

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

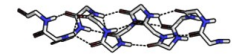
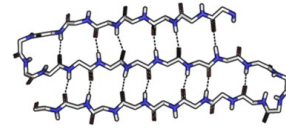
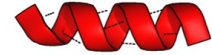
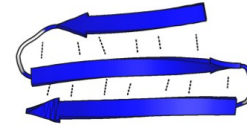


DNATCO Team:
Paulína Božíková
Michal Malý
Jakub Svoboda
Terezie Prchalová
Bohdan Schneider



“Secondary structure”

- Proteins
 - Pattern of backbone hydrogen bonds

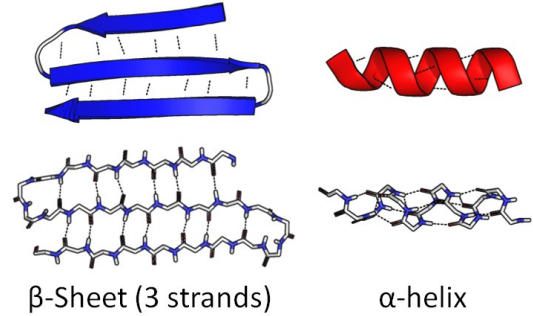


β -Sheet (3 strands)

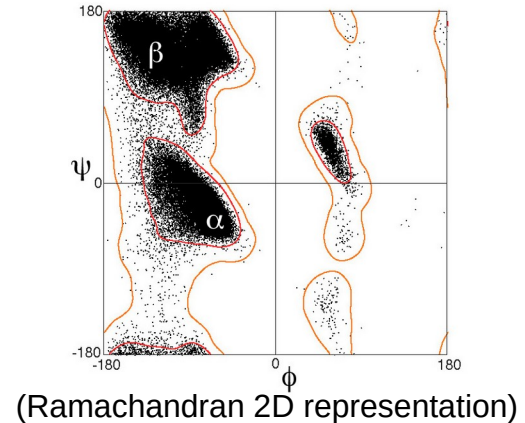
α -helix

“Secondary structure”

- Proteins
 - Pattern of backbone hydrogen bonds

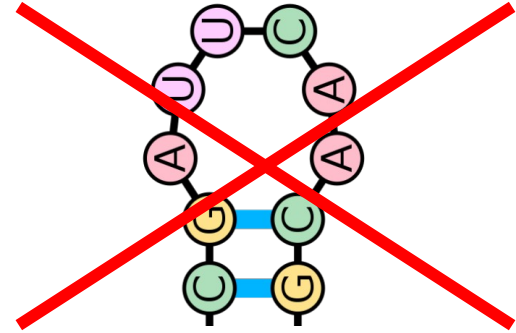


- Equivalent to ϕ/ψ backbone torsions



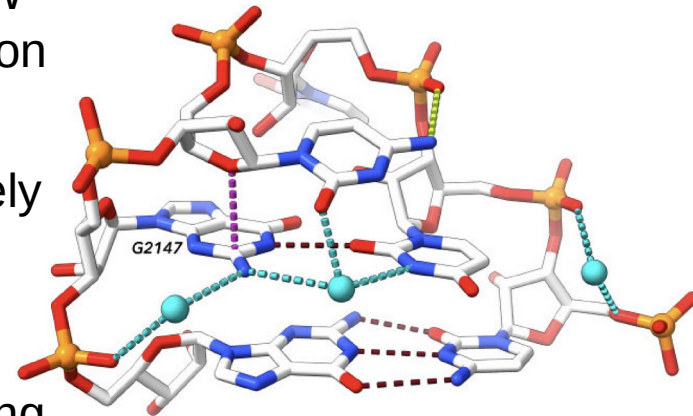
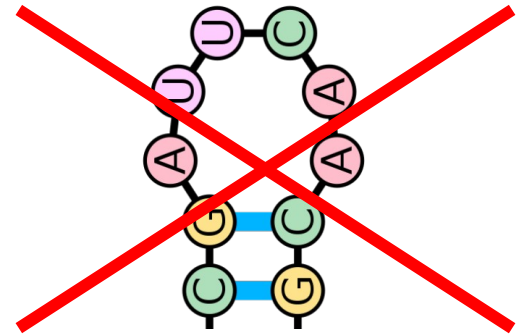
“Secondary structure”

- 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



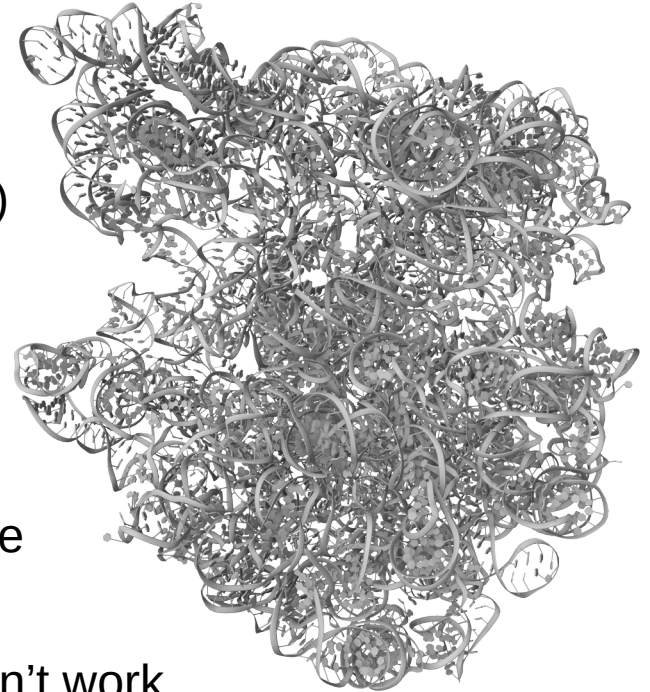
“Secondary structure”

