Full wwPDB X-ray Structure Validation Report (i)

Sep 16, 2024 – 10:06 AM EDT

PDB ID	:	8V35
Title	:	Crystal structure of HpsN from Cupriavidus pinatubonensis
Authors	:	Lee, M.
Deposited on	:	2023-11-27
Resolution	:	1.94 Å(reported)

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

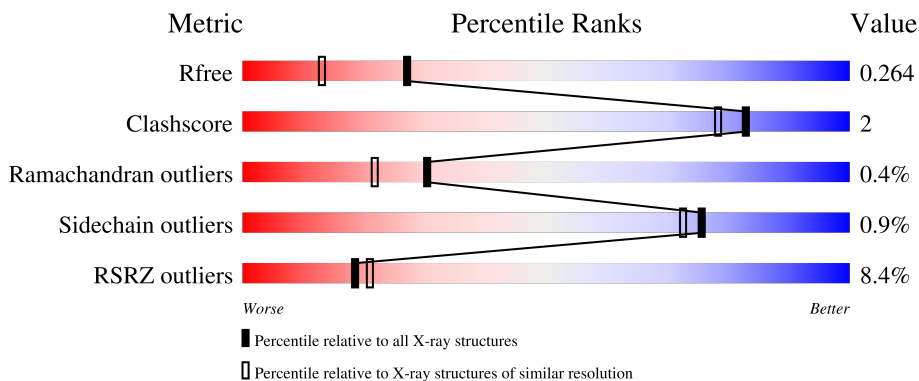|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.3

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 1.94 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		0.264
Clashscore		2
Ramachandran outliers		0.4%
Sidechain outliers		0.9%
RSRZ outliers		8.4%

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}	164625	1306 (1.94-1.94)
Clashscore	180529	1400 (1.94-1.94)
Ramachandran outliers	177936	1387 (1.94-1.94)
Sidechain outliers	177891	1387 (1.94-1.94)
RSRZ outliers	164620	1306 (1.94-1.94)

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	437	6% / 92% / 6% ·
1	B	437	10% / 84% / 7% / 9%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 6514 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 Sulfopropanediol 3-dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	431	Total 3262	C 2058	N 569	O 620	S 15	0	0	0
1	B	399	Total 3019	C 1907	N 522	O 575	S 15	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q46N53
B	0	GLY	-	expression tag	UNP Q46N53

