RNA in 3D: more than canonical pairs and “unpaired loops”

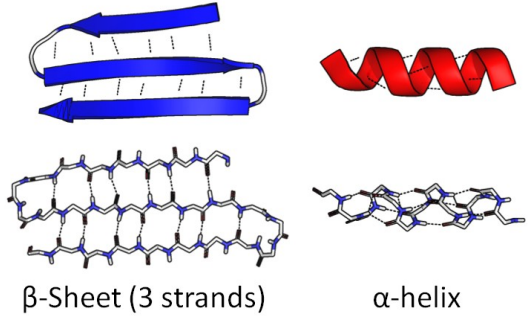
Jiří Černý
jiri.cerny@ibt.cas.cz

Laboratory of Structural Bioinformatics of Proteins
Institute of Biotechnology, Czech Academy of Sciences

DNATCO Team:
Paulína Božíková
Michal Malý
Jakub Svoboda
Terezie Prchalová
Bohdan Schneider

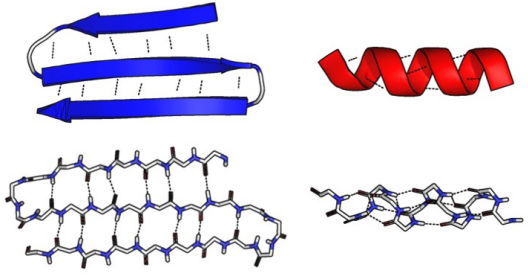
“Secondary structure”

- Proteins
  - Pattern of backbone hydrogen bonds

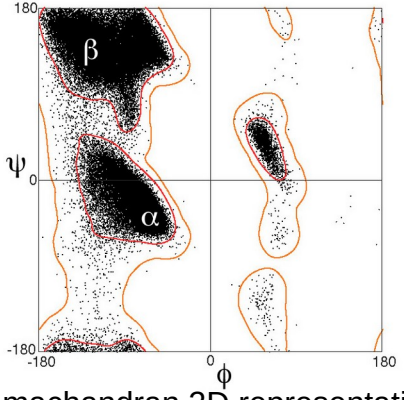


“Secondary structure”

- Proteins
  - Pattern of backbone hydrogen bonds



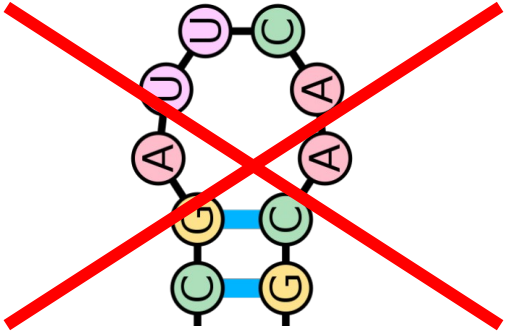
  - Equivalent to φ/ψ backbone torsions



(Ramachandran 2D representation)

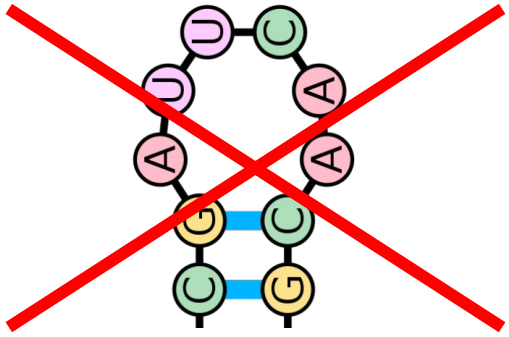
“Secondary structure”

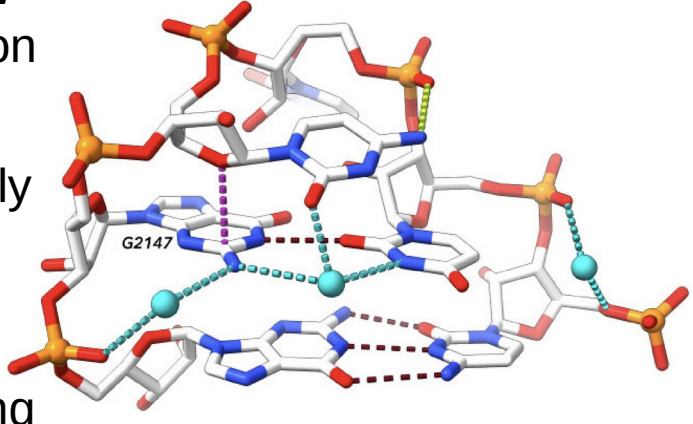
- Base pairing (hydrogen bond pattern) is only a small part of the structure
  - Using it as NA secondary structure is insufficient
  - In practice this is even worse because in most cases only canonical WC pairs are considered



“Secondary structure”

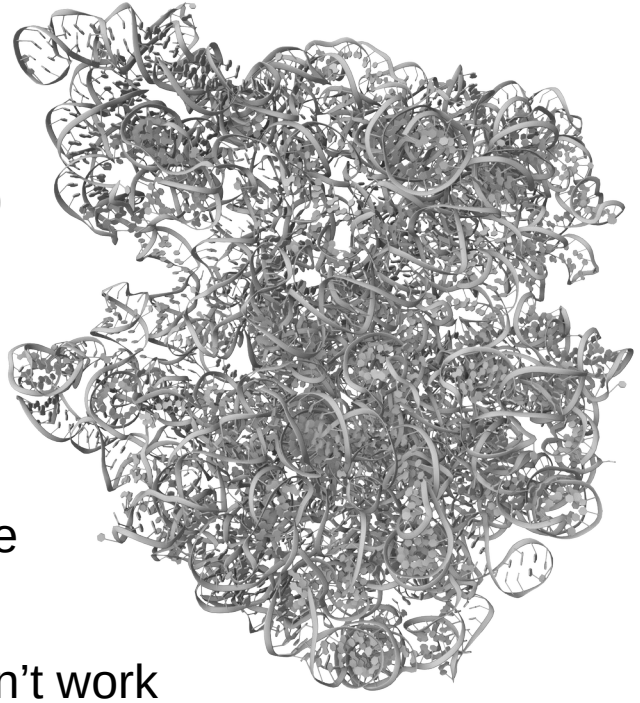
- Only a combination of much richer, more complicated **base interactions** (including non-canonical pairs) **and backbone conformations** can describe the full 3D structure in “2D”
  - NA bases are hetero-aromatic moieties with high dipoles, strong H-bonds and base stacking (vdW)
  - Stacking is responsible for the helicity, and vdW strength is comparable to H-bonding contribution
  - Sugar-phosphate backbone contains larger number of polar atoms (H-bonds) and negatively charged polarizable phosphodiester group
  - extra O2‘ OH in RNA adds further complexity
  - Water/ions/ligands play important role in defining the structure as well





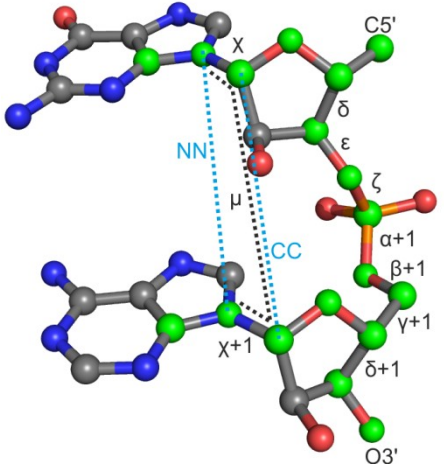
3D structures

- Much less structural data is available for Nucleic Acids
  - 200k protein structures in PDB
  - 16k NA structures in PDB
  - 42M AAs *vs.* 100k nucleotides (< 2Å resolution)
- No reliable algorithms for NA structure
  - Annotation and Validation
  - Modeling and Refinement
  - Including bond lengths and angles, not “only” the expected backbone torsions uncertainties
- Using protein tools and approaches for NAs just doesn’t work
- We are developing a set of tools, databases, and workflows available at the “database of molecular structures” datmos.org



