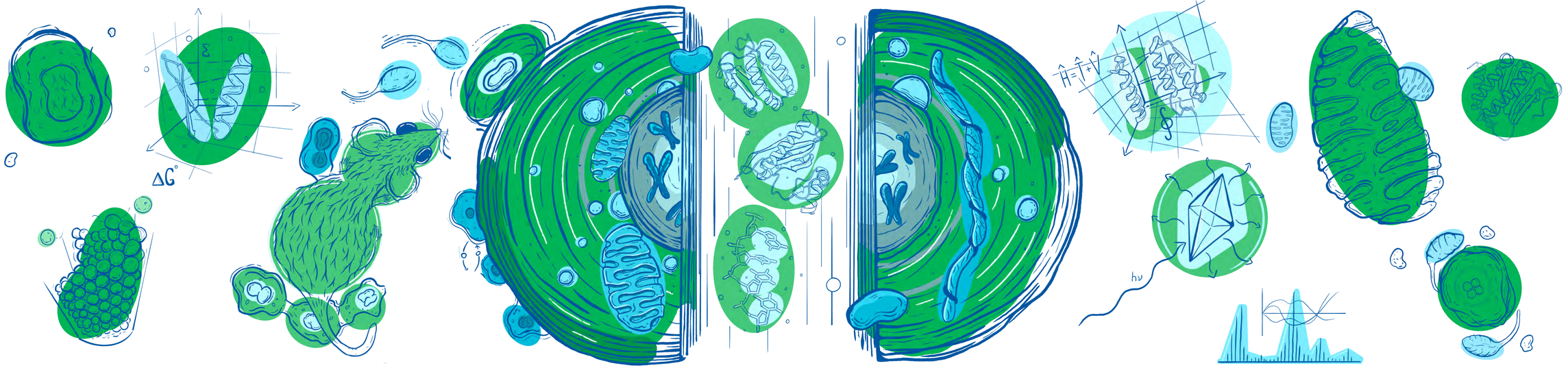


A Bottom-up Approach to RNA Architecture: Dinucleotide Conformers and Base Pairs

Bohdan Schneider

https://bs.structbio.org

RNA Structure & Function

Benasque, Spain

2026-07-06

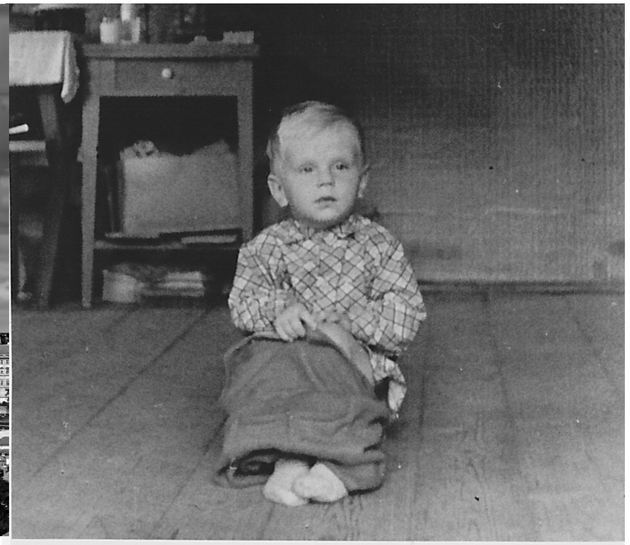
ibt

Czech Academy of Sciences

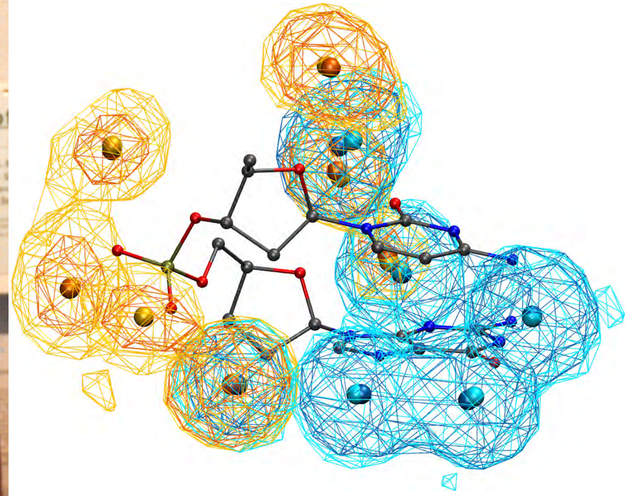
About me

- I'm a structural biologist, bioinformatician
- My scientific interests include
  - structure of DNA and RNA
  - hydration of NAs
  - structural databases
- I live, studied & did my PhD in Prague
- Postdoc & many visits at Rutgers
- Now I work at Institute of Biotechnology of the Czech Academy of Sciences

Prague



From my early-carrier biotechnology experiments



ibt

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Czech Academy of Sciences

About my workplace, Institute of Biotechnology of the Czech Academy of Sciences

Akademie věd České republiky
The Czech Academy of Sciences

- Curiosity-driven research and fundamental scientific questions
- … about 200 people
- Involved in European infrastructures
- Partner in BIOCEV Center
- Many scientific collaborations