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 4	C 2	O 2	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Zn 1	0	0
3	B	1	Total 1	Zn 1	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	126	Total 126	O 126	0	0
4	B	101	Total 101	O 101	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: Sulfopropanediol 3-dehydrogenase
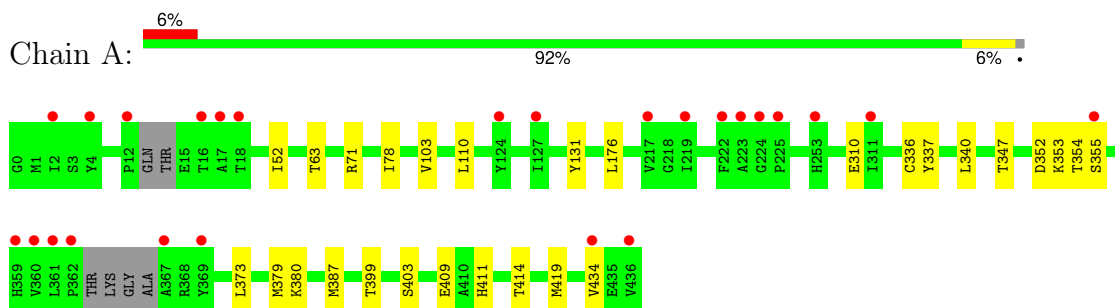
- Molecule 1: Sulfopropanediol 3-dehydrogenase
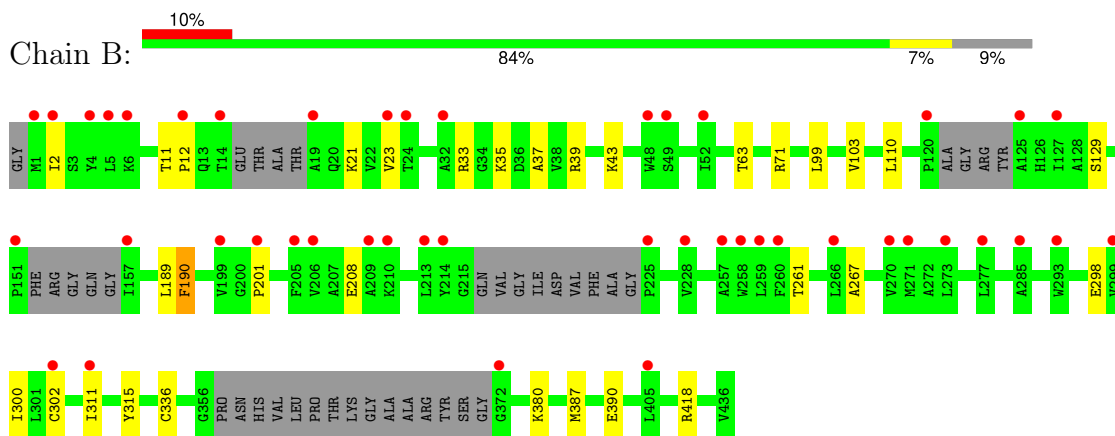
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	77.03Å 87.77Å 134.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.96 – 1.94 43.96 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.96-1.94) 99.9 (43.96-1.94)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.45 (at 1.94Å)	Xtriage
Refinement program	REFMAC 5.8.0415	Depositor
R, R_{free}	0.219 , 0.262 0.224 , 0.264	Depositor DCC
R_{free} test set	3480 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	43.5	Xtriage
Anisotropy	0.544	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 48.1	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	6514	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.40% 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: ZN, EDO

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	0.29	0/3329	0.52	0/4522
1	B	0.29	0/3079	0.54	0/4178
All	All	0.29	0/6408	0.53	0/8700

There are no bond length outliers.

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	3262	0	3225	16	0
1	B	3019	0	2986	16	0
2	A	4	0	6	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	126	0	0	1	0
4	B	101	0	0	1	0
All	All	6514	0	6217	29	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 (29) 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:336:CYS:SG	1:B:380:LYS:NZ	2.49	0.85
1:B:390:GLU:OE2	4:B:601:HOH:O	1.98	0.80
1:A:380:LYS:NZ	1:B:336:CYS:SG	2.58	0.76
1:B:33:ARG:HB2	1:B:37:ALA:HB2	1.80	0.64
1:B:302:CYS:SG	1:B:311:ILE:HD12	2.39	0.63
1:B:63:THR:O	1:B:71:ARG:NH2	2.38	0.55
1:B:300:ILE:HG21	1:B:311:ILE:HD13	1.91	0.51
1:B:129:SER:HB3	1:B:201:PRO:HG3	1.93	0.49
1:B:11:THR:HG22	1:B:12:PRO:HD2	1.95	0.48
1:A:409:GLU:HB2	1:A:434:VAL:HB	1.96	0.47
1:A:63:THR:O	1:A:71:ARG:NH2	2.48	0.46
1:A:419:MET:HA	1:A:419:MET:HE2	1.96	0.46
1:B:189:LEU:O	1:B:190:PHE:HB2	2.15	0.46
1:A:352:ASP:HB3	1:A:354:THR:O	2.17	0.45
1:A:78:ILE:HG12	1:A:131:TYR:CG	2.52	0.45
1:A:411:HIS:O	1:A:414:THR:OG1	2.23	0.43
1:A:337:TYR:CE1	1:A:340:LEU:HB2	2.54	0.43
1:B:2:ILE:HG21	1:B:311:ILE:HD11	2.00	0.43
1:A:52:ILE:HG23	1:A:176:LEU:HG	2.00	0.43
1:A:347:THR:HB	1:B:418:ARG:CZ	2.49	0.43
1:A:110:LEU:HD23	1:A:110:LEU:N	2.34	0.42
1:B:298:GLU:OE1	1:B:315:TYR:OH	2.25	0.42
1:A:399:THR:O	1:A:403:SER:OG	2.31	0.41
1:B:103:VAL:HG22	1:B:387:MET:HG2	2.02	0.41
1:B:261:THR:HG21	1:B:267:ALA:HB2	2.01	0.41
1:A:355:SER:HB2	1:A:373:LEU:HB3	2.02	0.41
1:A:103:VAL:HG22	1:A:387:MET:HG2	2.02	0.41
1:B:35:LYS:O	1:B:39:ARG:HG3	2.21	0.41
1:A:353:LYS:NZ	4:A:613:HOH:O	2.54	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	425/437 (97%)	413 (97%)	12 (3%)	0	100	100
1	B	388/437 (89%)	379 (98%)	6 (2%)	3 (1%)	16	7
All	All	813/874 (93%)	792 (97%)	18 (2%)	3 (0%)	30	22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	190	PHE
1	B	99	LEU
1	B	23	VAL