Nucleic acid backbone conformations

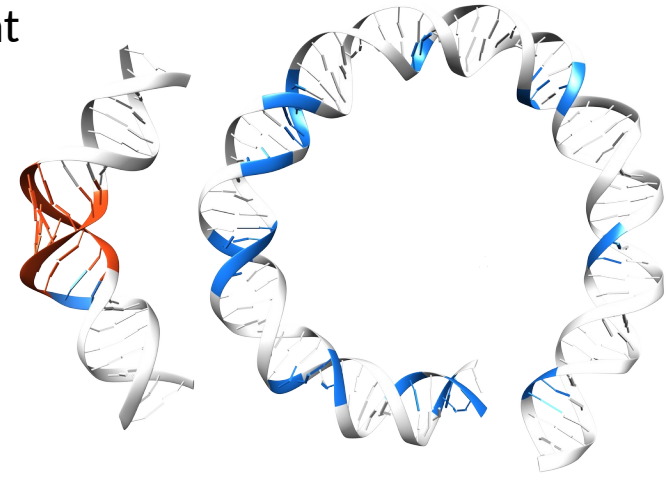
- Identification of important nucleic acid structure features at different levels of detail
  - Automated assignment (no expert knowledge needed)
  - Parameter free (based on/derived from data only)
- From atomic coordinates
  - Backbone conformations & Structural motifs
    - How to assign/annotate ***local NA conformations***?
      - Cartesian coordinates impractical
      - many degrees of freedom, no simple Ramachandran-like plot possible
      - ***working with the sugar phosphate backbone (torsions and distances) in a dinucleotide block***
      - clustering limited by (un)available structural data (tetranucleotide would be better)
      - (di)Nucleotide Conformations (***96+1 NtC classes***)



Dinucleotide block and its backbone parameters.

Nucleic acid backbone conformations

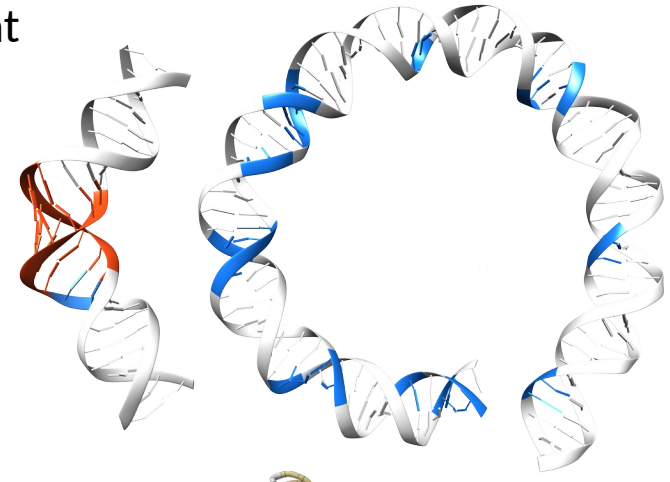
- Identification of important nucleic acid structure features at different levels of detail
  - Automated assignment (no expert knowledge needed)
  - Parameter free (based on/derived from data only)
- From atomic coordinates
  - Backbone conformations & Structural motifs
    - How to assign/annotate *local NA conformations*?
      - Cartesian coordinates impractical
      - many degrees of freedom, no simple Ramachandran-like plot possible
      - *working with the sugar phosphate backbone (torsions and distances) in a dinucleotide block*
      - clustering limited by (un)available structural data (tetranucleotide would be better)
      - (di)Nucleotide Conformations (***96+1 NtC classes***)

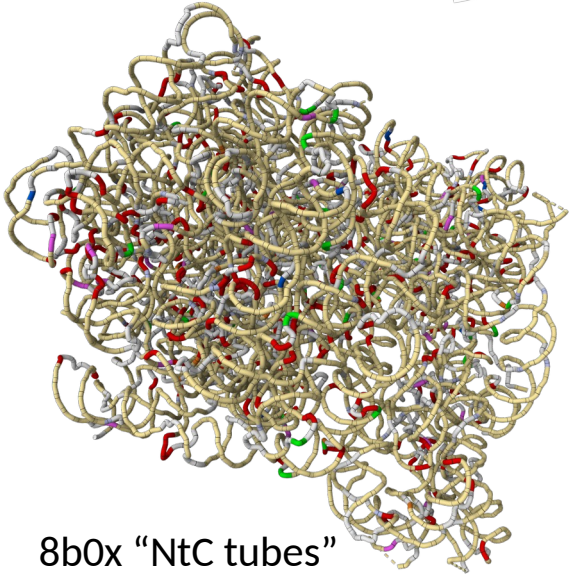


Rules of transcription factors and nucleosome interactions.

Nucleic acid backbone conformations

- Identification of important nucleic acid structure features at different levels of detail
  - Automated assignment (no expert knowledge needed)
  - Parameter free (based on/derived from data only)
- From atomic coordinates
  - Backbone conformations & Structural motifs
    - How to assign/annotate *local NA conformations*?
      - Cartesian coordinates impractical
      - many degrees of freedom, no simple Ramachandran-like plot possible
      - *working with the sugar phosphate backbone (torsions and distances) in a dinucleotide block*
      - clustering limited by (un)available structural data (tetranucleotide would be better)
      - (di)Nucleotide Conformations (***96+1 NtC classes***)





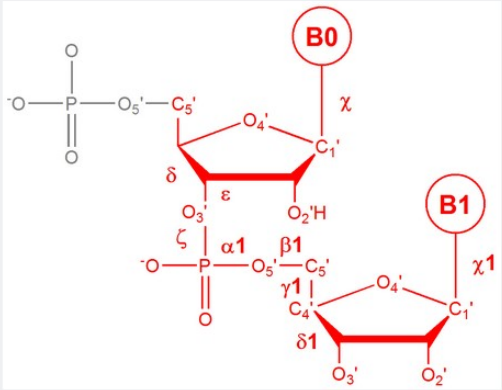
8b0x “NtC tubes”

2013: DNA only, 18 conformers (based on 2008 work)

Assignment of DNA conformers

The server assigns 18 conformers based on the values of their backbone torsion angles.

- Table defining the conformers is below and Cartesian coordinates of their representative samples can be downloaded.
- The conformers consist of three BI-forms, two BII-forms, three A-forms, three Z-forms, four mixed A/B forms, and three conformers with bases in the syn orientation.
- Conformationally extreme conformers are not assigned to any of the above; these steps formally represent the 19th conformer.
- More information about the conformers and the ways how the conformers were identified can be found in the paper by Svozil et al., Nucleic Acids Research, 36, 3690 (2008).
- Assignment of DNA conformers has been used in bioinformatic analysis of protein/DNA interfaces in the paper by Schneider et al., Nucleic Acids research, 42, 3381 (2014).



Please, read before you upload coordinates of your DNA:

- DNA steps are identified based on atom names as defined by the PDB format, version 3.1 or above (sugar atoms as O4' not O4*).
- Steps with non-standard or missing atoms that define torsions δ ... δ+1, χ, and χ+1 are not considered in the assignment process.
- Conformers are assigned for modified residues that contain standard names for atoms defining the step torsions between δ and δ+1 and χ and χ+1.

Browse for the PDB file

Browse... No file selected.

SUBMIT

or

Enter PDB ID

SUBMIT

Nucleotide conformers (ntC) used for the assignment of steps. Conformers are identified by numbers as in Svozil et al. Nucleic Acids Research, 36, 3690 (2008).

ntC	Description	δ	ε	ζ	α+1	β+1	γ+1	δ+1	χ	χ+1
8	"canonical" A-DNA	83	205	287	294	174	54	83	199	203
13	A-DNA, BI-like χ	89	201	275	294	162	54	89	244	244
19	A-DNA, α+1/γ+1 crank	84	194	290	149	192	182	88	205	188
41	A-to-B, δ>C3'-, δ+1 C2'-endo	90	196	280	299	179	55	142	222	256
32	BI-to-A, δ+1 O4'-endo	129	186	264	295	170	52	98	247	233
109	BII-to-A, δ+1>C3'-endo	142	213	181	297	139	52	90	273	207
110	as 109 plus α+1/γ+1 crank, high β+1	146	257	186	60	224	196	90	260	200
54	"canonical" BI	136	183	259	303	181	44	138	252	259
50	BI variant	129	181	265	300	177	50	123	246	245
86	BII variant	140	201	216	314	154	46	140	262	253
96	BII variant	143	245	170	297	141	46	141	271	257

https://dnatco.datmos.org/v1

2016: DNA, 44 conformers

Assignment of DNA conformers (DNATCO v2.3)

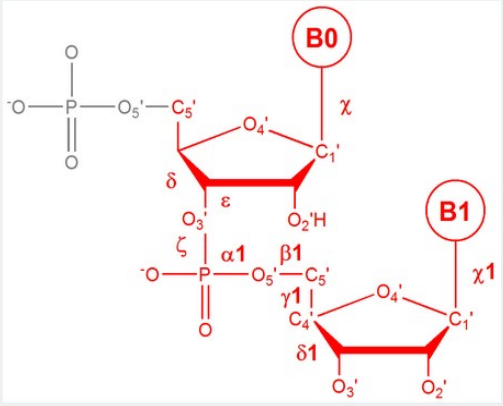
Tutorial Test run (PDB ID 1bna)

The server assigns 44 DNA conformers based on the values of their 9 backbone torsion angles.