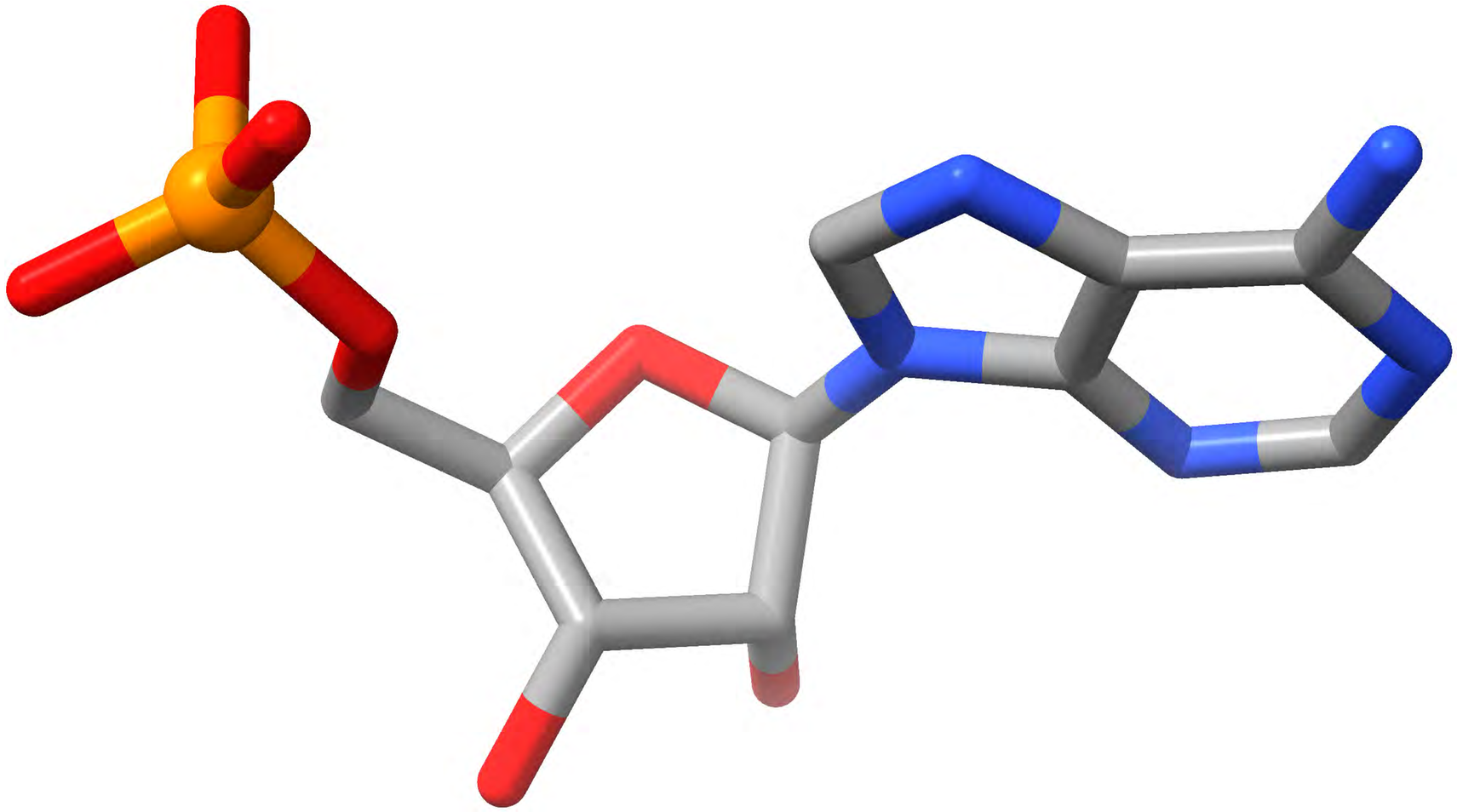
instruct ERIC

elixir

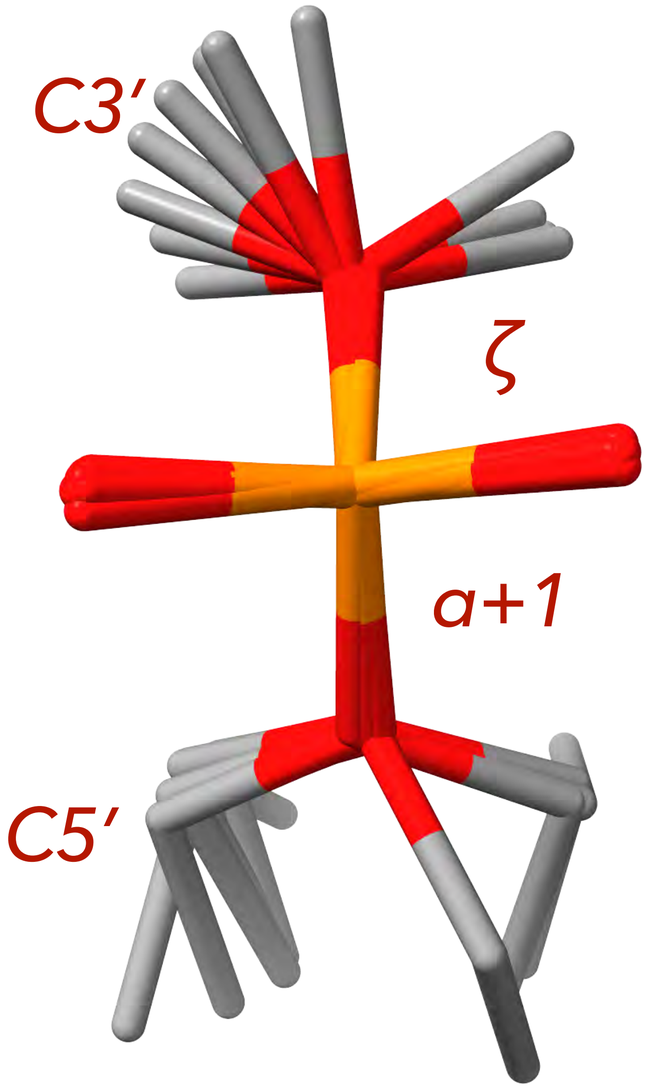
PARTNER
BIOCEV

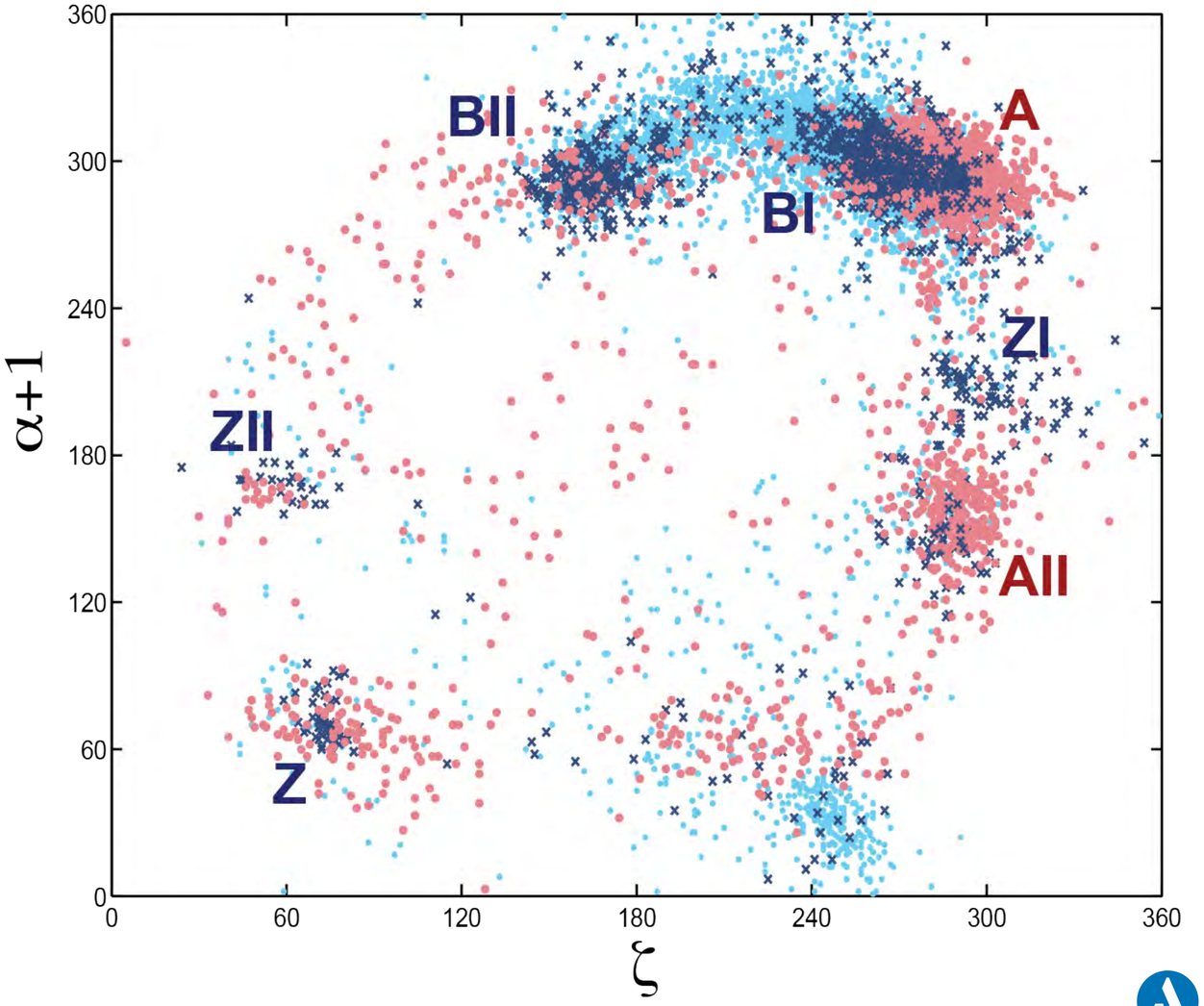
ibt

NA complexity stems from complexity of its components



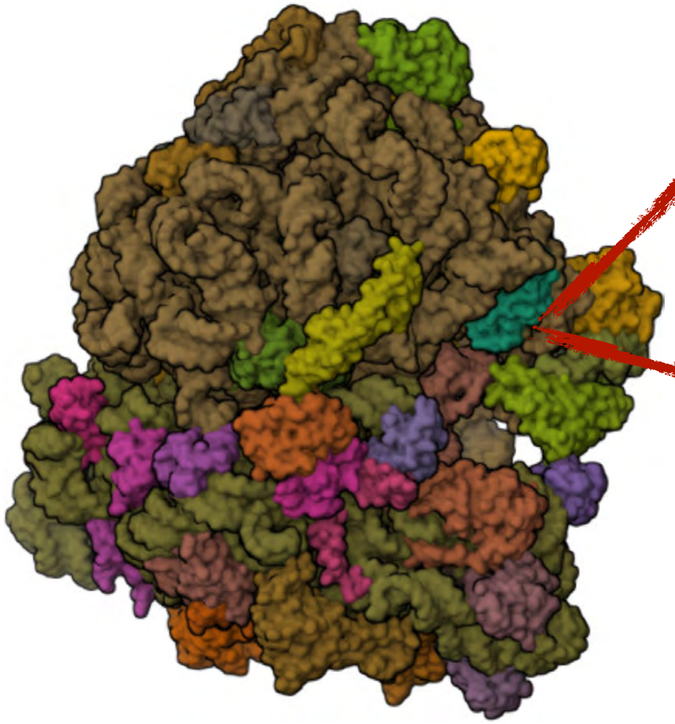
Phosphates: structural complexity



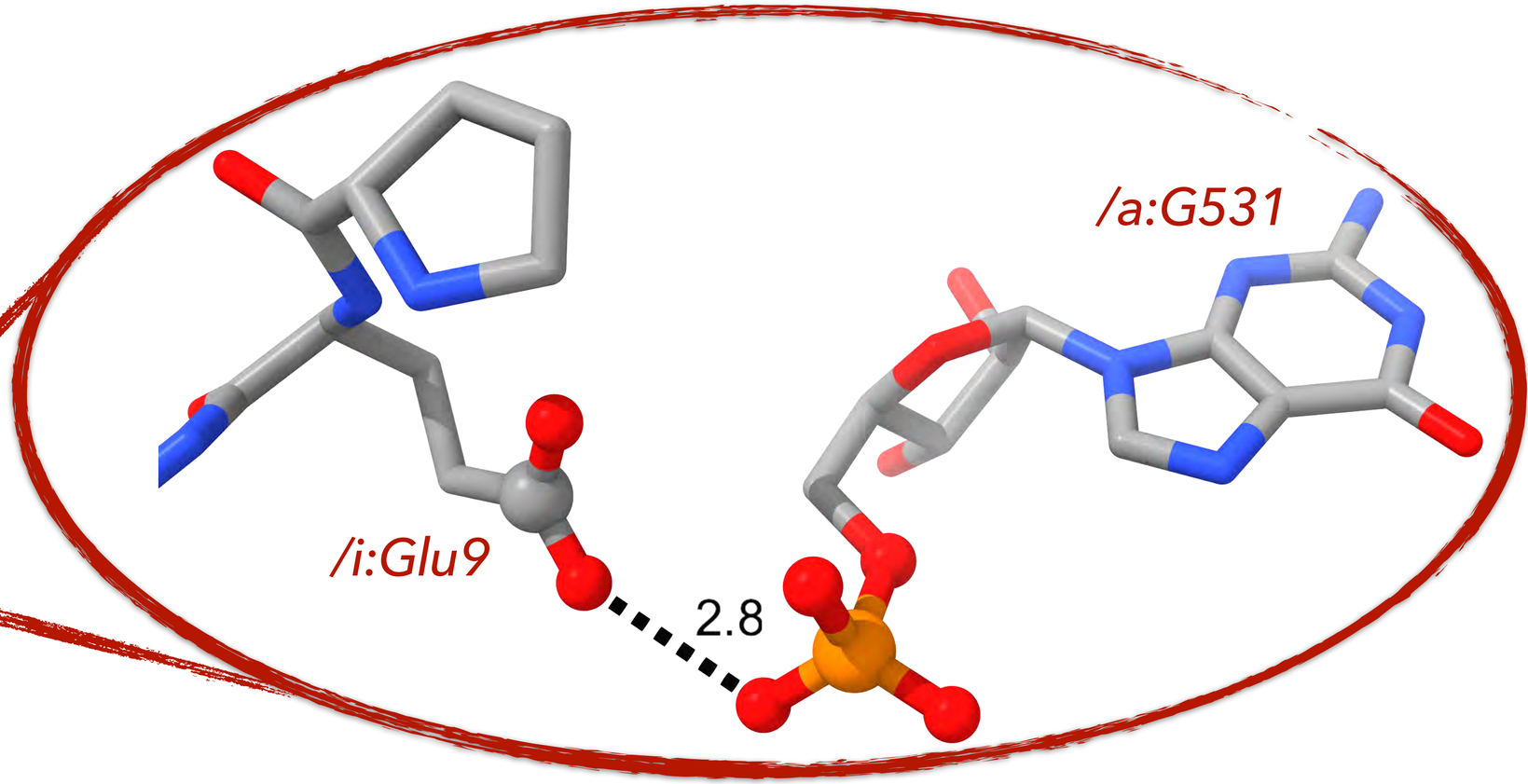


Czech Academy of Sciences

Phosphates: charge always -1?



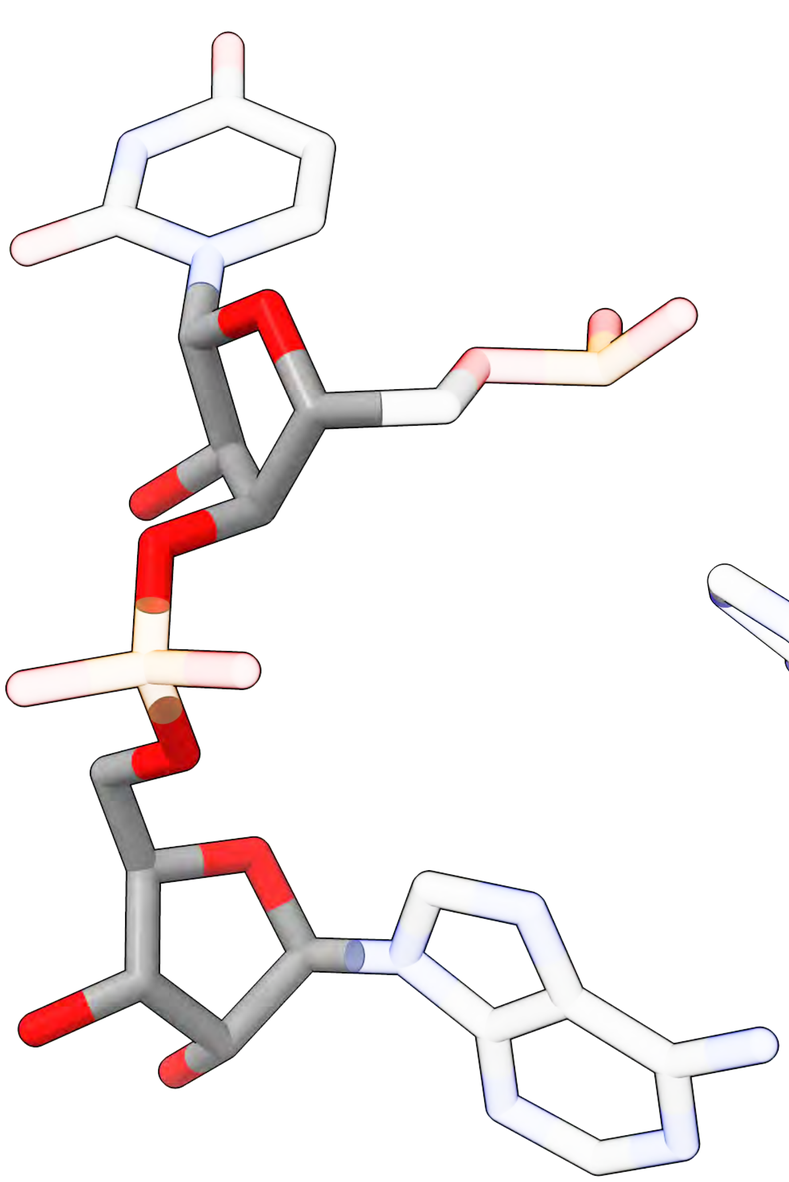
8b0x ribosome

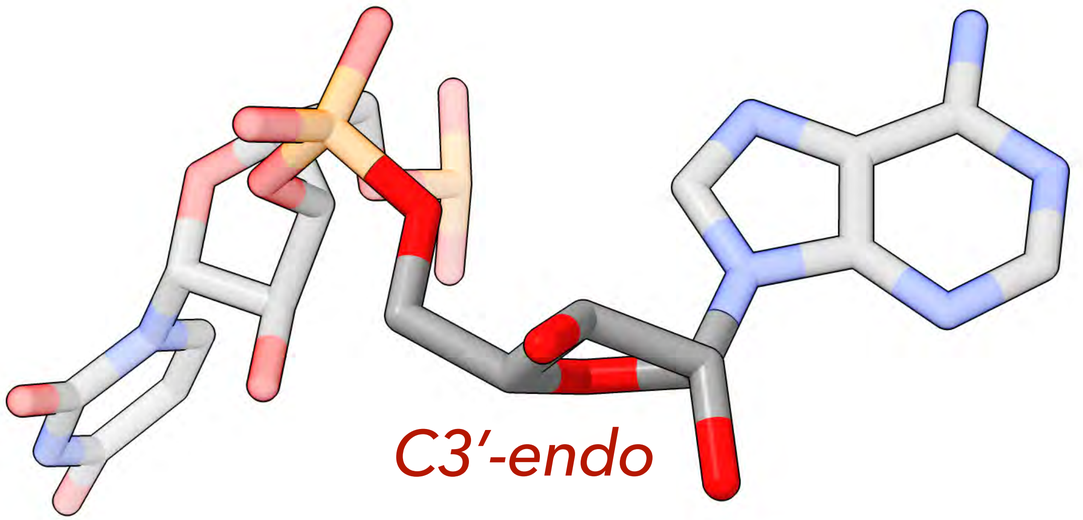


Czech Academy of Sciences

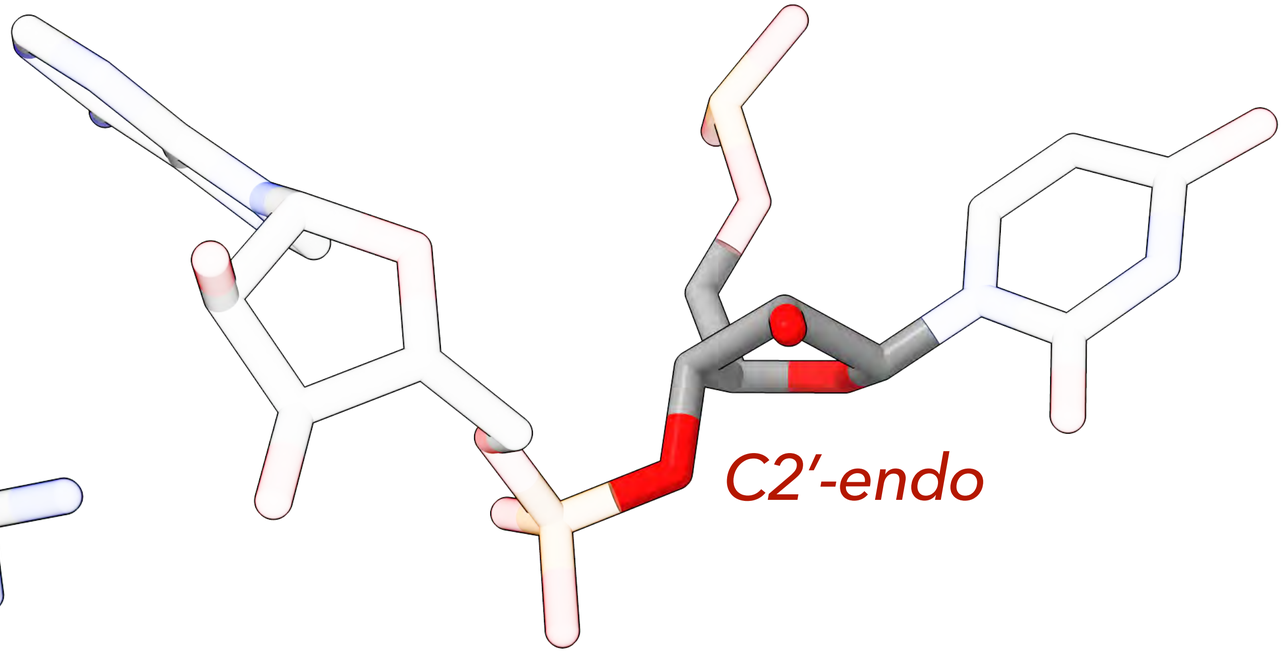
Ribose puckers:

