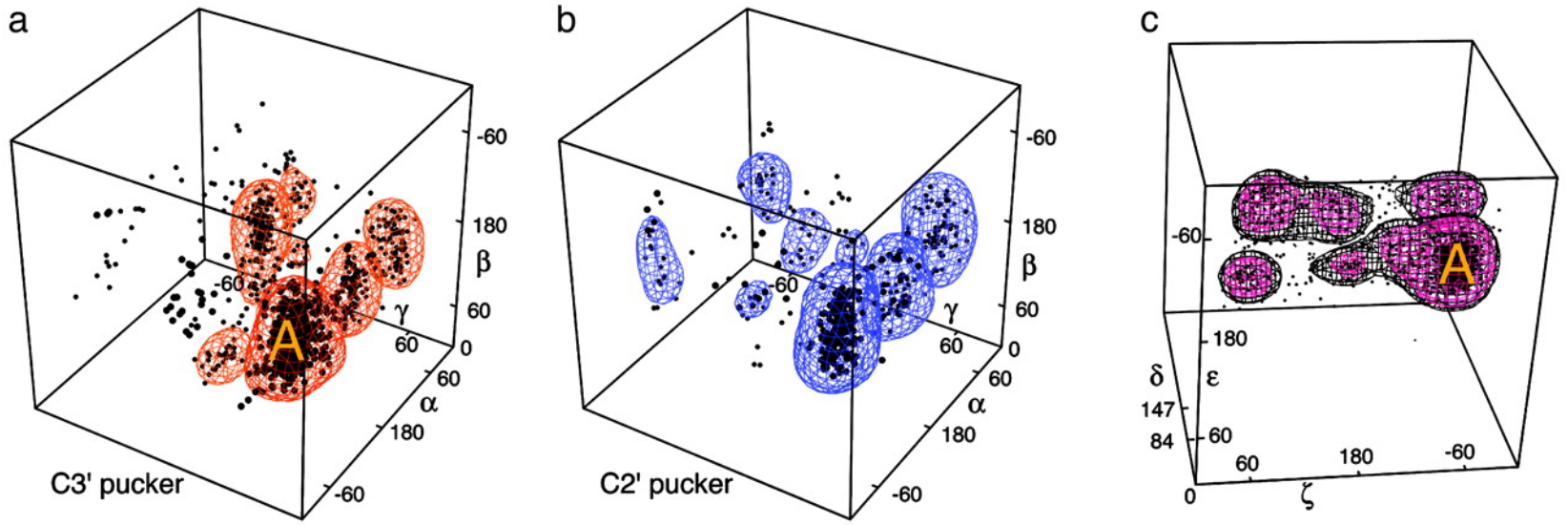


Figure 4-9a Principles of Biochemistry, 4/e

RNA backbone is rotameric



Plots of the heminucleotide angle triplets for data filtered by clashes and B > 60, with smoothed contours enclosing the top 7–10 peak clusters: (a) α–β–γ plot for adjacent sugar pucker C3′ endo.



Murray L J W et al. PNAS 2003;100:13904-13909

RNA backbone rotamers

	α δ / β ε / γ ζ	C3' e p	C3' e t	C3' e -140	C3' e m	C2' e p	C2' e t	C2' e m
C3' (2nd)	p t p	*		*	*	**	*	*
	p 110 t	*						
	p t t					*[a]		
	t t p				***[b]	*	**	*
	t 135 t				*			
	t t t				***[c]			
	m t p		**	**	**A**[d]		**[e]	***
	m -135 p				*			
	-110 80 t				**[f]			
	m t t							*
C2' (2nd)	p t p	*	*	*		**	*	
	p 110 t	*						
	p t m			*				
	t t p	**[g]			*[h]			
	t t t				*			
	m t p		*	*	***	*	*	*[i]
	m -135 p				***[j]			*
	m t m				*			*

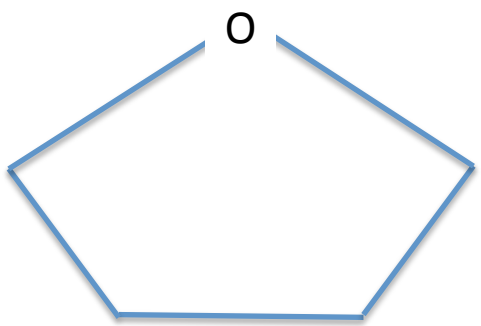
Conformer frequencies:

*	~1% of non-A
**	2-3% of non-A
***	4-20% of non-A

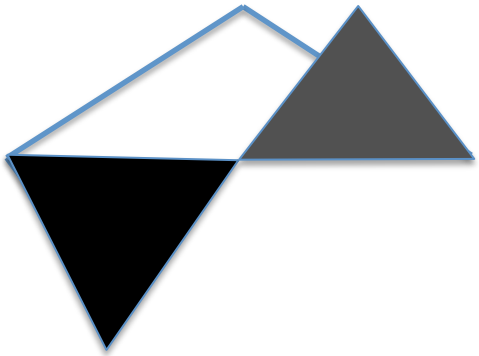
Angle codes:

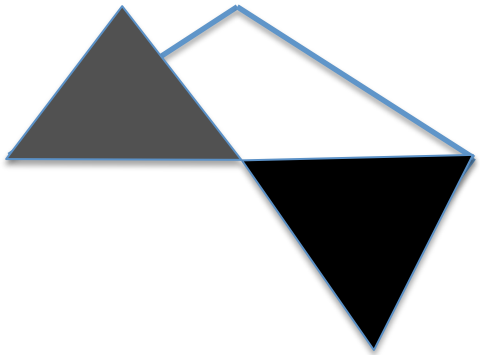
t	trans
m	gauche minus
p	gauche plus
e	eclipsed

Furanose ring

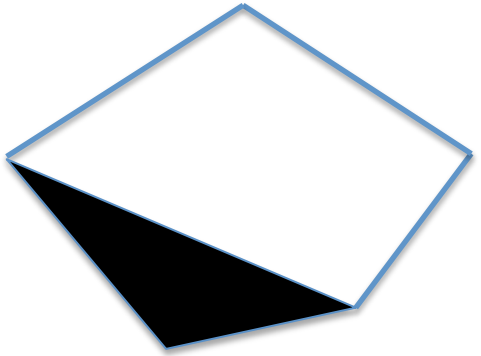


Twisted
(T)

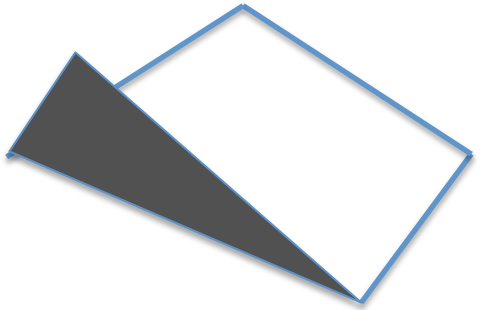




Envelope
(E)