- 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



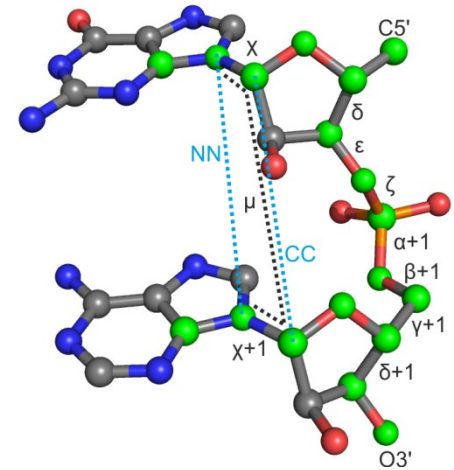
3D structures

- 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

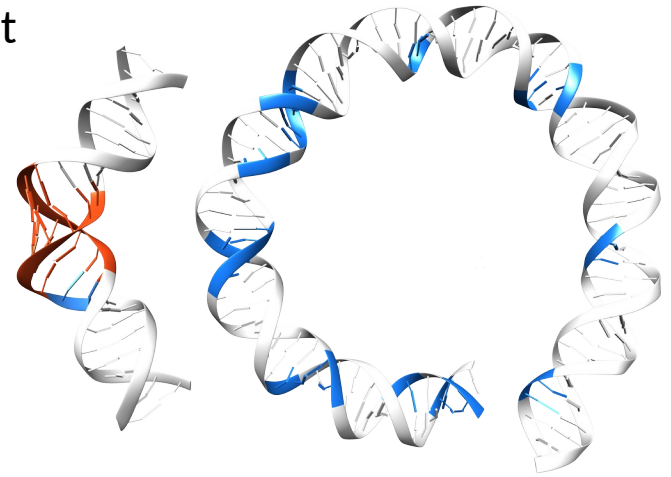
- Identification of important nucleic acid structure features at different levels of detail
 - Automated assignment (no expert knowledge needed)
 - Parameter free (based on/derived from data only)
- From atomic coordinates
 - Backbone conformations & Structural motifs
 - How to assign/annotate **local 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

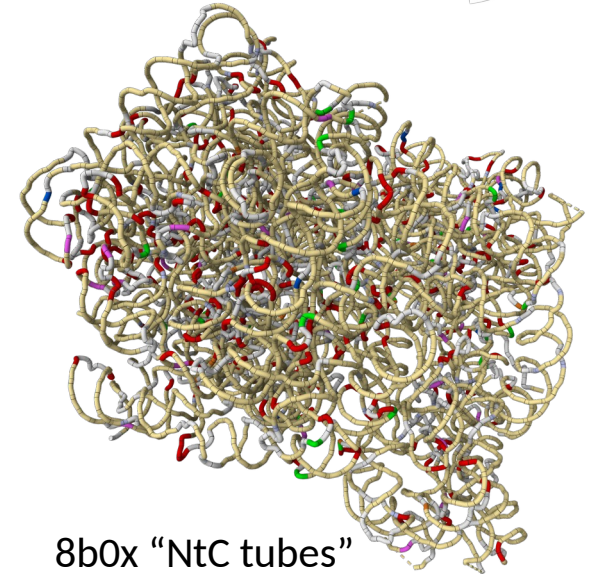
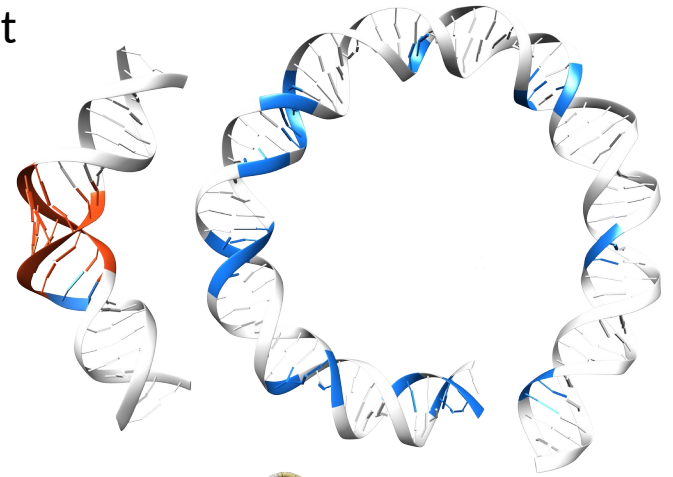
- Identification of important nucleic acid structure features at different levels of detail
 - Automated assignment (no expert knowledge needed)
 - Parameter free (based on/derived from data only)
- From atomic coordinates
 - Backbone conformations & Structural motifs
 - How to assign/annotate *local 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)

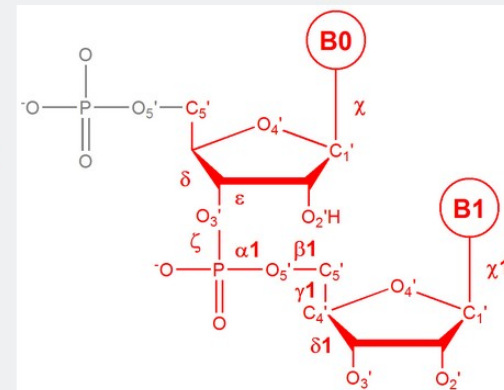
Assignment of DNA conformers

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 ... $\delta+1$, χ , and $\chi+1$ 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 $\delta+1$ and χ and $\chi+1$.



