Full wwPDB X-ray Structure Validation Report ⓘ

Oct 9, 2023 – 10:29 PM EDT

PDB ID	:	7MZ0
Title	:	Structure of human DNA polymerase beta complexed with dzA as the template base in a 1-nucleotide gapped DNA
Authors	:	Koag, M.-C.; Lee, S.
Deposited on	:	2021-05-23
Resolution	:	2.02 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

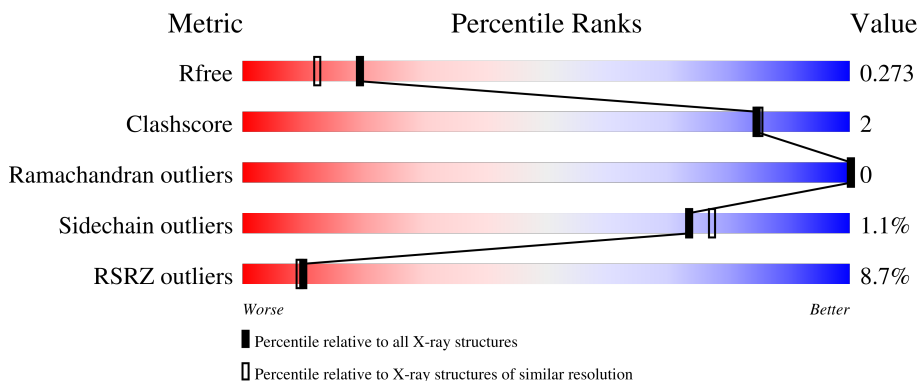|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
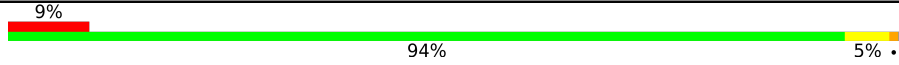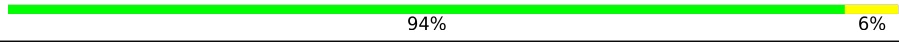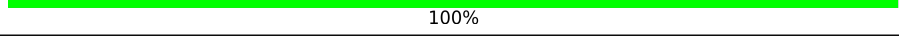*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.02 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.273
Clashscore		2
Ramachandran outliers		0
Sidechain outliers		1.1%
RSRZ outliers		8.7%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	10434 (2.04-2.00)
Clashscore	141614	11643 (2.04-2.00)
Ramachandran outliers	138981	11493 (2.04-2.00)
Sidechain outliers	138945	11492 (2.04-2.00)
RSRZ outliers	127900	10220 (2.04-2.00)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	329	9% / 94% / 5% ··
2	T	16	94% / 6%
3	P	10	100%
4	D	5	80% / 20%

WORLDWIDE wwPDB PROTEIN DATA BANK

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 3395 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called DNA polymerase beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	327	Total	C	N	O	S	0	0	0
			2593	1642	454	488	9			

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*CP*GP*AP*CP*(4DU)P*TP*CP*GP* CP*AP*TP*CP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	16	Total	C	N	O	P	0	0	0
			320	154	59	92	15			

- Molecule 3 is a DNA chain called DNA (5'-D(*GP*CP*TP*GP*AP*TP*GP*CP*GP*A)-3 ').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	P	10	Total	C	N	O	P	0	0	0
			205	98	40	58	9			

- Molecule 4 is a DNA chain called DNA (5'-D(P*GP*TP*CP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	5	Total	C	N	O	P	0	0	0
			106	49	20	32	5			

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Na	0	0
			2	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	118	Total 118	O 118	0	0
6	T	22	Total 22	O 22	0	0
6	P	13	Total 13	O 13	0	0
6	D	16	Total 16	O 16	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

• Molecule 1: DNA polymerase beta

Chain A:

• Molecule 2: DNA (5'-D(*CP*CP*GP*AP*CP*(4DU)P*TP*CP*GP*CP*AP*TP*CP*AP*GP*C)-3')

Chain T:

• Molecule 3: DNA (5'-D(*GP*CP*TP*GP*AP*TP*GP*CP*GP*A)-3')

Chain P:

There are no outlier residues recorded for this chain.