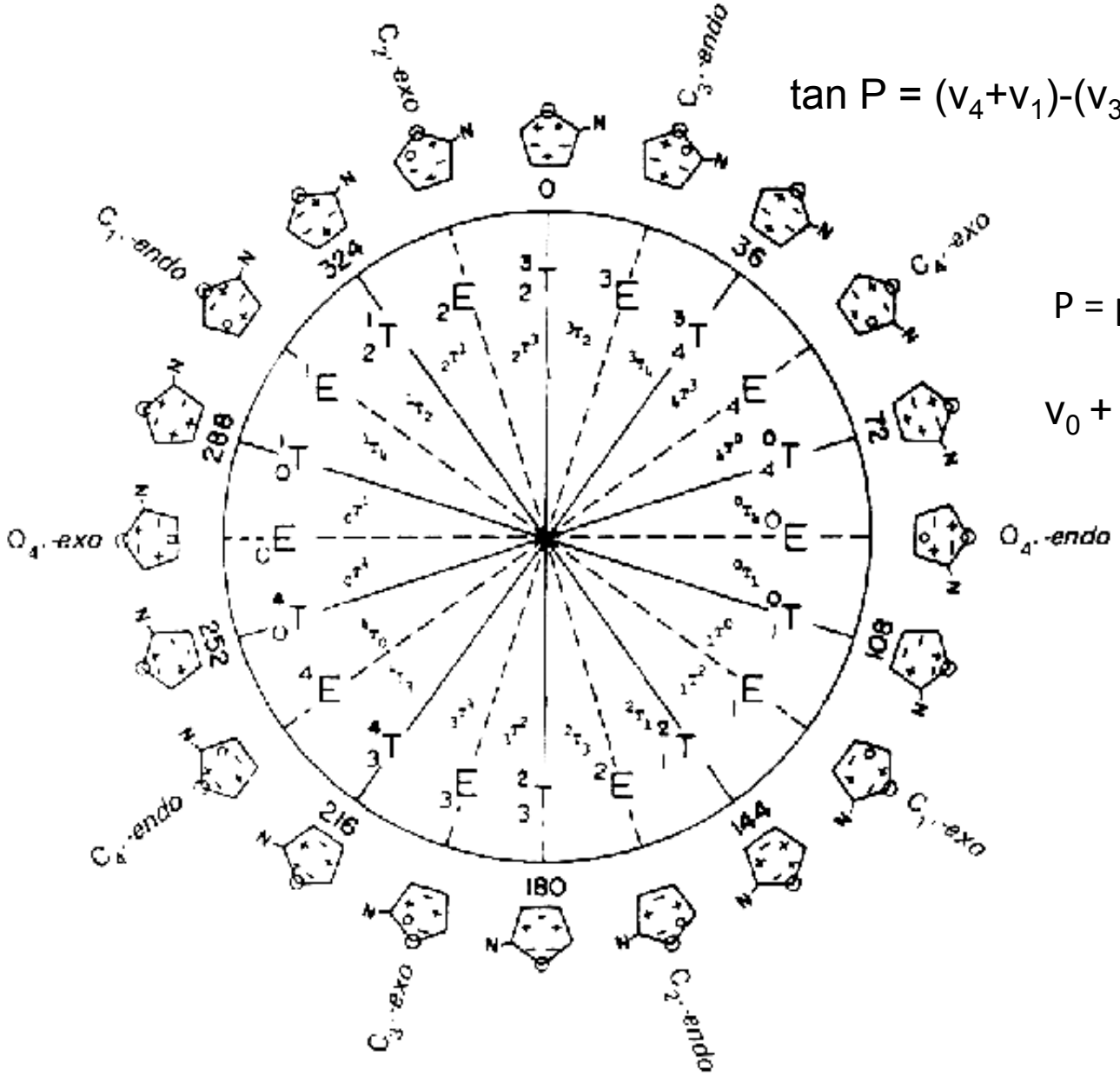


C3’ exo



C3’ endo

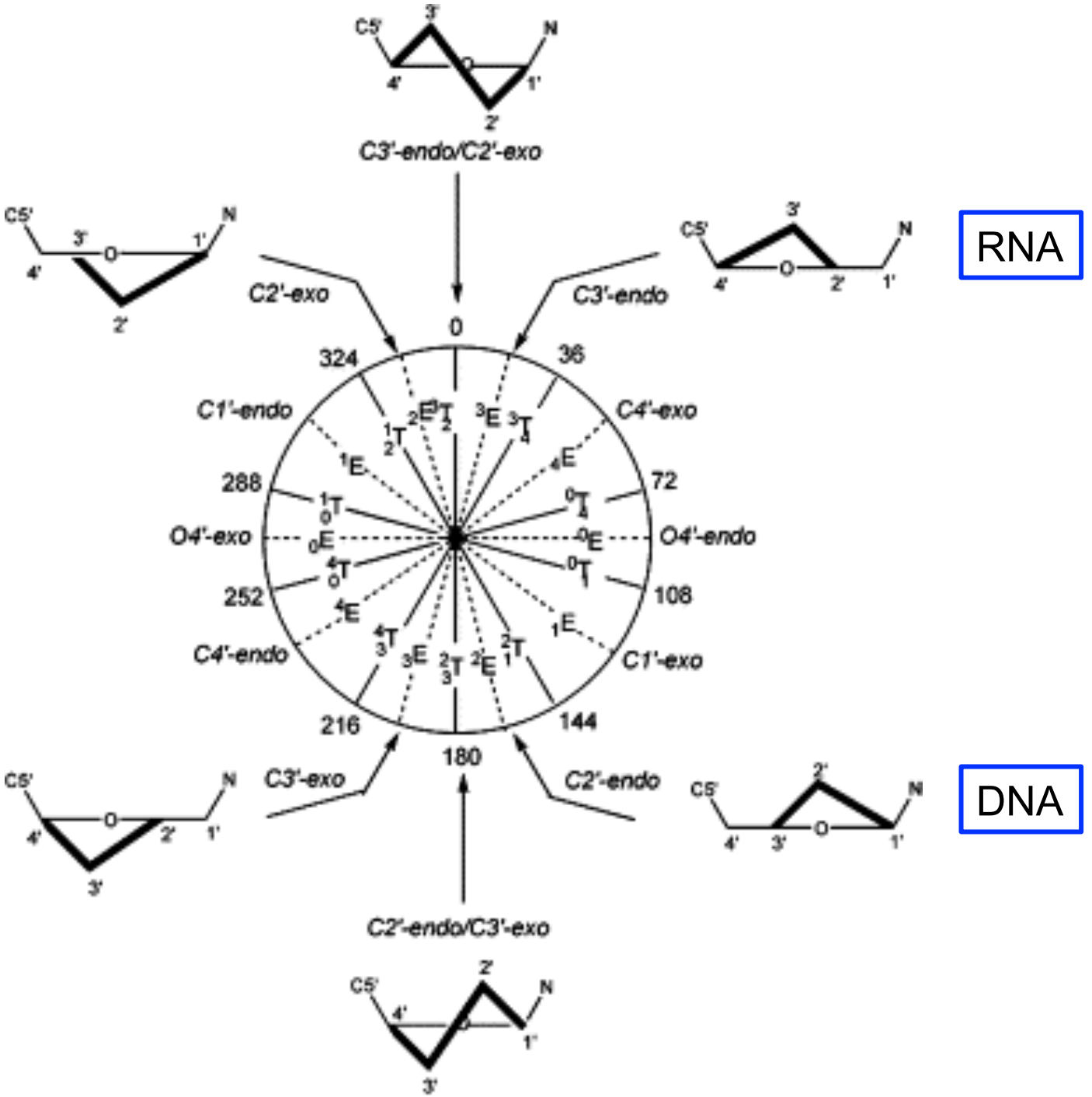
Sugar Puckering



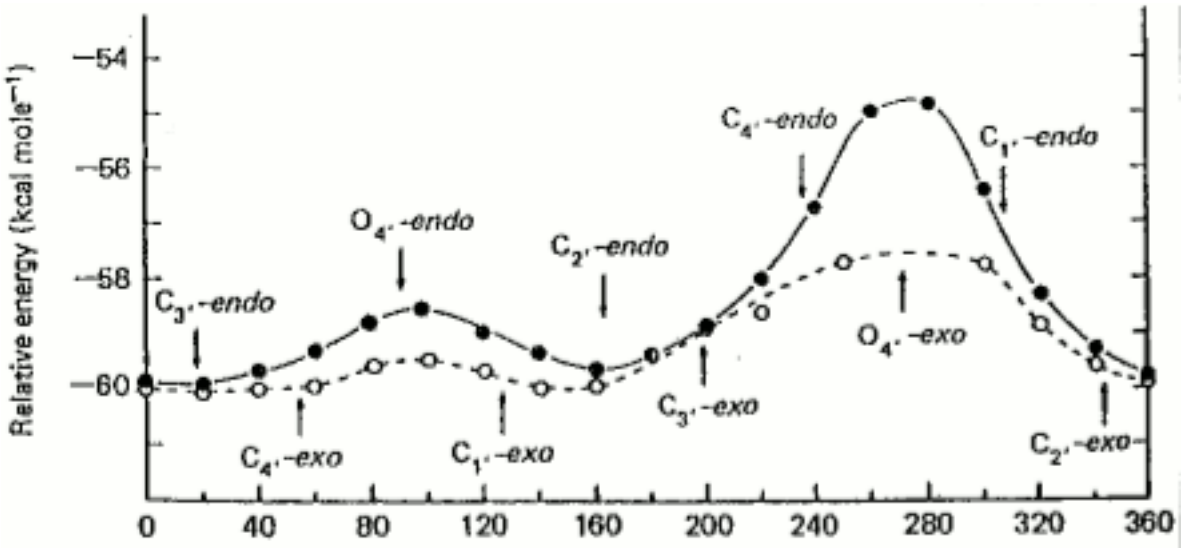
$\tan P = (v_4+v_1)-(v_3+v_0)/[2v_2(\text{sen}36^\circ + \text{sen}72^\circ)]$

P = pseudorotation angle

$v_0 + v_1 + v_2 + v_3 + v_4 = 0$

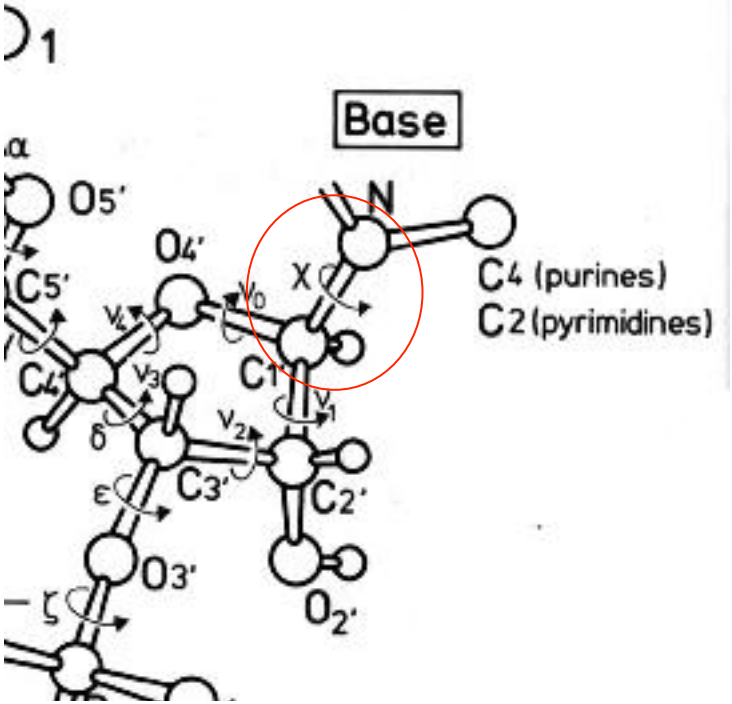

C3'-endo/C2'-exo
C2'-exo
C3'-endo
0
324
36
C1'-endo
C4'-exo
288
72
O4'-exo
O4'-endo
252
108
C4'-endo
C1'-exo
216
144
180
C3'-exo
C2'-endo
C2'-endo/C3'-exo
RNA
DNA

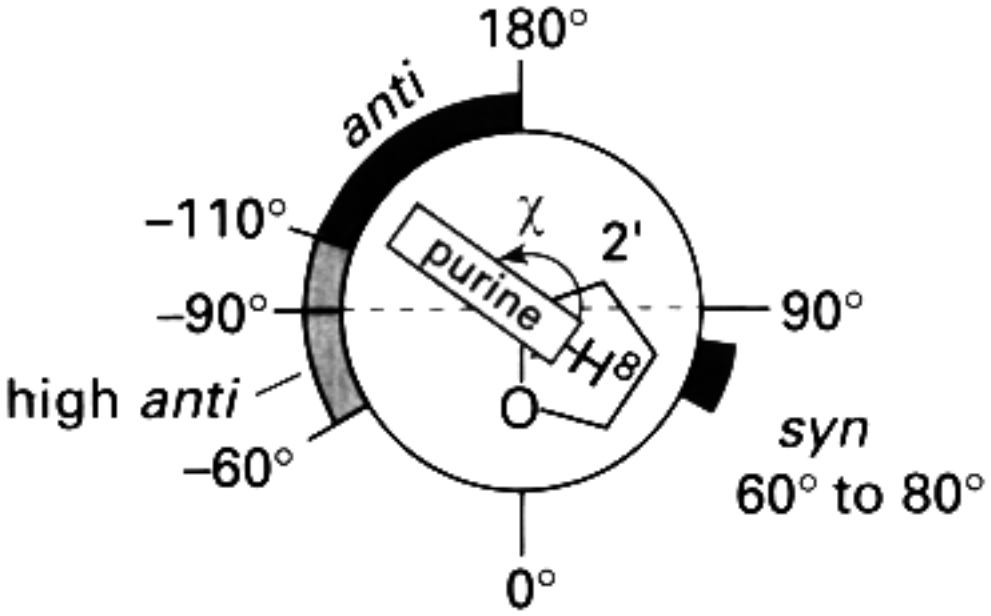
Energy variation related to *P*



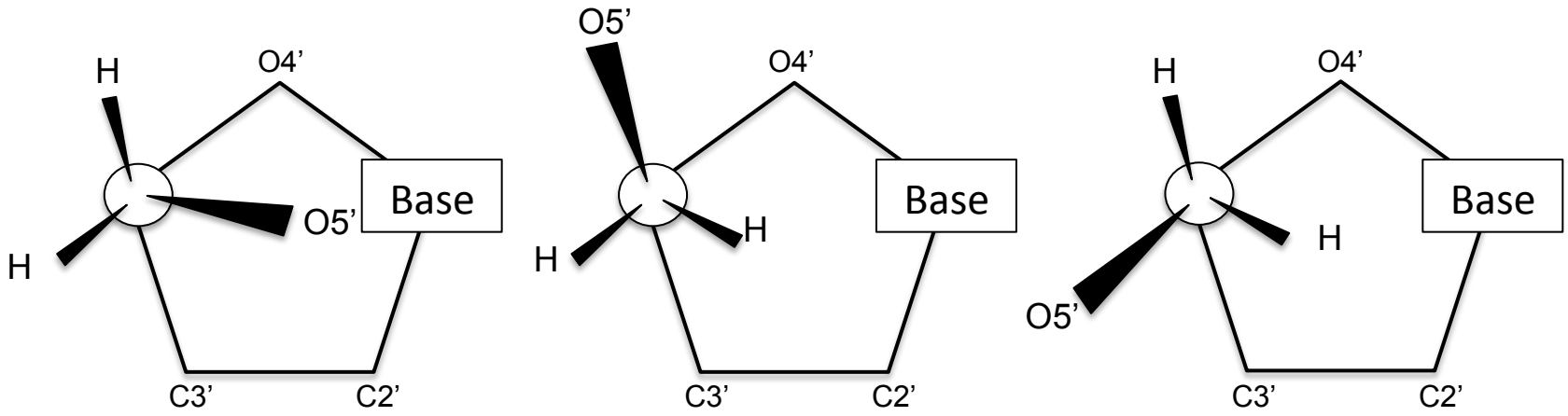
Pseudorotation phase angle *P* (degrees)