- Table defining the conformers is below; definition of the conformers, their esd values and Cartesian coordinates of their representative samples can be downloaded.
- Conformers are identified by four-letter symbols. "**A**", "**B**", "**Z**" letters imply stacked bases with first/second nucleotide in **A**, **B**, or **Z** like conformation. "**NS**" lables steps with **N**ot **S**tacked bases. "**S**" at 3rd or 4th position means that the 1st or 2nd base is in **syn** orientation.
- Conformationally extreme conformers are not assigned to any of the above; these steps formally represent the 45th conformer.
- How to cite:
  - The web service in version 2 is described in Černý et al., Nucleic Acids Research, 44, W284 (2016).
  - The conformers and the ways they were identified is described in the paper by Schneider et al., Acta Cryst D, 74, 52-64 (2018).
  - For example of application of the DNA Structural Alphabet see Schneider et al., Genes, 8(10), 278, (2017).

Please, read before you upload coordinates of your DNA:

- DNA steps are identified based on atom names as defined by the PDB format, version 3.1 or above (sugar atoms as O4' not O4*).
- Contact the authors for off-line analysis of non-standard or large structures (multiple NMR MODELs or MD simulation trajectory).
- Steps with non-standard or missing atoms that define torsions δ ... δ+1, χ, and χ+1 are not considered in the assignment process.
- Conformers are assigned for modified residues that contain standard names for atoms defining the step torsions between δ and δ+1 and χ and χ+1.



elixir CZECH REPUBLIC

Browse for the PDB file

Browse... No file selected.

SUBMIT

or

Enter PDB ID

SUBMIT

News and changes:

- v2.3:
  - Interactive 3D Complement pages for publications from 2017 (I3D) and 2018 (I3D) can be found at Proteopedia.
  - the improved assignment protocol involves known δ/pseudorotation angle correlation for detection of outliers
  - a reference (representative NtC structure) is superposed upon the selected step and cartesian RMSD for atoms defining the nine torsions is reported
  - protein/hetero/water groups from the original PDB can be displayed (P/H/W buttons)
  - introduced a tetrahedron centered at the phosphate:
    - its size represents the confal value
    - its user adjustable color allows simple recognition of A/BI/BII/miB/Z/NS/SYN character of a step
  - new buttons (D) reset to default colors; (Q) toggle quality, and (T) transparent tetrahedrons
- v2.2:

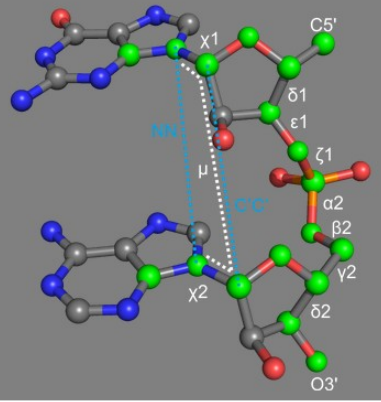
https://dnatco.datmos.org/v2.3

2019: 96 conformers, universal for DNA and RNA

HOME HELP TUTORIAL TABLE OF CONFORMERS BROWSE CONFORMERS ABOUT HOW TO CITE DOWNLOAD

DNATCO v3.2

Assignment of DNA and RNA conformers



Submit own PDB or CIF file SUBMIT

coordinates Browse... No file selected. (optional MAP file) Browse... No file selected.

Enter PDB ID (e.g. 1bna)

RCSB-PDB SUBMIT

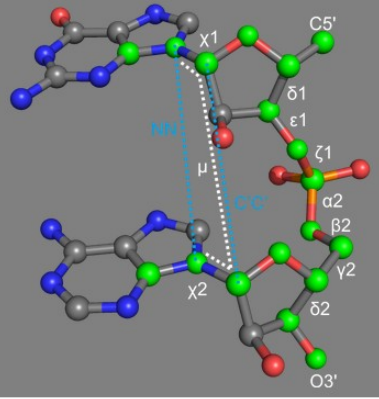
https://dnatco.datmos.org/v3.2

2021: using Mol*

HOME HELP TUTORIAL TABLE OF CONFORMERS BROWSE CONFORMERS ABOUT HOW TO CITE DOWNLOAD

DNATCO[?] v4.1

Assignment of DNA and RNA conformers



Submit own PDB or CIF file[?] SUBMIT

Coordinates Browse... No file selected.

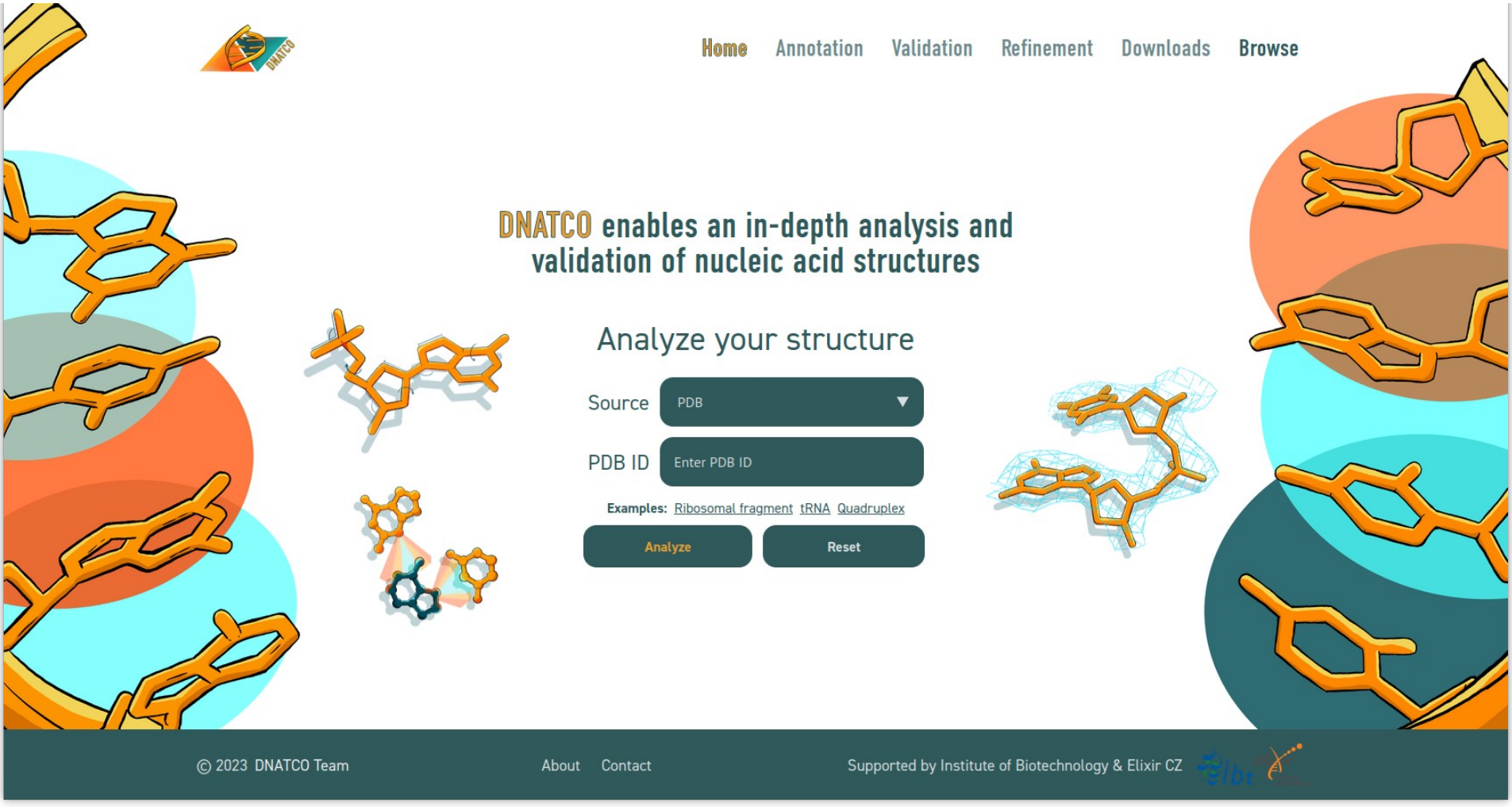
Electron density map (optional) Browse... No file selected.