conformer OP11 at 8b0x ribosome: U1168-A1169



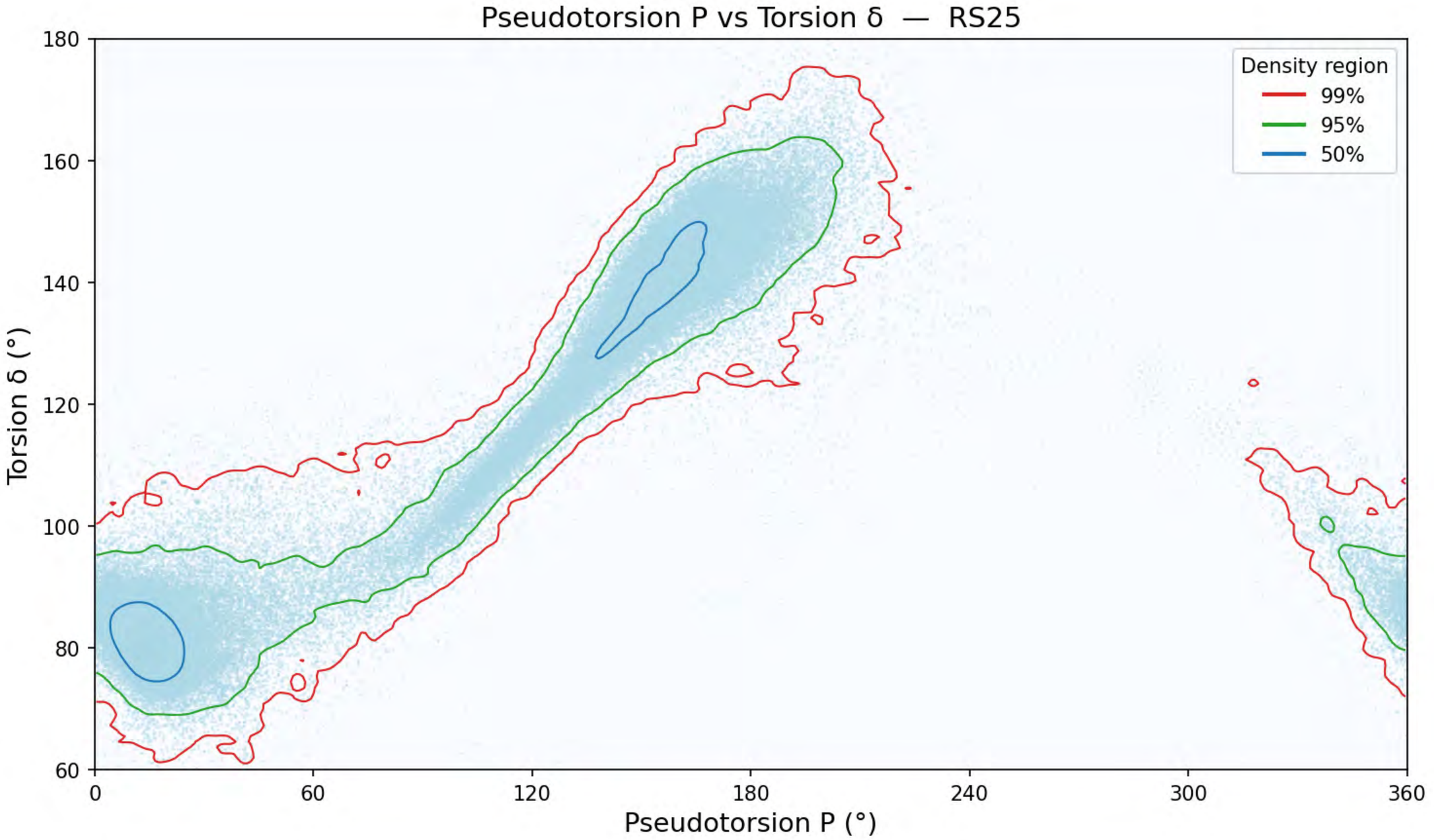


C3′-endo

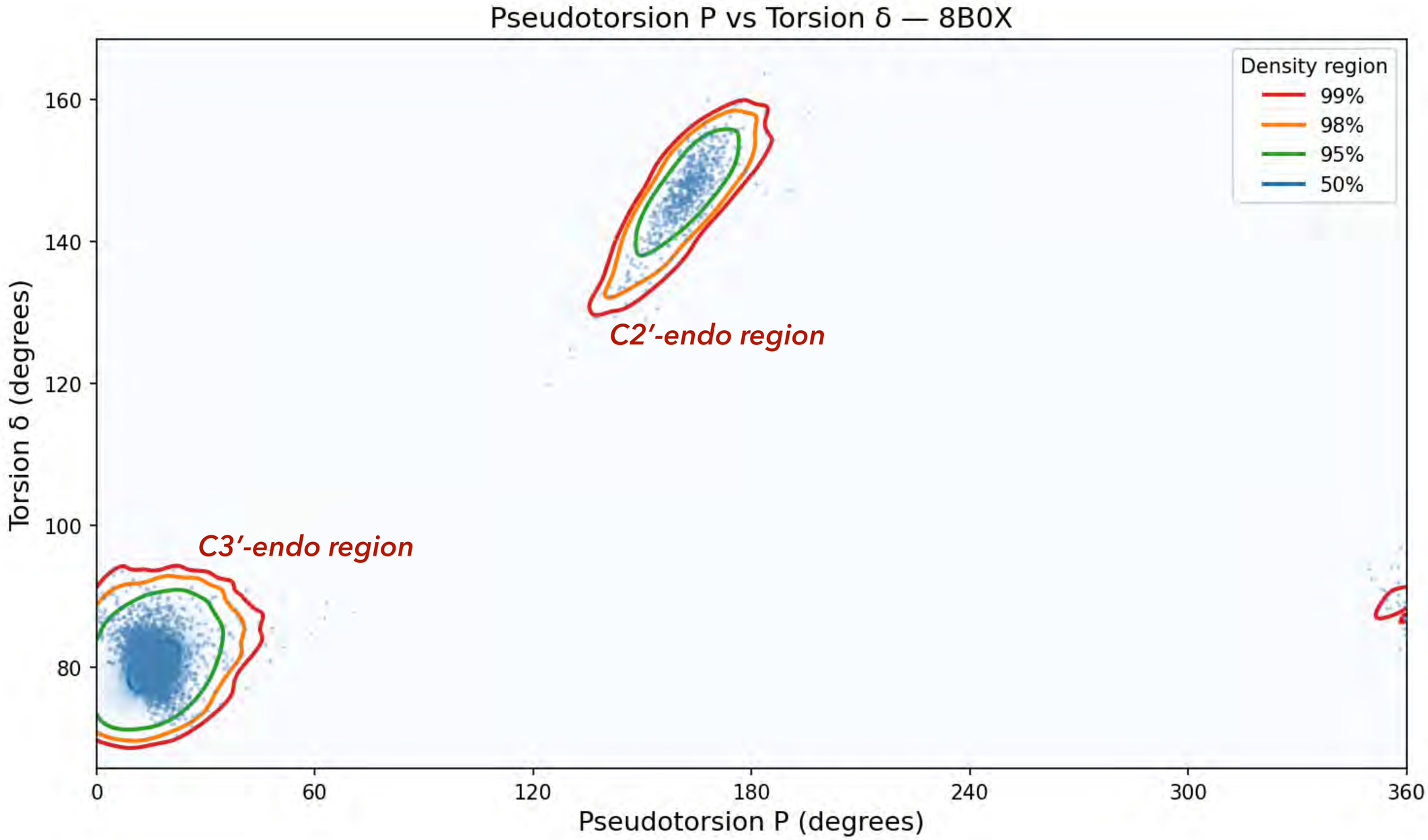


C2′-endo

Overall distribution of ribose puckers

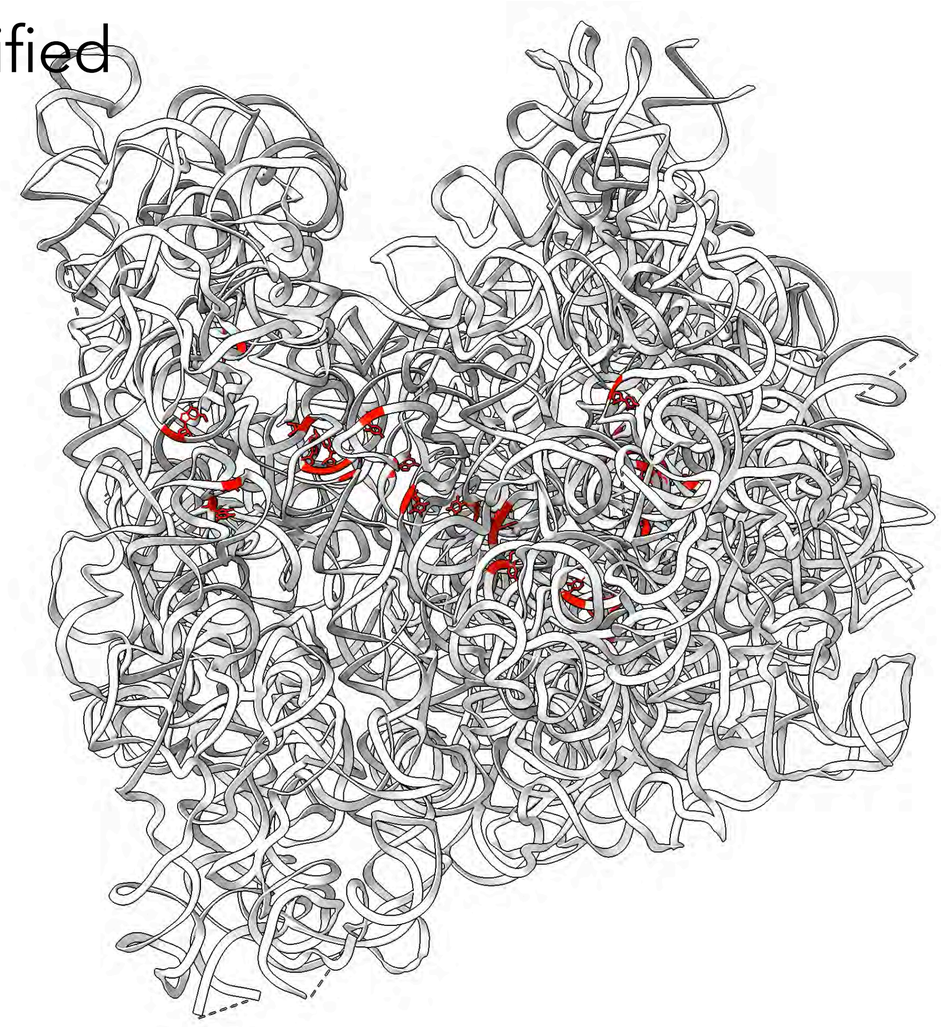


Overall distribution of ribose puckers



Bases

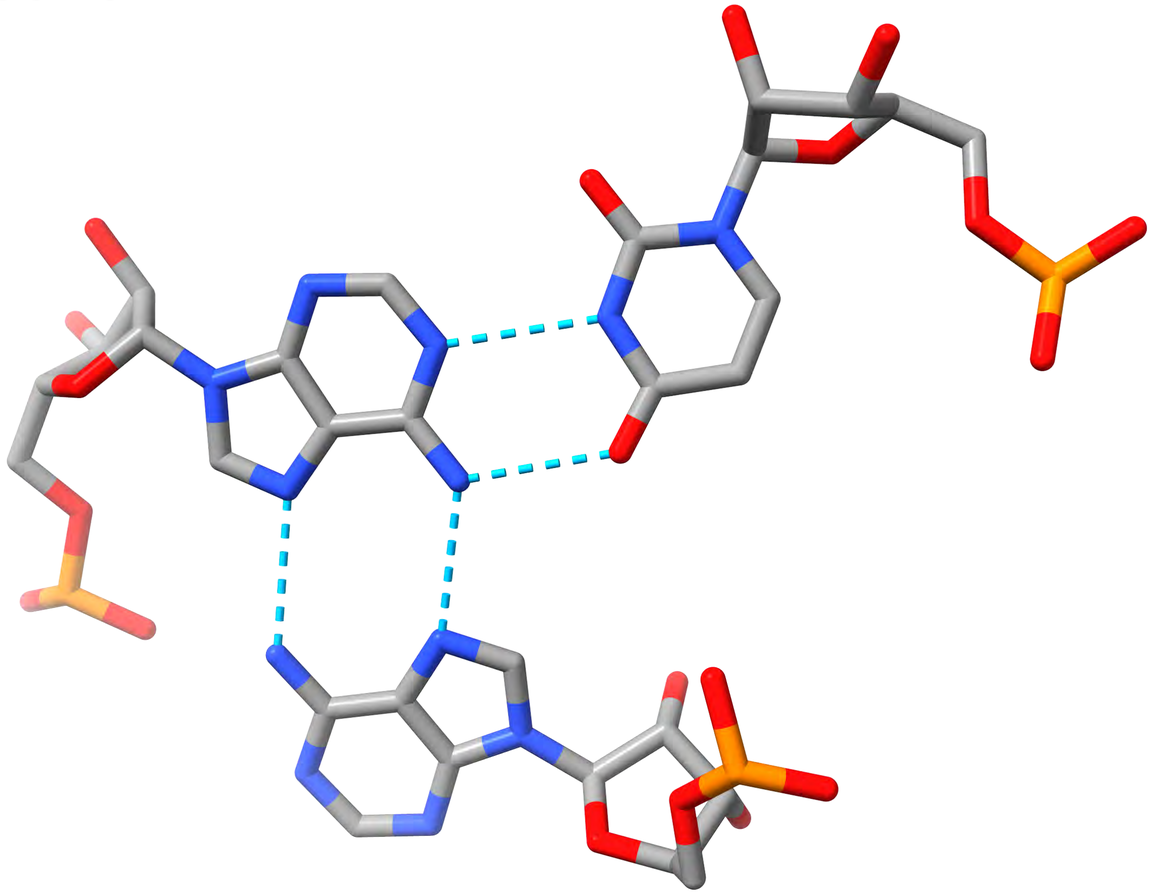
- Aromatic nitrogenous compounds stacking by their π-electrons and forming hydrogen bonds by their ***edges***
- Four standard complemented by many modified
  - *OMG, OMC, OMU, 5MU, PSU …*