• Molecule 4: DNA (5'-D(P*GP*TP*CP*GP*G)-3')

Chain D:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	54.65Å　　79.17Å　　54.82Å 90.00°　　106.27°　　90.00°	Depositor
Resolution (Å)	19.79　－　2.02 19.79　－　2.02	Depositor EDS
% Data completeness (in resolution range)	95.7 (19.79-2.02) 95.4 (19.79-2.02)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.02Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.221　,　0.276 0.226　,　0.273	Depositor DCC
R_{free} test set	1493 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	22.9	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 37.6	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.027 for l,-k,h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	3395	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.20% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: NA, 4DU

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$ >5	RMSZ	#$|Z|$ >5
1	A	0.26	0/2642	0.41	0/3552
2	T	0.58	0/333	0.81	0/508
3	P	0.53	0/230	0.88	0/354
4	D	1.09	1/118 (0.8%)	0.86	0/179
All	All	0.39	1/3323 (0.0%)	0.54	0/4593

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1	DG	OP3-P	-10.56	1.48	1.61

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2   Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2593	0	2595	14	0
2	T	320	0	181	0	0
3	P	205	0	114	0	0
4	D	106	0	57	0	0
5	A	2	0	0	0	0
6	A	118	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	D	16	0	0	0	0
6	P	13	0	0	0	0
6	T	22	0	0	0	0
All	All	3395	0	2947	14	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

All (14) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:324:GLN:HG3	1:A:324:GLN:O	1.84	0.77
1:A:289:LYS:HD3	1:A:324:GLN:HG2	1.71	0.72
1:A:289:LYS:HE3	1:A:324:GLN:NE2	2.15	0.61
1:A:304:THR:HG23	1:A:305:GLY:N	2.18	0.59
1:A:36:TYR:CZ	1:A:40:ARG:HD2	2.42	0.54
1:A:112:ARG:NH2	6:A:503:HOH:O	2.40	0.53
1:A:304:THR:CG2	1:A:305:GLY:N	2.75	0.50
1:A:36:TYR:CE1	1:A:40:ARG:HD2	2.47	0.49
1:A:302:GLY:HA3	1:A:306:VAL:HG13	1.94	0.49
1:A:306:VAL:O	1:A:306:VAL:HG22	2.13	0.49
1:A:123:GLU:HG3	1:A:126:ARG:NH2	2.28	0.48
1:A:67:THR:O	1:A:71:GLU:HG3	2.18	0.43
1:A:170:ASP:HB3	1:A:173:TYR:CD2	2.54	0.43
1:A:230:LYS:NZ	6:A:501:HOH:O	2.32	0.41

There are no symmetry-related clashes.

5.3　Torsion angles ⓘ

5.3.1　Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	323/329 (98%)	310 (96%)	13 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2   Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	280/290 (97%)	277 (99%)	3 (1%)	73	77

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	289	LYS
1	A	306	VAL
1	A	325	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	31	GLN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

5.4   Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the

expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
2	4DU	T	6	2	19,23,24	2.53	9 (47%)	20,33,36	1.99	2 (10%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	4DU	T	6	2	-	0/3/21/22	0/3/3/3

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	6	4DU	C4-C5	-4.62	1.30	1.40
2	T	6	4DU	P-O5'	-4.56	1.49	1.62
2	T	6	4DU	C3-C4	-4.17	1.32	1.41
2	T	6	4DU	O4'-C1'	-3.92	1.33	1.42
2	T	6	4DU	C6-C5	-3.00	1.32	1.43
2	T	6	4DU	C1'-N9	-2.79	1.41	1.49
2	T	6	4DU	C8-N7	-2.72	1.29	1.34
2	T	6	4DU	C5-N7	-2.42	1.30	1.39
2	T	6	4DU	O3'-C3'	-2.09	1.38	1.43

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	T	6	4DU	C3-C2-N1	-7.32	116.60	123.81
2	T	6	4DU	C5-C4-N9	2.30	108.57	105.72

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å²)	Q<0.9
1	A	327/329 (99%)	0.33	31 (9%)	8	8	12, 24, 59, 70	0
2	T	15/16 (93%)	-0.27	0	100	100	17, 27, 38, 43	0
3	P	10/10 (100%)	-0.23	0	100	100	16, 25, 34, 36	0
4	D	5/5 (100%)	-0.33	0	100	100	17, 18, 28, 34	0
All	All	357/360 (99%)	0.28	31 (8%)	10	9	12, 24, 58, 70	0