Enter PDB ID (e.g. 1bna)

RCSB-PDB SUBMIT

https://dnatco.datmos.org/v4.1

2023: client-side web application



https://dnatco.datmos.org/v5.0

Annotation

Home | Annotation | Validation | Refinement | Downloads | Browse

Main features
Assigned NtCs
Structure Info
Downloads
Help

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)

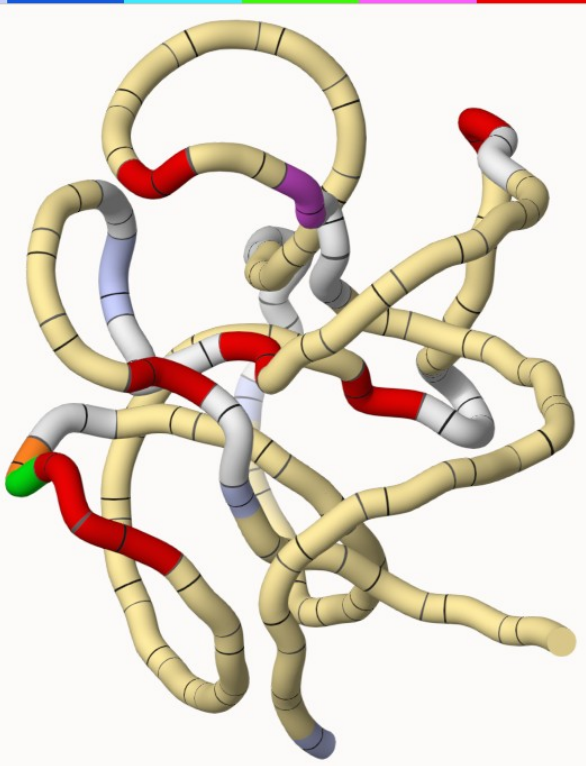
Main features

Counts of NtC

NtC	Count
AA00	40
AA01	4
AB05	3
NANT	14
OP13	2
AA08	8
BB16	1
OP24	1
OP10	1
OP09	1
OP20	1
ZZ1S	1
OP04	1
IC01	1

Counts of CANA

A | BI | BII | miB | Z | IC | OPN | SYN | N



https://dnatco.datmos.org/v5.0

Annotation

Home Annotation Validation Refinement Downloads Browse

Main features
Assigned NtCs
Structure Info
Downloads
Help

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)

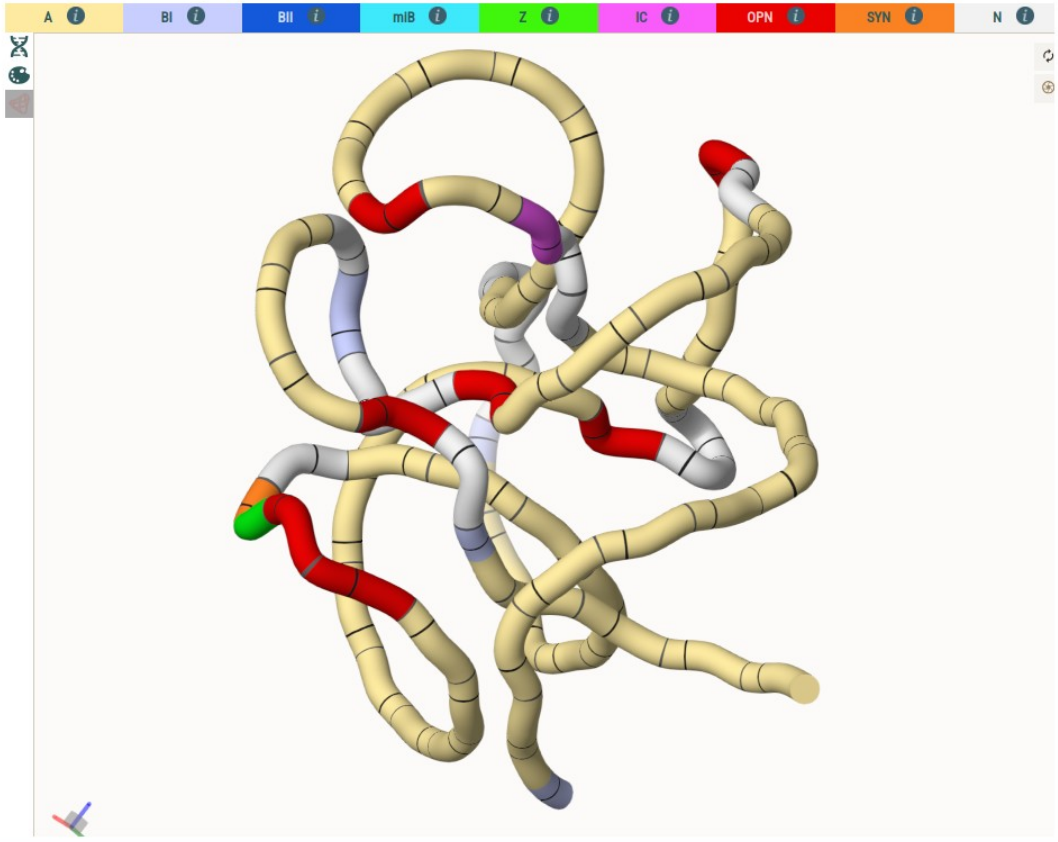
Number of paired bases

Data provided by FR3D

Type of BP	Count
cHS	1
cHW	1
cSH	1
cWH	1
cWW	46
tHH	2
tHS	4
tHW	3
tSH	4
tSW	3
tSS	6
tWH	3
tWS	3

Counts of Nucleotide in polymer entity

Nucleotide	Count
G	32
A	13
U	12
C	23



How to cite Help Contact

Supported by Institute of Biotechnology & Elixir CZ

https://dnatco.datmos.org/v5.0

Annotation

Home Annotation Validation Refinement Downloads Browse

Main features
Assigned NtCs
Structure Info
Downloads
Help

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)

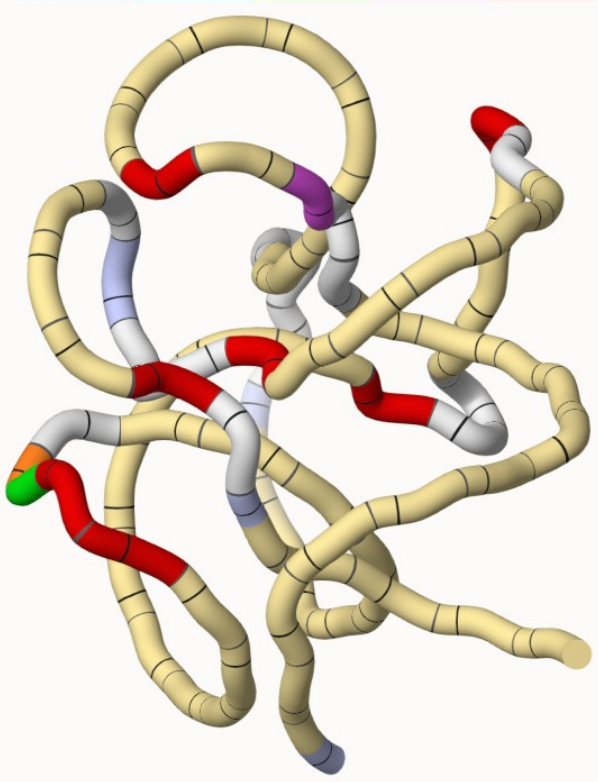
Assigned NtC

Chain All NAs

Table of assigned dinucleotide NtC classes

Chain	Dinucleotide	NtC	CANA
B	G2105 G2106	AA00	AAA
B	G2106 G2107	AA00	AAA
B	G2107 A2108	AA01	AAw
B	A2108 U2109	AA00	AAA
B	U2109 G2110	AB05	A-B
B	G2110 C2111	OP20	OPN
B	C2111 G2112	OP13	OPN
B	G2112 U2113	AA08	AAA
B	U2113 A2114	AA08	AAA
B	A2114 G2115	AA00	AAA
B	G2115 G2116	OP17	OPN
B	G2116 A2117	BB16	BBw
B	A2117 U2118	OP14	OPN
B	U2118 A2119	OP13	OPN
B	A2119 G2120	OP13	OPN
B	G2120 G2121	AA00	AAA
B	G2121 U2122	AA08	AAA
B	U2122 G2123	AA00	AAA
B	G2123 G2124	AA00	AAA
B	G2124 G2125	AA08	AAA