Orientation around glycosidic angle



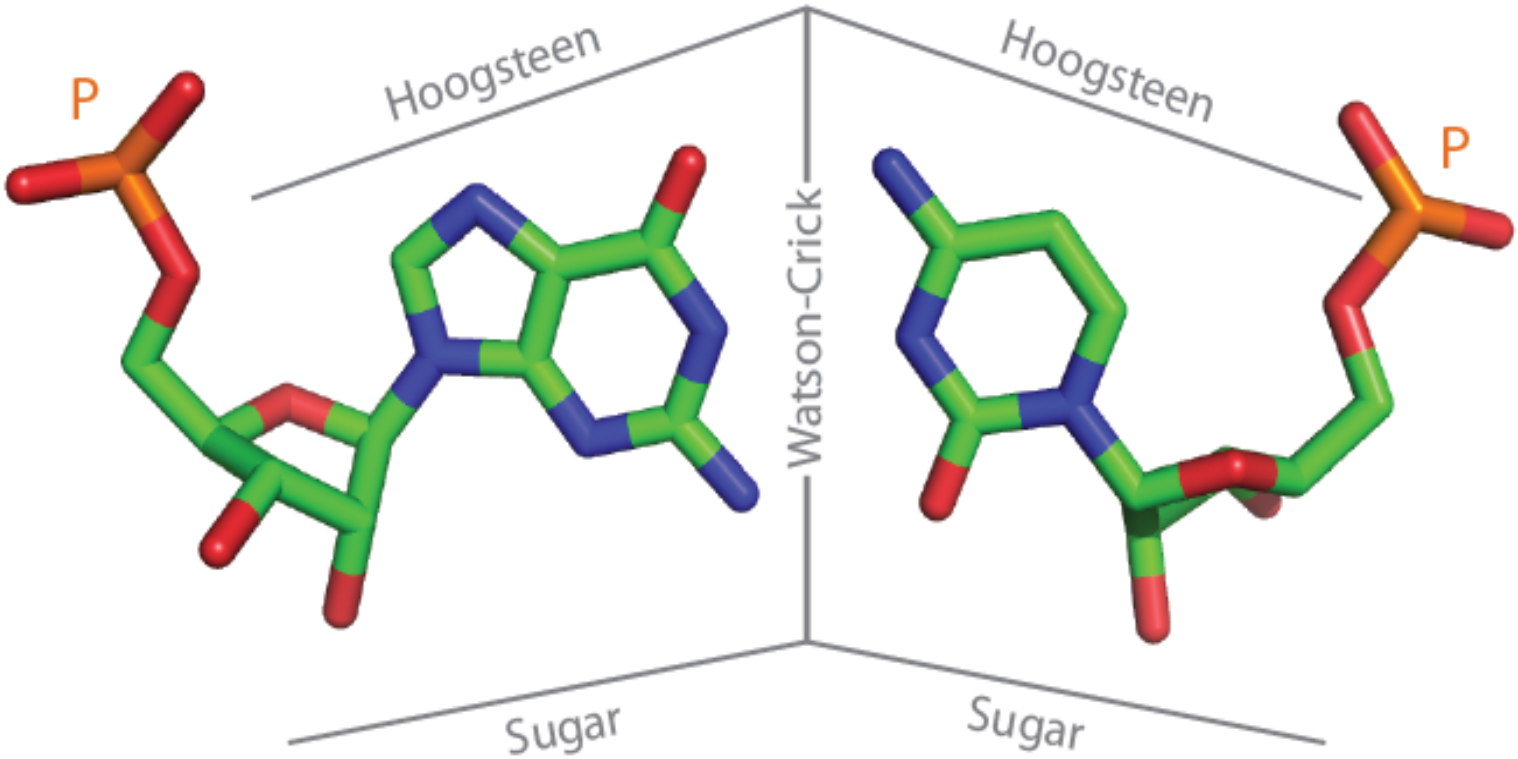


C4’-C5’ torsion angle

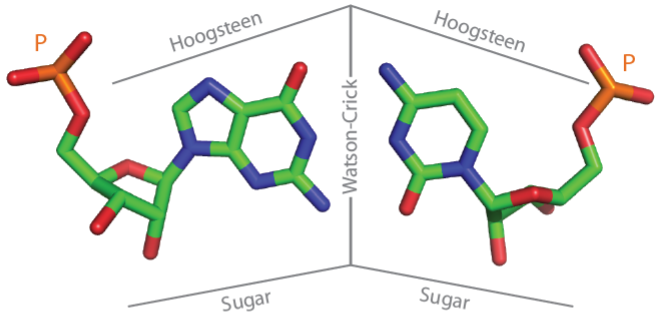


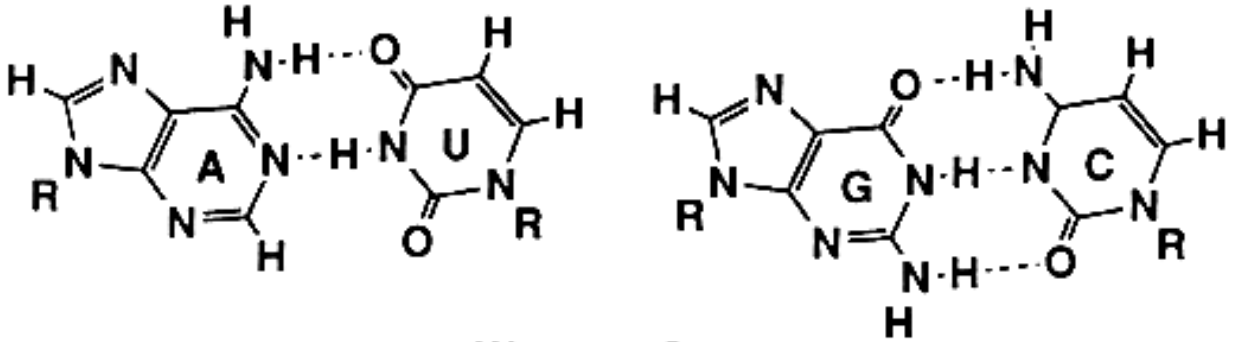
+sc ap -sc

Base pairs

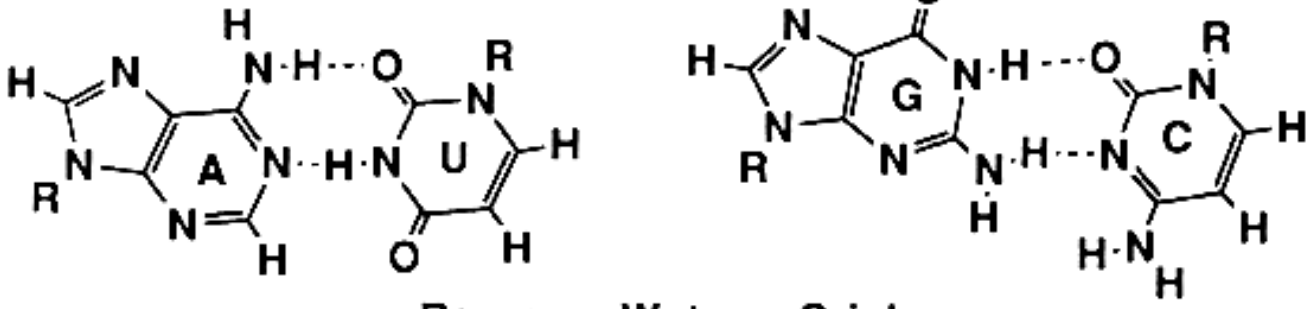


Base pairs



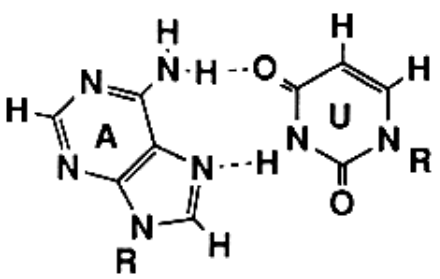


Watson-Crick

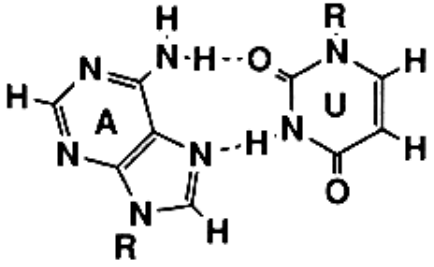


Reverse Watson-Crick

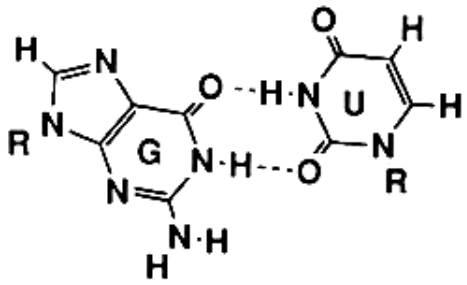
Hoogsteen, Wobble



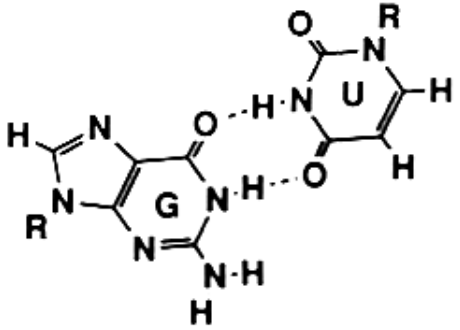
A•U Hoogsteen



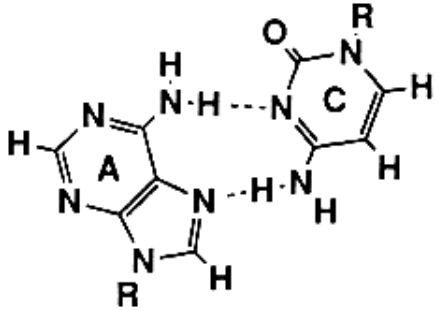
A•U Reverse Hoogsteen



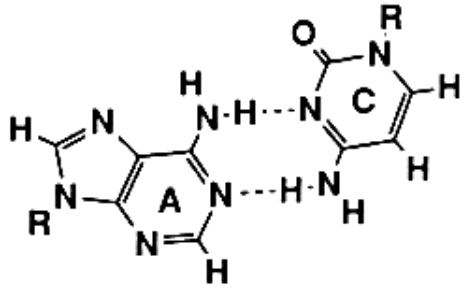
G•U Wobble



G•U Reverse Wobble

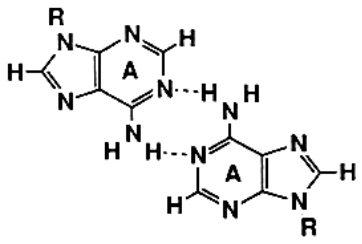


A•C Reverse Hoogsteen

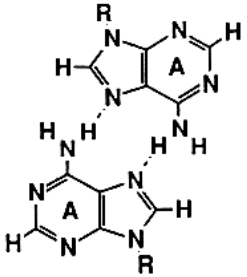


A•C Reverse Wobble

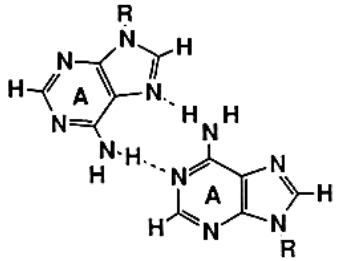
Homopurines



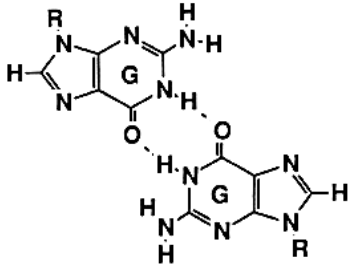
A•A N1-amino, symmetric



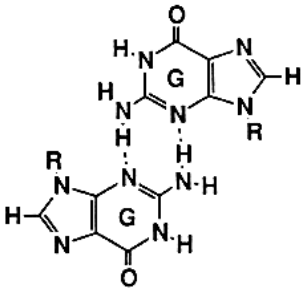
A•A N7-amino, symmetric



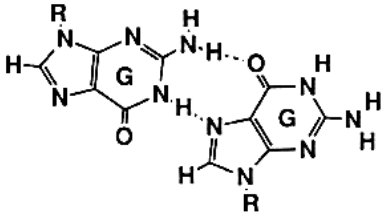
A•A N1-amino, N7-amino



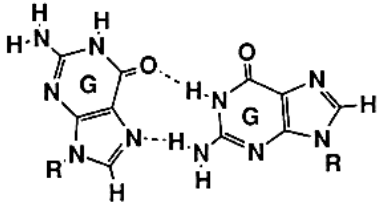
G•G N1-carbonyl, symmetric



G•G N3-amino, symmetric

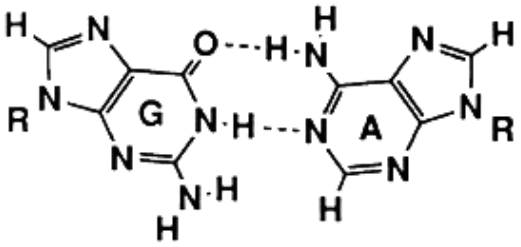


G•G N7-N1, carbonyl-amino

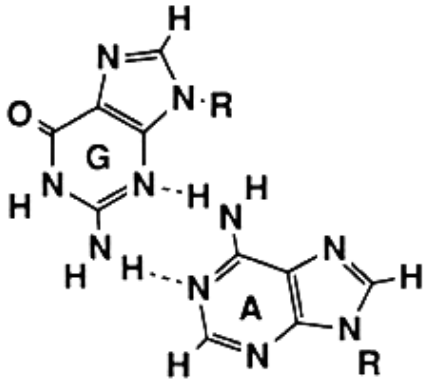


G•G N1-carbonyl, N7-amino

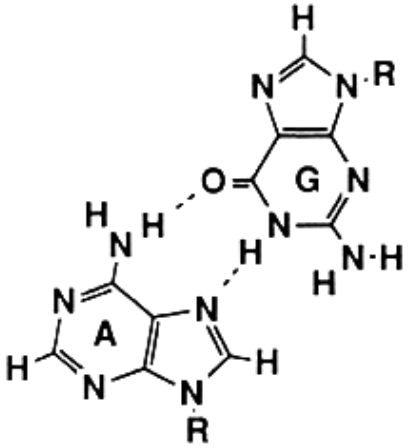
Heteropurines



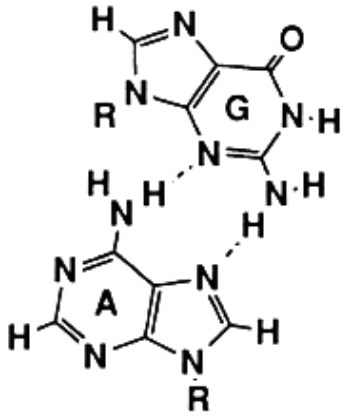