Browse for the PDB file

No file selected.

Enter PDB ID

or

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	$\alpha+1$	$\beta+1$	$\gamma+1$	$\delta+1$	χ	$\chi+1$
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, $\alpha+1/\gamma+1$ crank	84	194	290	149	192	182	88	205	188
41	A-to-B, $\delta>C3'$, $\delta+1 C2'$ -endo	90	196	280	299	179	55	142	222	256
32	BI-to-A, $\delta+1 O4'$ -endo	129	186	264	295	170	52	98	247	233
109	BII-to-A, $\delta+1>C3'$ -endo	142	213	181	297	139	52	90	273	207
110	as 109 plus $\alpha+1/\gamma+1$ crank, high $\beta+1$	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

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" labels steps with Not Stacked 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 $\delta \dots \delta+1$, χ , and $\chi+1$ 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 $\delta+1$ and χ and $\chi+1$.

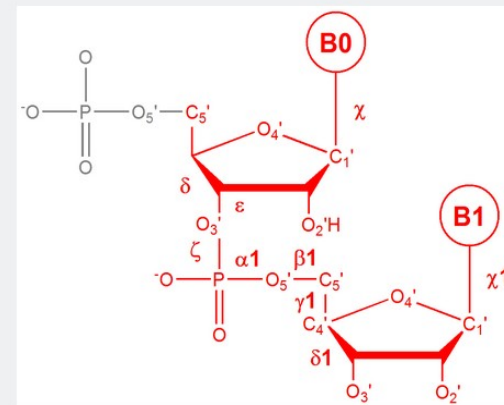


Browse for the PDB file

No file selected.

or

Enter PDB ID



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:
 - new conformer identification protocol was introduced. The confal value (Complement quality) was changed from 1 to 100 (lower is better fit to the reference)

2019: 96 conformers, universal for DNA and RNA



HOME

HELP

TUTORIAL

TABLE OF CONFORMERS

BROWSE CONFORMERS

ABOUT

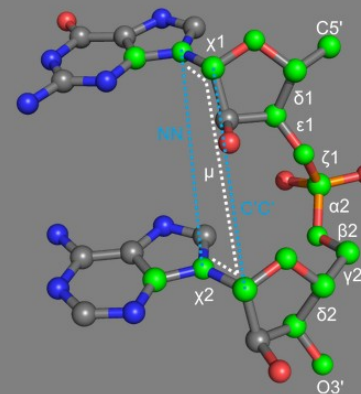
HOW TO CITE

DOWNLOAD



DNATCO v3.2

Assignment of DNA and RNA conformers



Submit own PDB or CIF file

coordinates No file selected.

(optional MAP file) No file selected.

Enter PDB ID (e.g. 1bna)

2021: using Mol*



HOME

HELP

TUTORIAL

TABLE OF CONFORMERS

BROWSE CONFORMERS

ABOUT

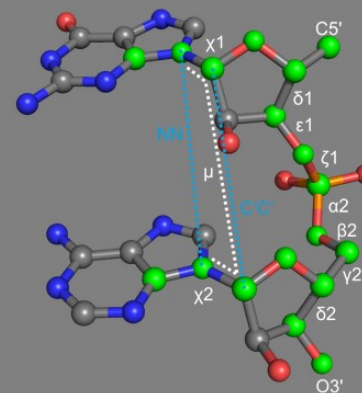
HOW TO CITE

DOWNLOAD



DNATCO^[?] v4.1

Assignment of DNA and RNA conformers



Submit own PDB or CIF file[¶]

SUBMIT

Coordinates No file selected.

Electron density map (optional) No file selected.

Enter PDB ID (e.g. 1bna)

RCSB-PDB ▾

SUBMIT

2023: client-side web application



[Home](#) [Annotation](#) [Validation](#) [Refinement](#) [Downloads](#) [Browse](#)

DNATCO enables an in-depth analysis and validation of nucleic acid structures

Analyze your structure

Source

PDB

PDB ID

Enter PDB ID

Examples: [Ribosomal fragment](#) [tRNA](#) [Quadruplex](#)

Analyze

Reset



Annotation



Home **Annotation** Validation Refinement Downloads Browse

Main features

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)