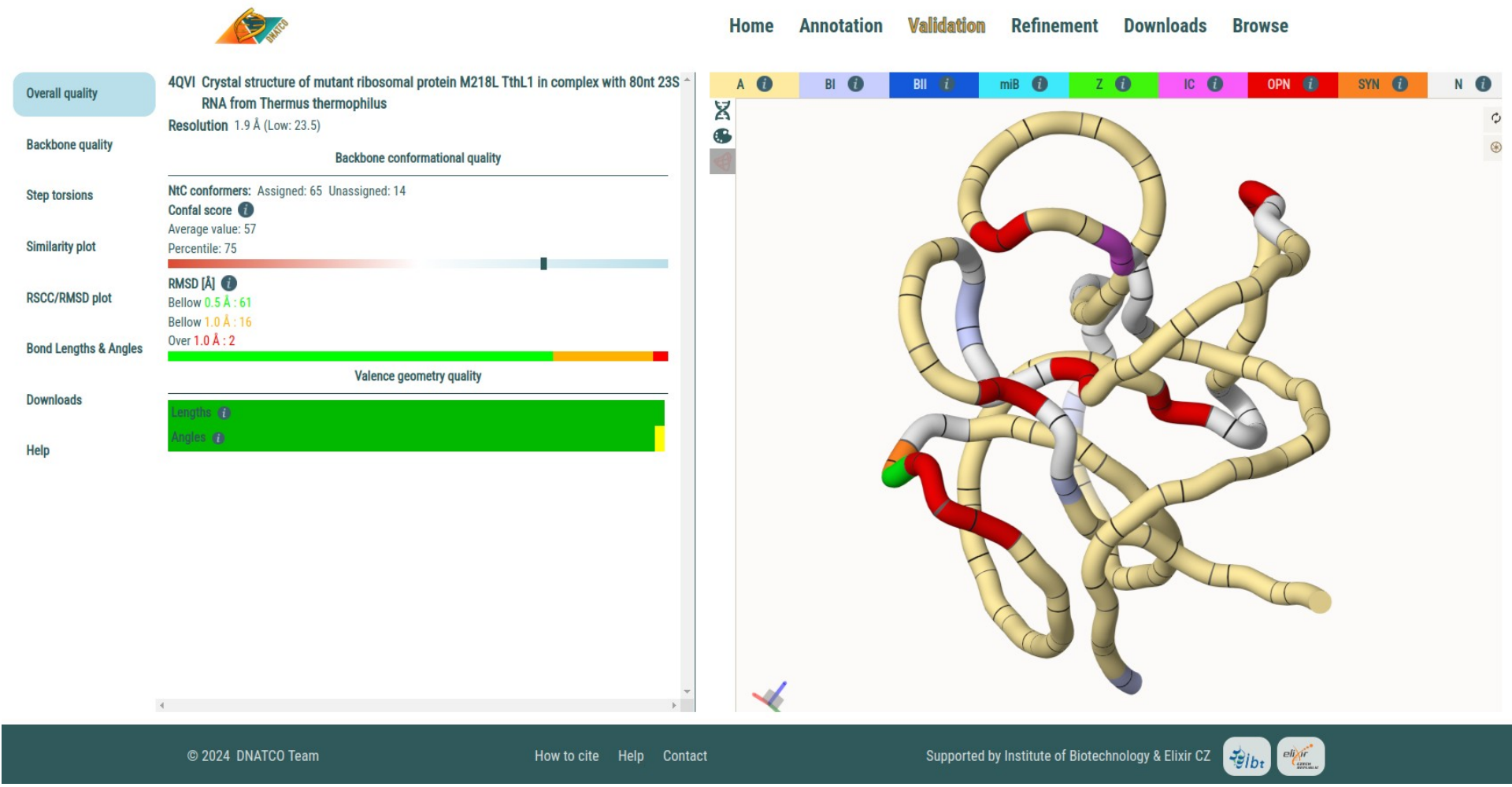
A BI BII mIB Z IC OPN SYN N



How to cite Help Contact
Supported by Institute of Biotechnology & Elixir CZ

https://dnatco.datmos.org/v5.0

Validation



https://dnatco.datmos.org/v5.0

Validation

Home Annotation Validation Refinement Downloads Browse

- Overall quality
- Backbone quality
- Step torsions
- Similarity plot
- RSCC/RMSD plot
- Bond Lengths & Angles
- Downloads
- Help

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

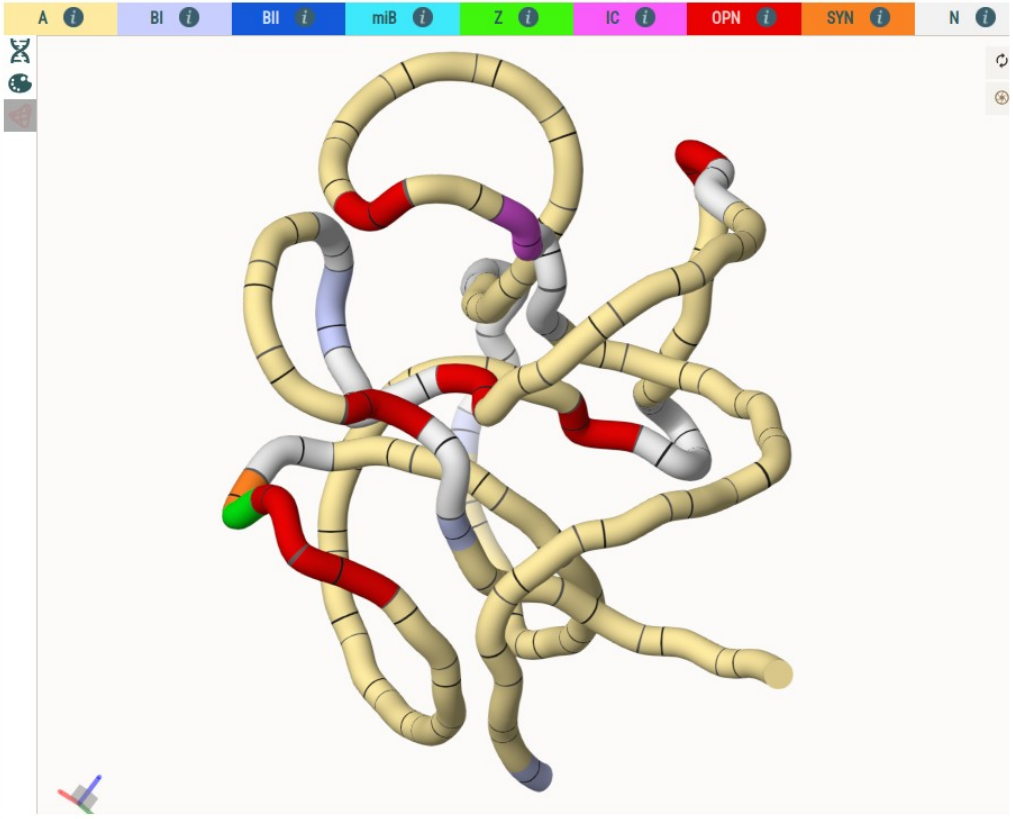
Resolution 1.9 Å (Low: 23.5)

