

Strukturní bioinformatika – nukleové kyseliny

Olomouc, 15.11.2023

RNDr. Miroslav Krepl, Ph.D.

Nukleové kyseliny - historie

- Proteiny (**1838**); Lipidy (**1815**); Sacharidy (~**1800**) – *exaktní popis, izolace*

➤ **Používané tisíce let**



Nukleové kyseliny - historie

- NK objev v roce **1889**, jádro buňky obsahuje látky „kyselé povahy“ – *velmi malé množství*
- **Žádné komerční využití, problém izolovat → ignorovány**
- **1919**, složení (cukr, báze, fosfát)
- **1944**, transformace živých bakterií pomocí nukleové kyseliny z usmrcených buněk
- **1952**, role DNA jako nositelky dědičné informace, Chargaffovo pravidlo

Nukleové kyseliny - historie

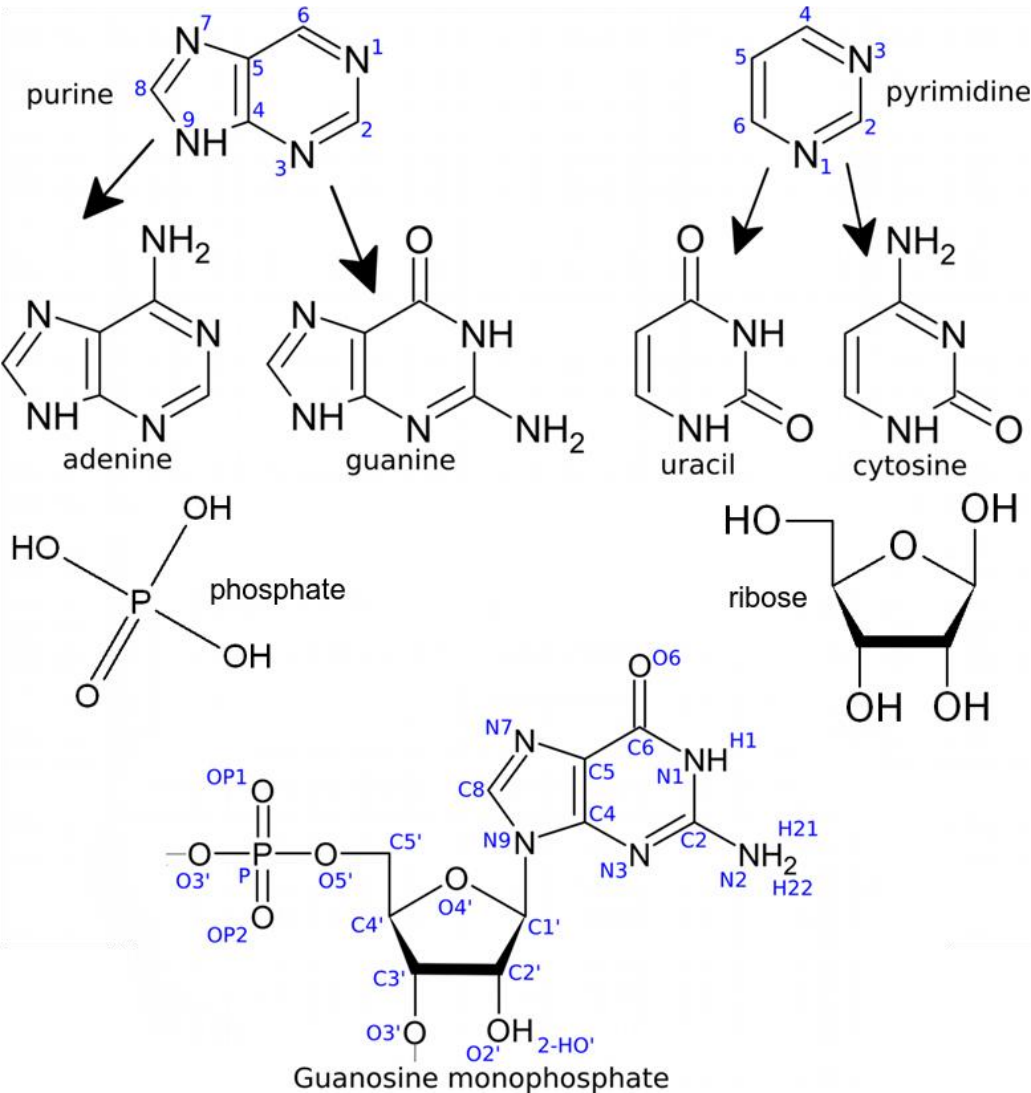
- **1953**, struktura DNA
- **1957**, centrální dogma molekulární biologie
 - Problémy se syntézou a purifikací NA
 - Nebylo automaticky přijímáno...
- Difrakce vláken; krystaly dinukleotidů; plný helix DNA až **1980**, struktura tRNA **1974**



Nukleové kyseliny - složení

- „Nukleová kyselina“ obvykle značí polymer (polynukleotid) – nezahrnuje ATP, GTP, UTP apod.
- Základní jednotka nukleotid (báze, cukr, fosfát)
- Cukr: ribosa, 2-dexoxyribosa (RNA vs. DNA)
- Báze variabilní – adenin (A), guanin (G), cytosin (C), uracil (U, RNA), thymin (T, DNA)
- Formy zápisu: $5'-ATGCAT-3'$; $5'-ApTpGpCpApT-3'$;
(syntéza od 5' konce)
- **duplex:** $(ATGCAT)_2$ **DNA:** $d(ATGCAT)_2$ **RNA:** $r(AUGCAU)_2$

Nukleové kyseliny - složení



- Základní báze odvozeny z pyrimidinu a purinu (dusíkaté heterocykly)

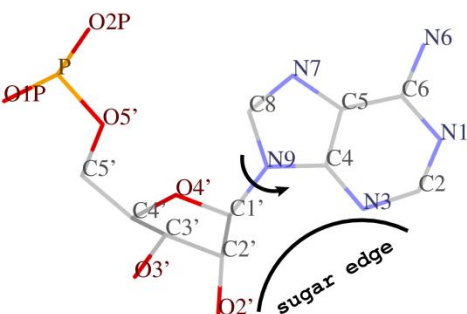
- Rozdíl v exocyklických chemických skupinách

- N-glykosidická (cukr/báze), esterová (cukr/fosfát), fosfodiesterová (nukleotid/nukleotid) vazba

- Chemicky jednodušší oproti proteinům (4 dusíkaté báze vs. 20 jedinečných aminokyselin)

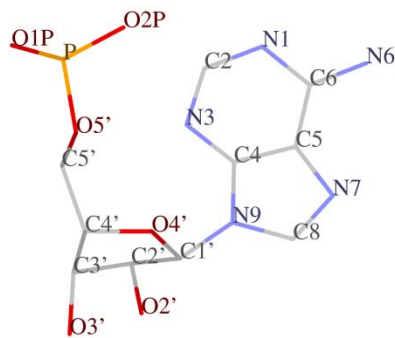
Nukleové kyseliny - struktura

- Geometricky složitý popis (zejména páteře)
- Báze planární, ale schopné rotace
- Rotace báze kolem glykosidické vazby – úhel χ

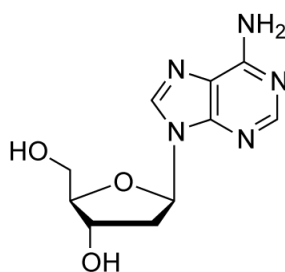


- Chi úhel definován atomy O4'-C1'-N9(N1)-C4(C2)
- **Anti** ($150^\circ - 300^\circ$) a **syn** ($-30^\circ - 150^\circ$) oblasti
- Syn výskyt u purinů $\rightarrow$ Syn/Anti bias ve strukturách

anti

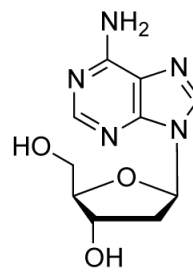


syn

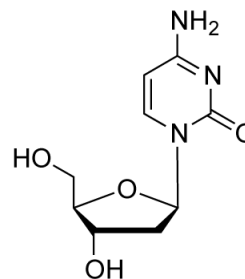


anti

deoxyadenosine (a purine nucleoside)

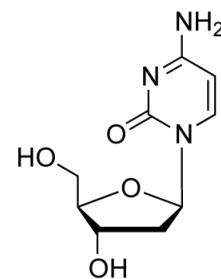


syn



anti

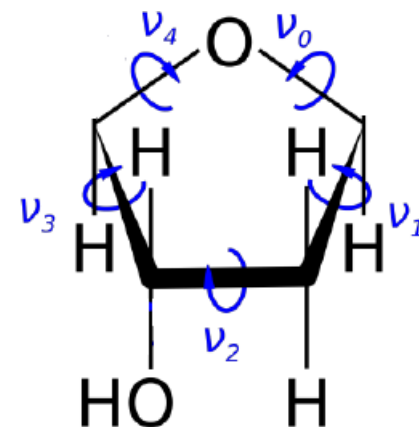
deoxycytidine (a pyrimidine nucleoside)



syn

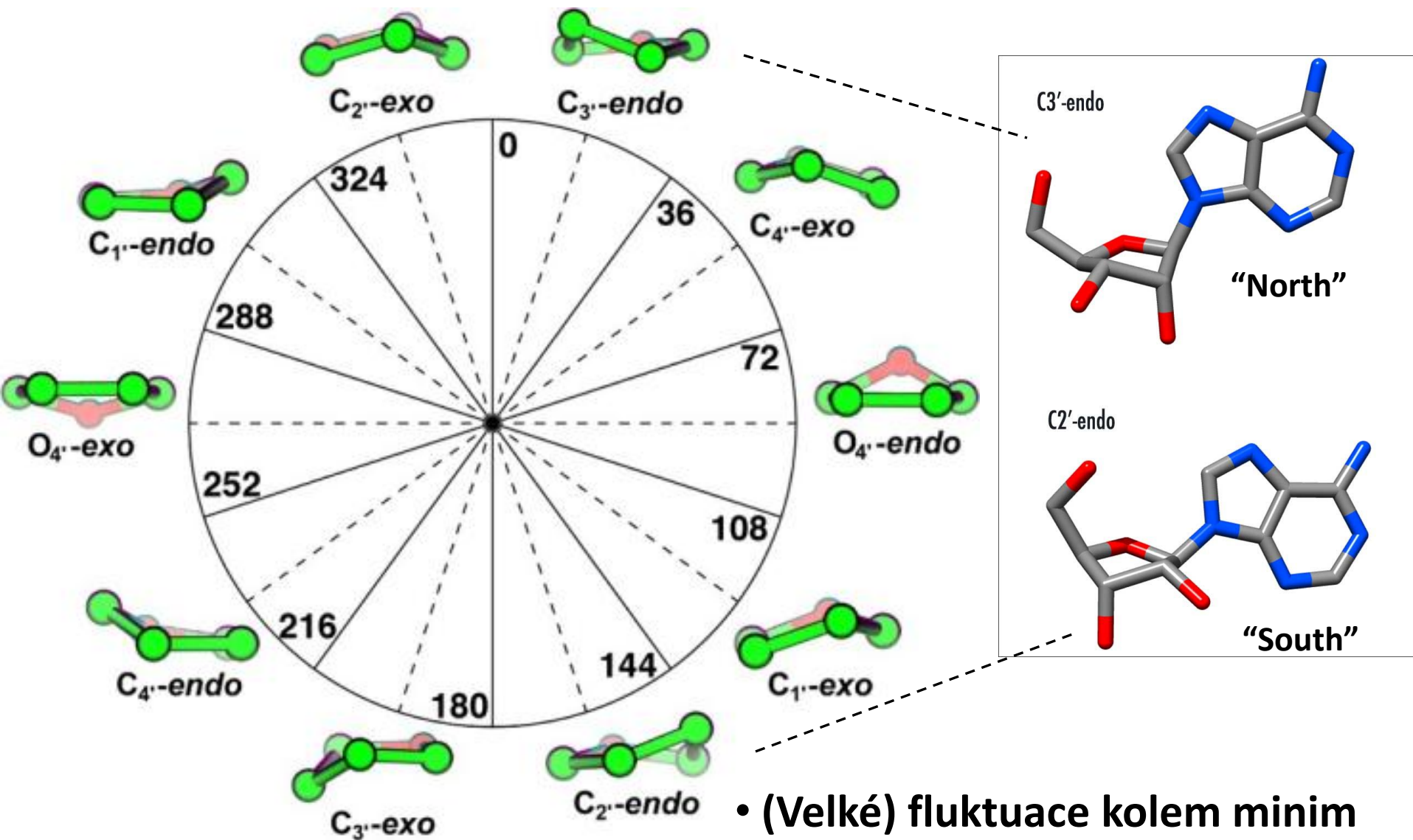
Nukleové kyseliny - struktura

- Ribosa a 2-deoxyribosa jsou cyklické ale jejich 5-členný kruh není planární – popis vyžaduje 4 dihedrální úhly
- Popis pomocí **puckeru** – definuje, který atom kruhu se nachází mimo pomyslnou rovinu a kterým směrem
- Matematické vyjádření pomocí pseudorotačního úhlu
- V praxi jsou nejčastější C3'-endo (RNA) a C2'-endo (DNA) puckery jež mají nejnižší energii