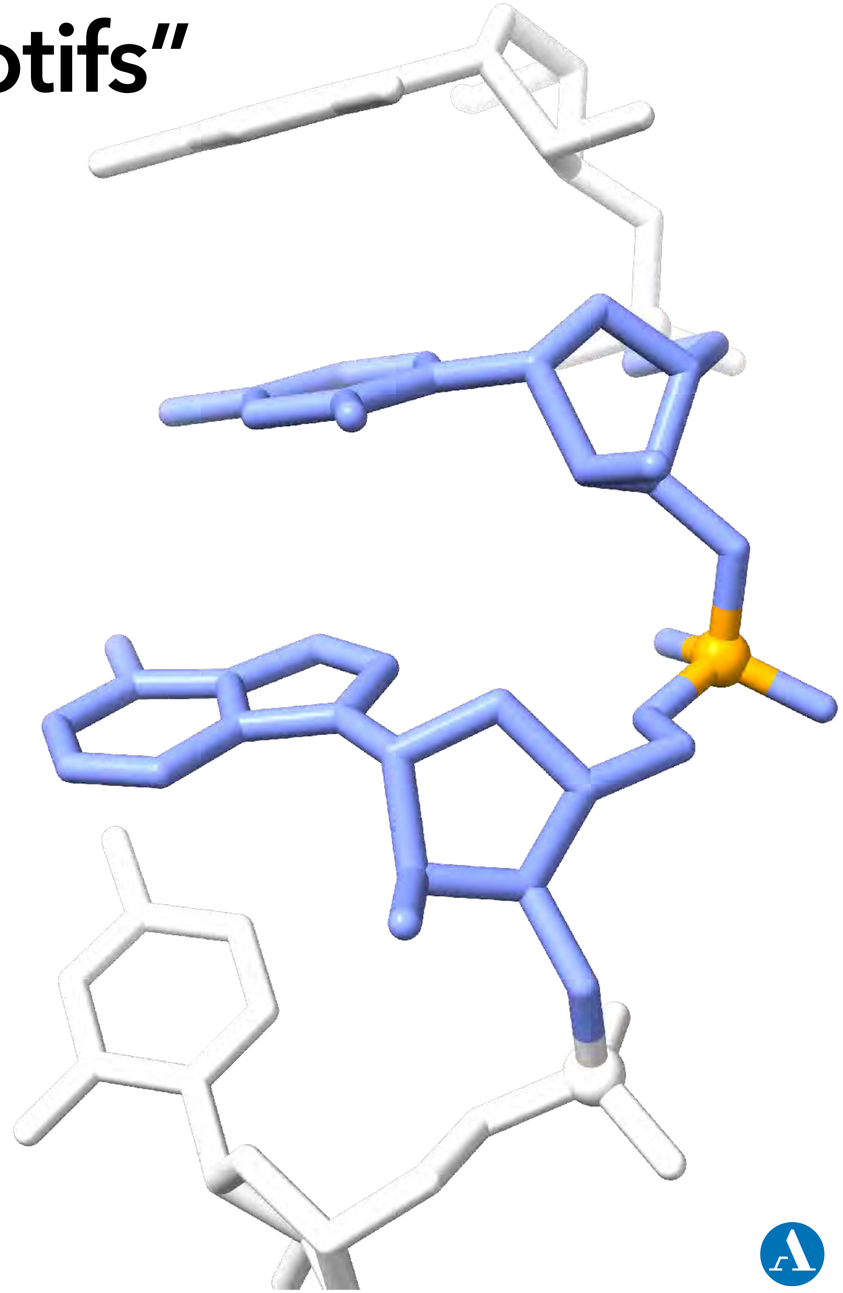


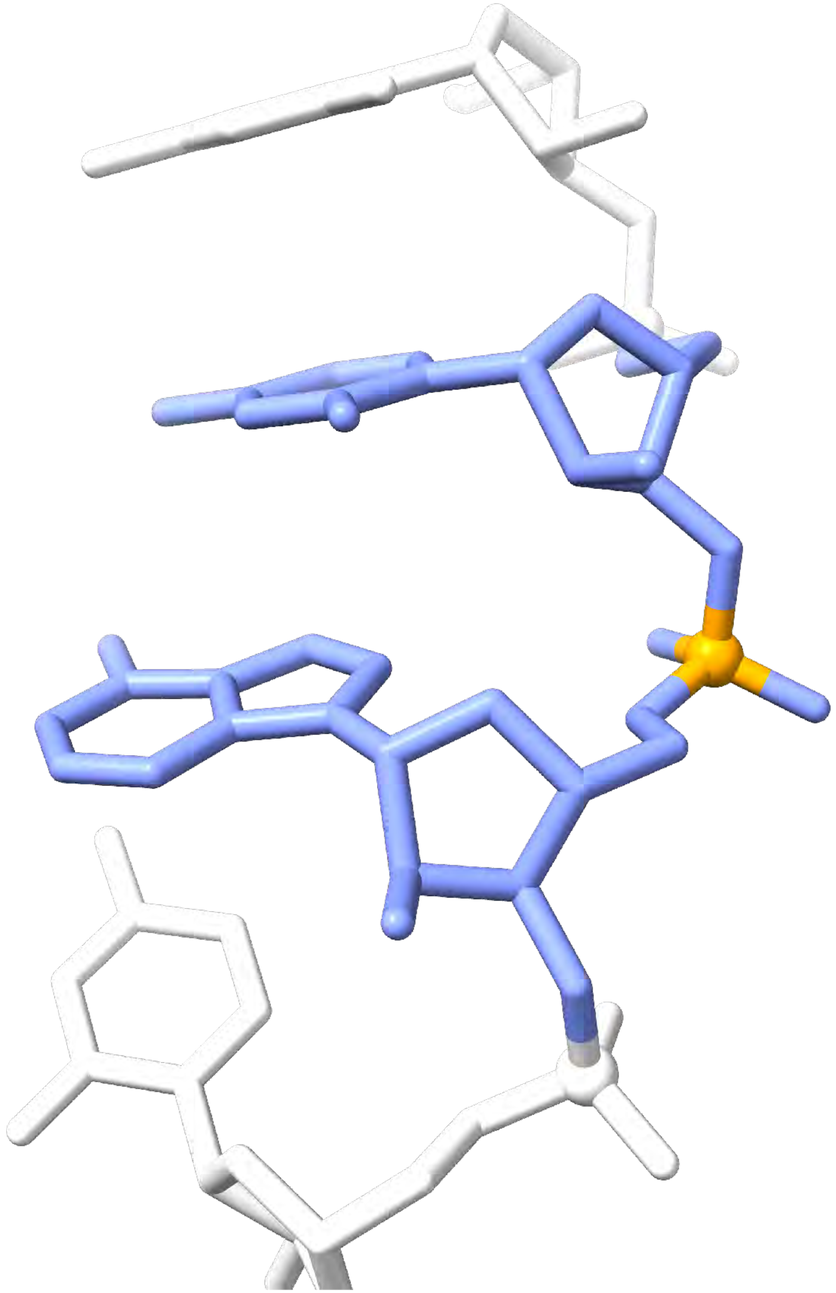
8b0x modified its in red

Dinucleotides: the smallest RNA “motifs”

- Dinucleotides in sequence
- Nucleotides bound into (multiple) base pairs







NA backbone geometry: often ignored … to the detriment of the field

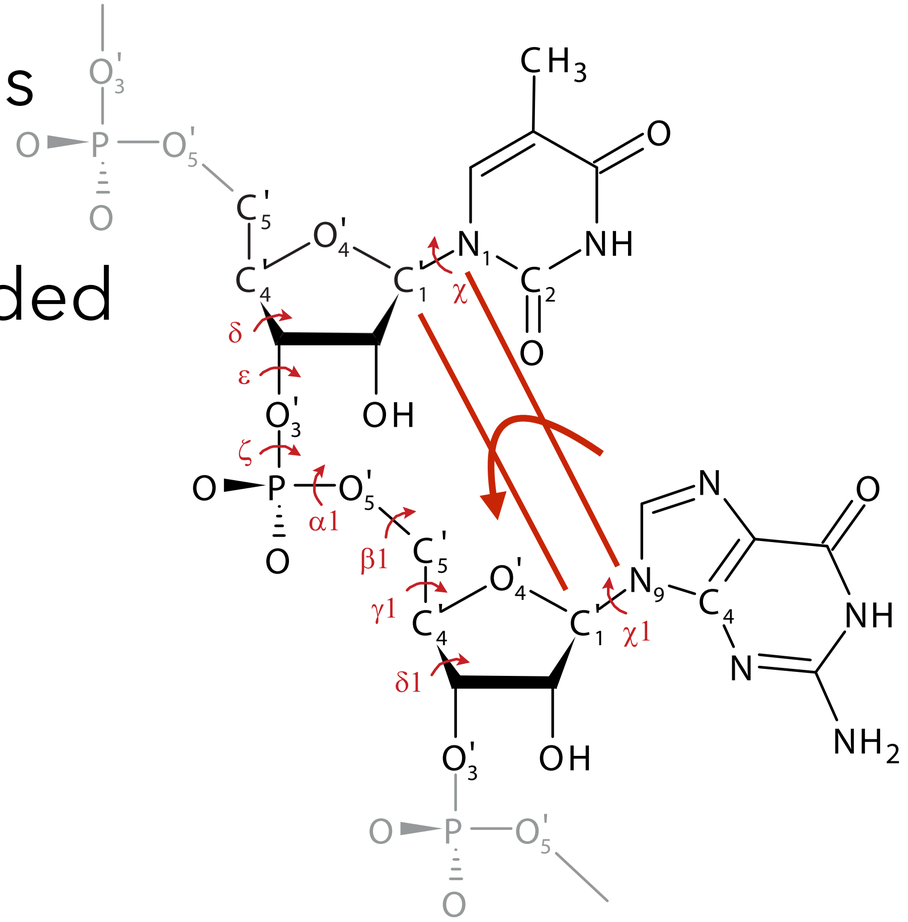
- NA backbone: analysis of correlated torsions

  analyses of 2 or 3 isolated torsions are misleading

- Analysis of the dinucleotide fragment provided

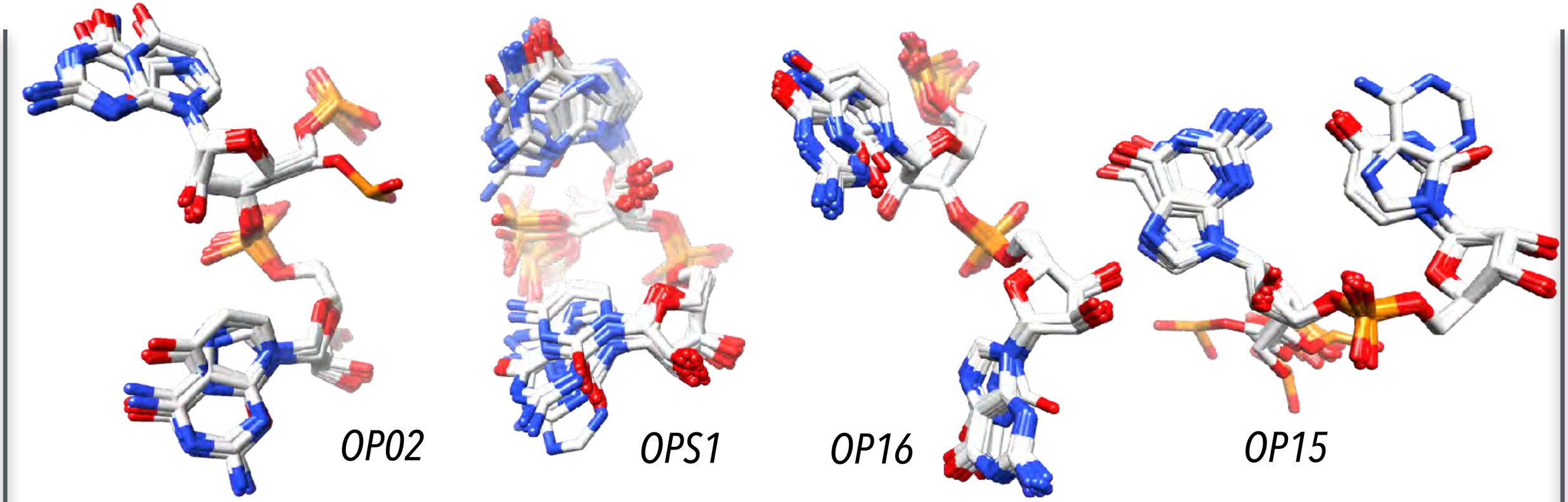
  96 NtC classes

  - analysis in 12D torsion space
  - cluster analysis + empirical rules



Černý et al.: Nucleic Acids Research **48**: 6367 (2020).

Two levels of analysis:
NtC = geometry assignment; CANA = symbolic annotation



NtCs *OPxx grouped to* ***CANA*** *"letter" OPN*

NtC conformational classes at *dnatco.datmos.org*

… annotate, validate, refine …

Home | Annotation | Validation | Refinement | Downloads | Browse

4QVI Crystal structure of mutant ribosomal protein M218L TthL1 in complex with 80nt 23S RNA from Thermus thermophilus

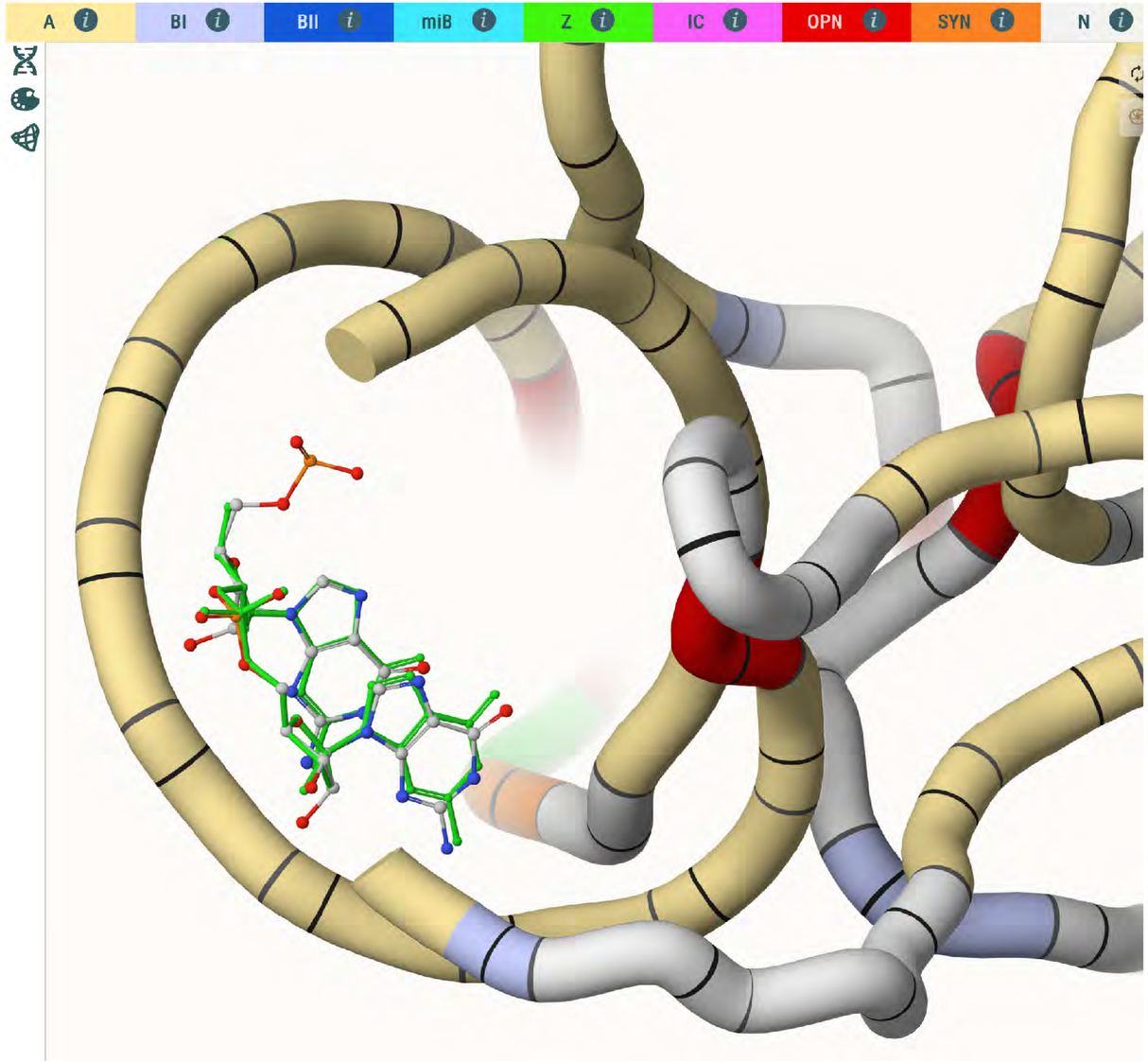
Resolution 1.9 Å (Low: 23.5)