Backbone conformational quality

Chain: All NAs

Table of assigned dinucleotide NtC classes

Chain	Dinucleotide	NtC	CANA	Confal score	RMSD
B	G2105 G2106	AA00	AAA	83	0.164
B	G2106 G2107	AA00	AAA	84	0.190
B	G2107 A2108	AA01	AAw	60	0.240
B	A2108 U2109	AA00	AAA	41	0.445
B	U2109 G2110	AB05	A-B	69	0.256
B	G2110 C2111	OP20	OPN	0	0.615
B	C2111 G2112	OP13	OPN	41	0.607
B	G2112 U2113	AA08	AAA	55	0.314
B	U2113 A2114	AA08	AAA	77	0.217
B	A2114 G2115	AA00	AAA	30	0.450
B	G2115 G2116	OP17	OPN	0	2.277
B	G2116 A2117	BB16	BBw	41	0.412
B	A2117 U2118	OP14	OPN	0	0.572
B	U2118 A2119	OP13	OPN	0	0.905
B	A2119 G2120	OP13	OPN	72	0.564
B	G2120 G2121	AA00	AAA	46	0.281



How to cite Help Contact

Supported by Institute of Biotechnology & Elixir CZ

https://dnatco.datmos.org/v5.0

Validation

Home Annotation Validation Refinement Downloads Browse

- Overall quality
- Backbone quality
- Step torsions
- Similarity plot
- RSCC/RMSD plot
- Bond Lengths & Angles
- Downloads
- Help

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

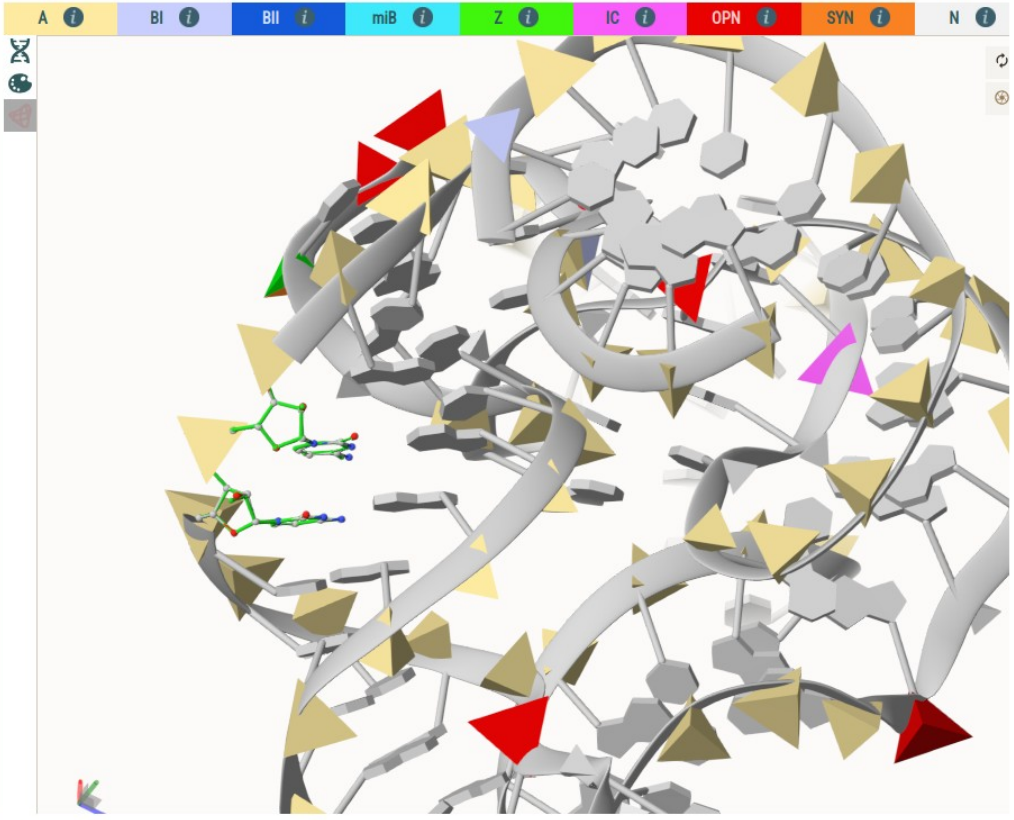
Resolution 1.9 Å (Low: 23.5)

Backbone conformational quality

Chain: All NAs

Table of assigned dinucleotide NtC classes

B	G2125 A2126	AA07	AAu	0	0.600
B	A2126 G2127	OP24	OPN	88	0.204
B	G2127 C2128	AA00	AAA	67	0.248
B	C2128 C2129	AA00	AAA	92	0.164
B	C2129 U2130	AA00	AAA	79	0.206
B	U2130 G2131	BB2S	SYN	0	1.268
B	G2131 U2132	OP21	OPN	0	0.708
B	U2132 G2133	OP20	OPN	0	0.513
B	G2133 A2134	OP10	OPN	97	0.118
B	A2134 A2135	AA00	AAA	78	0.253
B	A2135 C2136	AA08	AAA	47	0.477
B	C2136 C2137	AA00	AAA	70	0.127
B	C2137 C2138	AA00	AAA	82	0.149
B	C2138 C2139	AA00	AAA	87	0.177
B	C2139 C2140	AA00	AAA	77	0.094
B	C2140 G2141	AA00	AAA	83	0.148
B	G2141 C2142	AA08	AAA	82	0.167
B	C2142 C2143	AA00	AAA	90	0.158