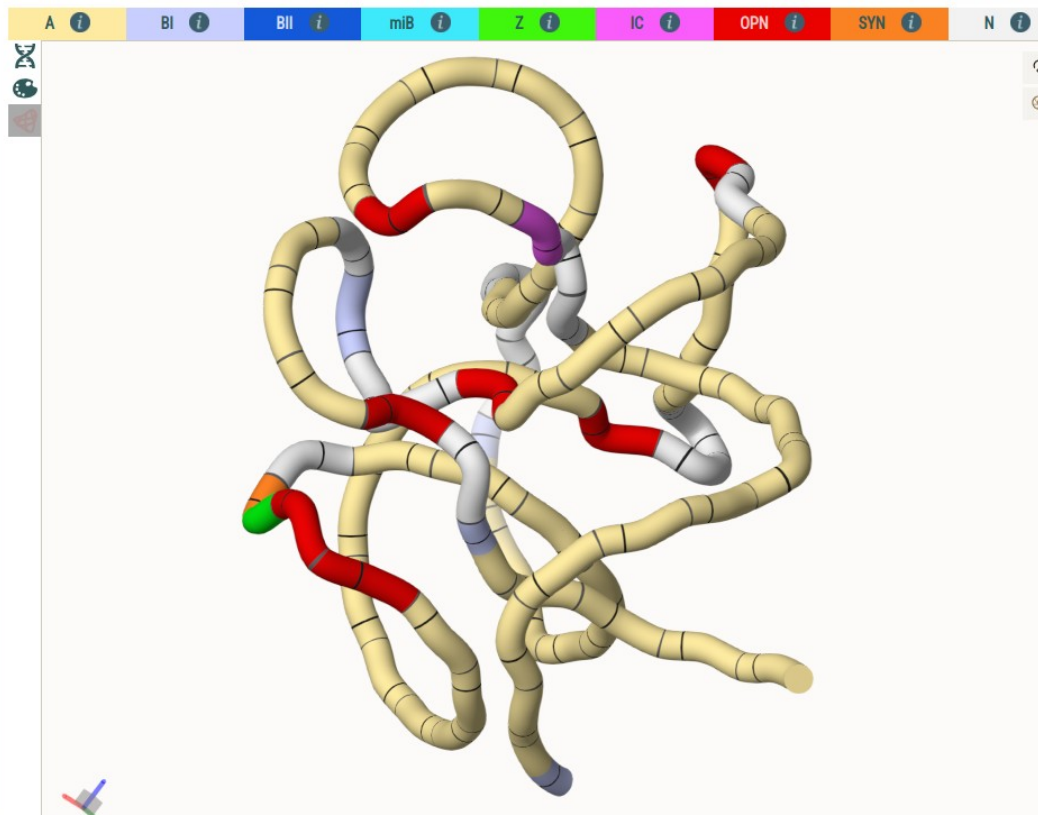
Assigned NTCs

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 ⓘ



Annotation



Home **Annotation** Validation Refinement Downloads Browse

Main features

4QV1 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 NTCs

Structure Info

Downloads

Help

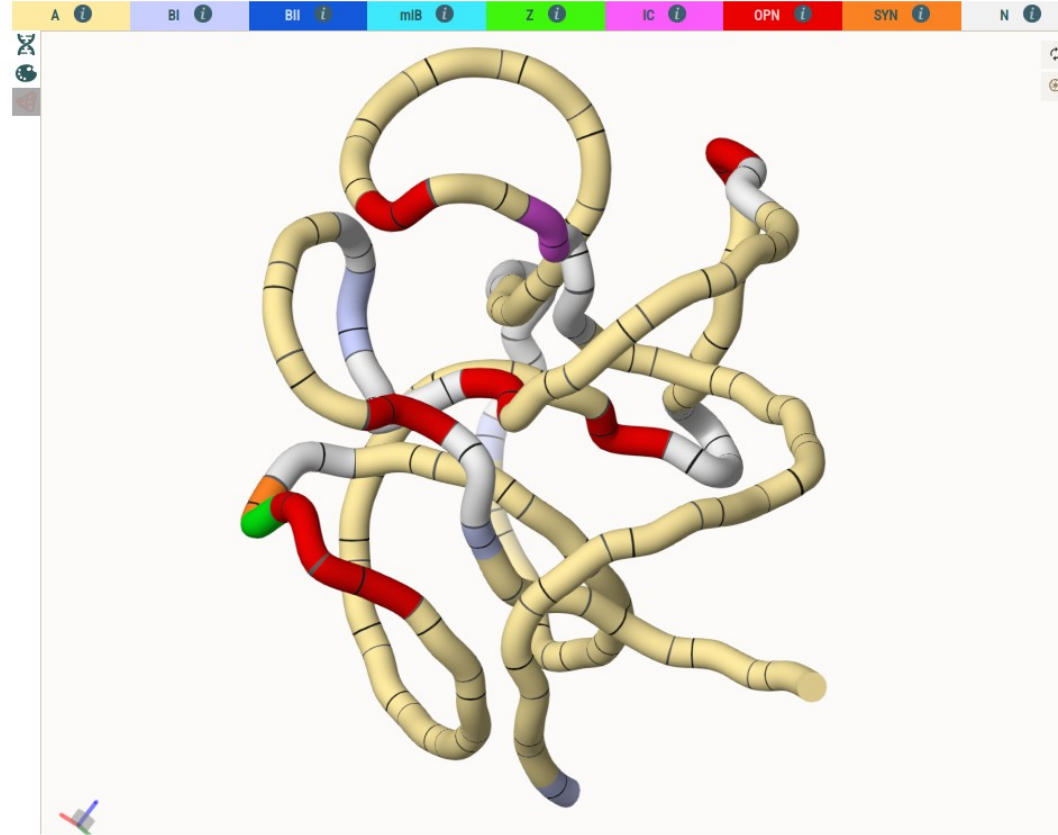
Number of paired bases

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



Annotation



Home **Annotation** Validation Refinement Downloads Browse

Main features

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 NTCs

Assigned NTC

Structure Info

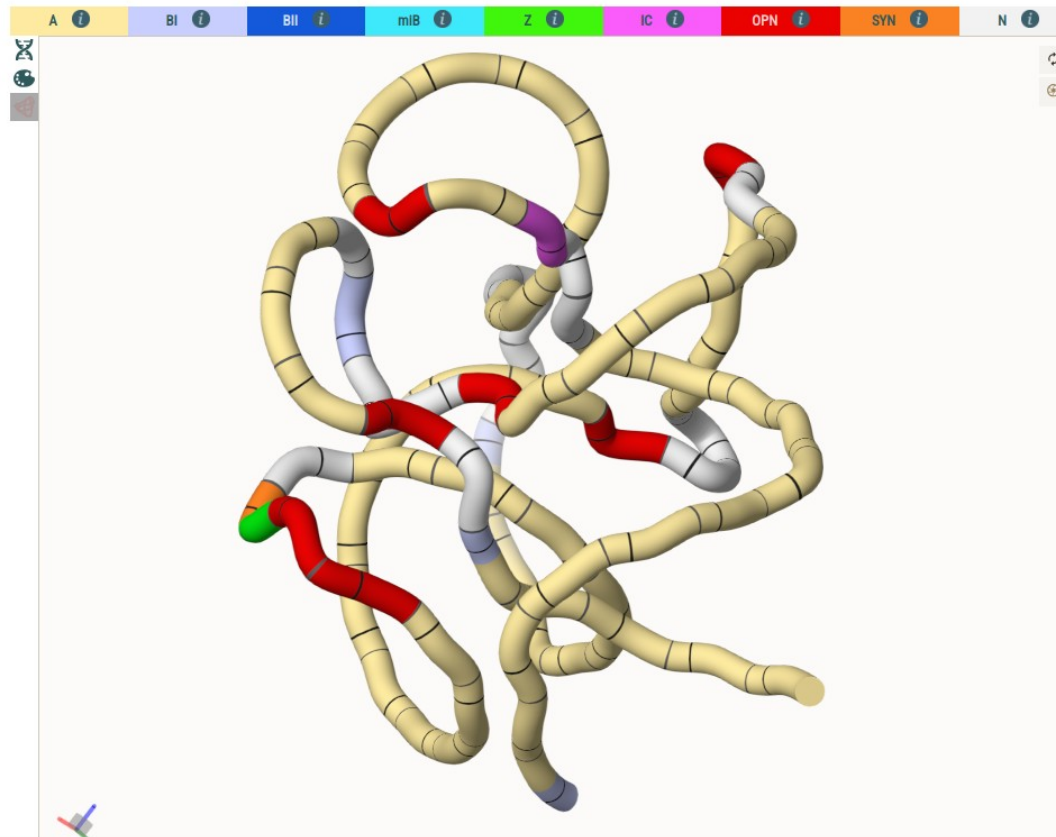