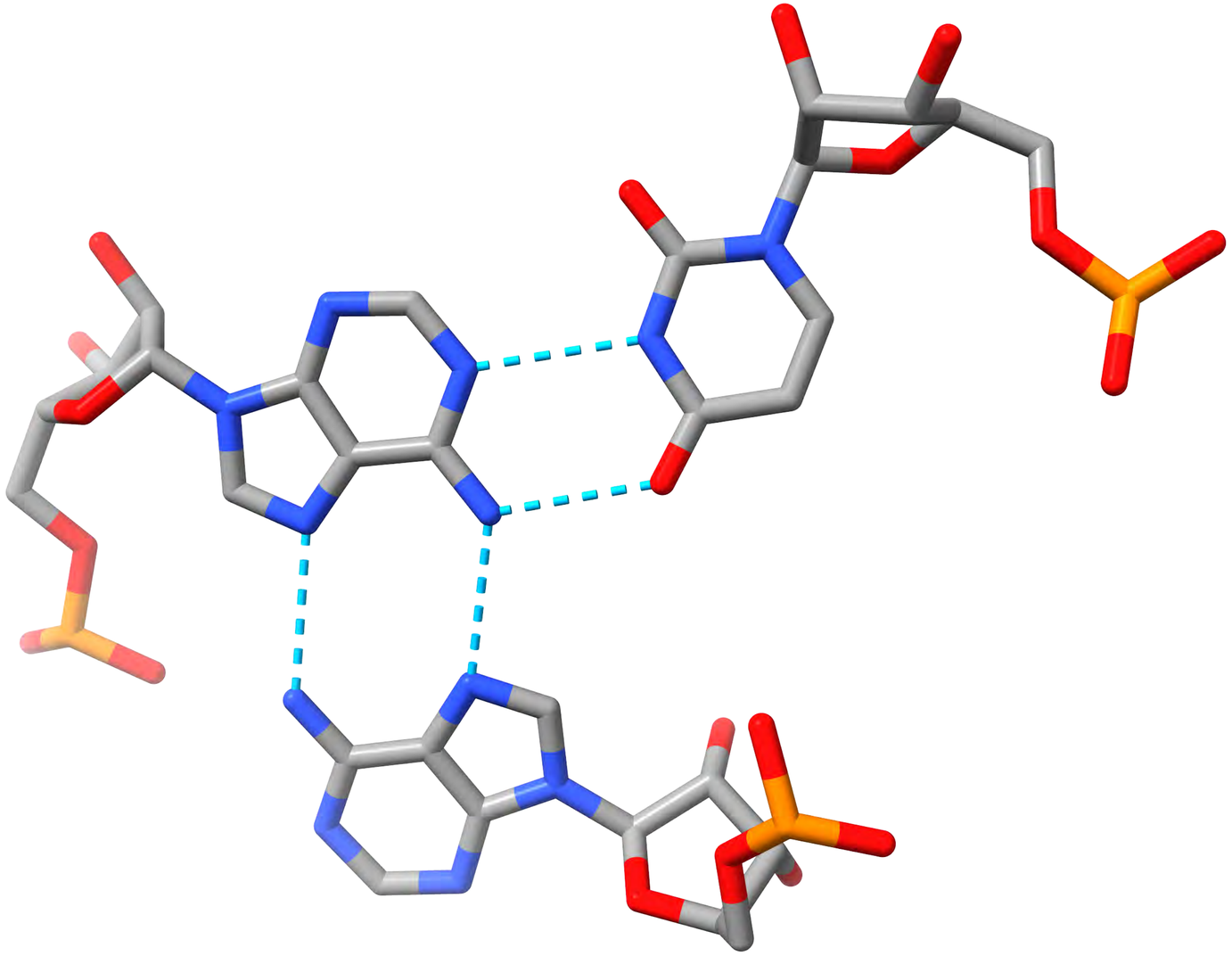
Backbone conformational quality

Chain: All NAs

Table of assigned dinucleotide NtC classes

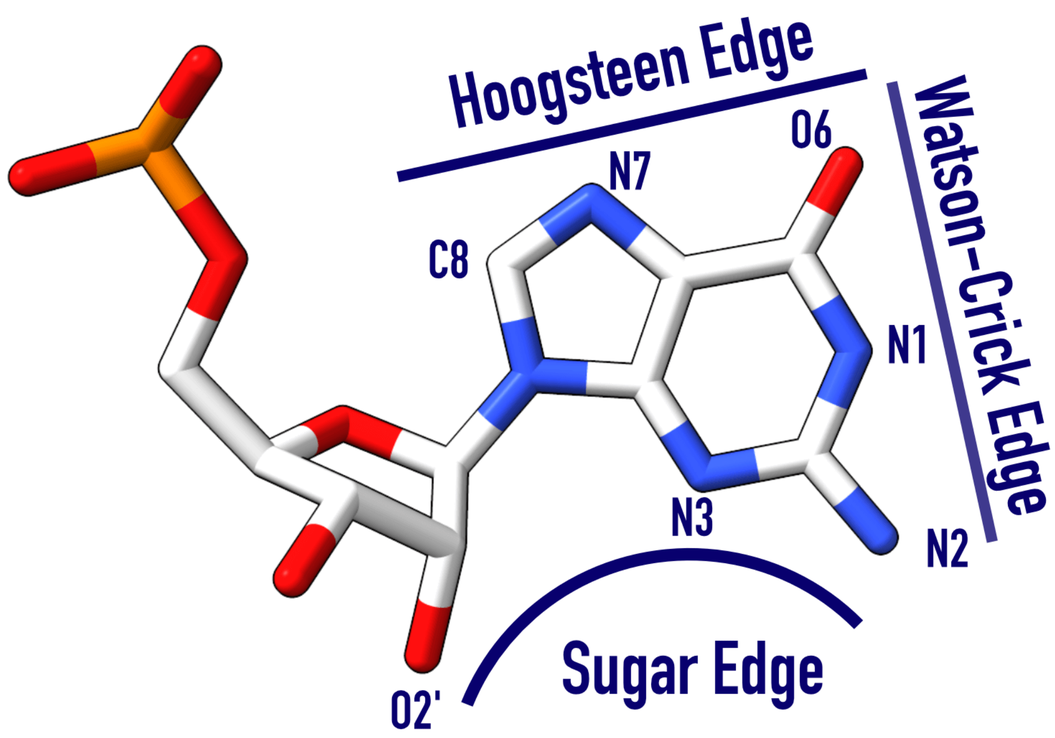
B	C2146 G2147	ZZ1S	ZZZ	66	0.212
B	G2147 G2148	AA05	AAw	0	0.524
B	G2148 G2149	AA00	AAA	87	0.171
B	G2149 U2150	AA00	AAA	71	0.205
B	U2150 G2151	AA00	AAA	86	0.171
B	G2151 G2152	AA00	AAA	54	0.198
B	G2152 G2153	AA00	AAA	69	0.139
B	G2153 G2154	AA00	AAA	72	0.229
B	G2154 G2155	AA01	AAw	94	0.147
B	G2155 G2156	AA00	AAA	19	0.503
B	G2156 G2157	AB05	A-B	27	0.475
B	G2157 A2158	OP22	OPN	0	0.770
B	A2158 G2159	BA17	B-A	0	0.805
B	G2159 G2160	AA11	AAw	0	0.640
B	G2160 C2161	AA00	AAA	54	0.312
B	C2161 G2162	AA00	AAA	65	0.293
B	G2162 C2163	AA00	AAA	54	0.357
B	C2163 C2164	AA00	AAA	36	0.446
B	C2164 G2165	AA00	AAA	47	0.391
B	G2165 G2166	AA00	AAA	68	0.301
B	G2166 U2167	AA00	AAA	68	0.218



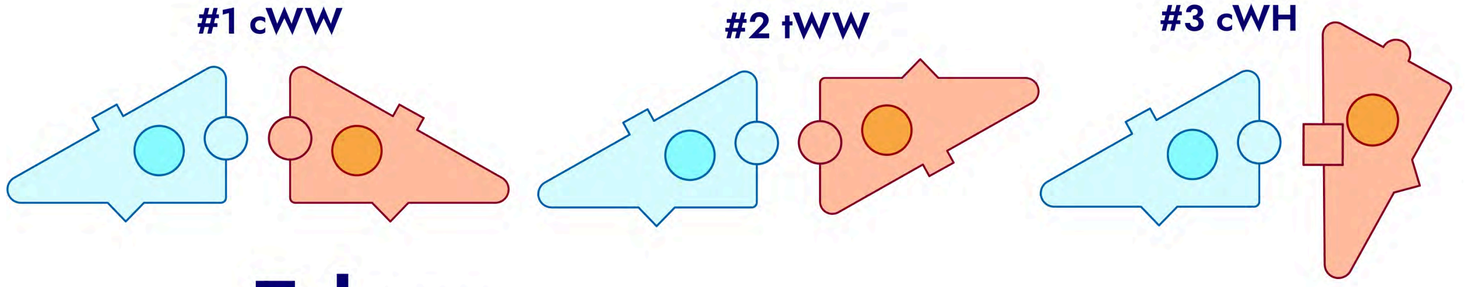


How are base pairs classified?

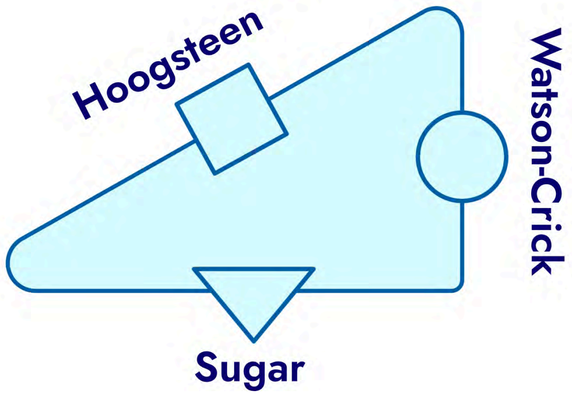
We use ***Leontis-Westhof classification*** based on three base "***edges***" leads to twelve base pair ***families*** which form 152 possible ***classes***.

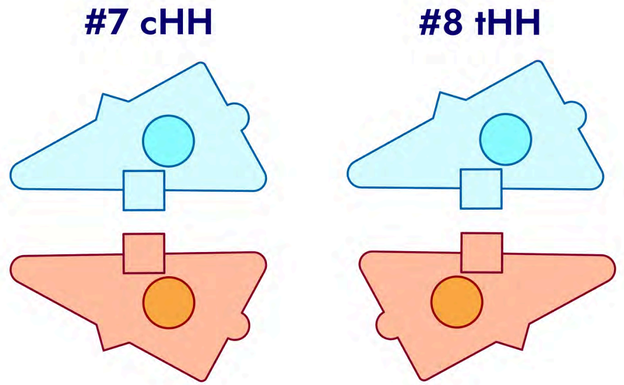


Families



Edges

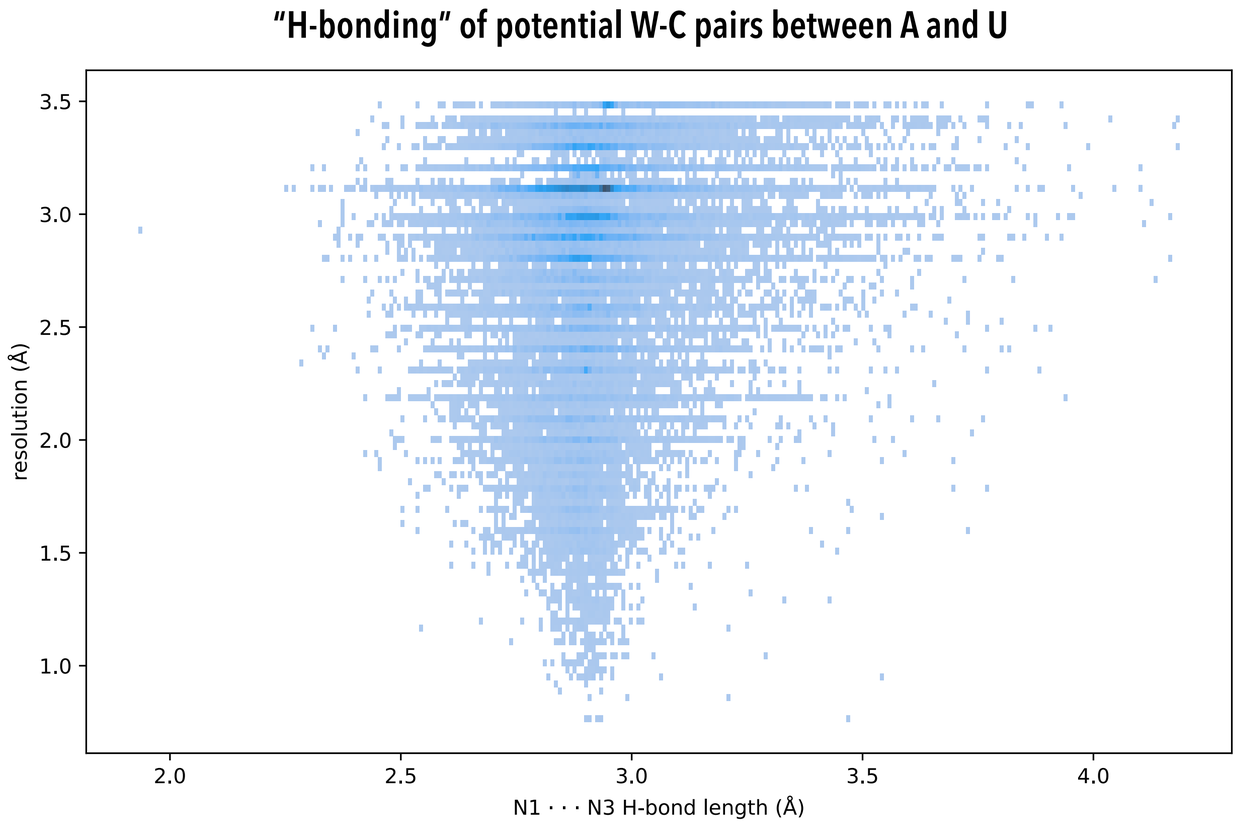




Let's Make Base Pairing Great Again!