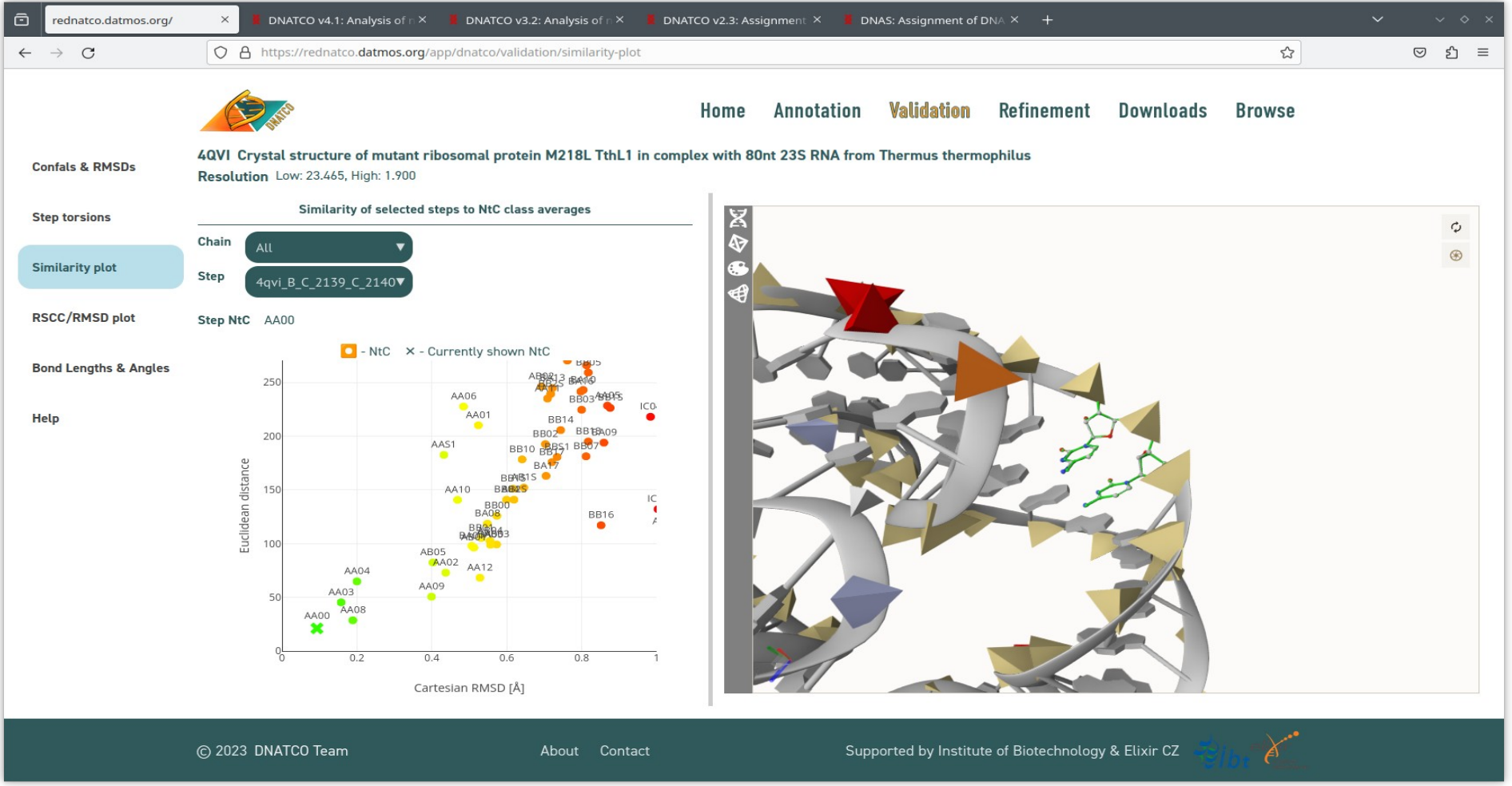


How to cite Help Contact

Supported by Institute of Biotechnology & Elixir CZ

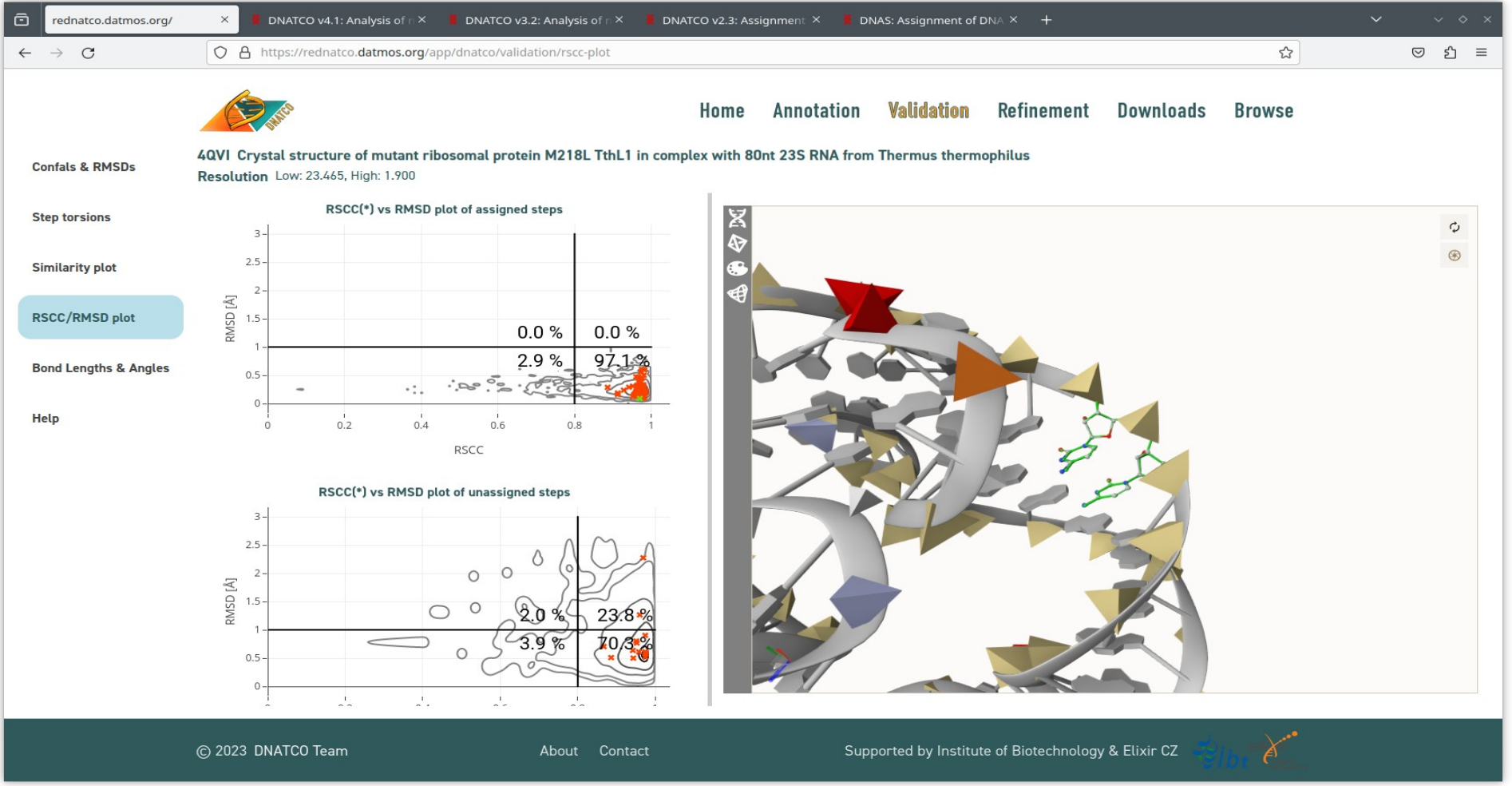
https://dnatco.datmos.org/v5.0

Validation



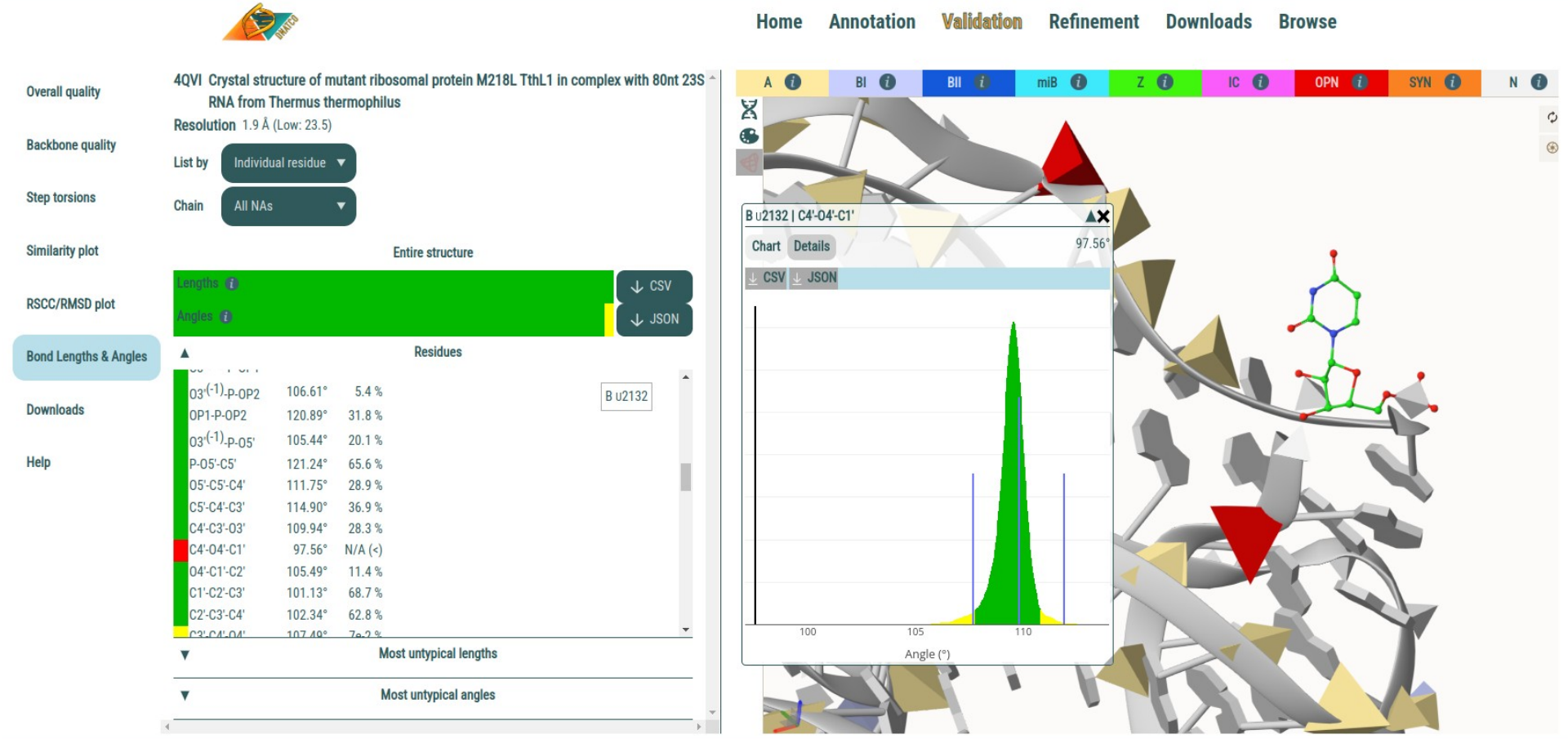
https://dnatco.datmos.org/v5.0

Validation



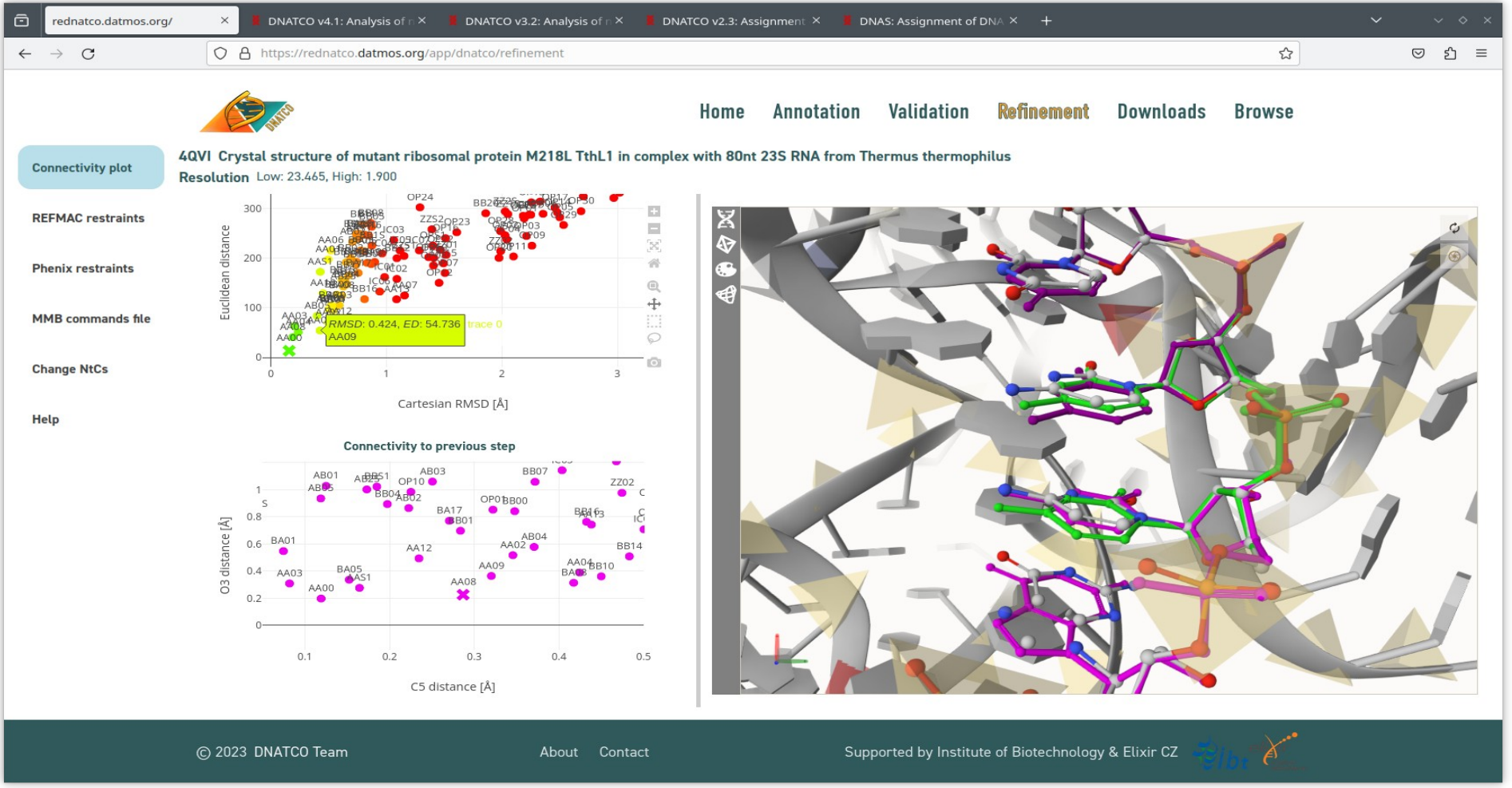
https://dnatco.datmos.org/v5.0

Validation



https://dnatco.datmos.org/v5.0

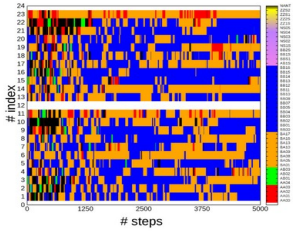
Refinement and Modeling with *NtCs*

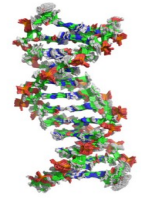


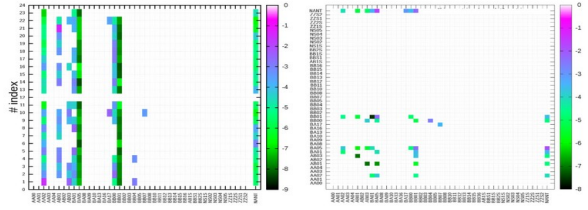
https://dnatco.datmos.org/v5.0

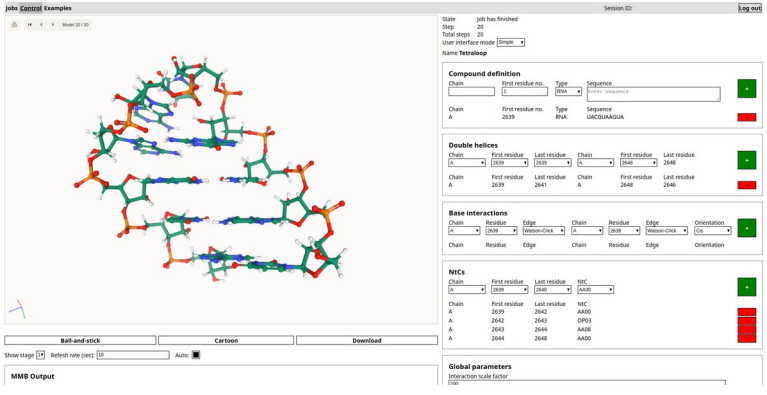