All (31) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	304	THR	5.2
1	A	245	ASN	4.7
1	A	306	VAL	4.0
1	A	324	GLN	3.5
1	A	208	PRO	3.5
1	A	288	GLU	3.5
1	A	301	LEU	3.4
1	A	287	LEU	3.3
1	A	302	GLY	3.2
1	A	204	SER	3.1
1	A	335	GLU	3.1
1	A	7	PRO	2.9
1	A	300	PRO	2.9
1	A	284	ALA	2.9
1	A	10	THR	2.9
1	A	319	ILE	2.9
1	A	325	TRP	2.8
1	A	244	LYS	2.8
1	A	285	HIS	2.7
1	A	270	LEU	2.7
1	A	282	MET	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	292	THR	2.6
1	A	248	LYS	2.5
1	A	298	ILE	2.5
1	A	280	LYS	2.5
1	A	307	ALA	2.4
1	A	309	GLU	2.3
1	A	8	GLN	2.2
1	A	323	ILE	2.1
1	A	303	VAL	2.1
1	A	326	LYS	2.1

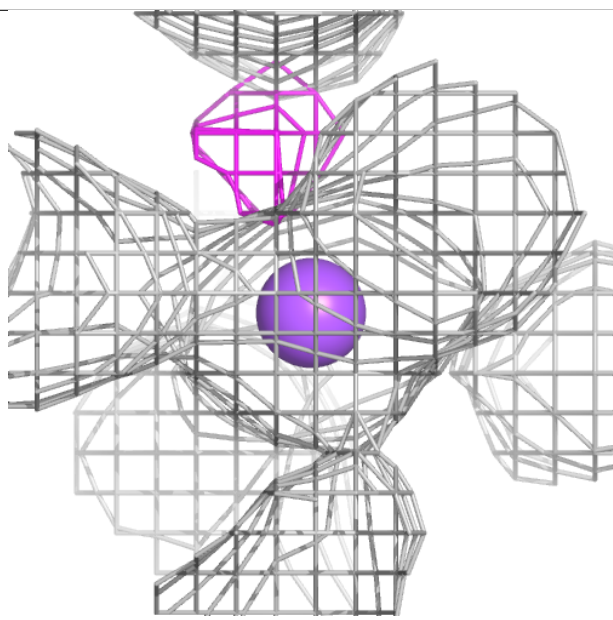## 6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	4DU	T	6	21/22	0.92	0.13	29,40,46,49	0

6.3   Carbohydrates ⓘ

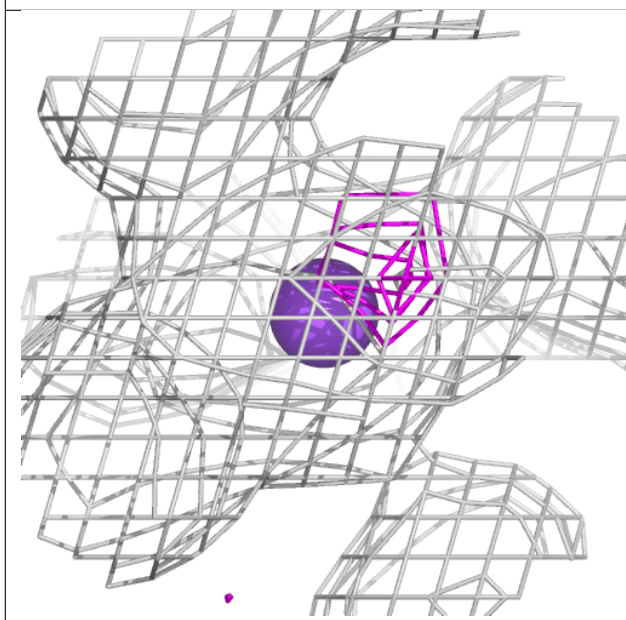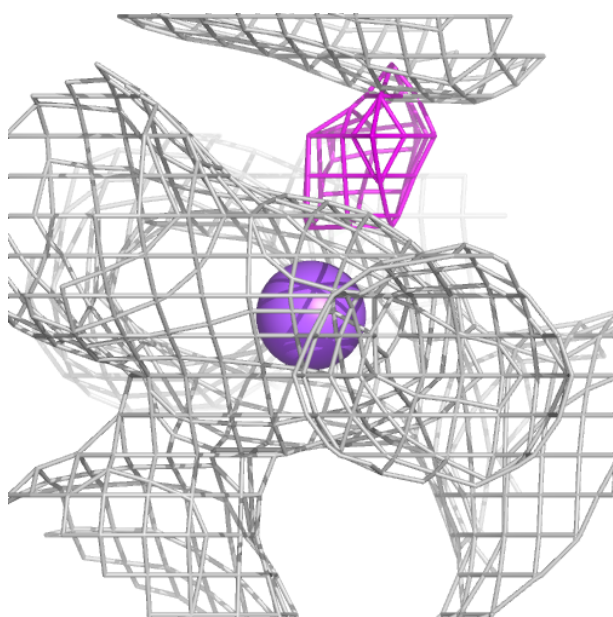There are no monosaccharides in this entry.