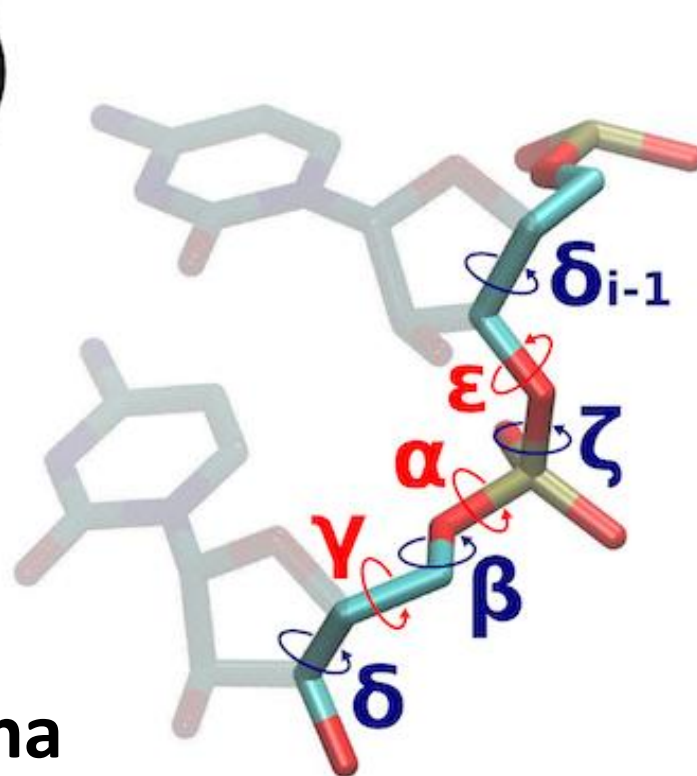
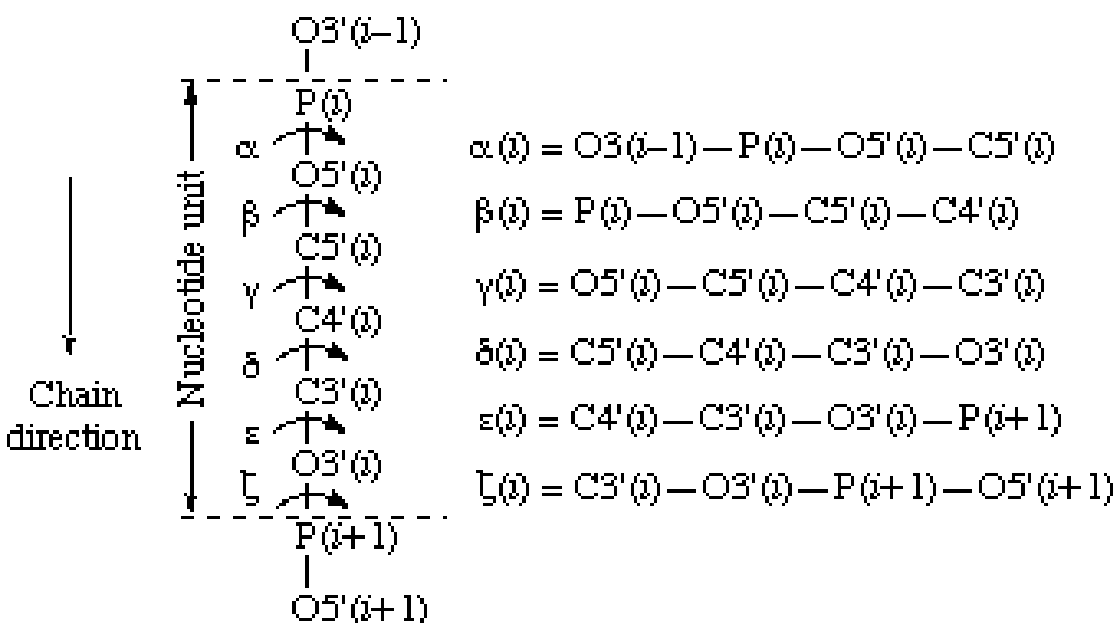
$$P = \text{atan} \frac{(\nu_2 + \nu_4) - (\nu_1 + \nu_3)}{2\nu_0(\sin 36 + \sin 72)}$$

Nukleové kyseliny - struktura



Nukleové kyseliny - struktura

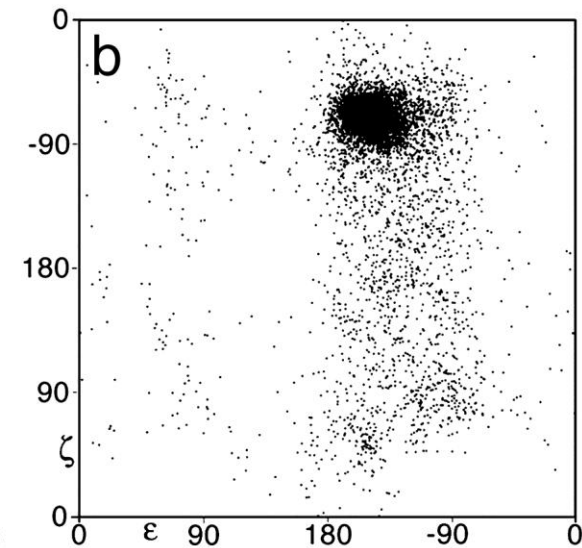
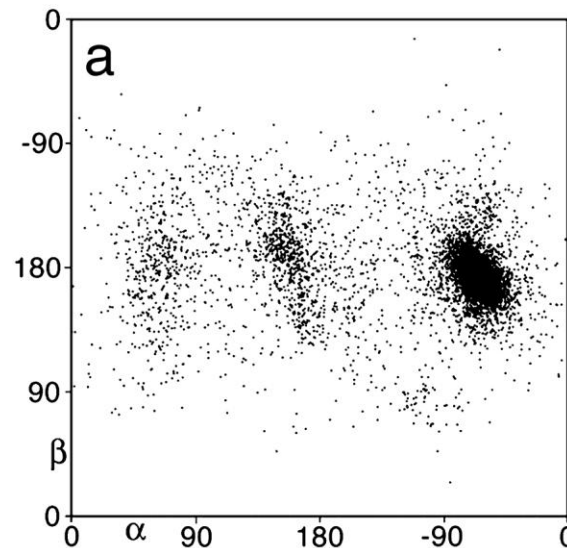
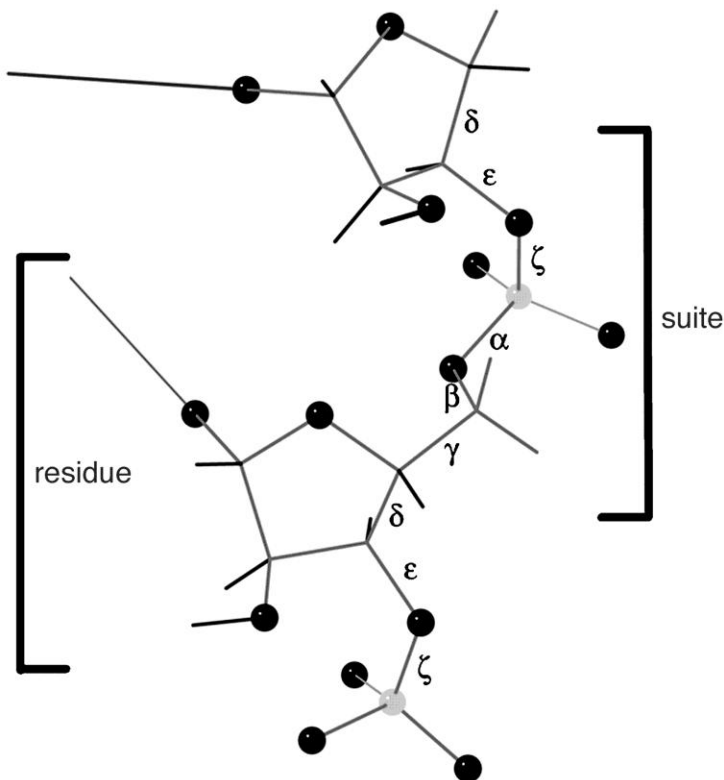
- Cukr-fosfátová páteř popsána 6 dihedrálními úhly – $\alpha, \beta, \gamma, \delta, \epsilon, \zeta$



- Pokračuje po celé délce vlákna**
- Oproti proteinům (phi, psi, omega), úplný popis páteře NA ohrožuje duševní zdraví**

Nukleové kyseliny - struktura

- Úhly jsou silně korelovány mezi sebou i mezi nukleotidy
- Definice páteřních „suites“ podle Richardsonové



- Jednotlivé kombinace páteřních úhlů tvoří jednotlivé suity

Nukleové kyseliny - struktura

Table 4. The main DNA conformational classes identified in the present work

Description	<i>N</i>	Clustered torsions										Cluster number
		γ	δ	ϵ	ζ	$\alpha + 1$	$\beta + 1$	$\gamma + 1$	$\delta + 1$	χ	$\chi + 1$	
'Canonical' A-form, labeled AI	192	54	82	205	285	294	172	55	83	201	202	8
AII, A-form with an $\alpha + 1/\gamma + 1$ switch	44	52	82	195	291	149	194	182	87	204	188	19
A with δ , $\delta + 1$ close to O4'-endo	9	44	101	192	281	297	182	44	99	210	211	25
AI-BI, with δ C3'-, $\delta + 1$ C2'-endo	32	54	86	194	281	301	179	55	142	214	251	41
AI-BI, with δ O4'-, $\delta + 1$ C2'-endo	34	54	99	186	274	297	178	51	141	235	264	47
BI-AI, with δ O4'-endo	100	51	130	183	267	297	171	51	106	250	239	32
BII-AI, with an $\alpha + 1/\gamma + 1$ switch, high $\beta + 1$	9	49	146	257	186	60	224	196	90	260	200	110
BI variation in complexes	412	45	137	178	255	304	187	45	139	252	256	58
'Canonical' B-form, labeled BI	1,531	47	136	184	262	302	179	45	138	251	260	54
BII variation in complexes	269	43	140	201	216	314	156	46	140	261	253	86
BII-form	340	46	143	245	172	297	142	46	141	269	259	96
BI, with an $\alpha + 1/\gamma + 1$ switch	109	46	139	195	245	32	196	296	150	252	253	116
BI, 3'-mismatches with an $\chi + 1$ syn, $\alpha + 1/\gamma + 1$ switch	8	50	137	196	225	33	187	295	145	257	70	122
AI-BI, 3'-mismatches with $\chi + 1$ syn	14	58	91	214	280	295	176	56	139	238	67	121
Z-form, Y-R step	21	54	147	264	76	66	186	179	95	205	61	123
ZI-form, R-Y step	40	177	96	242	292	210	233	54	144	63	205	124
ZII-form, R-Y step	18	179	95	187	63	169	162	44	144	58	213	126

'Description' is a short annotation of the conformation, '*N*' is the number of dinucleotides which define the conformation, 'Clustered Torsions' are the arithmetic means calculated for the torsions used in the clustering process, with the torsions being defined in Figure 1. Bold font is used merely to indicate the three most important DNA forms.