G•A N1-N1,
carbonyl-amino



G•A N3-amino,
amino-N1

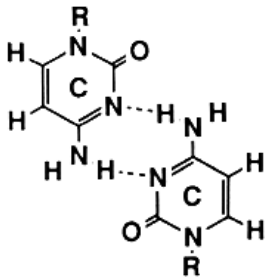


A•G N7-N1,
amino-carbonyl

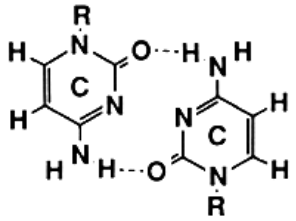


A•G N7-amino,
amino-N3

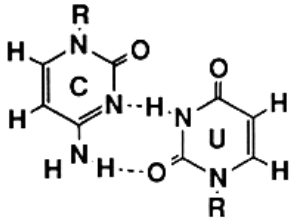
Pyrimidines



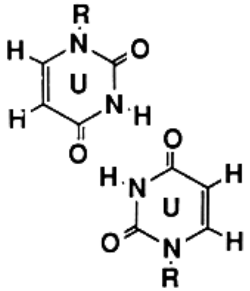
C•C N3-amino, symmetric



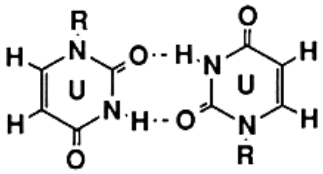
C•C carbonyl-amino, symmetric



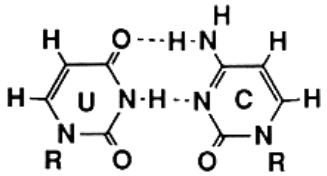
C•U N3-N3, 2-carbonyl-amino



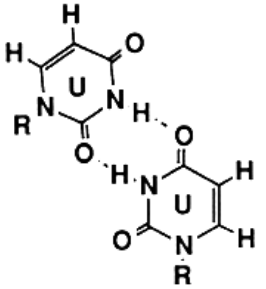
U•U 4-carbonyl-N3, symmetric



U•U 2-carbonyl-N3, symmetric

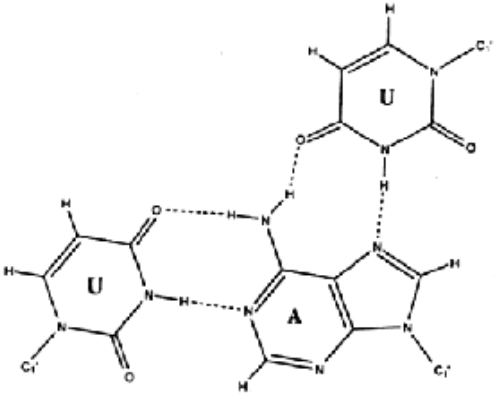


U•C N3-N3, 4-carbonyl-amino

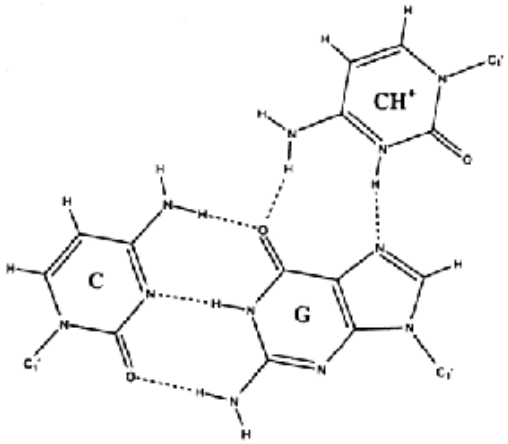


U•U 2-carbonyl-N3, 4-carbonyl-N3

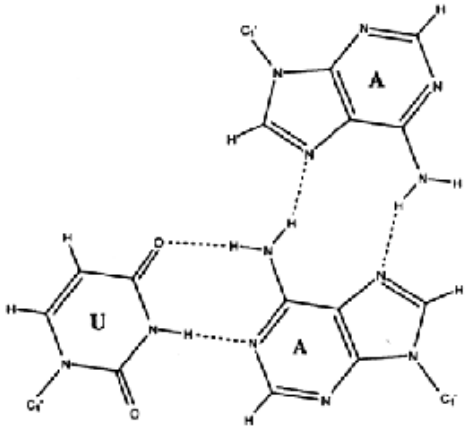
Base triples



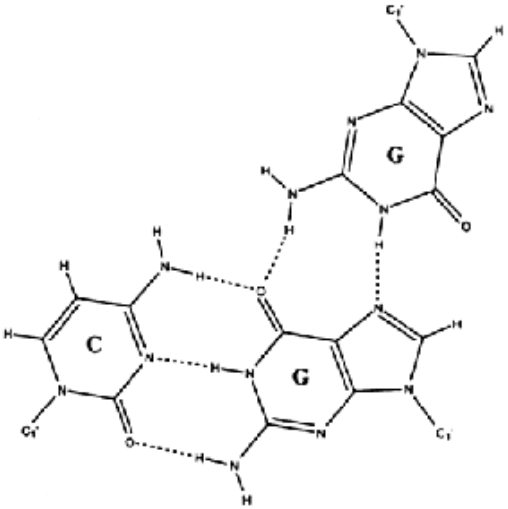
U•A•U



C•G•CH+

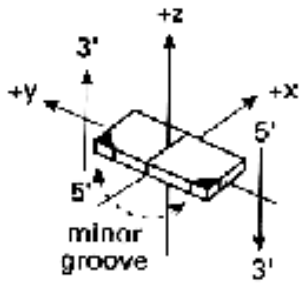


U•A•A

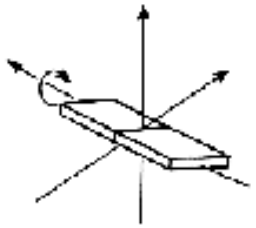


C•G•G

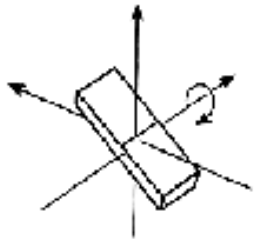
Movements of base pairs



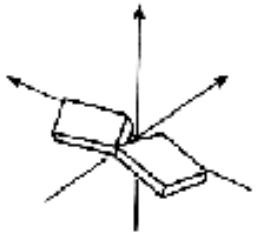
co-ordinate frame



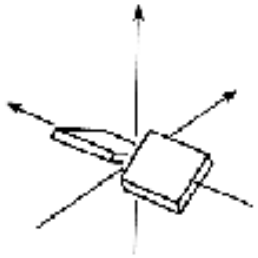
tip (θ)



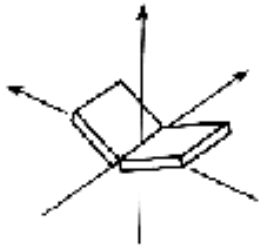
inclination (η)



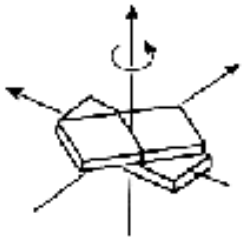
opening (σ)



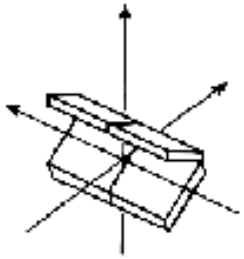
propeller (π)



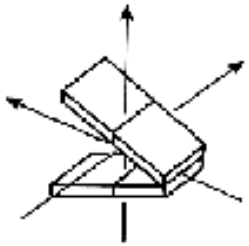
buckle (κ)