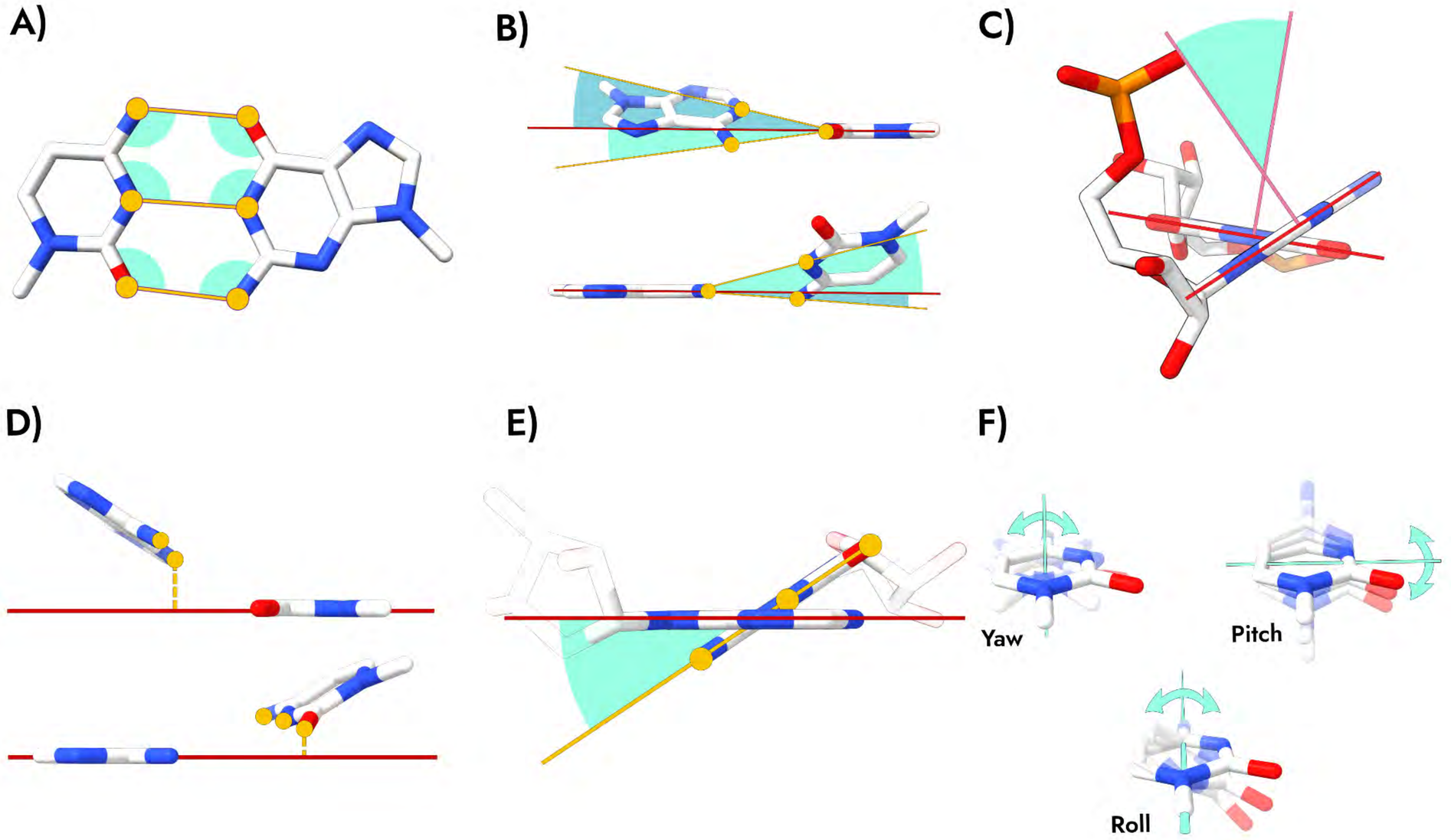
NAPAIR base pair assignment

1. Find any base – base atoms closer than 4 Å to get base – base *contacts*
2. Filter the base – base contacts by a set of parameters to get *potential pairs*
   - Parameters must have wide distributions



NAPAIR parameters

A) atom-atom distances and angles; B) distance-to-plane angle; C) coplanarity angle; D) coplanarity shift; E) coplanarity edge angle; F) aircraft angles yaw, pitch and roll



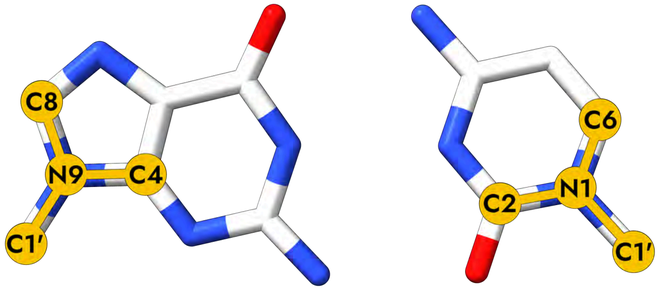
Czech Academy of Sciences

Let's Make Base Pairing Great Again!

NAPAIR base pair assignment II

3. Calculate rmsd between the tested pair and all members of the *Curated set*, pairs with good geometry.

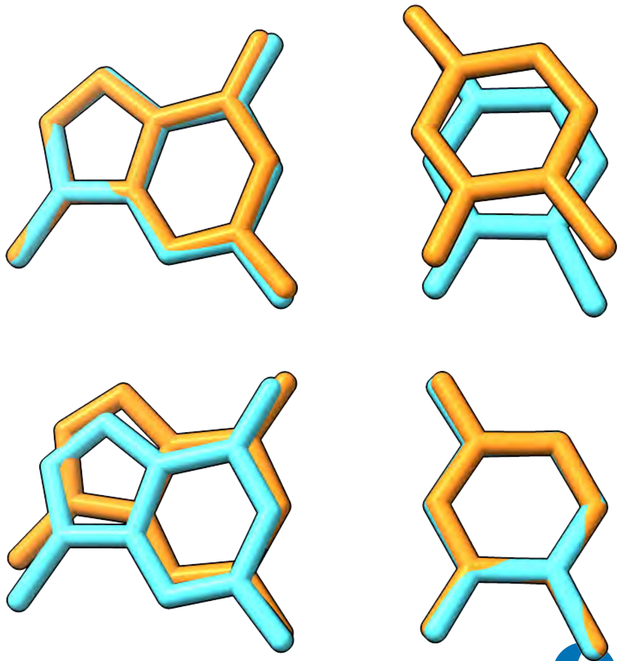
The coordinate frame to calculate the rmsd values



4. *Validate* pair quality by *5NN* (*Nearest Neighbor*) metric.

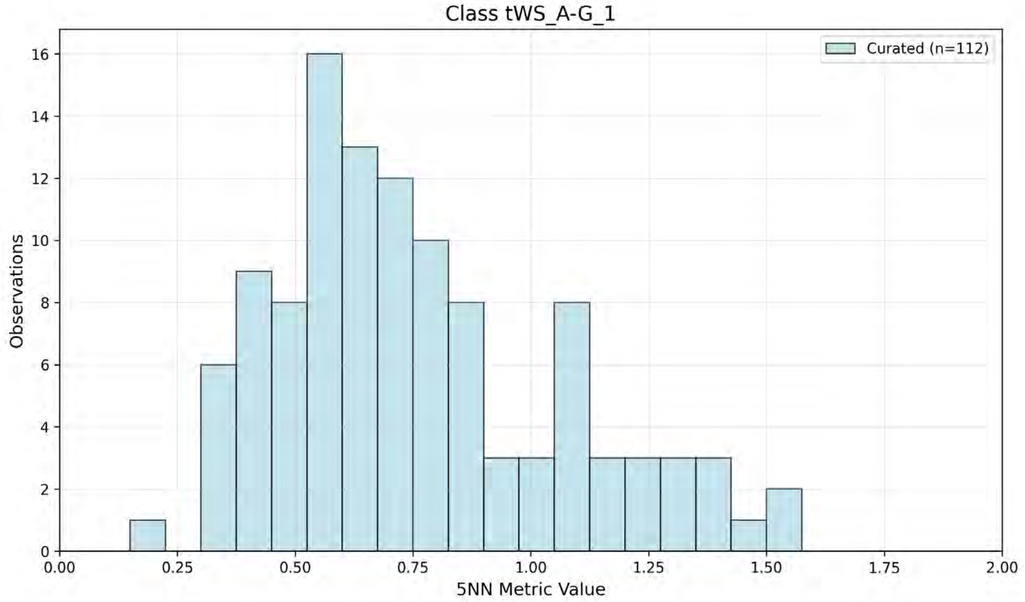
$$5NN = \frac{1}{\sqrt{\mathrm{RMSD}_1^2 + \mathrm{RMSD}_2^2 + \mathrm{RMSD}_3^2 + \mathrm{RMSD}_4^2 + \mathrm{RMSD}_5^2}}$$

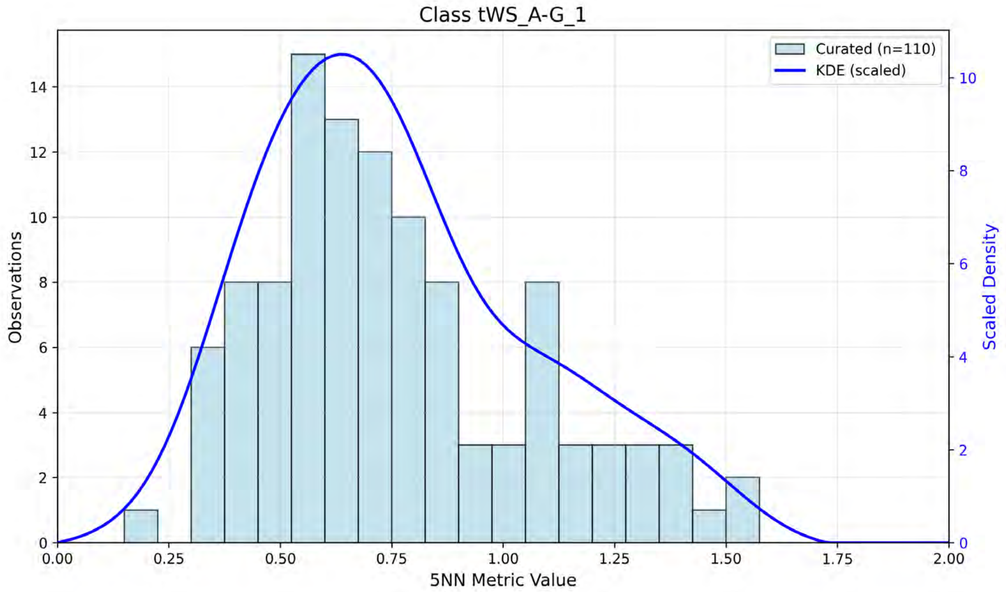
NAPASCO=CDF(5NN)[1-CI(5NN)]