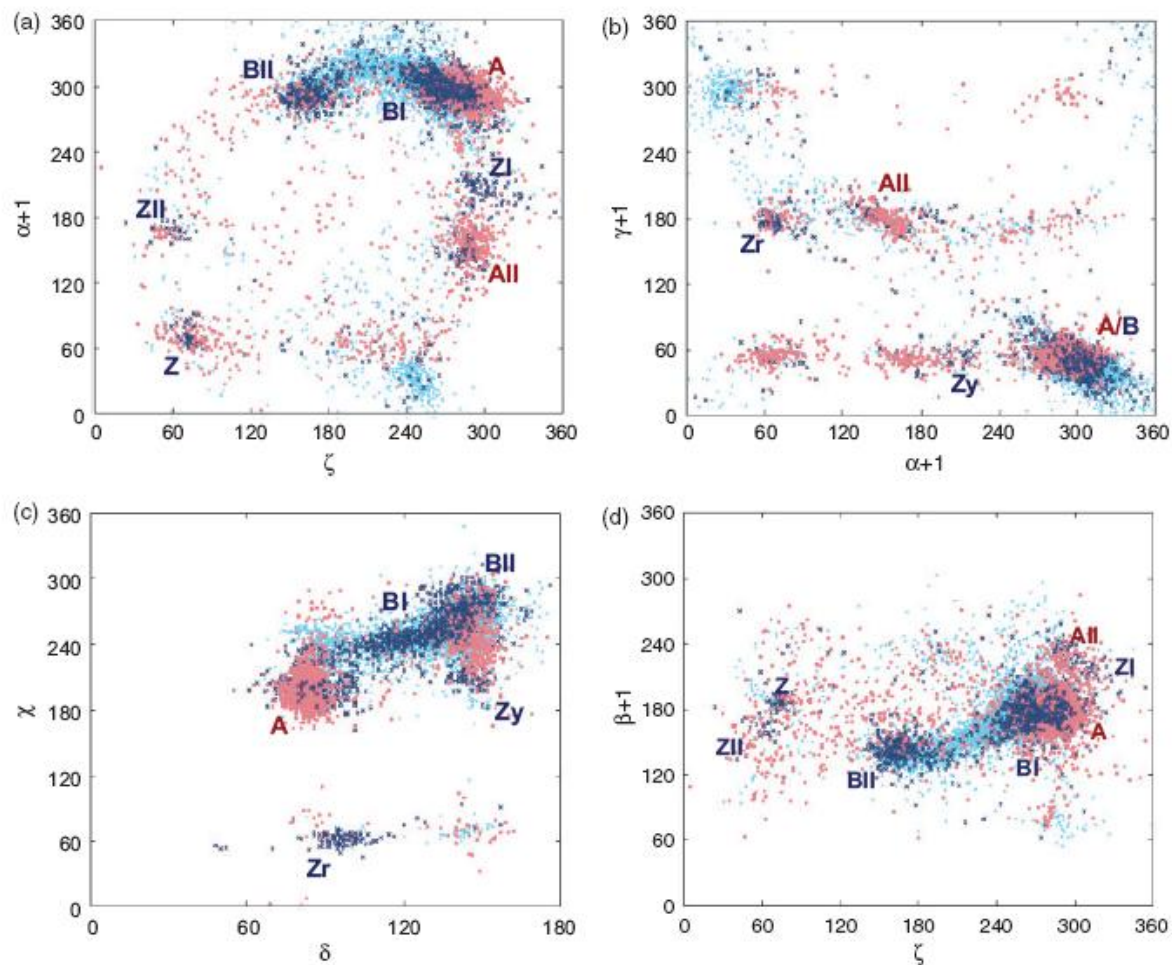
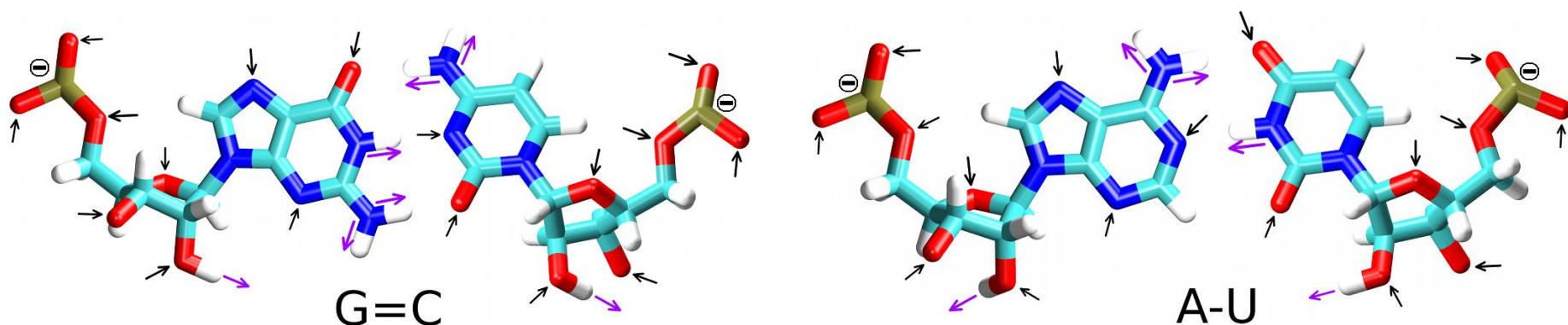


Figure 2. Two-dimensional scattergrams of torsion angles in naked DNA (Dataset 2 from Table 1, dark blue) and in DNA from complexes (Dataset 1, cyan). A, B and Z are the respective double-helical forms, r stands for purines and y for pyrimidines. The conformations of almost 4000 RNA dinucleotides are plotted as pink dots for comparison.

Nukleové kyseliny - struktura

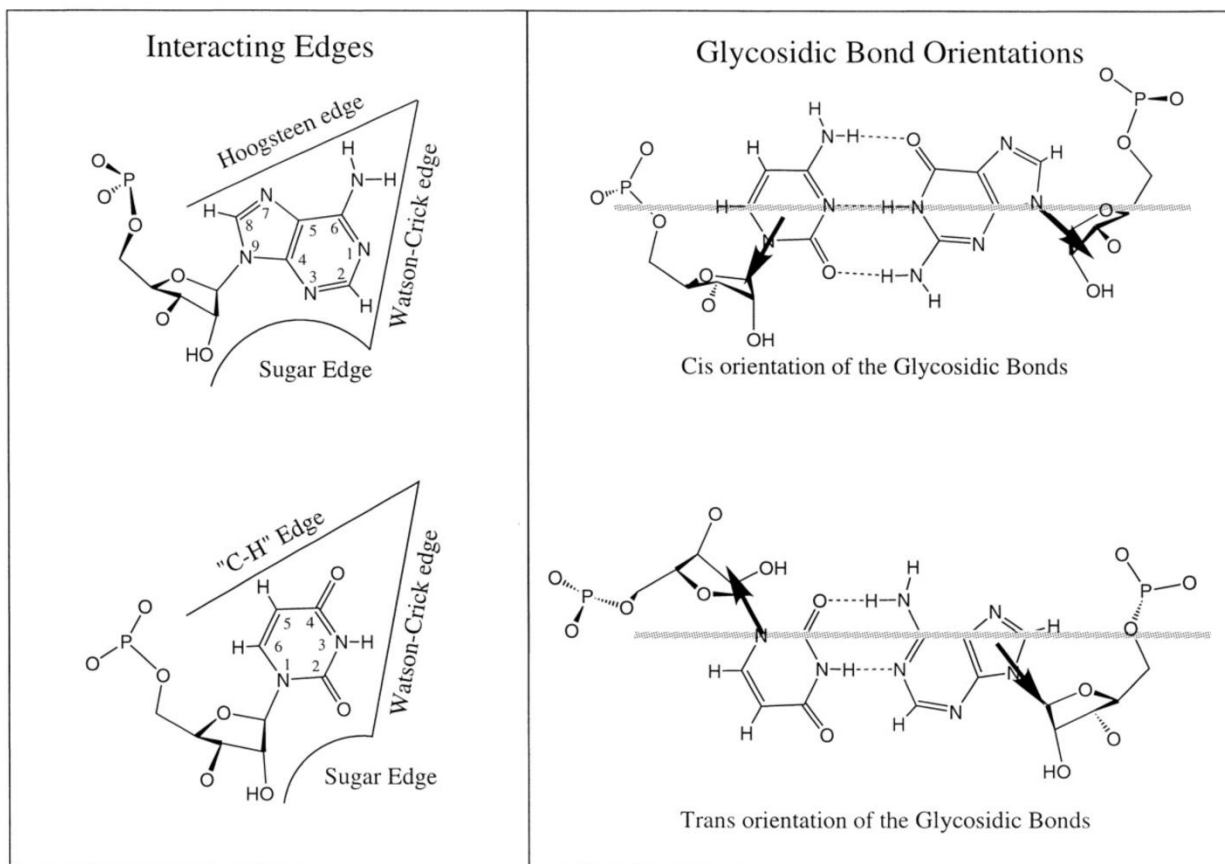
- Báze jsou chemicky téměř identické
- Jedinečnost dána odlišným potenciálem pro strukturní interakce tvořené vodíkovými vazbami (typicky párování báze/báze, ale i báze/cukr, báze/fosfát)
- Základní (kanonické) párování G/C, A/U (A/T)



- Kanonické párování (SŠ učebnice); typické hlavně pro DNA helixy

Nukleové kyseliny - struktura

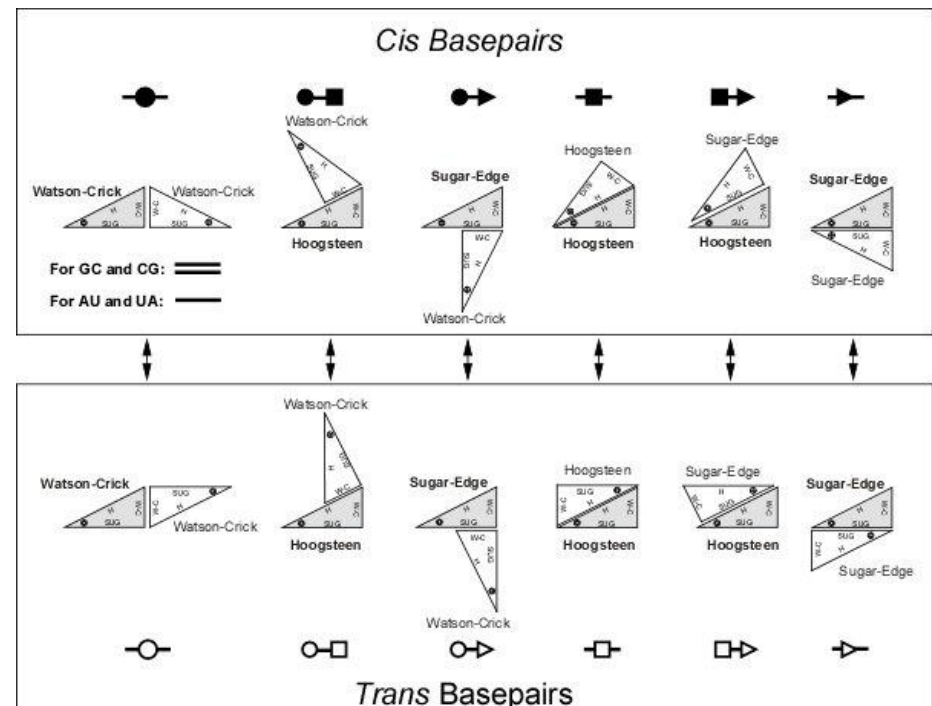
- 3 hrany nukleových kyselin – Watson-Crickova (WC), Hoogsteenova (H), cukerná (S)
- 2 různé orientace glykosidických vazeb – cis, trans



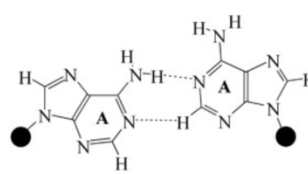
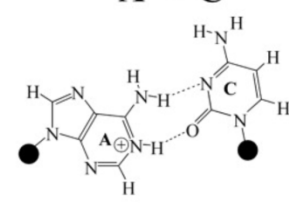
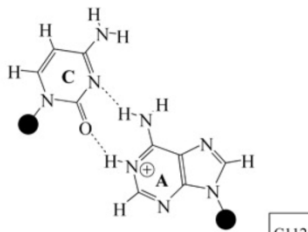
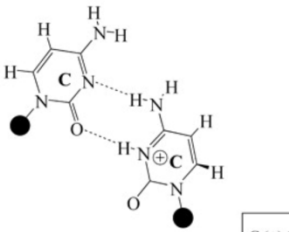
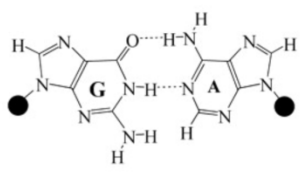
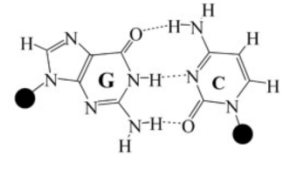
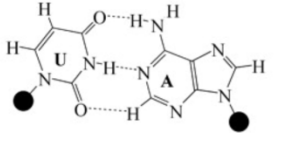
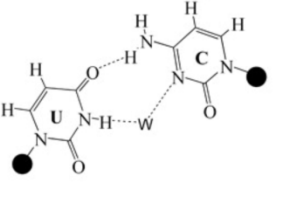
- Rozšířené u RNA
- „Nekanonické“ páry bází
- Výrazně zvyšuje schopnost nukleových kyselin tvořit rozmanité struktury