Chain **All NAs**

Downloads

Help

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



Validation



Overall quality

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 quality

Backbone conformational quality

Step torsions

NTC conformers: Assigned: 65 Unassigned: 14

Confal score ⓘ

Average value: 57

Similarity plot

Percentile: 75

RSCC/RMSD plot

RMSD [Å] ⓘ

Bellow 0.5 Å : 61

Bellow 1.0 Å : 16

Over 1.0 Å : 2

Bond Lengths & Angles

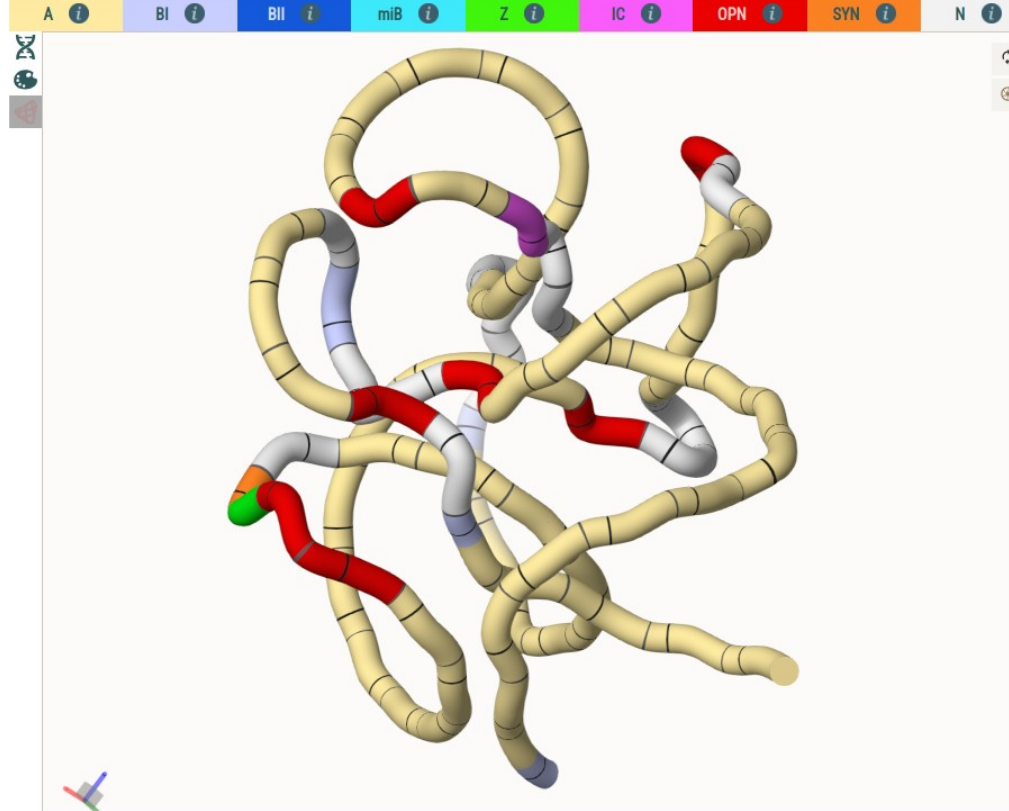
Valence geometry quality

Downloads

Lengths ⓘ

Angles ⓘ

Help



Validation



Home Annotation **Validation** Refinement Downloads Browse