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	333/337 (99%)	331 (99%)	2 (1%)	84	83
1	B	309/337 (92%)	305 (99%)	4 (1%)	65	58
All	All	642/674 (95%)	636 (99%)	6 (1%)	75	72

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	310	GLU
1	A	379	MET
1	B	21	LYS
1	B	43	LYS
1	B	110	LEU
1	B	208	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	166	GLN
1	B	323	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

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	EDO	A	501	-	3,3,3	0.27	0	2,2,2	0.37	0

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	EDO	A	501	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

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(Å^2)	Q<0.9
1	A	431/437 (98%)	0.53	25 (5%)	30	33	34, 52, 83, 115	0
1	B	399/437 (91%)	0.89	45 (11%)	11	13	34, 59, 110, 132	1 (0%)
All	All	830/874 (94%)	0.70	70 (8%)	18	21	34, 55, 103, 132	1 (0%)

All (70) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	206	VAL	6.2
1	B	19	ALA	4.5
1	B	214	TYR	4.2
1	A	367	ALA	4.1
1	B	1	MET	4.0
1	B	4	TYR	3.9
1	A	361	LEU	3.9
1	A	124	TYR	3.7
1	B	201	PRO	3.7
1	B	125	ALA	3.6
1	A	2	ILE	3.6
1	A	223	ALA	3.5
1	B	151	PRO	3.5
1	B	293	TRP	3.3
1	B	257	ALA	3.3
1	A	4	TYR	3.3
1	B	23	VAL	3.2
1	B	205	PHE	3.2
1	A	217	VAL	3.1
1	B	2	ILE	3.1
1	A	369	TYR	3.1
1	B	120	PRO	3.1
1	A	222	PHE	3.1
1	B	209	ALA	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	434	VAL	3.0
1	B	12	PRO	3.0
1	B	210	LYS	2.9
1	B	270	VAL	2.8
1	A	360	VAL	2.7
1	B	228	VAL	2.7
1	B	258	TRP	2.6
1	A	253	HIS	2.6
1	A	362	PRO	2.6
1	B	225	PRO	2.6
1	A	359	HIS	2.6
1	B	405	LEU	2.5
1	A	219	ILE	2.5
1	B	157	ILE	2.5
1	B	277	LEU	2.5
1	B	372	GLY	2.5
1	B	127	ILE	2.4
1	B	24	THR	2.4
1	B	5	LEU	2.4
1	A	436	VAL	2.4
1	A	16	THR	2.4
1	A	225	PRO	2.4
1	B	14	THR	2.4
1	B	299	VAL	2.3
1	A	127	ILE	2.3
1	B	48	TRP	2.3
1	A	355	SER	2.3
1	A	12	PRO	2.2
1	A	18	THR	2.2
1	B	199	VAL	2.2
1	B	259	LEU	2.2
1	B	266	LEU	2.2
1	B	273	LEU	2.2
1	B	285	ALA	2.2
1	B	49	SER	2.1
1	A	224	GLY	2.1
1	B	6	LYS	2.1
1	B	260	PHE	2.1
1	A	311	ILE	2.1
1	B	311	ILE	2.1
1	B	302	CYS	2.0
1	A	17	ALA	2.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	32	ALA	2.0
1	B	52	ILE	2.0
1	B	271	MET	2.0
1	B	213	LEU	2.0