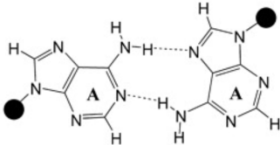
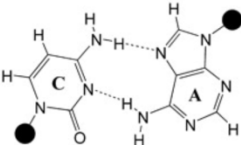
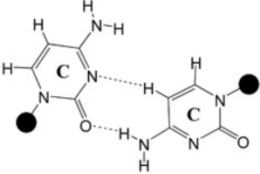
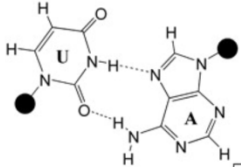
S. No	Interacting Edges	Glycosidic Bond Orientation	Local Strand Orientation	Symbolic Representation
1	Watson-Crick/ Watson-Crick	Cis	Anti-parallel	
2	Watson-Crick/ Watson-Crick	Trans	Parallel	
3	Watson-Crick/ Hoogsteen	Cis	Parallel	
4	Watson-Crick/ Hoogsteen	Trans	Anti-parallel	
5	Watson-Crick/ Sugar	Cis	Anti-parallel	
6	Watson-Crick/ Sugar	Trans	Parallel	
7	Hoogsteen/ Hoogsteen	Cis	Anti-parallel	
8	Hoogsteen/ Hoogsteen	Trans	Parallel	
9	Hoogsteen/ Sugar	Cis	Parallel	
10	Hoogsteen/ Sugar	Trans	Anti-parallel	
11	Sugar/ Sugar	Cis	Anti-parallel	
12	Sugar/ Sugar	Trans	Parallel	

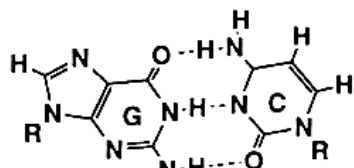
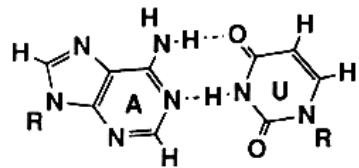
- Klasifikace podle Leontise
- Dané párování není možné pro každou dvojici bází (donory/akceptory)
- Známá četnost výskytu v databázi



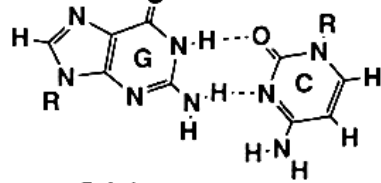
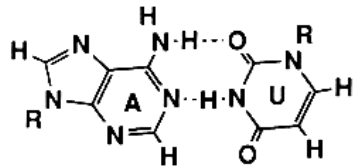
1 - Cis Watson-Crick/Watson-Crick

A A → A  A1912 - A1927 RR0033 resolution = 2.3 Å C1'-C1' distance = 12.3 Å 1-1-0	C A → C  A (+) 105 - C112 AR0001 resolution = 2.3 Å C1'-C1' distance = 10.4 Å 2-0-0
C C → A  C112 - A (+) 105 AR0001 resolution = 2.3 Å C1'-C1' distance = 10.4 Å 2-0-0	C → C  C (+) 170 - C30 URX035 resolution = 3.1 Å C1'-C1' distance = 8.5 Å 2-0-0
G G → A  G215 - A105 URX053 resolution = 2.5 Å C1'-C1' distance = 12.5 Å 2-0-0	G = C  G4 - C21 URL050 resolution = 2.4 Å C1'-C1' distance = 10.3 Å 3-0-0
U U - A  U14 - A19 AR0006 resolution = 1.9 Å C1'-C1' distance = 10.3 Å 2-1-0	U → C  U7 - C8 AR0005 resolution = 1.8 Å C1'-C1' distance = 11.8 Å 1-0-1

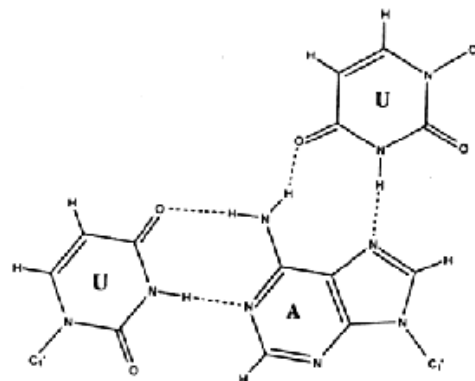
	A	C
A	A $\circ \square$ A  A7 - A6 URL051 resolution = 2.3 Å C1'-C1' distance = 12.5 Å	
C	C $\circ \square$ A  C163 - A148 PR0021 resolution = 1.8 Å C1'-C1' distance = 11.1 Å	C $\circ \square$ C  C1834 - C1841 RR0033 resolution = 2.3 Å C1'-C1' distance = 10.3 Å
G		
U	U $\circ \square$ A  U103 - A73 URL064 resolution = 1.5 Å C1'-C1' distance = 9.8 Å	



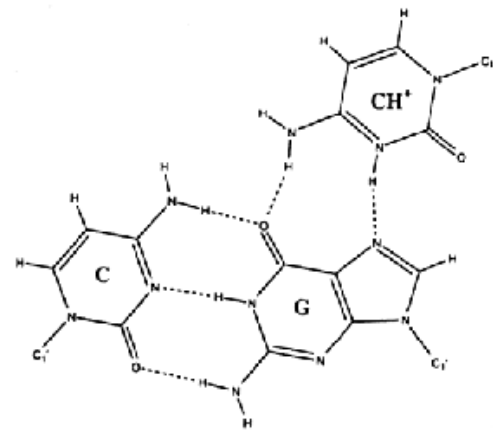
Watson-Crick



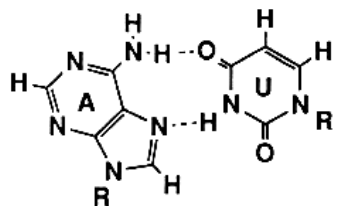
Reverse Watson-Crick



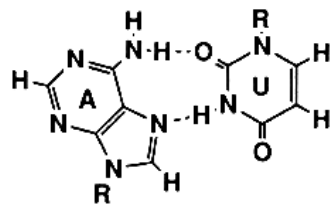
U-A-U



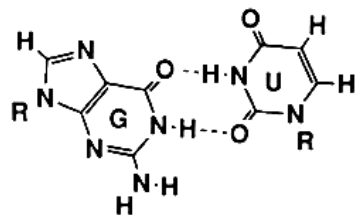
C-G-CH⁺



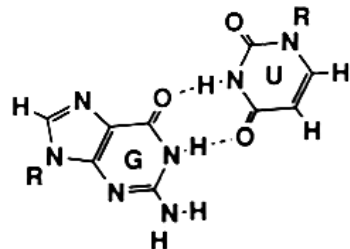
A-U Hoogsteen



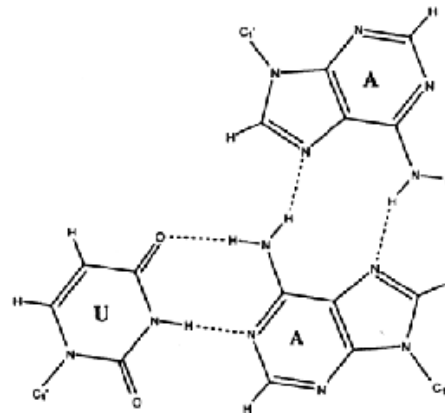
A-U Reverse Hoogsteen



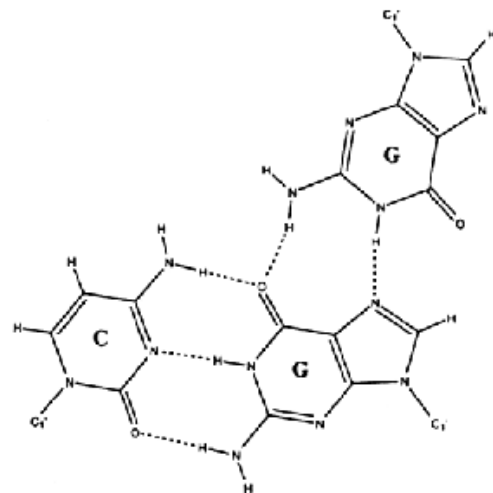
G-U Wobble



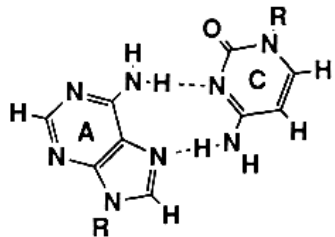
G-U Reverse Wobble



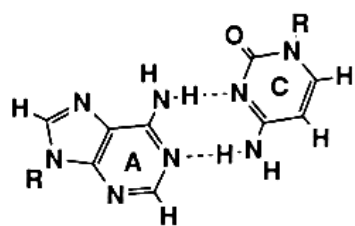
U-A-A



C-G-G



A-C Reverse Hoogsteen



A-C Reverse Wobble

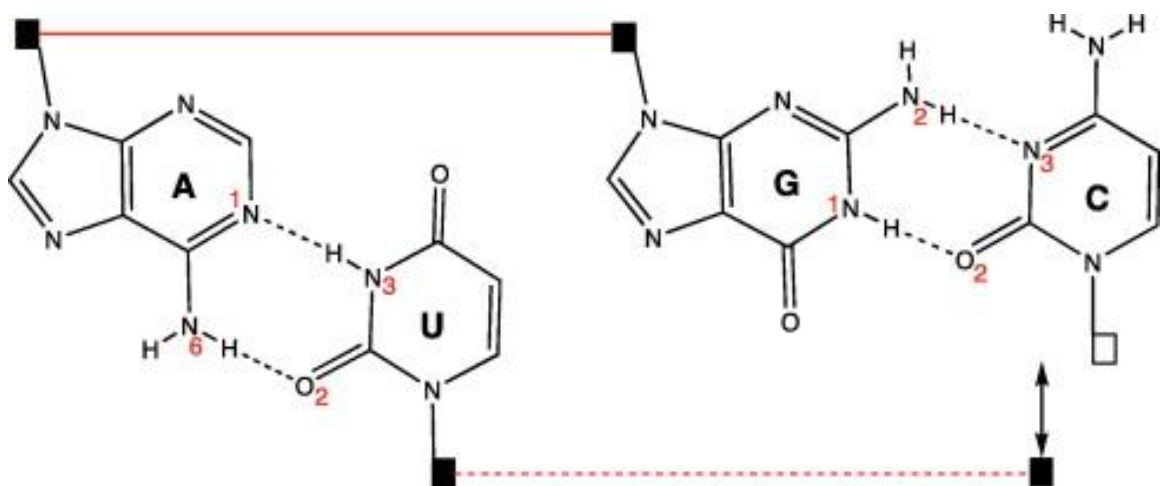
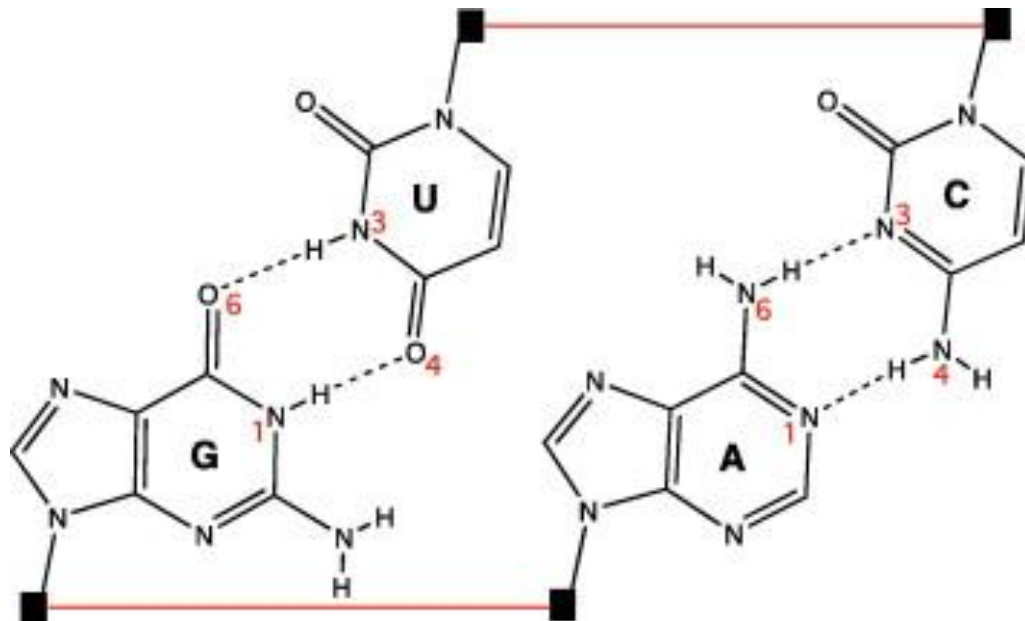
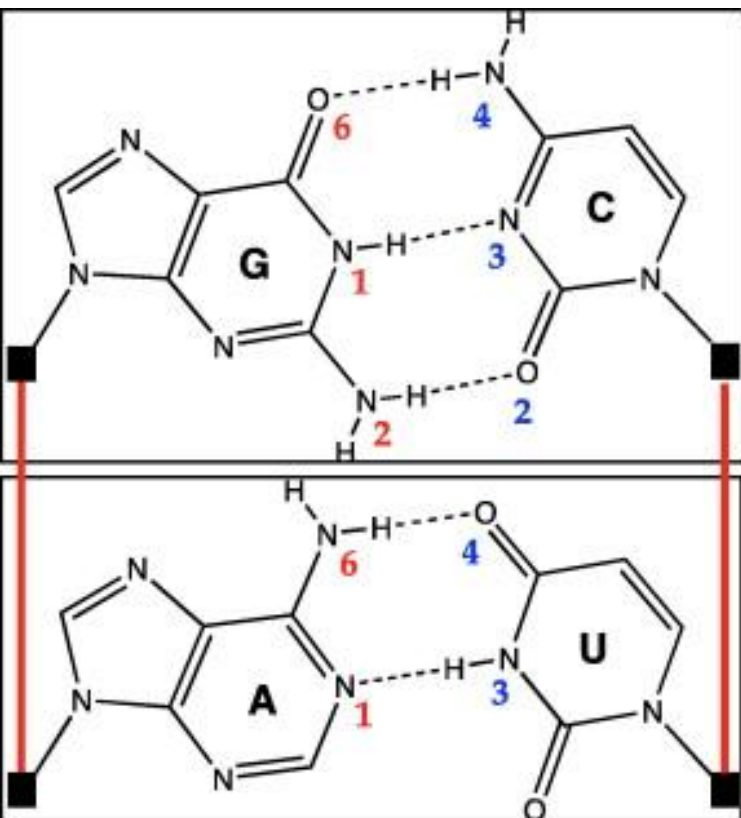
Databáze struktur NA