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

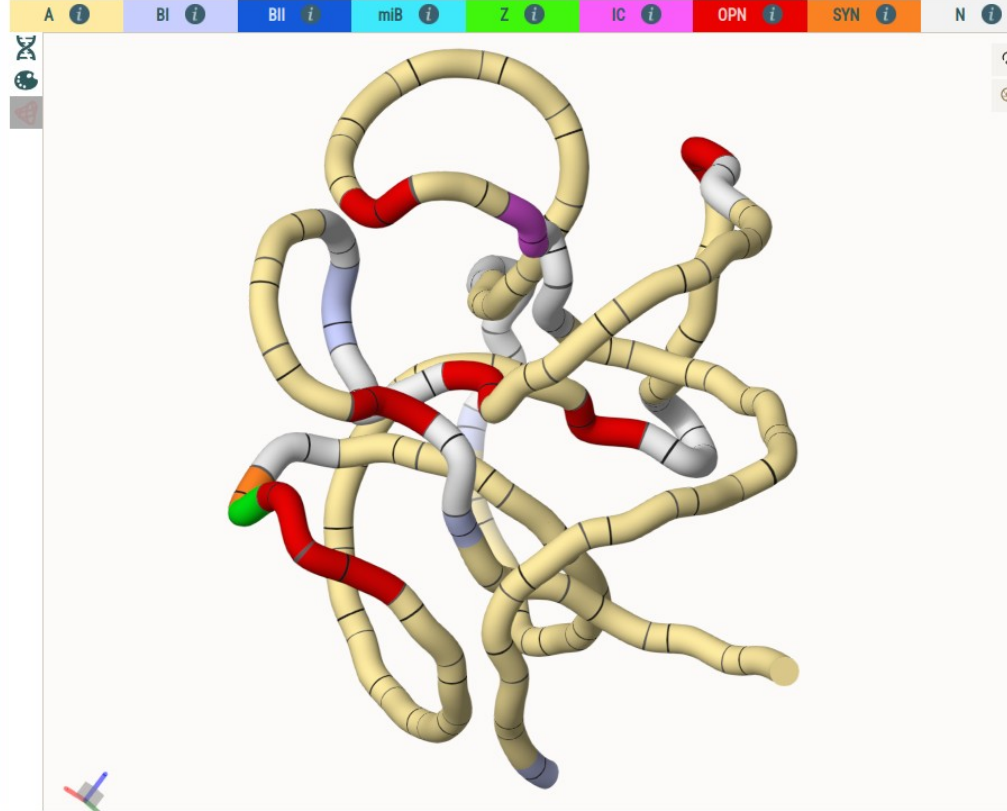
Resolution 1.9 Å (Low: 23.5)

Backbone conformational quality

Chain All NAs

Table of assigned dinucleotide Ntc classes

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



Validation



Home Annotation **Validation** Refinement Downloads Browse

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

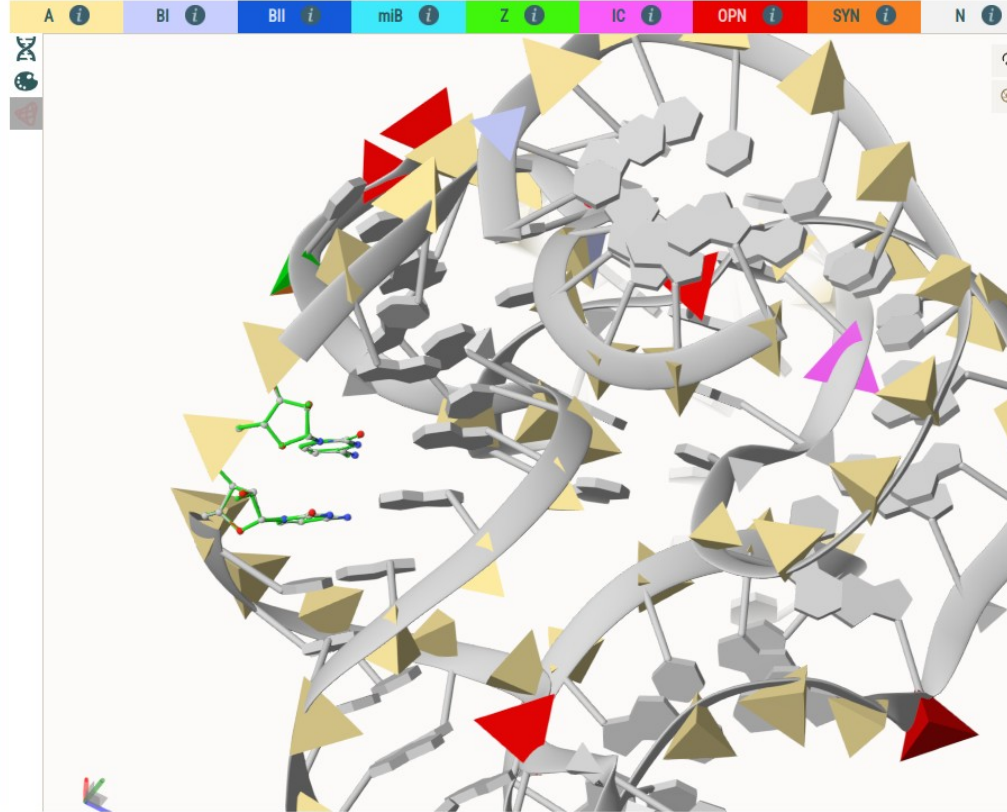
Resolution 1.9 Å (Low: 23.5)

Backbone conformational quality

Chain All NAs

Table of assigned dinucleotide NTC classes

B	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
R	C2142	C2143	AA00	AAA	90	0.158