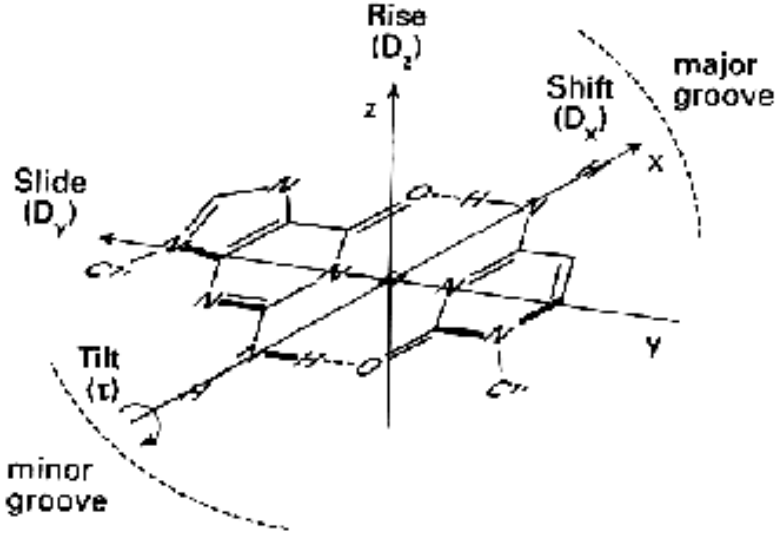
twist (Ω)



roll (P)



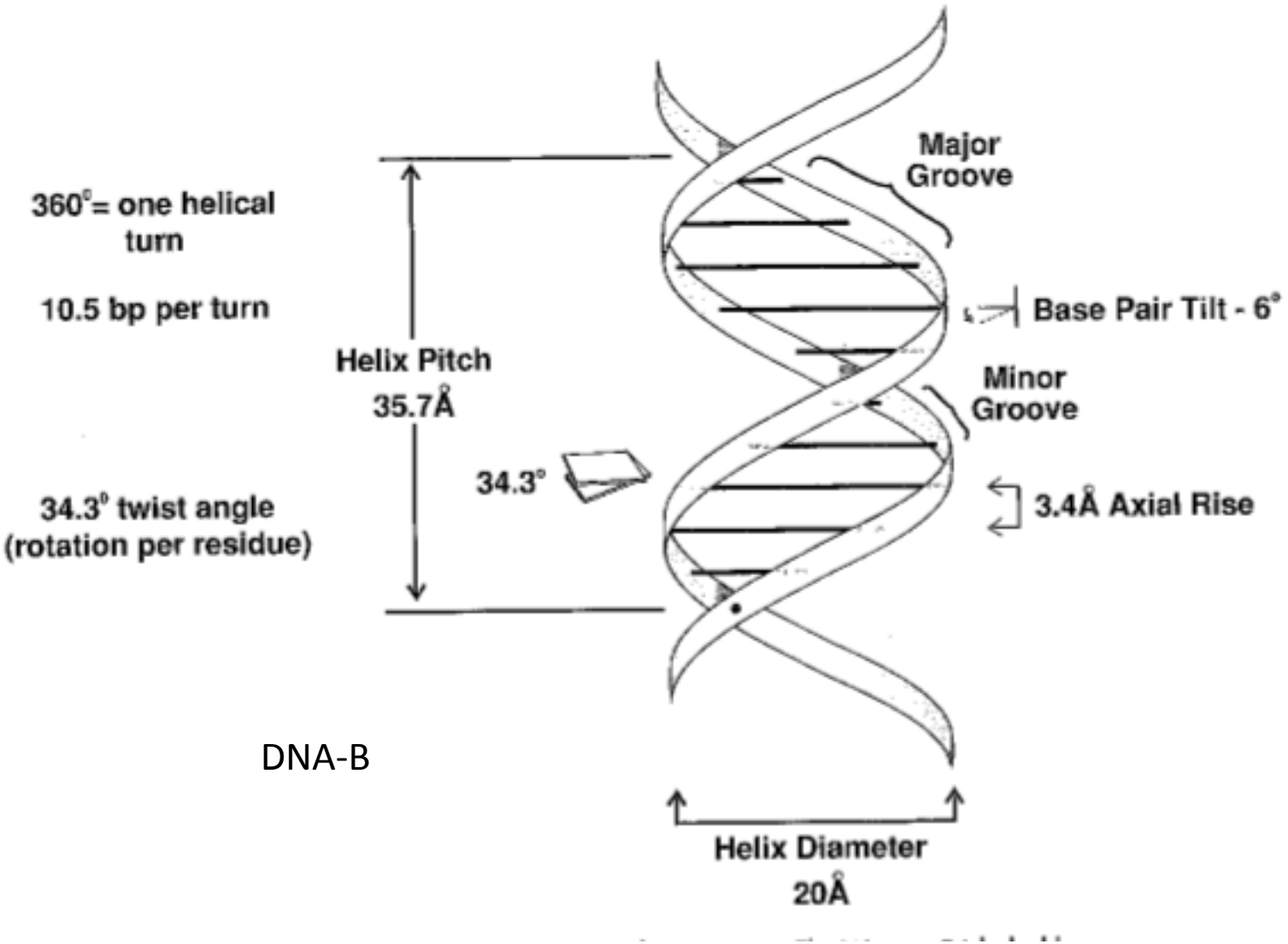
tilt (τ)



Helical parameters

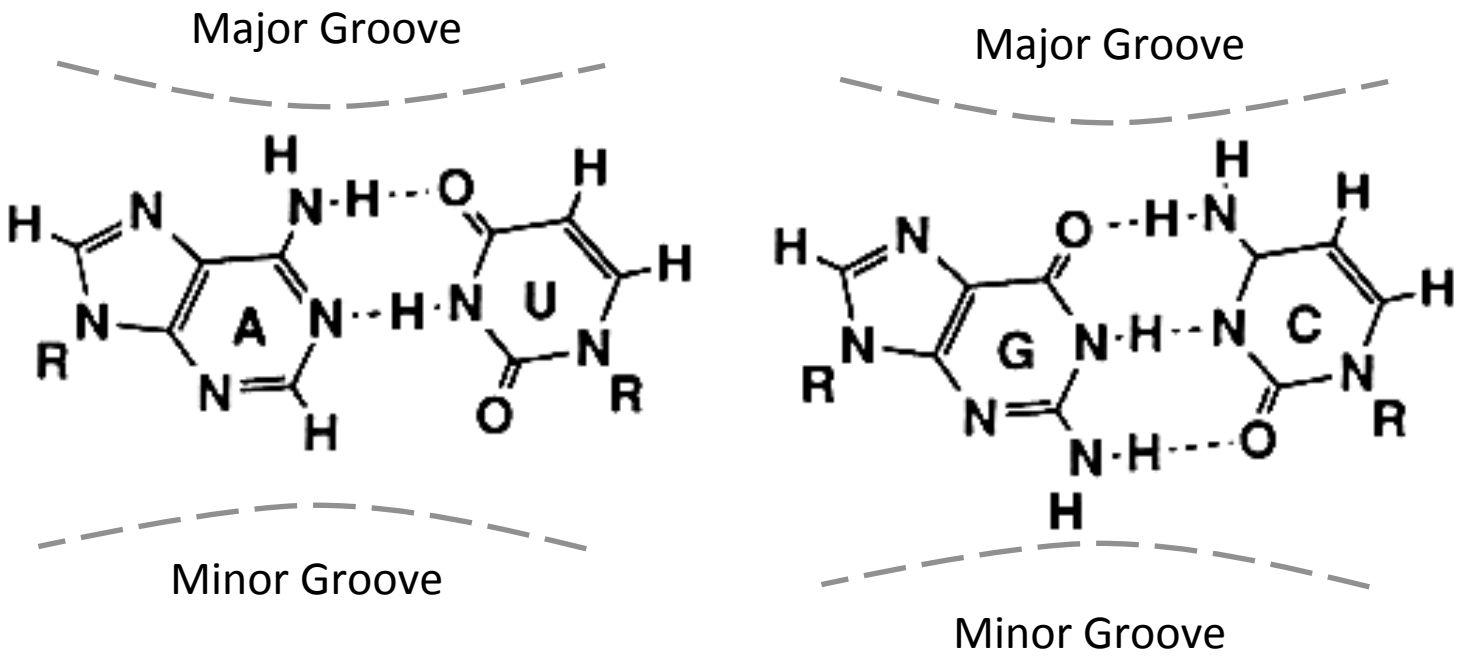
- Sense. Helical rotation of the double helix. Clock- or anticklockwise
- Residues per turn. Number of residues in one helical turn.
- Axial rise. Distance between adjacent planar bases.
- Helix pitch. Length of one helical turn.
- Diameter of the helix. Width across the helix.
- Rotation per residue or Twist angle. Angle between two adjacent base pairs.

Helical parameters



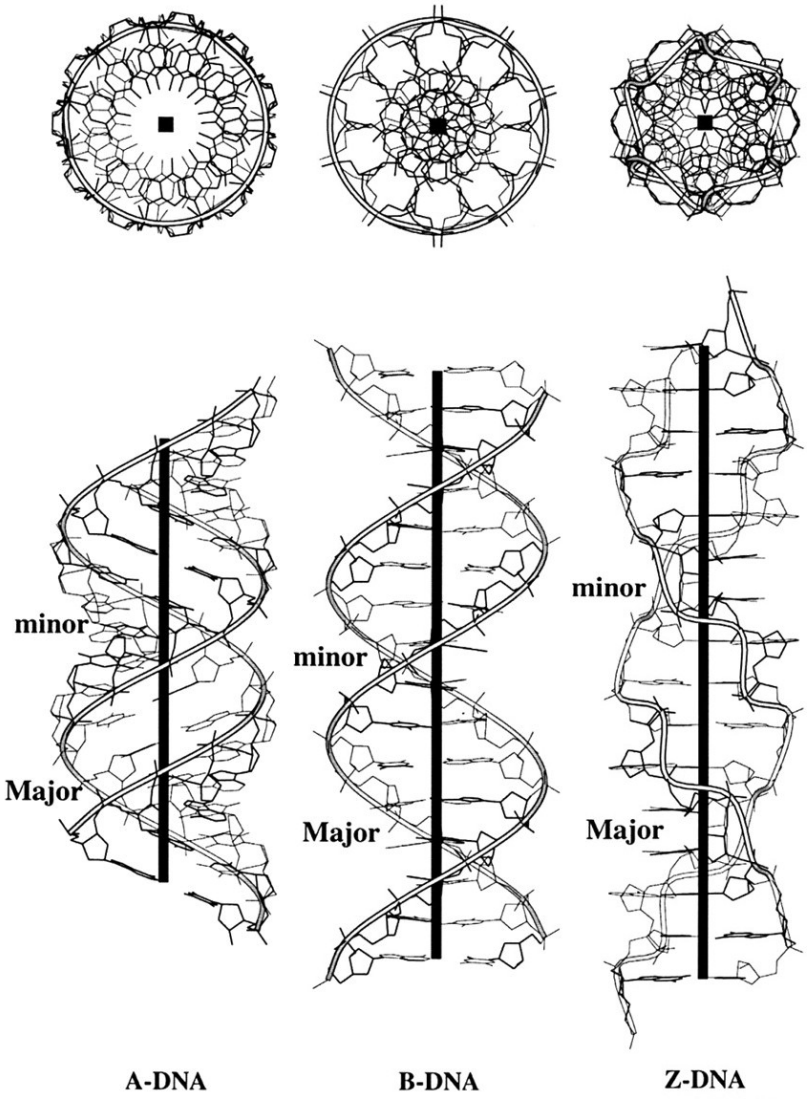
DNA-B

Major VS Minor Groove



Nucleic Acid Helical Structures

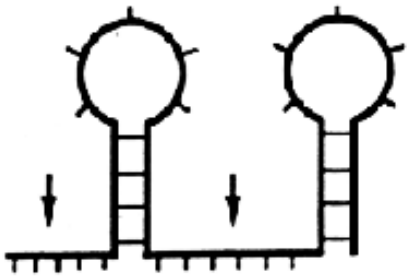
Geometry attribute:	A-form	B-form	Z-form
Helix sense	right-handed	right-handed	left-handed
Repeating unit	1 bp	1 bp	2 bp
Rotation/bp	33.6°	35.9°	60°/2
Mean bp/ turn	11	10.5	12
Inclination of bp to axis	+19°	−1.2°	−9°
Rise/bp along axis	2.4 Å (0.24 nm)	3.4 Å (0.34 nm)	3.7 Å (0.37 nm)
Rise/turn of helix	24.6 Å (2.46 nm)	33.2 Å (3.32 nm)	45.6 Å (4.56 nm)
Mean propeller twist	+18°	+16°	0°
Glycosyl angle	anti	anti	pyrimidine: anti,
			purine: syn
Sugar pucker	C3'-endo	C2'-endo	C: C2'-endo,
			G: C2'-exo
Diameter	23 Å (2.3 nm)	20 Å (2.0 nm)	18 Å (1.8 nm)