- **Nucleic Acid Database (ndbserver.rutgers.edu/)**
 - Speciální odnož PDB databáze zaměřená na nukleové kyseliny (čisté NA i jejich komplexy)
 - poskytuje informace unikátní pro nukleovou kyselinu (např. dihedralní úhly páteře; helikální parametry, atd.)
 - Obsahuje list užitečných bioinformatických nástrojů pro nukleové kyseliny
 - Programy pro analýzu struktury helixu – **X3DNA**, **Curves+**

Nukleové kyseliny - struktura

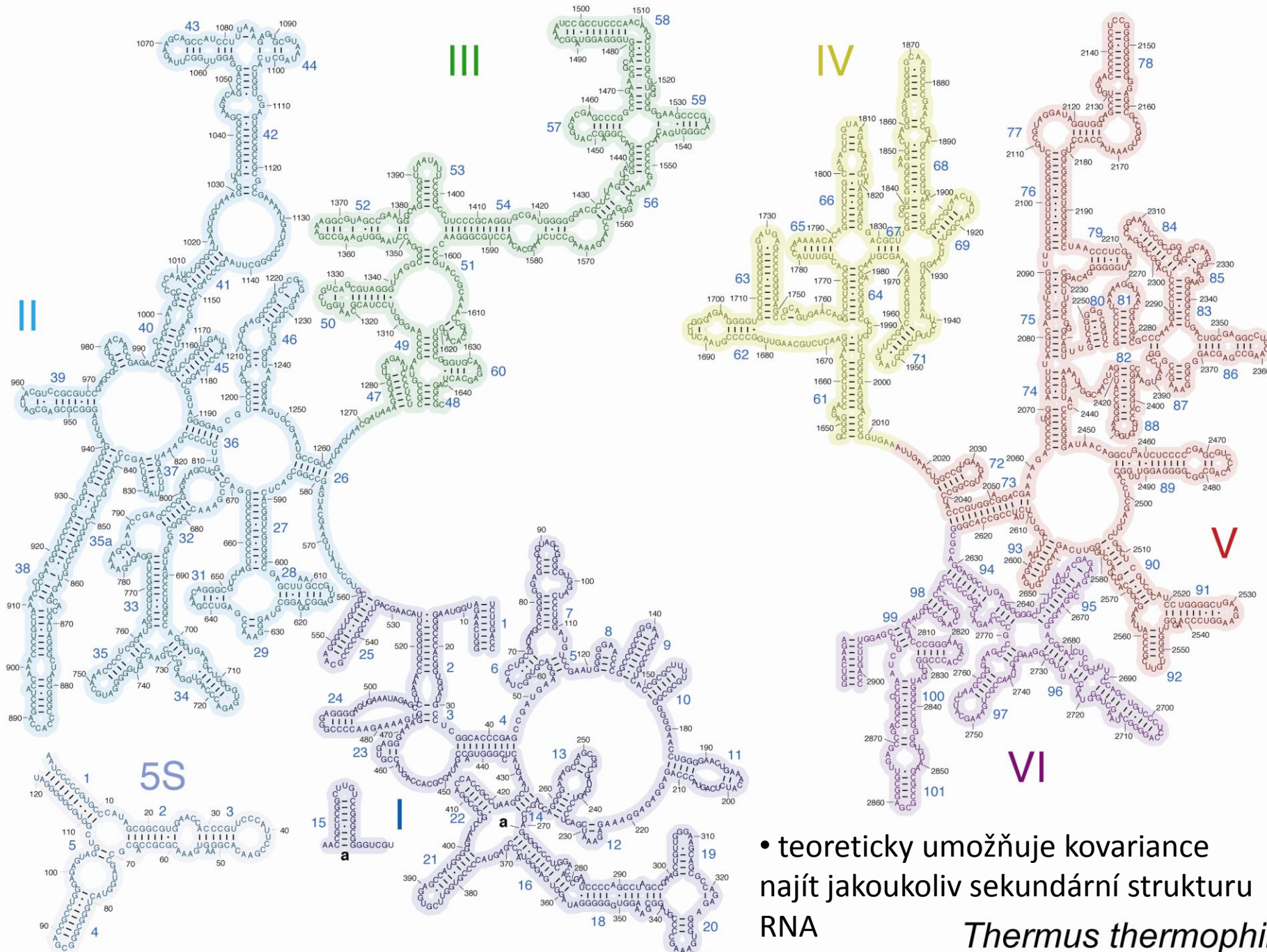
- **Isostericita** párů bází
- Dva páry bází jsou navzájem zcela isosterické pokud jejich záměna (substituce) není spjata se změnami 3D struktury
- U velmi malé změny se hovoří o „near-isostericity“
- Jedná se o zcela klíčový princip v evoluci nukleových kyselin - **kovariance**; nutno uvažovat při studiu sekvenčních variací, atd.

- **Isostericita párů bází**



Nukleové kyseliny - struktura

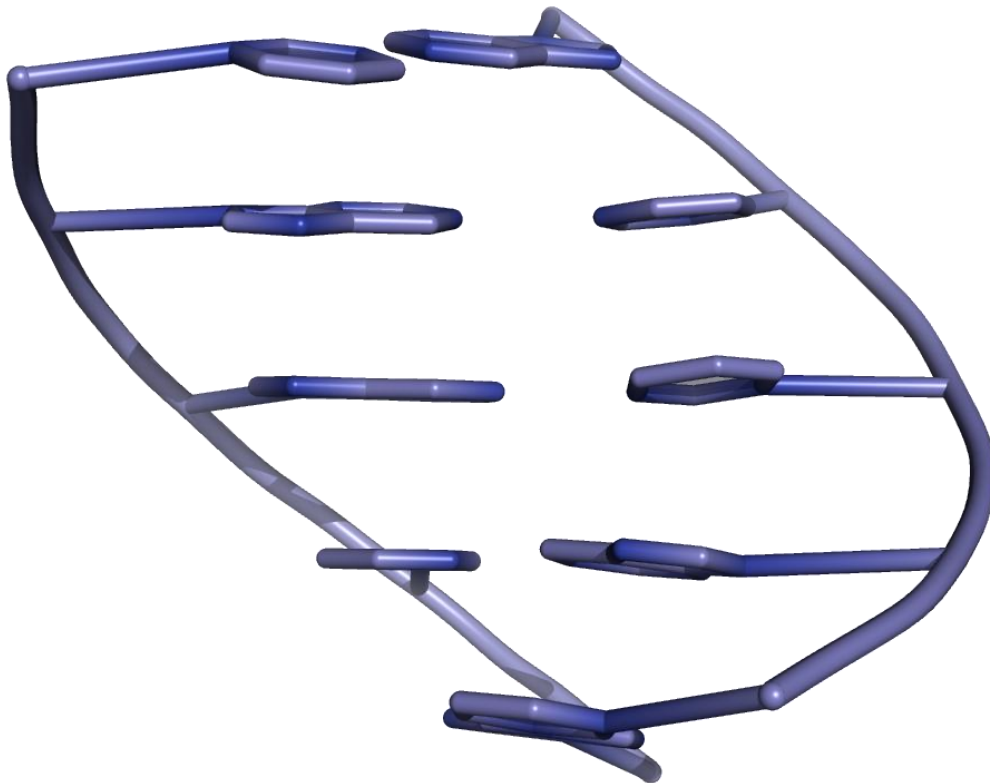
- **Isostericita** párů bází
- Definována jako vzdálenost C1'-C1' atomů a ekvivalentní párování funkčních skupin
- V evoluci RNA funkčních celků (např. ribosom), jsou isosterické a near-isosterické substituce běžné → silně konzervovaná struktura, slabě konzervovaná sekvence
- Neisosterické substituce (měnící strukturu) jsou obvykle letální a v evoluci vzácné



Nespecifické interakce

Nukleové kyseliny - struktura

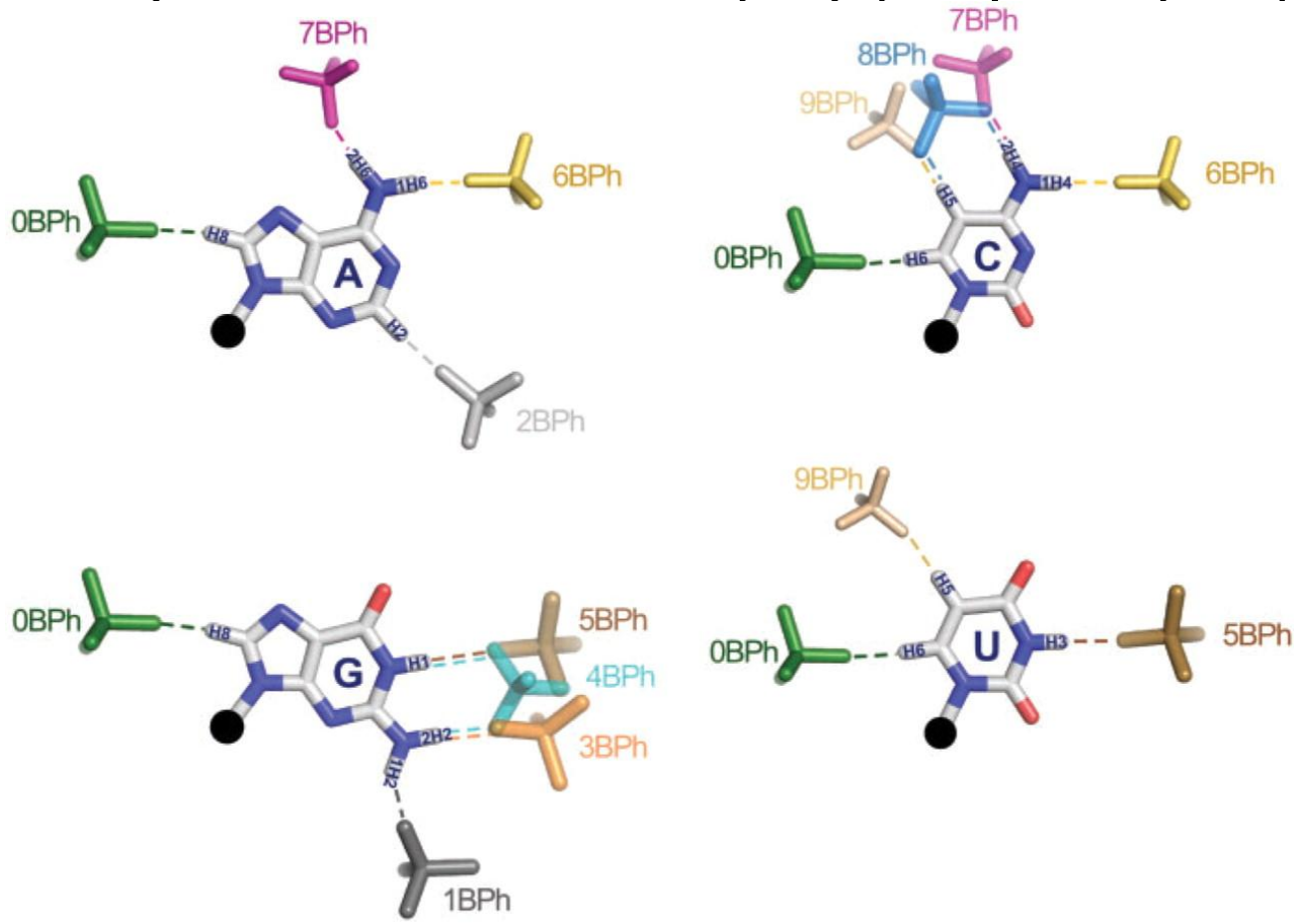
- **Stacking** bází
- Nespecifická interakce