Validation

rednatco.datmos.org/ | DNATCO v4.1: Analysis of | DNATCO v3.2: Analysis of | DNATCO v2.3: Assignment | DNAS: Assignment of DNA

https://rednatco.datmos.org/app/dnatco/validation/similarity-plot

DNATCO 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 Low: 23.465, High: 1.900

Confals & RMSDs

Step torsions

Similarity plot

RSQC/RMSD plot

Bond Lengths & Angles

Help

Chain: All

Step: 4qvi_B_C_2139_C_2140

Step NtC: AA00

Similarity of selected steps to NtC class averages

Legend: ■ - NtC × - Currently shown NtC

Euclidean distance

Cartesian RMSD [Å]

Validation

rednatco.datmos.org/ | DNATCO v4.1: Analysis of | DNATCO v3.2: Analysis of | DNATCO v2.3: Assignment | DNAS: Assignment of DNA

https://rednatco.datmos.org/app/dnatco/validation/rssc-plot

DNATCO 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 Low: 23.465, High: 1.900

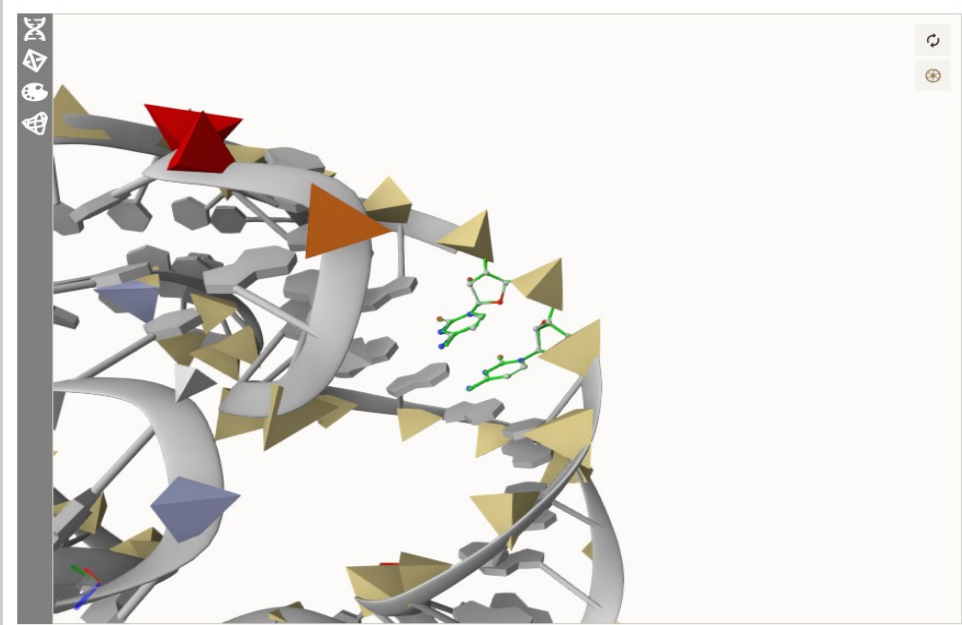
Confals & RMSDs
Step torsions
Similarity plot
RSSC/RMSD plot
Bond Lengths & Angles
Help

RSSC(*) vs RMSD plot of assigned steps

RMSD [Å]	RSSC < 0.8	RSSC > 0.8
0.0 - 1.0	2.9 %	97.1 %
1.0 - 3.0	0.0 %	0.0 %

RSSC(*) vs RMSD plot of unassigned steps

RMSD [Å]	RSSC < 0.8	RSSC > 0.8
0.0 - 1.0	3.9 %	70.3 %
1.0 - 3.0	2.0 %	23.8 %



Validation



Home Annotation **Validation** Refinement Downloads Browse

Overall quality

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)