6.4   Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.
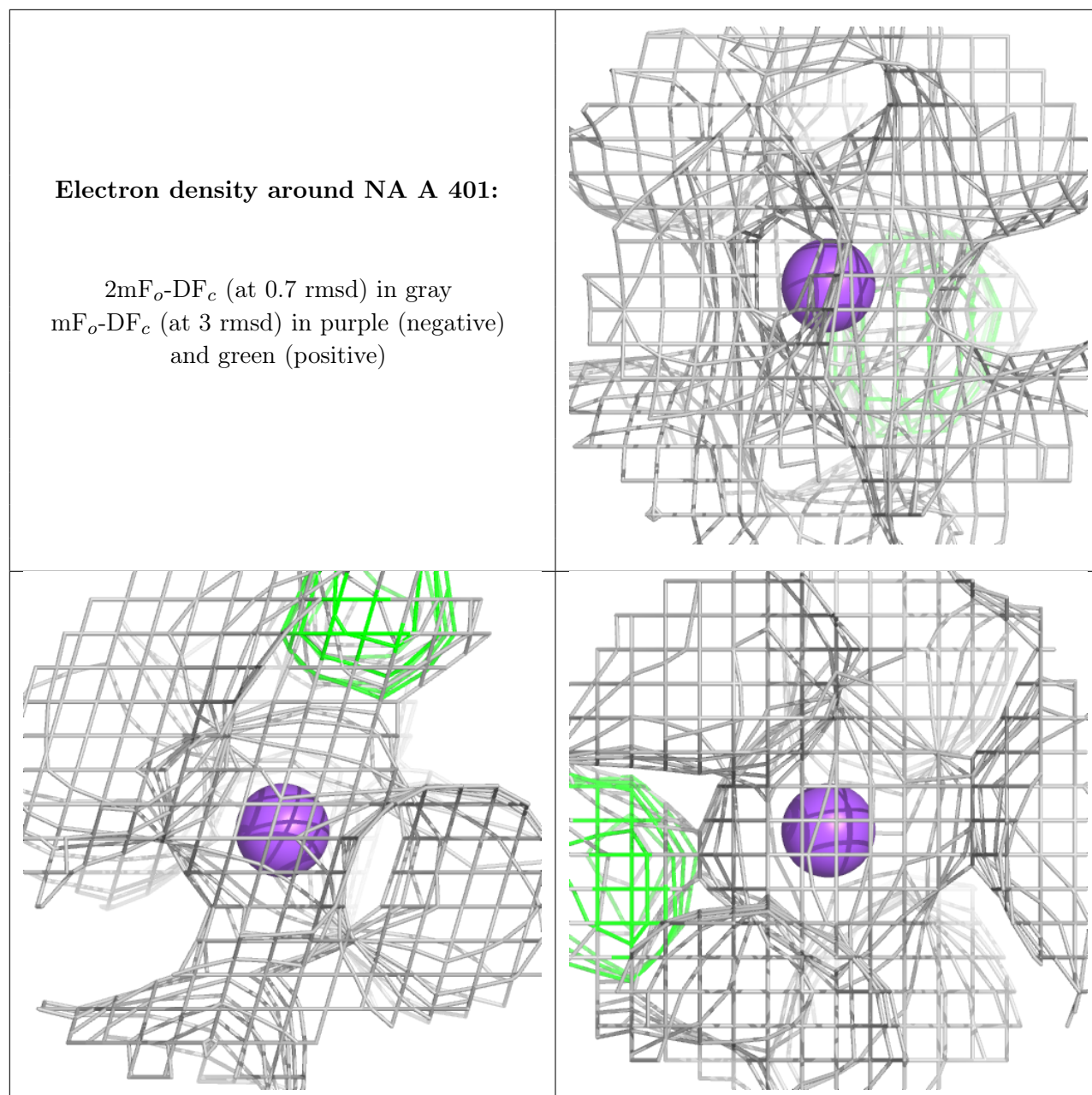
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	NA	A	402	1/1	0.90	0.14	25,25,25,25	0
5	NA	A	401	1/1	0.98	0.06	10,10,10,10	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around NA A 402:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around NA A 401:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)

6.5 Other polymers ⓘ

There are no such residues in this entry.