- Velikost překryvu
- Klasifikace podle toho která „tvář“ (face) báze interaguje – s53, s33, s55

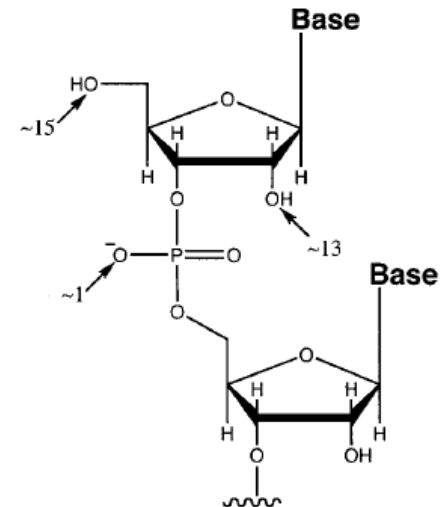
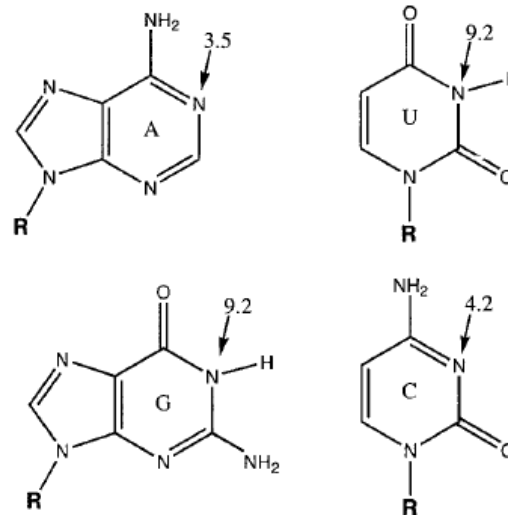
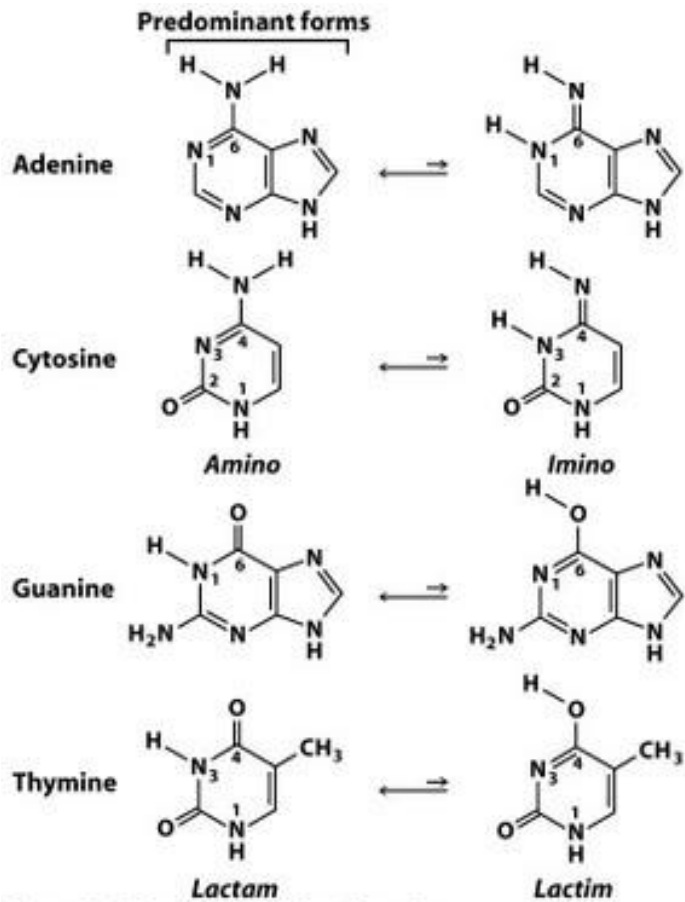
Nukleové kyseliny - struktura

- Interakce báze/fosfát
- Interakce mezi donorem vodíkové vazby (báze) a OP1/OP2 kyslíkem fosfátové skupiny (vždy akceptor)



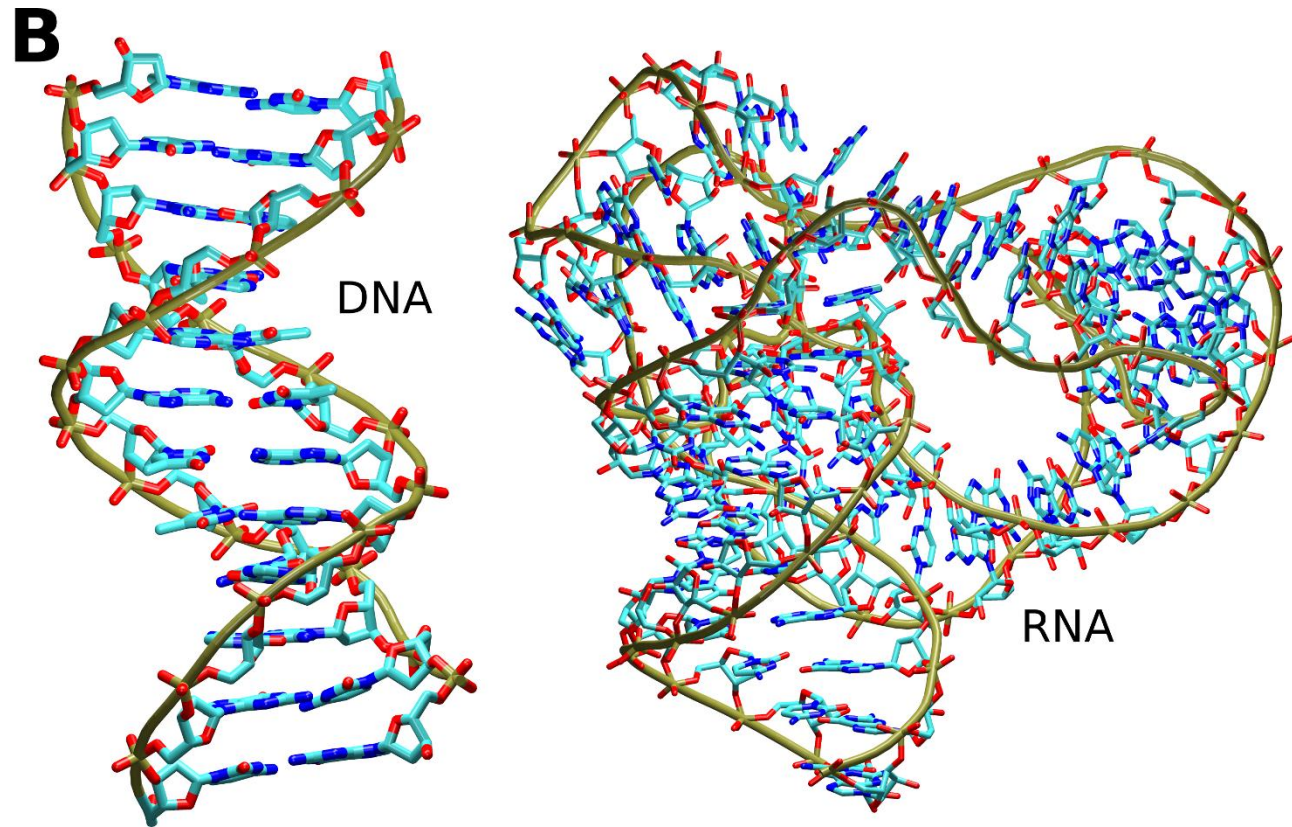
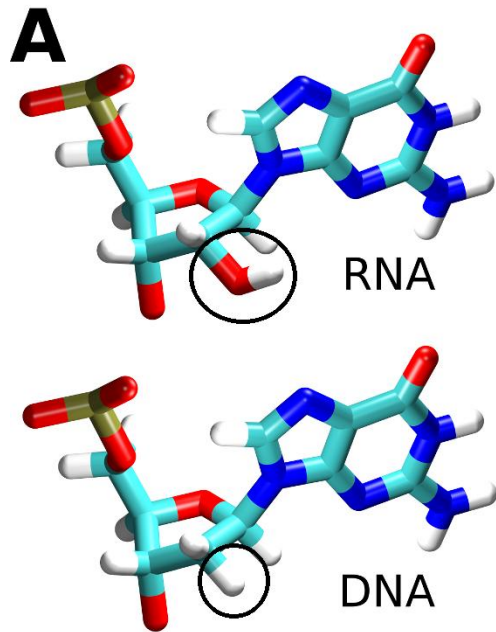
Nukleové kyseliny - struktura