Secondary structures in RNA

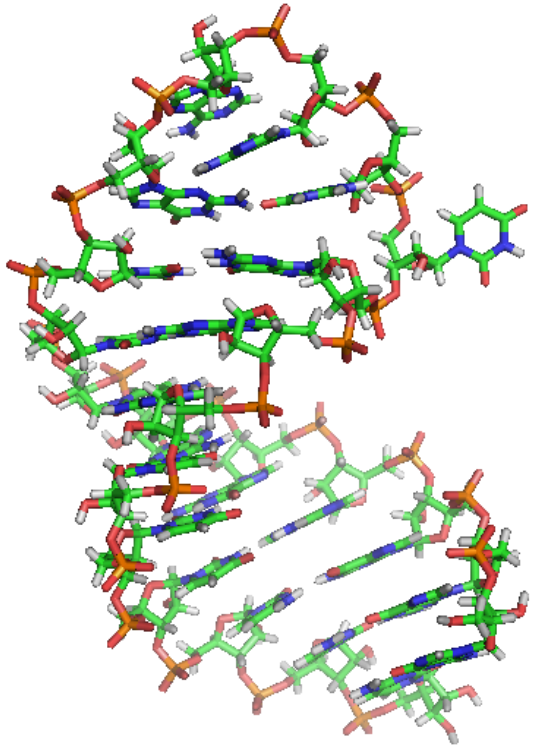
a. DUPLEXES

b. SINGLE STRANDED REGIONS



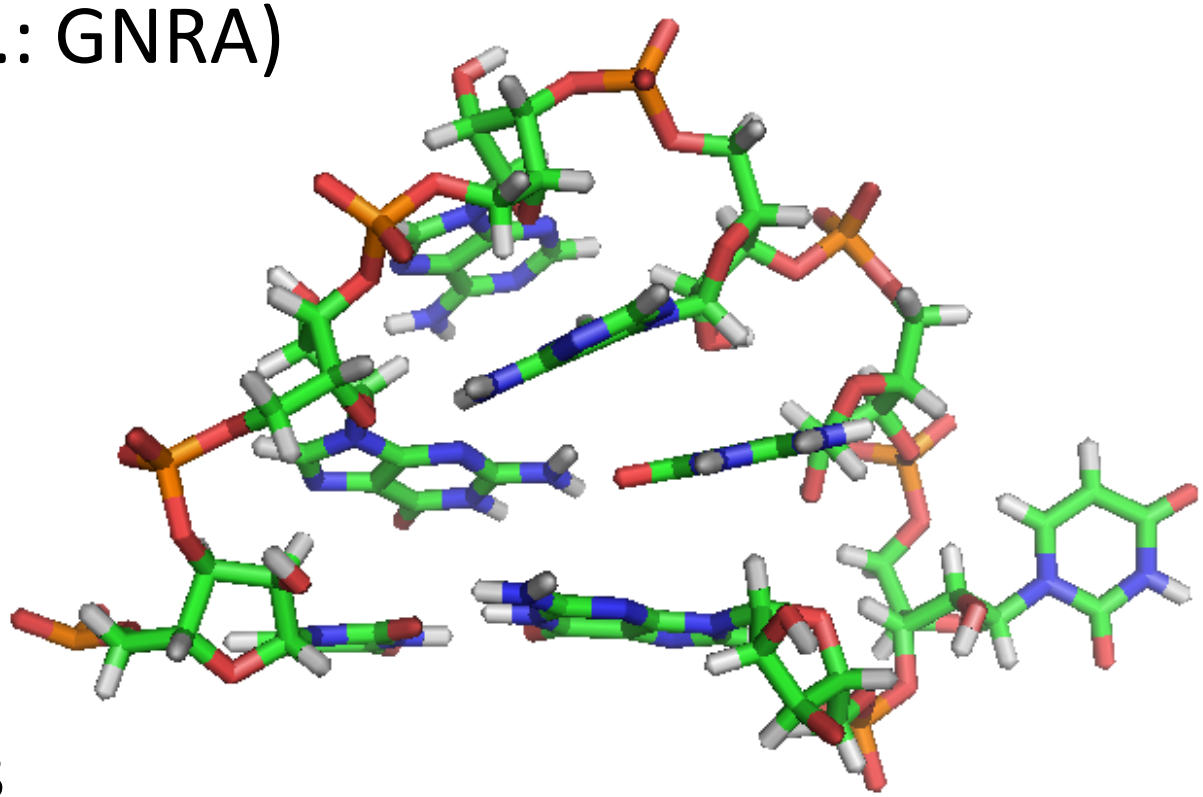
Apical loops (hairpins)

c. HAIRPINS

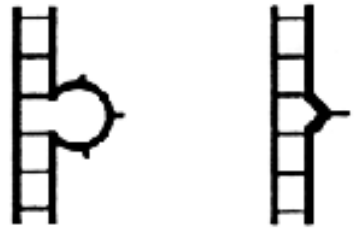


Several kinds of apical loops

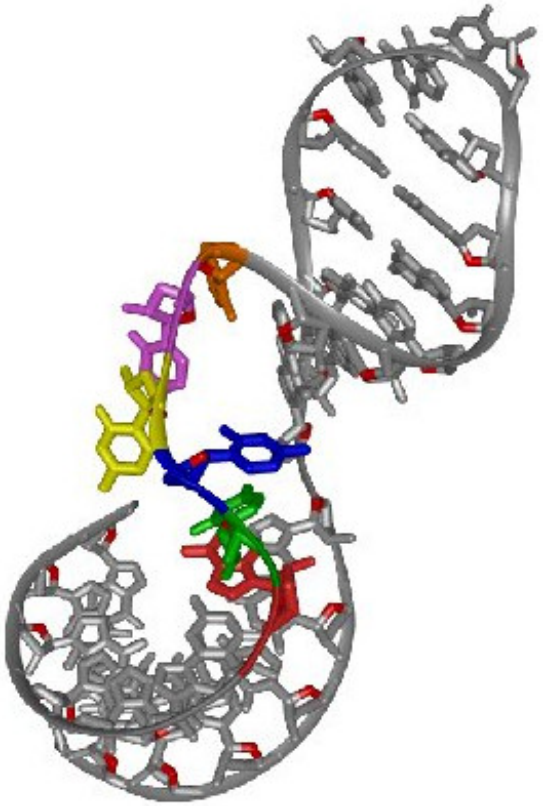
- Minimum: triloops
- Tetraloops (e.g.: GNRA)
- Pentaloops
- Hexaloops
- .
- .
- .
- whatever loops