Refinement and Modeling with *NtCs*

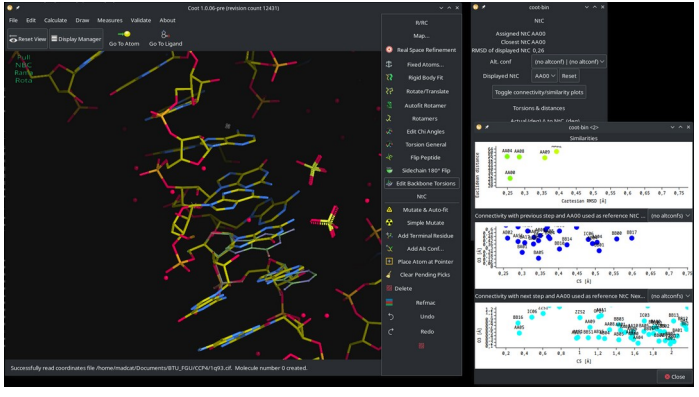
- Restraints for xray and cryoEM refinement
  - Phenix and REFMAC
- Enhanced sampling MD in gromacs
  - PMF-based enhanced sampling MD
  - tens of ns superior to µs classical MD
- Modeling and density fitting
  - MMB (MacroMoleculeBuilder)
  - Coot











RNA in 3D: more than canonical pairs and “unpaired loops”

Jiří Černý
jiri.cerny@ibt.cas.cz

Laboratory of Structural Bioinformatics of Proteins
Institute of Biotechnology, Czech Academy of Sciences

DNATCO Team:
Paulína Božíková
Michal Malý
Jakub Svoboda
Terezie Prchalová
Bohdan Schneider