- Speciální stavy bází – **protonace** a **tautomerizace**



Nukleové kyseliny – vyšší strukturní celky

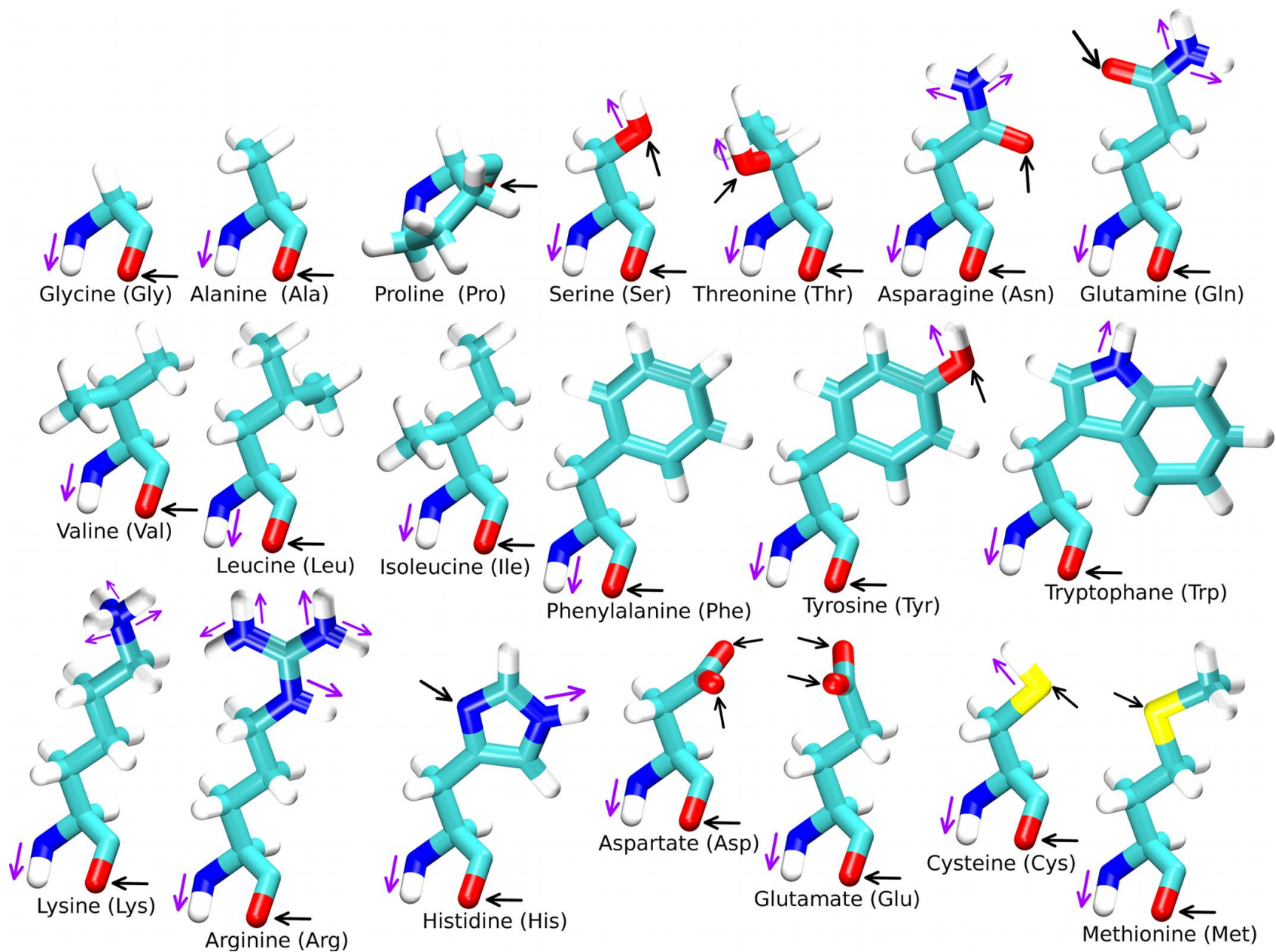
DNA vs. RNA

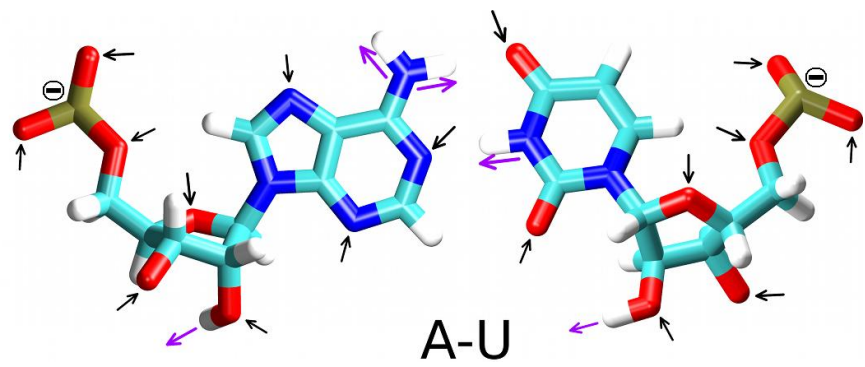
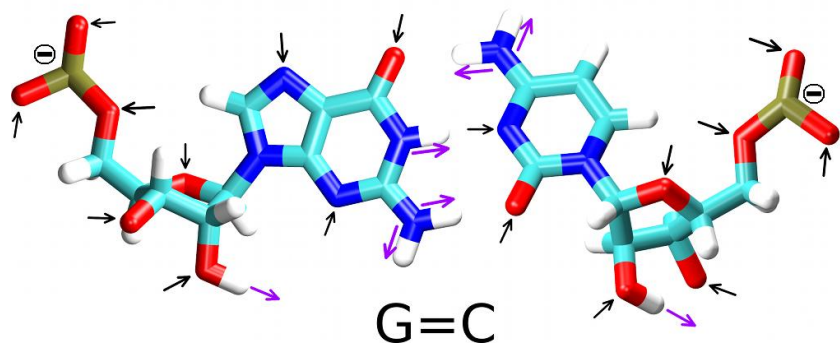


DNA vs. RNA vs. protein

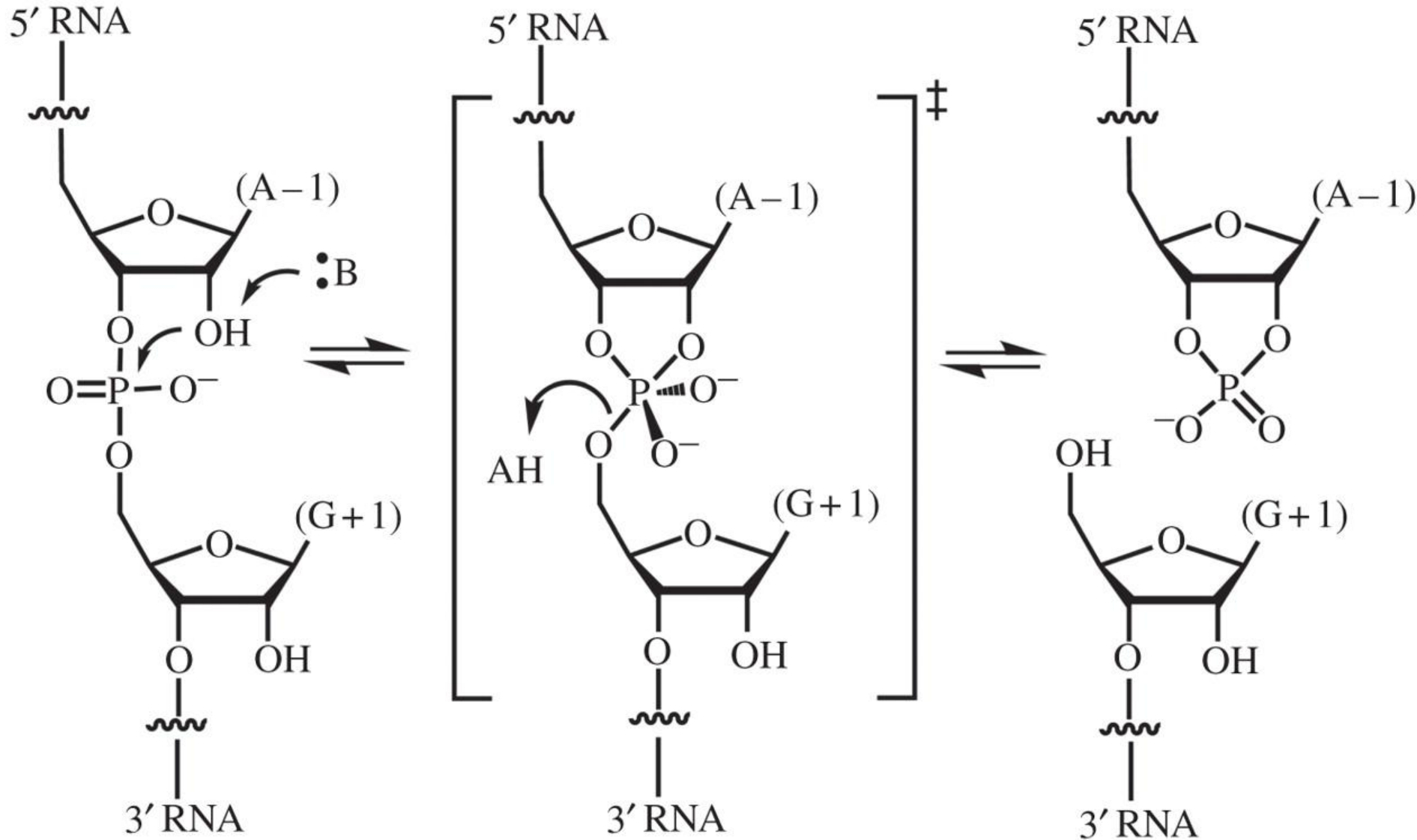
Molekula:	DNA	RNA	protein
Nese genetickou informaci?	Ano	Ano	Ne
Katalyzuje chemické reakce?	Ne*	Ano	Ano
Celkový náboj	Negativní	Negativní	Negativní/neutrální/ pozitivní

- RNA (pravděpodobně) evolučně nejstarší biomolekula
- umí nést genetickou informaci (ale DNA to umí lépe...)
- umí katalyzovat chemické reakce (ale proteiny to umí lépe...)
 - odstranění 2-hydroxylové skupiny (RNA→DNA) výrazně stabilizuje fosfodiesterovou vazbu
 - širší spektrum aminokyselin (bočních řetězců) umožňuje více chemických reakcí

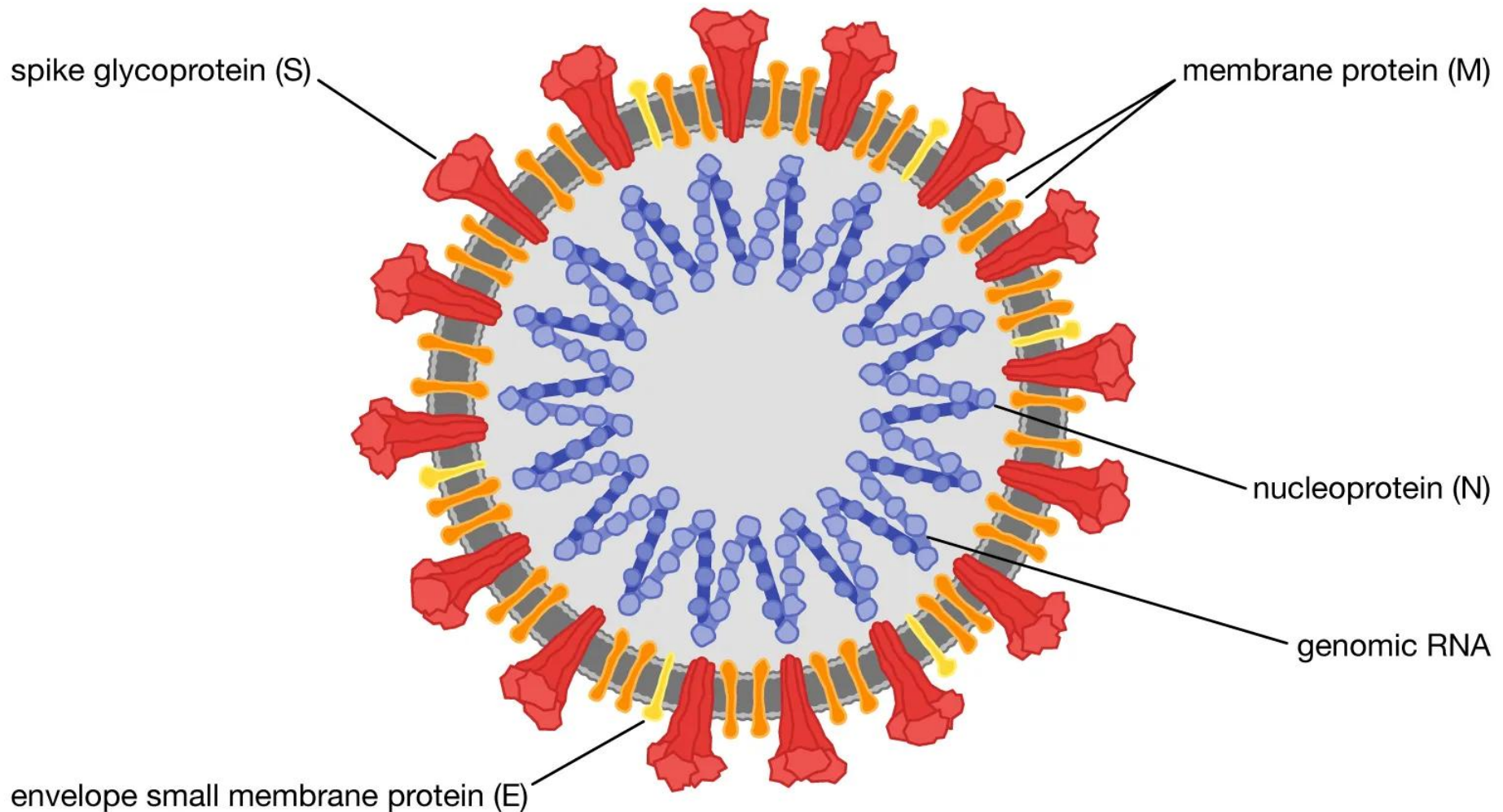




RNA autohydrolýza

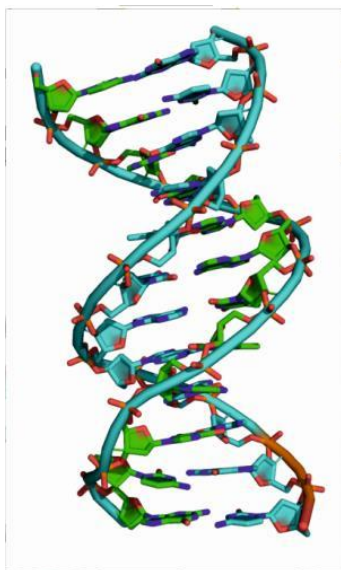


Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)