d. **BULGES**

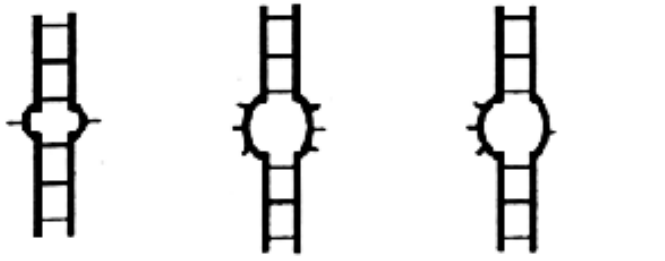


BULGE SINGLE-BASE BULGE



e. INTERNAL LOOPS

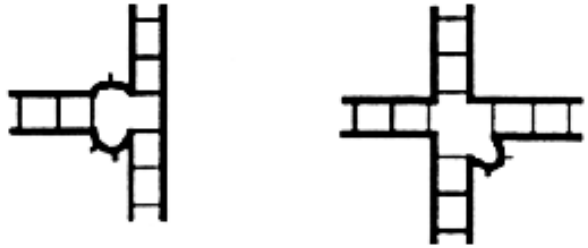
f. JUNCTIONS



MISMATCH

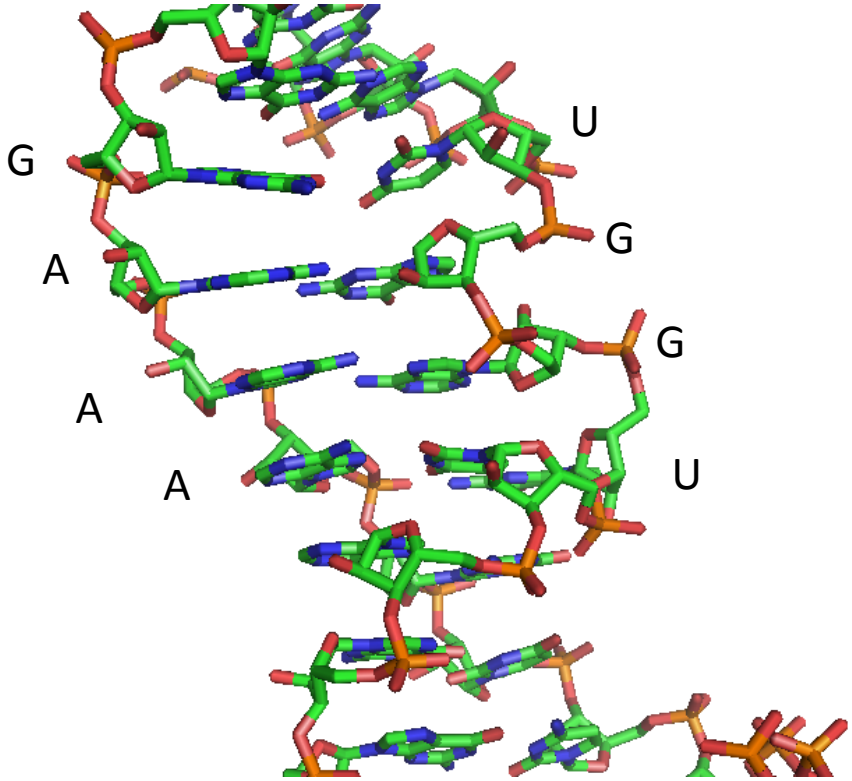
SYMMETRIC INTERNAL LOOP

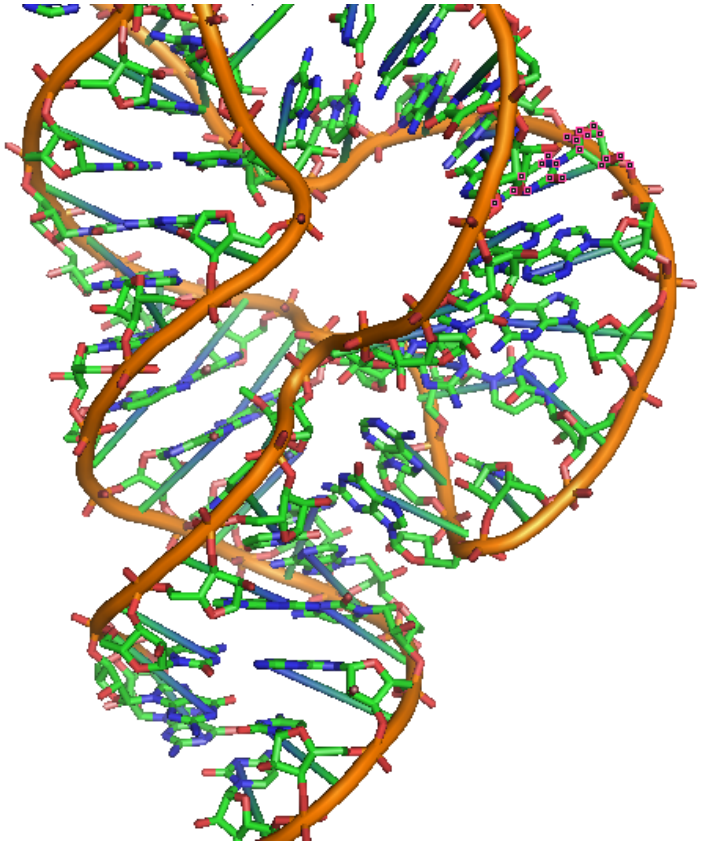
ASYMMETRIC INTERNAL LOOP



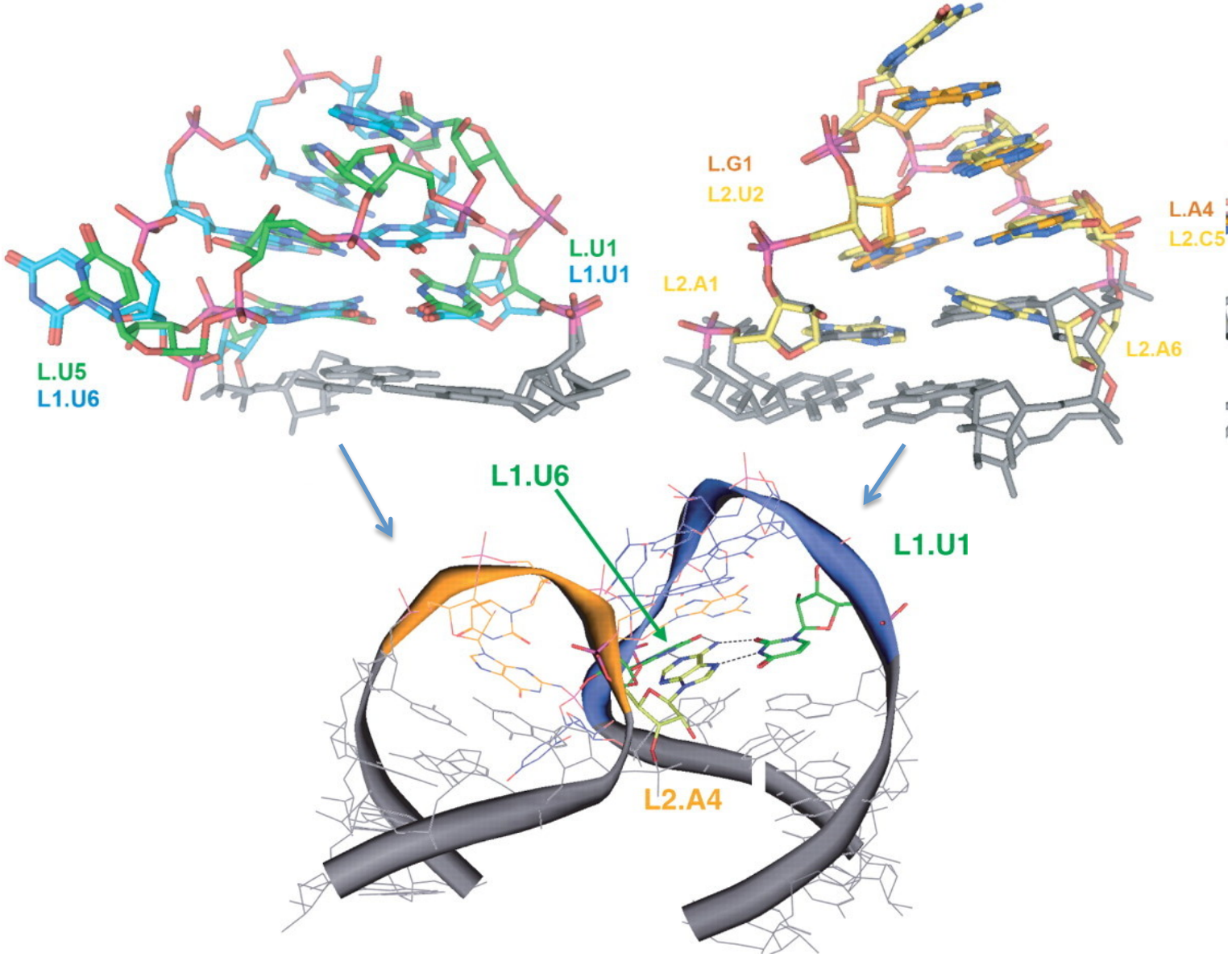
THREE STEM

FOUR STEM

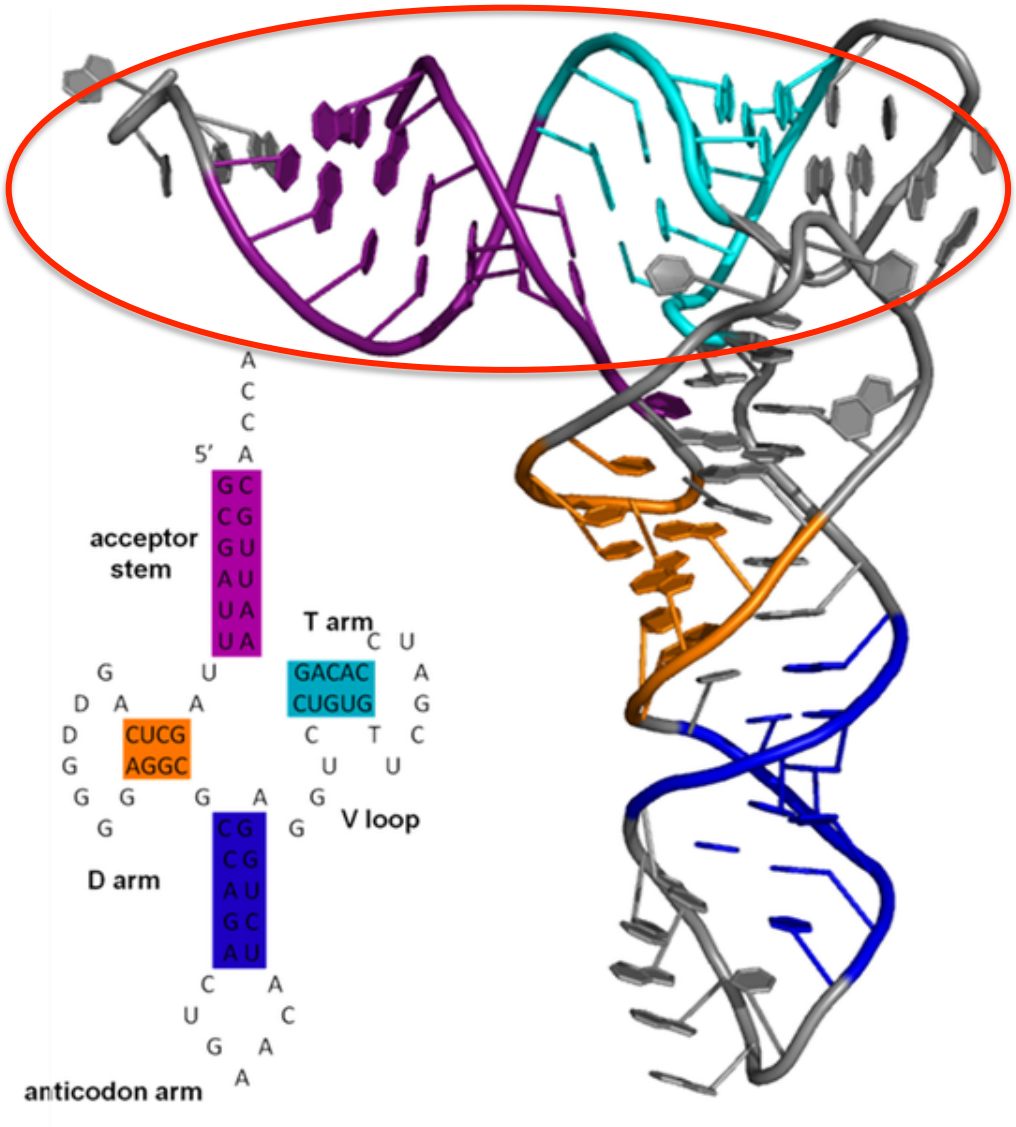




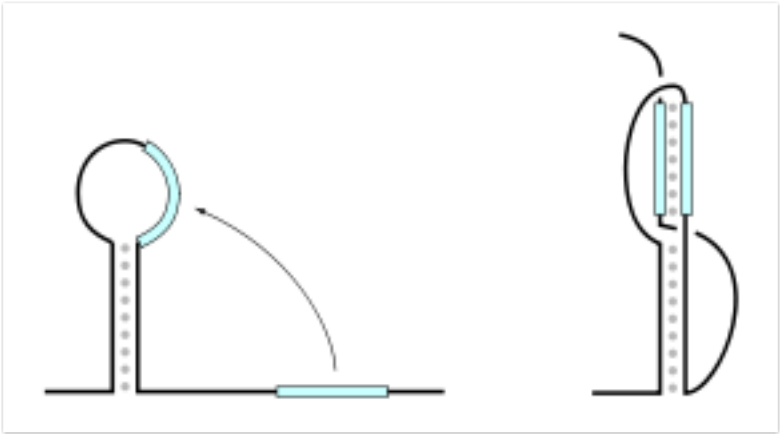
Loop-loop interactions

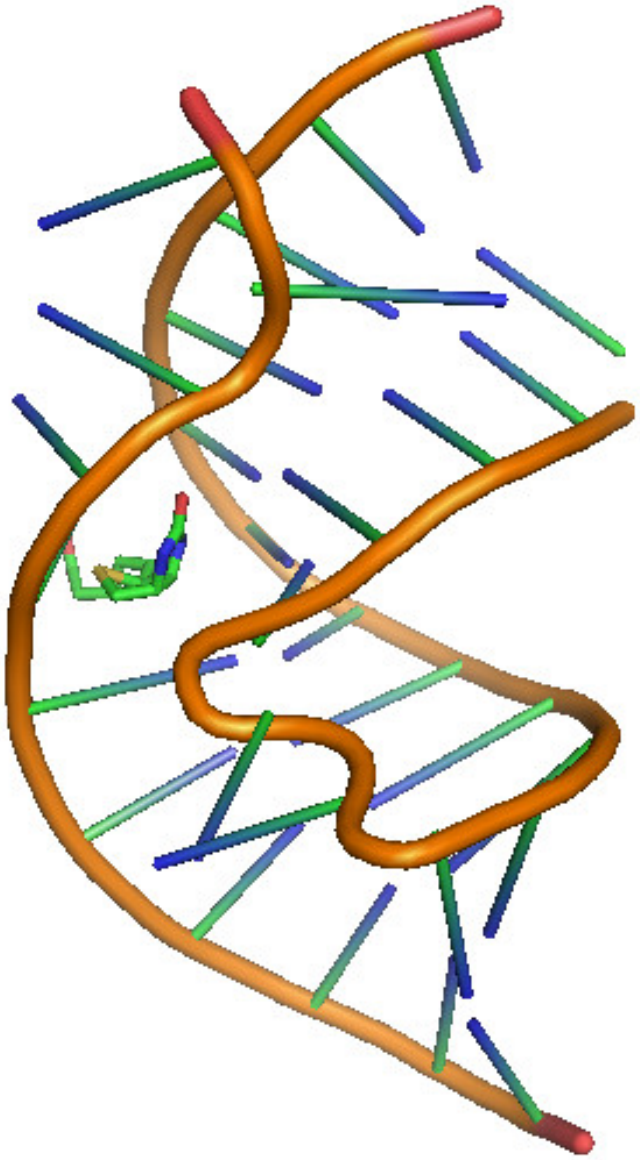


Coaxial stacking



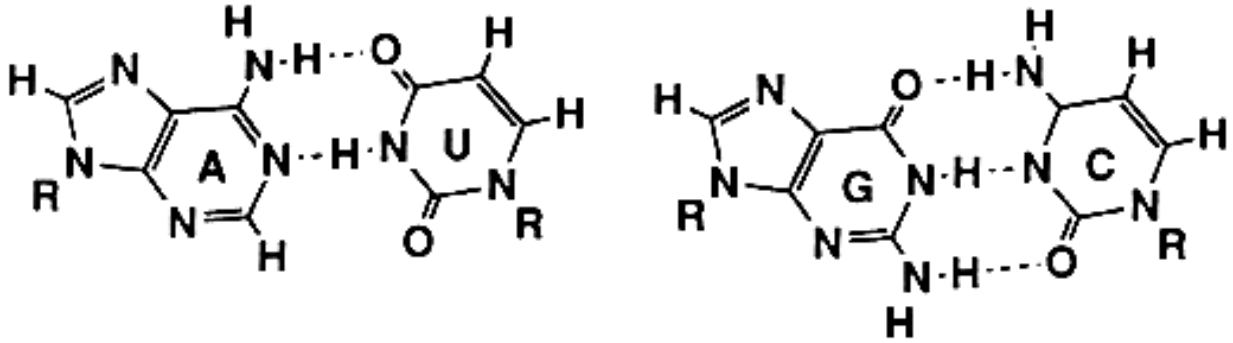
Pseudoknots





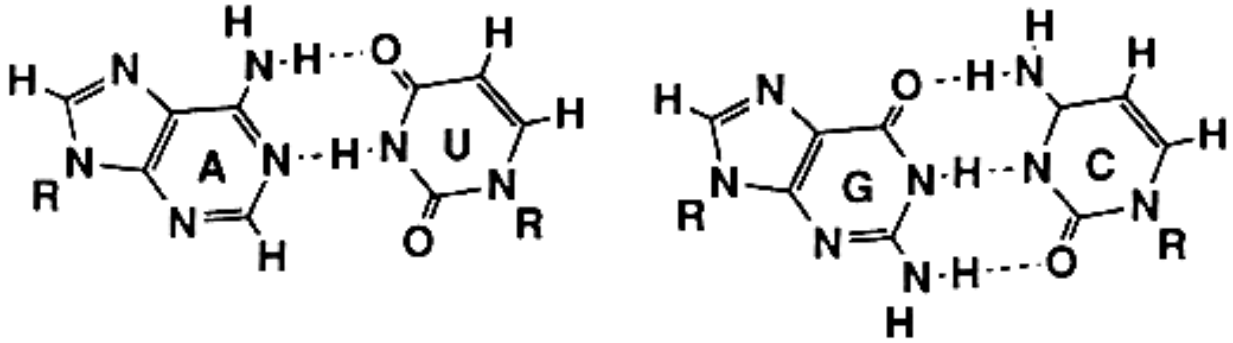
Forces that drive RNA folding

- 1) Hydrogen bonds

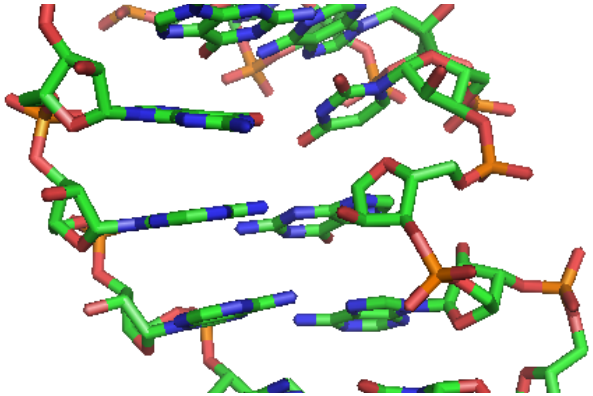


Forces that drive RNA folding

- 1) Hydrogen bonds

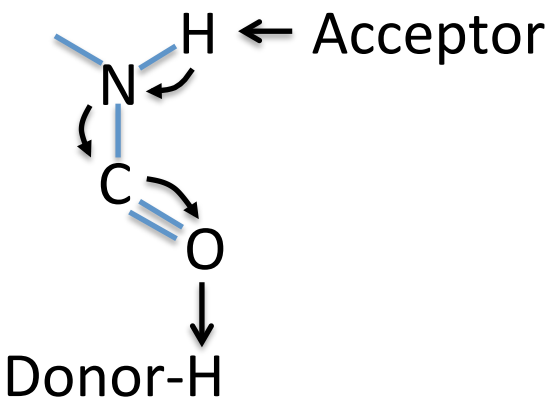


- 2) Stacking interactions



Hydrogen Bonding

- Short, noncovalent, highly directional between an electronegative atom (-C=O, -N:) and a H atom bound to an electropositive atom (-N-H).
- 20-30 times weaker compared to covalent bonds (3-7 kcal/mol VS 80-100 kcal/mol)
- Additive and cooperative



Base Stacking

- Base stacking = Van der Waals forces + Hydrophobic interactions
- 4-15 kcal/dinucleotide
- Van der Waals = dipole-dipole interactions + London dispersion forces

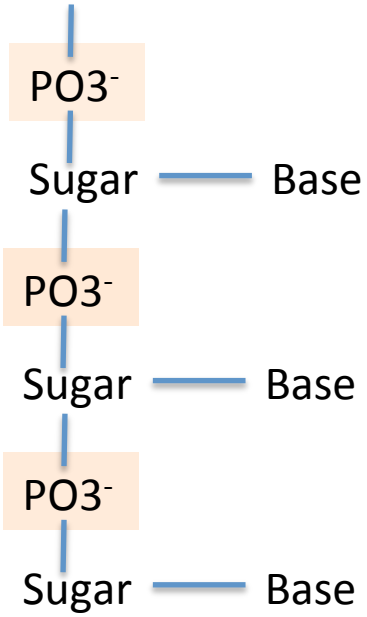
Dinucleotide base pairs	Stacking energies (kcal/mol/stacked pair)
GC · GC	-14.59
AC · GT	-10.51
TC · GA	-9.81
CG · CG	-9.69
GG · CC	-8.26
AT · AT	-6.57
TG · CA	-6.57
AG · CT	-6.78
AA · TT	-5.37
TA · TA	-3.82