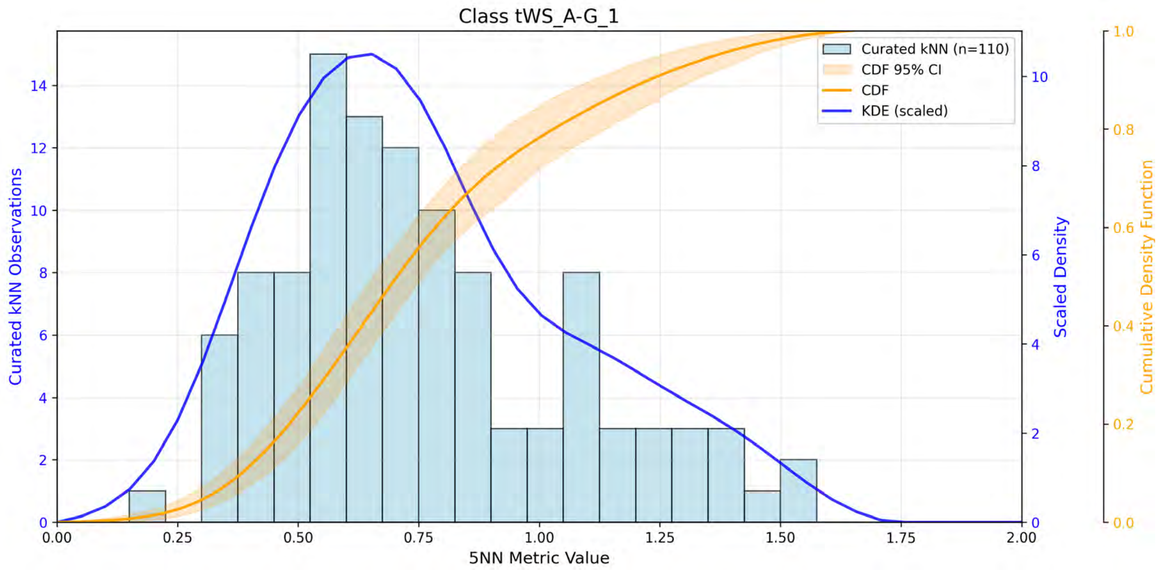


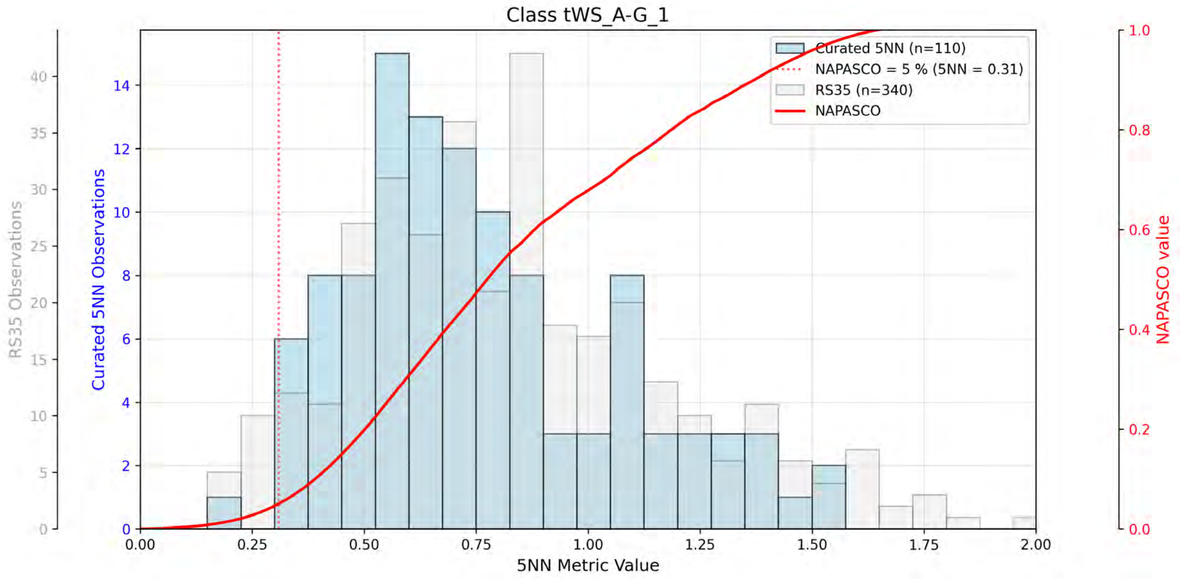
Czech Academy of Sciences

NAPAIR validation based on pairs in the Curated Sets: *NAPASCO*





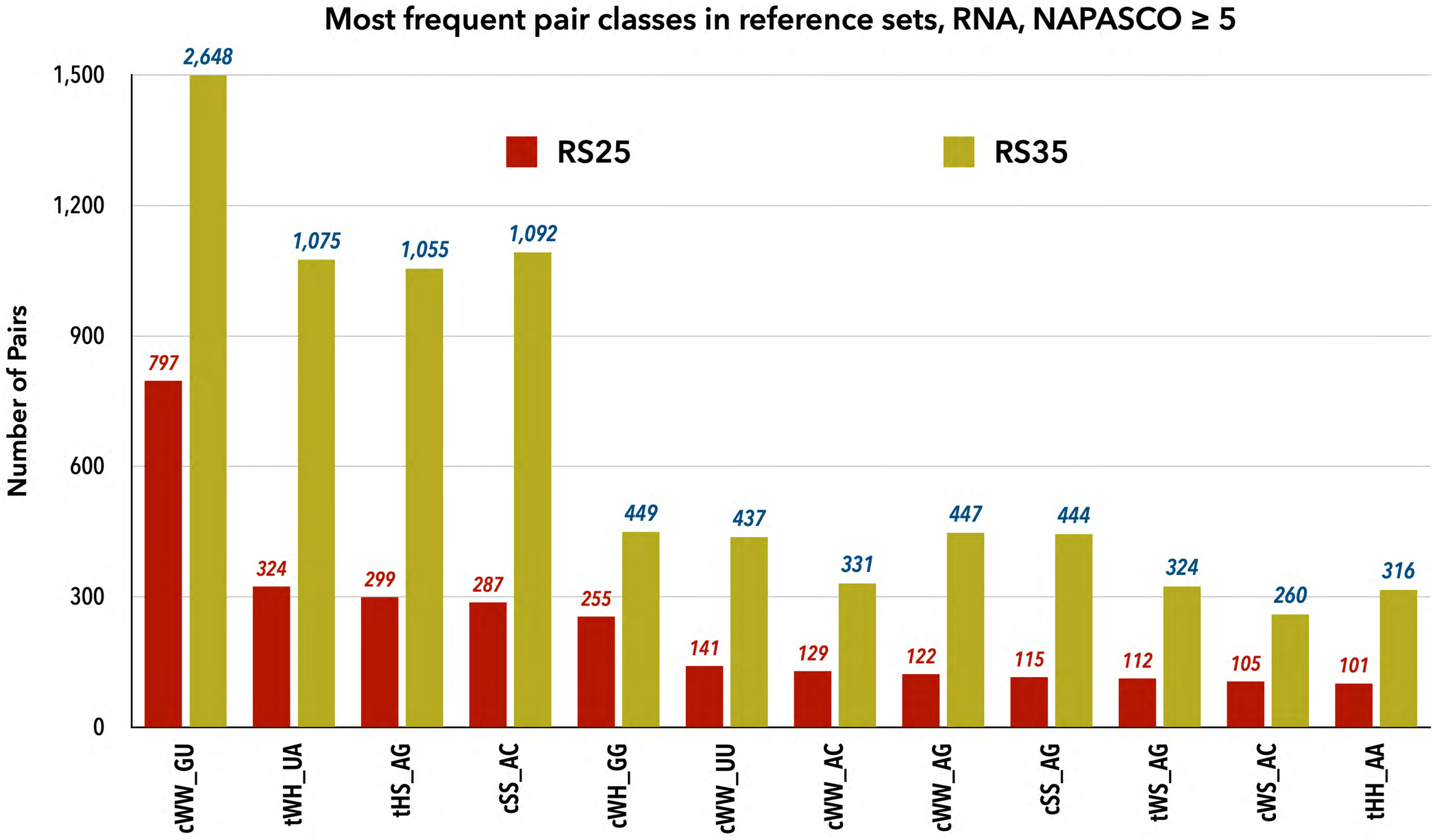




Czech Academy of Sciences

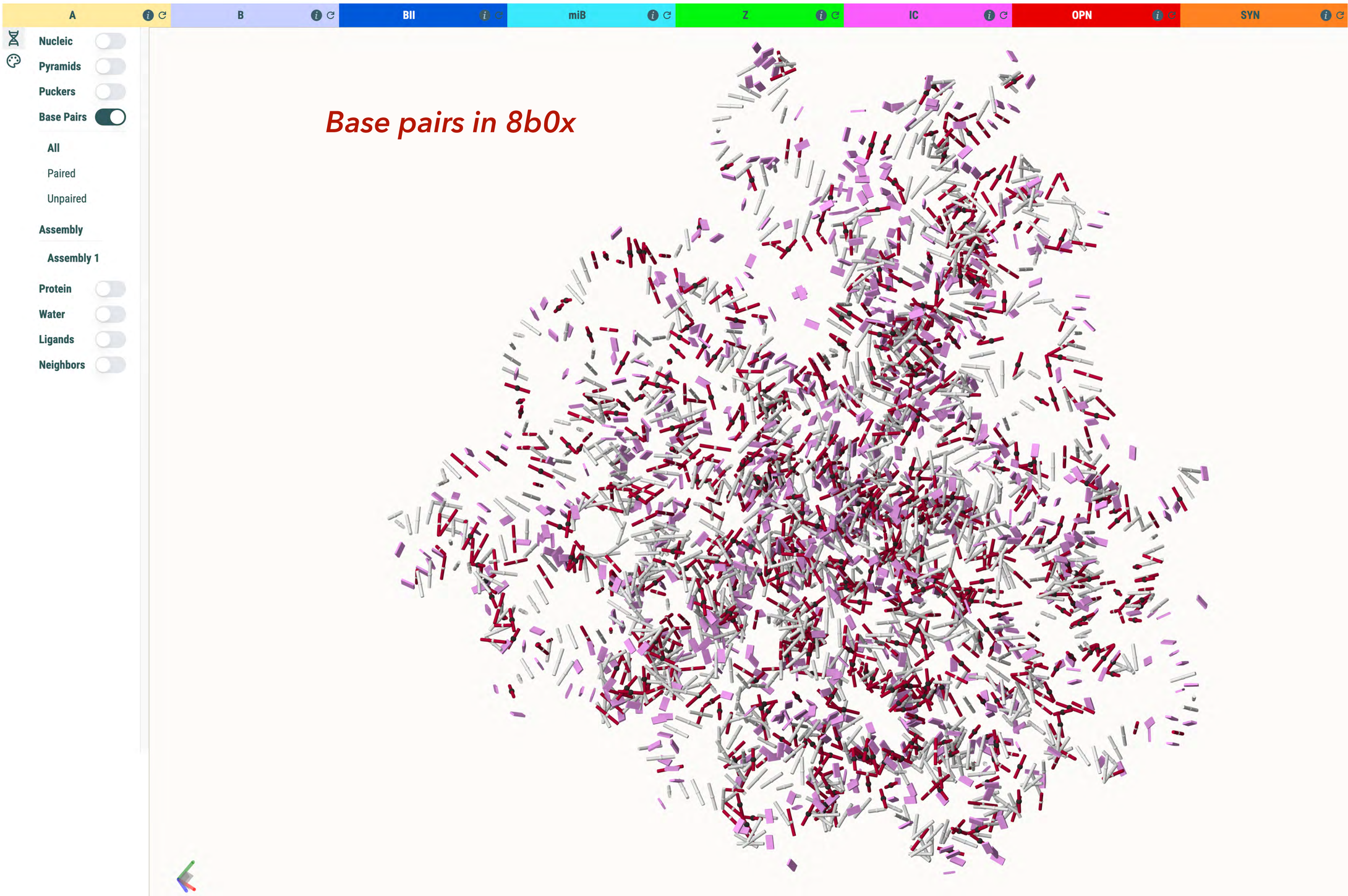
NAPAIR assignment: most frequent L-W classes in RNA

cWW_G-C 52 %, cWW_A-U 21 %, rest 27 %

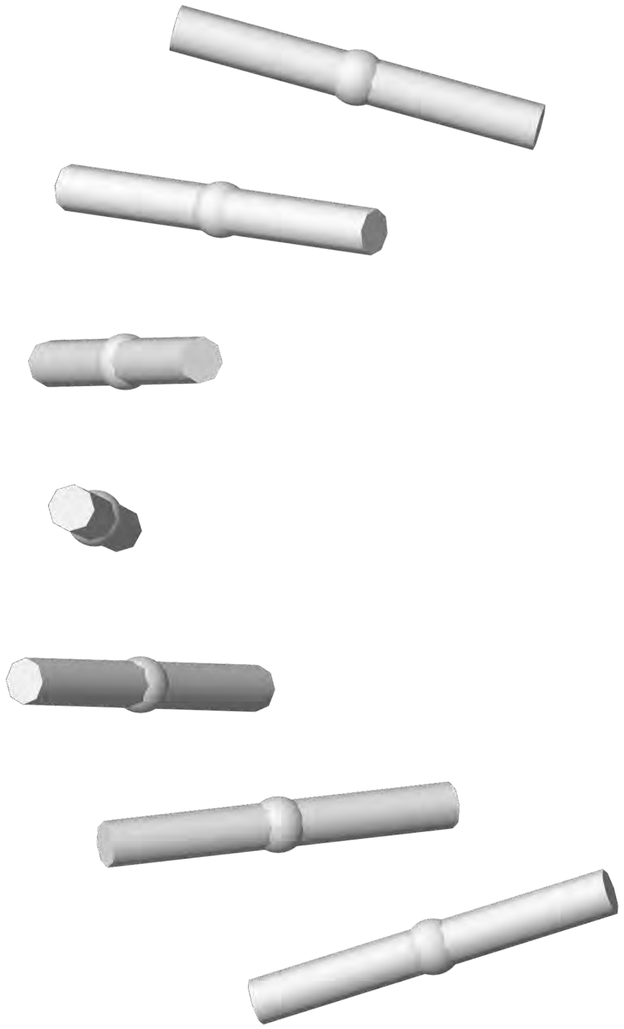


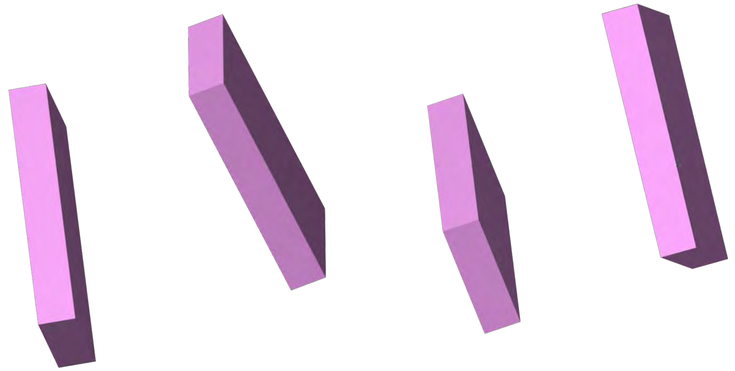
Czech Academy of Sciences

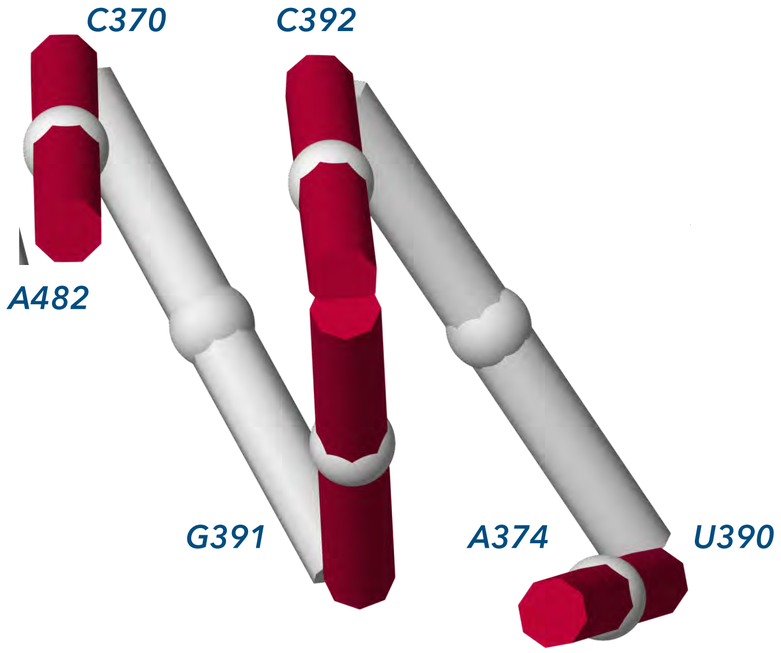
Base pair visualization @ dnatco.datmos.org