List by Individual residue

Chain All NAs

Similarity plot Entire structure

RSCC/RMSD plot

Bond Lengths & Angles

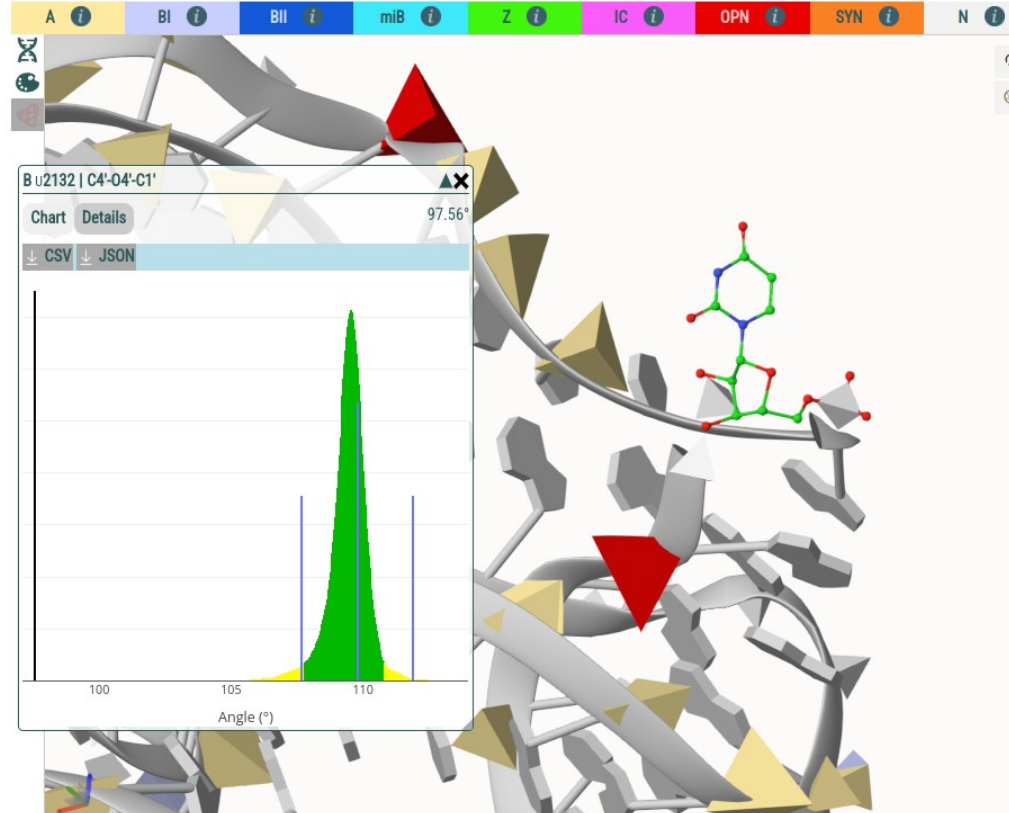
Lengths CSV

Angles JSON

Residues	Angle (°)	%
O3 ⁽⁻¹⁾ -P-OP2	106.61°	5.4 %
OP1-P-OP2	120.89°	31.8 %
O3 ⁽⁻¹⁾ -P-O5'	105.44°	20.1 %
P-O5'-C5'	121.24°	65.6 %
O5'-C5'-C4'	111.75°	28.9 %
C5'-C4'-C3'	114.90°	36.9 %
C4'-C3'-O3'	109.94°	28.3 %
C4'-O4'-C1'	97.56°	N/A (<)
O4'-C1'-C2'	105.49°	11.4 %
C1'-C2'-C3'	101.13°	68.7 %
C2'-C3'-C4'	102.34°	62.8 %
C3'-C4'-O4'	107.40°	76.2 %

Most untypical lengths

Most untypical angles



Refinement and Modeling with NtCs

rednatco.datmos.org/ | DNATCO v4.1: Analysis of | DNATCO v3.2: Analysis of | DNATCO v2.3: Assignment | DNAS: Assignment of DNA

https://rednatco.datmos.org/app/dnatco/refinement

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 Low: 23.465, High: 1.900

Connectivity plot

REFMAC restraints

Phenix restraints

MMB commands file

Change NtCs

Help

Euclidean distance

Cartesian RMSD [Å]

RMSD: 0.424, ED: 54.736, trace 0

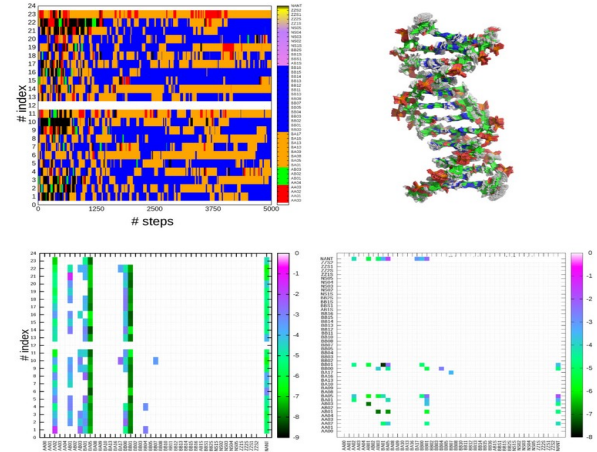
Connectivity to previous step

O3 distance [Å]

C5 distance [Å]

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 μ s classical MD
- Modeling and density fitting
 - MMB (MacroMoleculeBuilder)
 - Coot



The screenshot shows the MacroMoleculeBuilder (MMB) software interface. The main window displays a 3D ball-and-stick model of a protein structure. The interface includes several panels:

- Job Controls:** Shows job status (Finished), steps (20), and user interface mode (Simple).
- Compound definitions:** A table with columns for Chain, First residue no., Type, and Sequence. It lists a single entry for Chain A, residues 2039-2046, of type RNA with sequence URGCUAGUAAU.
- Double helices:** A table with columns for Chain, First residue, Last residue, Chain, First residue, Last residue. It lists a single entry for Chain A, residues 2039-2041 and 2046-2046.
- Base Intersections:** A table with columns for Chain, Residue, Edge, Chain, Residue, Edge, Orientation. It lists a single entry for Chain A, residues 2039 and 2041, with edge CA and orientation Cis.
- NtCs:** A table with columns for Chain, First residue, Last residue, NtC. It lists a single entry for Chain A, residues 2039-2046, with NtC values ranging from 2042 to 2046.
- Global parameters:** Shows interaction scale factor (1.0).

The screenshot shows the Coot software interface. The main window displays a 3D ball-and-stick model of a protein structure. The interface includes several panels:

- File, Edit, Calculate, Draw, Measures, Validate, About:** Standard menu options.
- Real Space Refinement:** A panel with options for Fixed Atoms, Add/Remove, Displayed NtC, Rigid Body FE, Rotate/Translate, As-built Residue, Reference, Edit On Angles, Torsion General, Flip Peptide, Selection 180° Flip, and Edit Backbone Torsion.
- Model & AutoFit:** A panel with options for Simple Model, Add Terminal Residue, Add All Conf., Tweak Atomic Number, and Clear Pending Picks.
- Connectivity with 2D maps:** A panel showing a scatter plot of connectivity values for residues 1-10, with a color scale from 0 (blue) to 10 (red).
- Connectivity with real data and AAD used in reference NtC:** A panel showing a scatter plot of connectivity values for residues 1-10, with a color scale from 0 (blue) to 10 (red).

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