Base Stacking

- Isodesmic reaction
- Additive
- Diffusion controlled reaction
- Weak force controlled process
- $E_{tot} = E_d + E_p + E_{el}$
- E_{tot} = Total energy
- E_d = Dispersion energy
- E_p = Polarization energy
- E_{el} = Electrostatic energy ($\cong$80% E_{tot})

Other interactions

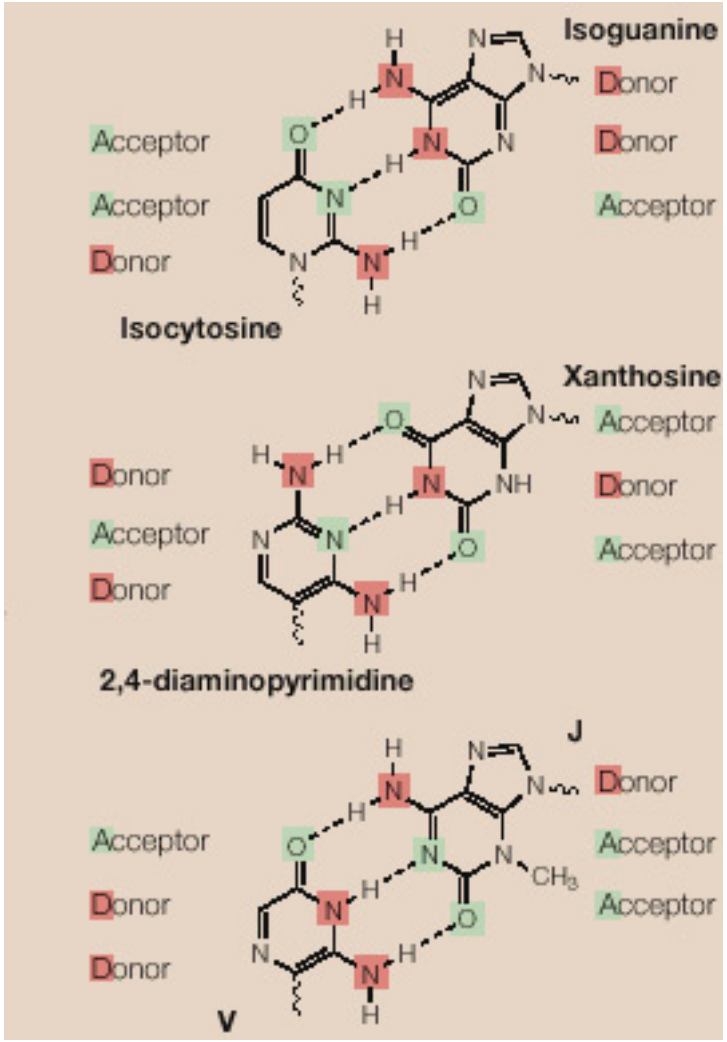
- RNA is a polianion



- Metal ion coordination ($Mg\ 2^{+}$)
  - Diffusion controlled
  - Though water or direct contact
  - $PO3^{-}$ or bases

Bonus track

- Why are there four letters in the genetic alphabet?



NDB

http://ndbserver.rutgers.edu/

ndb

Site Index

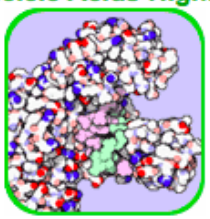
WELCOME TO THE
NUCLEIC ACID DATABASE
a repository of three-dimensional structural information about nucleic acids

- Atlas
- Deposit Data
- Download Data
- Search
- Education
- Standards
- Tools
- Links

Number of Released Structures:
6425 Structures
Last Update: 20-Feb-2013

Search the NDB by ID
Enter an NDB ID or PDB ID
Search
Search for Released Structures

Nucleic Acids Highlight



About NDB

The NDB follows the dictionaries and formats used by the Worldwide Protein Data Bank. Please see www.wwpdb.org for format announcements and documentation.

Archive of NDB newsletters

The NDB is supported by funds from the National Science Foundation and the Department of Energy.

In citing the NDB please refer to: H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. (1992) The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids. Biophys. J., 63, 751-759.

ndbadmin@ndbserver.rutgers.edu