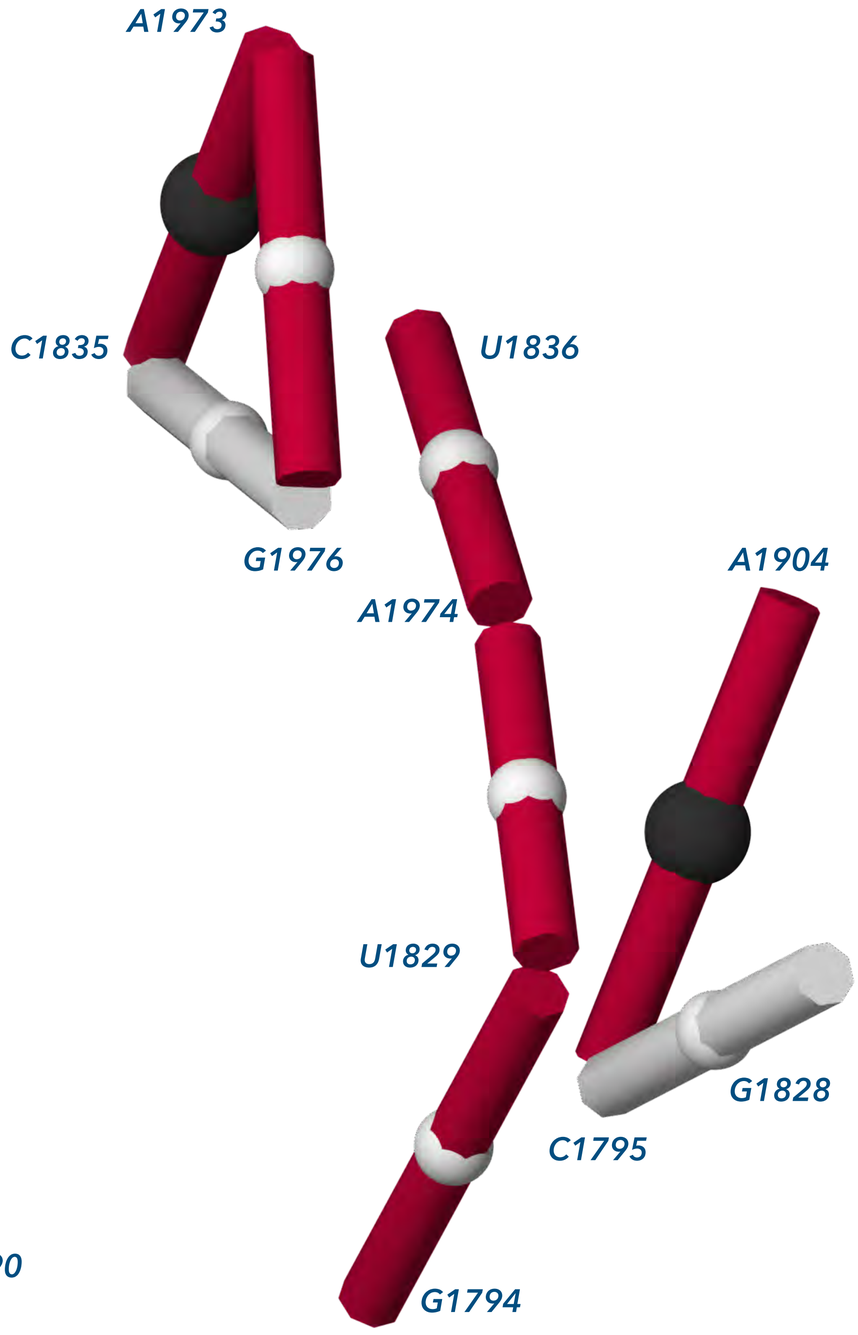


Base pairs motifs in 8b0x

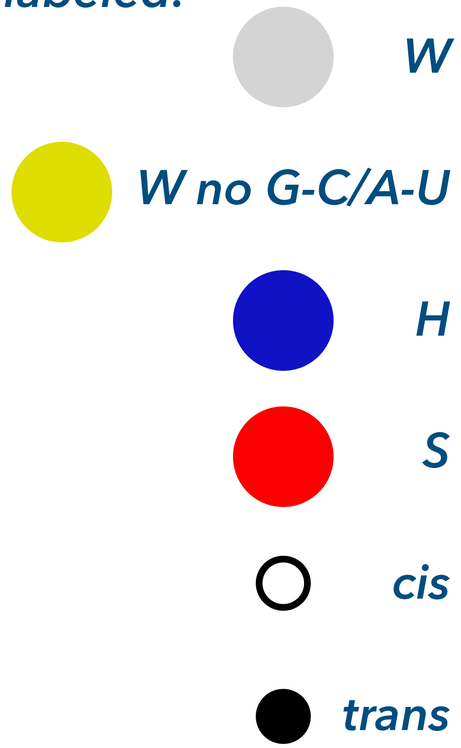


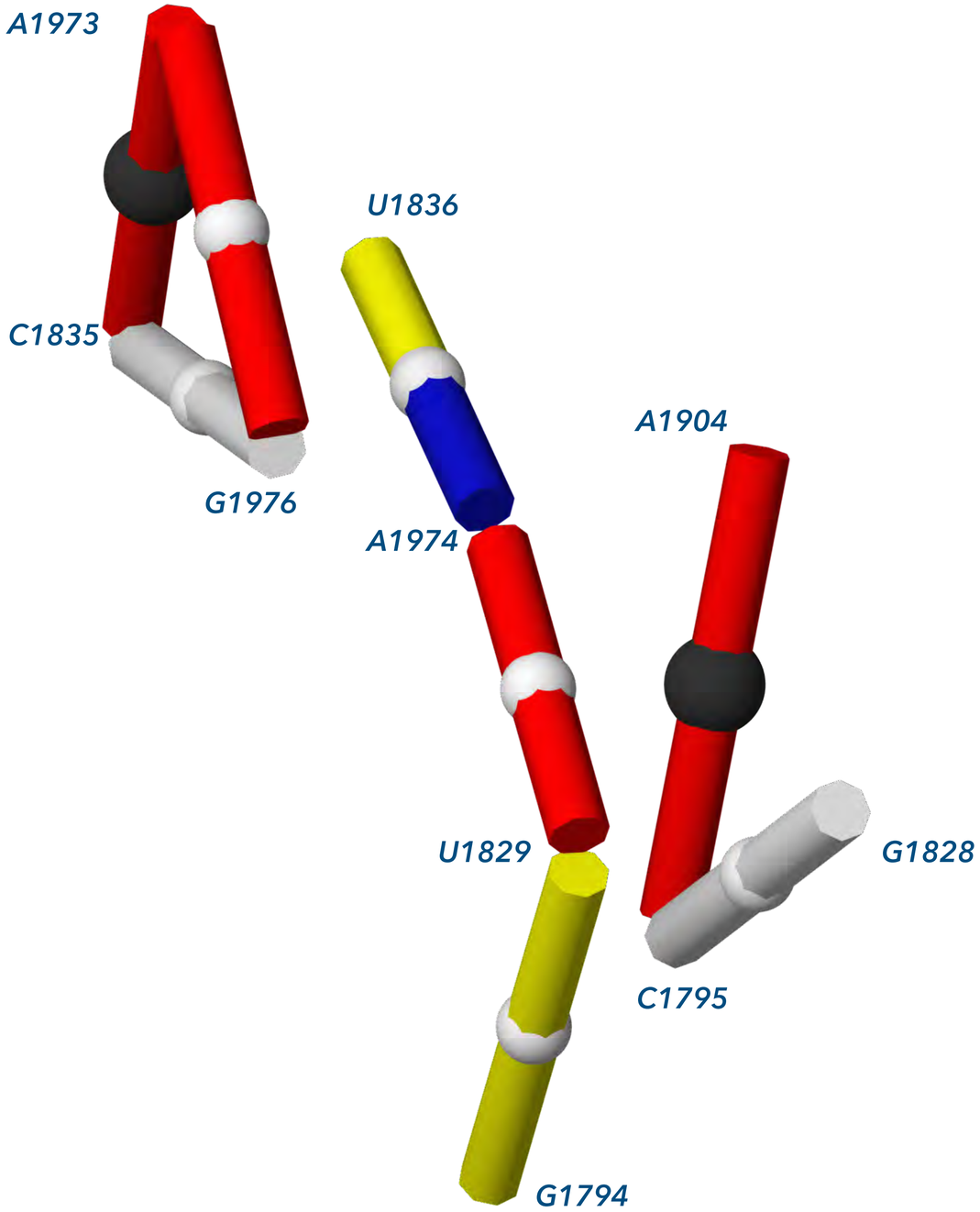






… and a funny 8b0x motif with edges labeled:





Does it matter what data we use to classify structures?

- PDB is
  - highly redundant
  - full of low-resolution *i.e.*, unreliable structures
- Quality filtered subsets of the public archive are a must
- We work with two reference sets:

  RS35 quality-filtered ensemble of X-ray structures with resolution ≤ 3.5 Å

  … and its subset RS25 of "high-resolution" structures, resolution ≤ 2.5 Å
  - clustered by sequence identity
  - the highest quality *chain* selected based on the BGSU quality metric extended for DNA.
    Metric has been used in NAVAL project

  RS25: 36,010 RNA, 39,856 DNA nucleotides

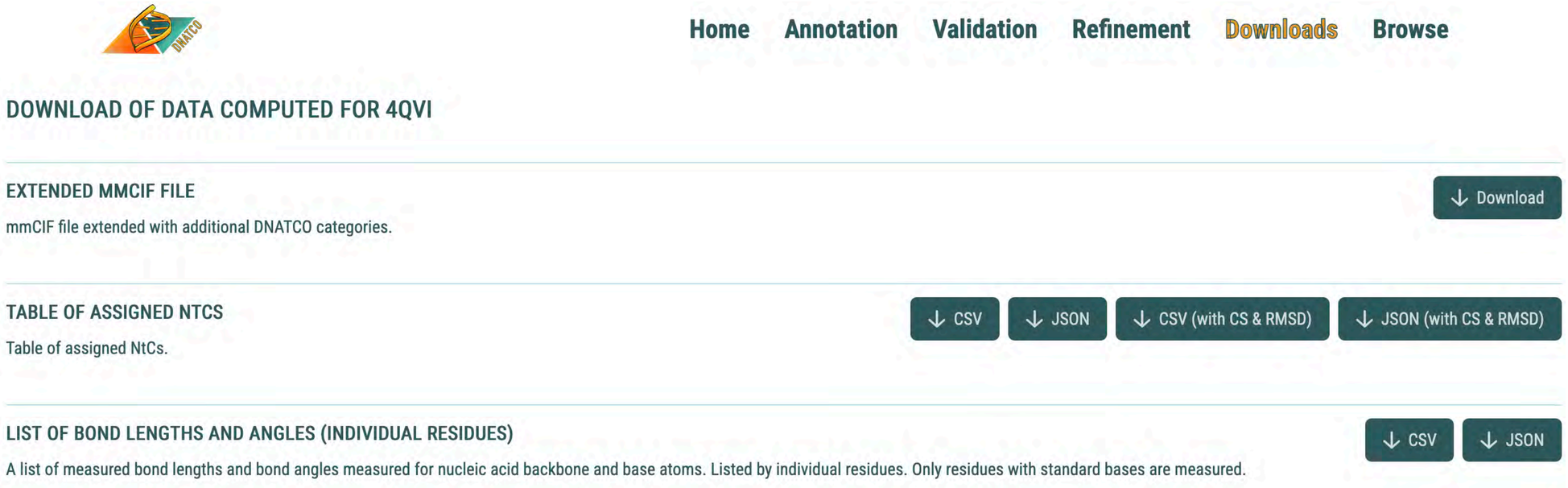
  RS35: 120,179 RNA, 86,220 DNA nucleotides

  the whole PDB: 10,267,270 RNA, 963,416 DNA nucleotides

… and how about the interoperability?

- … all outputs from *dnatco.datmos.org* are based on mmCIF categories
  - full mmCIF files complemented by NA-specific categories can be downloaded
- mmCIF categories describing base pairing

```
_ndb_base_pair_list
_ndb_base_pair_annotation
_ndb_base_pair_validation
```



Public archive describes basics for proteins

```
loop_
_struct_conf.conf_type_id
_struct_conf.id
_struct_conf.pdbx_PDB_helix_id
_struct_conf.beg_label_comp_id
_struct_conf.beg_label_asym_id
_struct_conf.beg_label_seq_id
_struct_conf.pdbx_beg_PDB_ins_code
_struct_conf.end_label_comp_id
_struct_conf.end_label_asym_id
_struct_conf.end_label_seq_id
_struct_conf.pdbx_end_PDB_ins_code
_struct_conf.beg_auth_comp_id
_struct_conf.beg_auth_asym_id
_struct_conf.beg_auth_seq_id
_struct_conf.end_auth_comp_id
_struct_conf.end_auth_asym_id
_struct_conf.end_auth_seq_id
_struct_conf.pdbx_PDB_helix_class
_struct_conf.details
_struct_conf.pdbx_PDB_helix_length
HELX_P HELX_P1 AA1 THR A 52  ? GLU A 62  ? THR AAA 52  GLU AAA 62  1 ? 11
HELX_P HELX_P2 AA2 SER A 96  ? GLN A 103 ? SER AAA 96  GLN AAA 103 1 ? 8
HELX_P HELX_P3 AA3 SER A 105 ? ASP A 125 ? SER AAA 105 ASP AAA 125 1 ? 21
#
```

```
loop_
_struct_sheet_order.sheet_id
_struct_sheet_order.range_id_1
_struct_sheet_order.range_id_2
_struct_sheet_order.offset
_struct_sheet_order.sense
AA1 1 2 ? parallel
AA1 2 3 ? anti-parallel
AA1 3 4 ? anti-parallel
AA2 1 2 ? parallel
AA2 2 3 ? anti-parallel
AA2 3 4 ? parallel
AA2 4 5 ? anti-parallel
AA2 5 6 ? anti-parallel
#
 …
```

Czech Academy of Sciences

… and for nucleic acids …

Summary

NAPAIR: validated base pairs

All outputs (NAPAIR, conformational NtC classes) interoperable, based on mmCIF

All “dinucleotide” NA features represented in tables and 3D graphics at dnatco.datmos.org

This research was funded by
Czech Academy of Sciences,
grant RVO 86652036

ELIXIR CZ from
MEYS Czech Republic
grant LM2023055 to

The DNATCO team

Jiří Černý
Michal Malý
Paulína Božíková
Terezie Prchalová
Jakub Svoboda
Lada Biedermannová
Bohdan Schneider

The NAPAIR team

Jakub Svoboda
Stanislav Lukeš
Jiří Černý
Helen Berman
Lada Biedermannová
Paulína Božíková
Robbie Joosten
Cathy Lawson
Marcin Magnus
Brinda Vallat
Craig Zirbel &
Eric Westhof
Bohdan Schneider

bohdan.schneider@ibt.cas.cz

Czech Academy
of Sciences

Thank you for your attention!

ibt

Czech Academy of Sciences