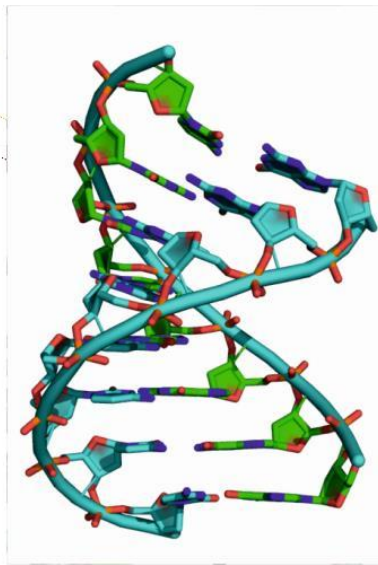


RNA viruses up to 33 000 nucleotides (Covid 30k)
“Minimal” DNA cellular genome ~100 000

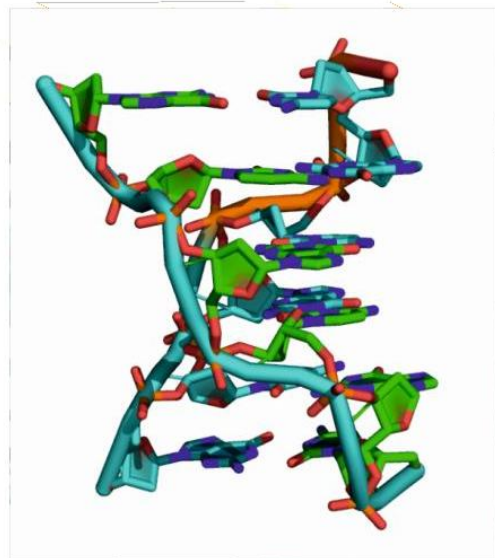
DNA struktury, helikální formy



B-DNA



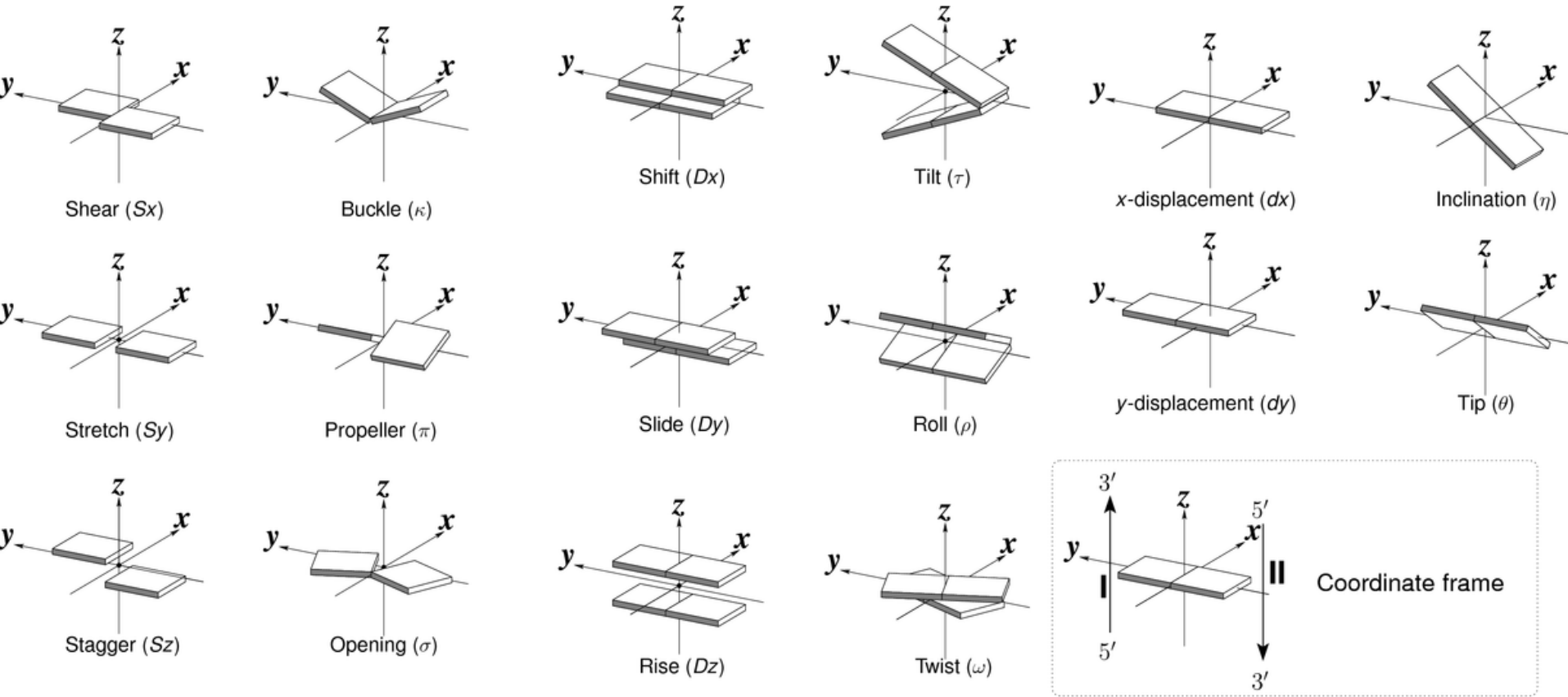
A-DNA



Z-DNA

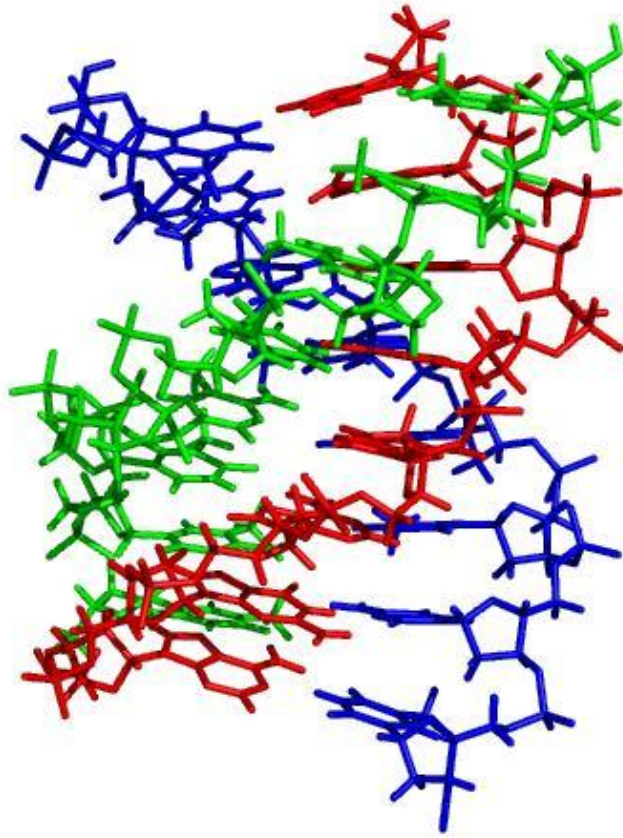
Geometry attribute:	A-form	B-form	Z-form
Helix sense	right-handed	right-handed	left-handed
Repeating unit	1 bp	1 bp	2 bp
Rotation/bp	33.6°	35.9°	60°/2
Mean bp/turn	11	10.5	12
Inclination of bp to axis	+19°	-1.2°	-9°
Rise/bp along axis	2.4 Å (0.24 nm)	3.4 Å (0.34 nm)	3.7 Å (0.37 nm)
Rise/turn of helix	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)

Parametry párů bází – X3DNA

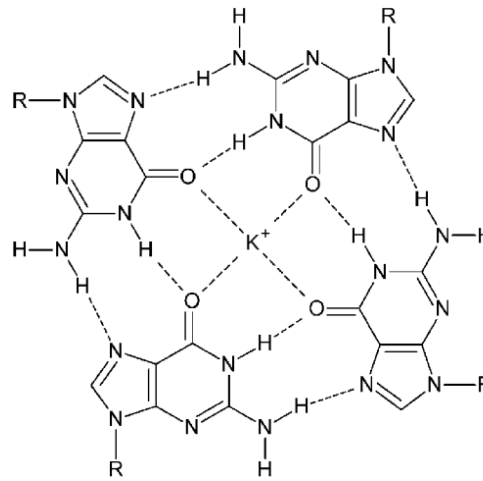
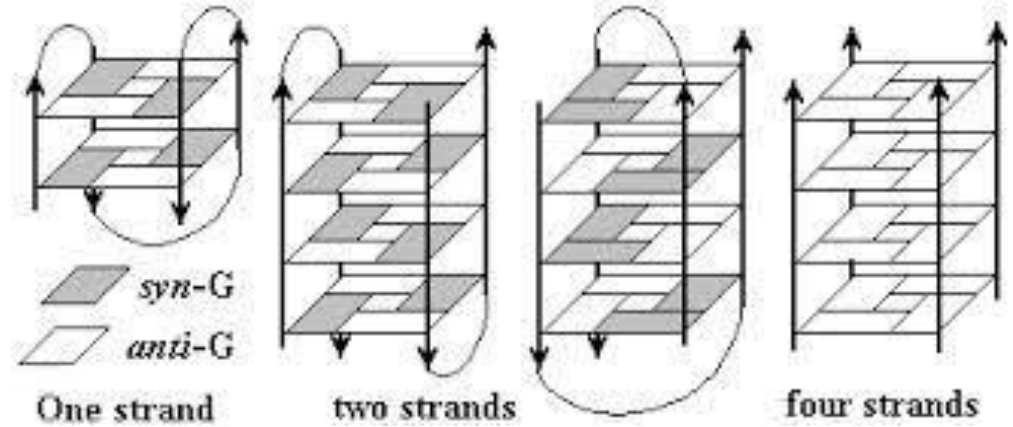


- velký žlábek (major groove)
- malý žlábek (minor groove)

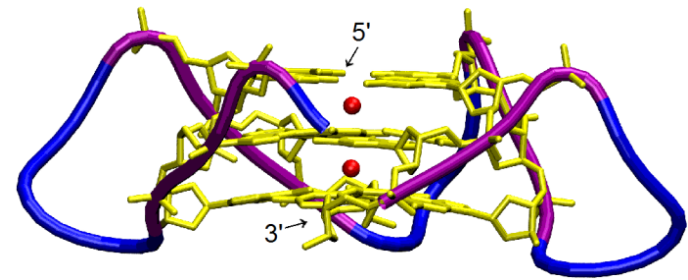
DNA struktury - další

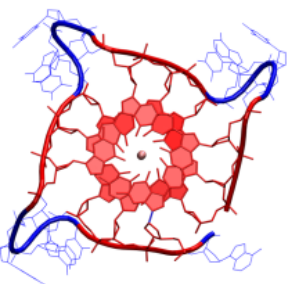


DNA triplex

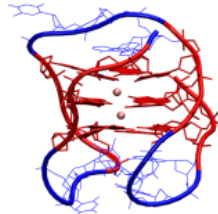


DNA G-kvadruplex

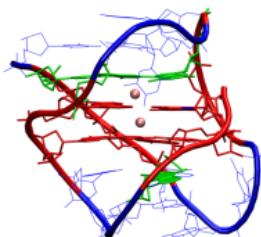




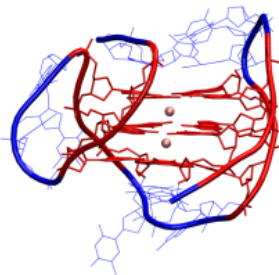
parallel



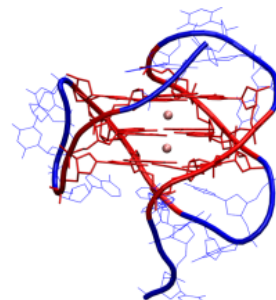
basket



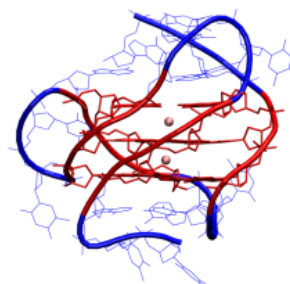
2-layer basket



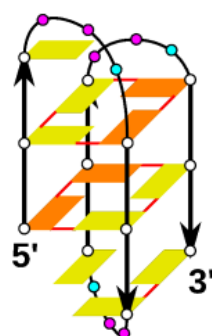
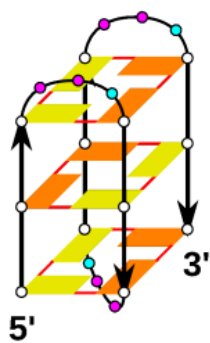
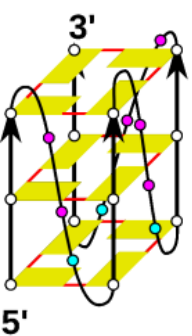
hybrid-1



hybrid-2



antiparallel with
a propeller loop



 *anti* G

 *syn* G

 A

 T