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

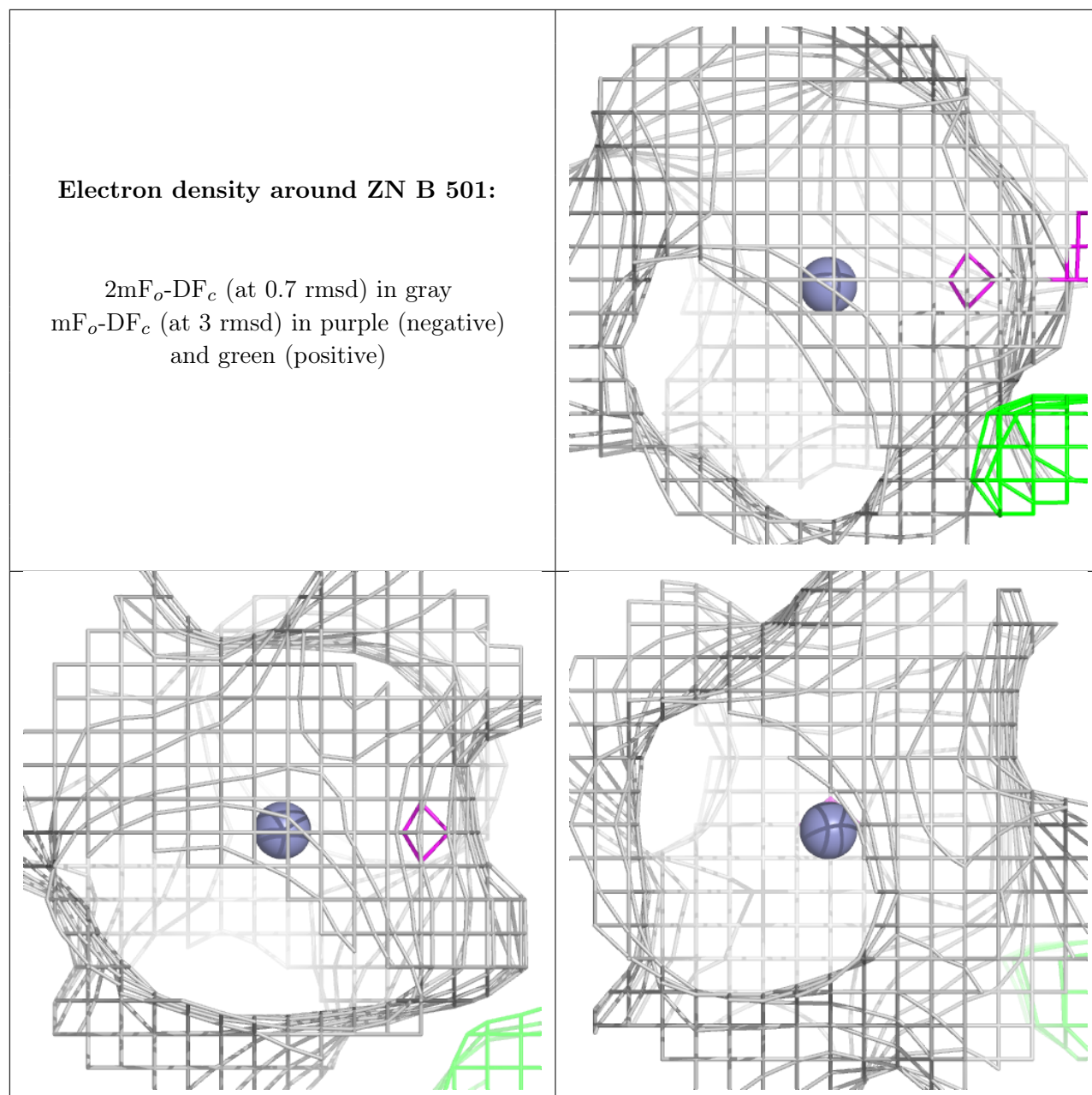There are no non-standard protein/DNA/RNA residues in this entry.

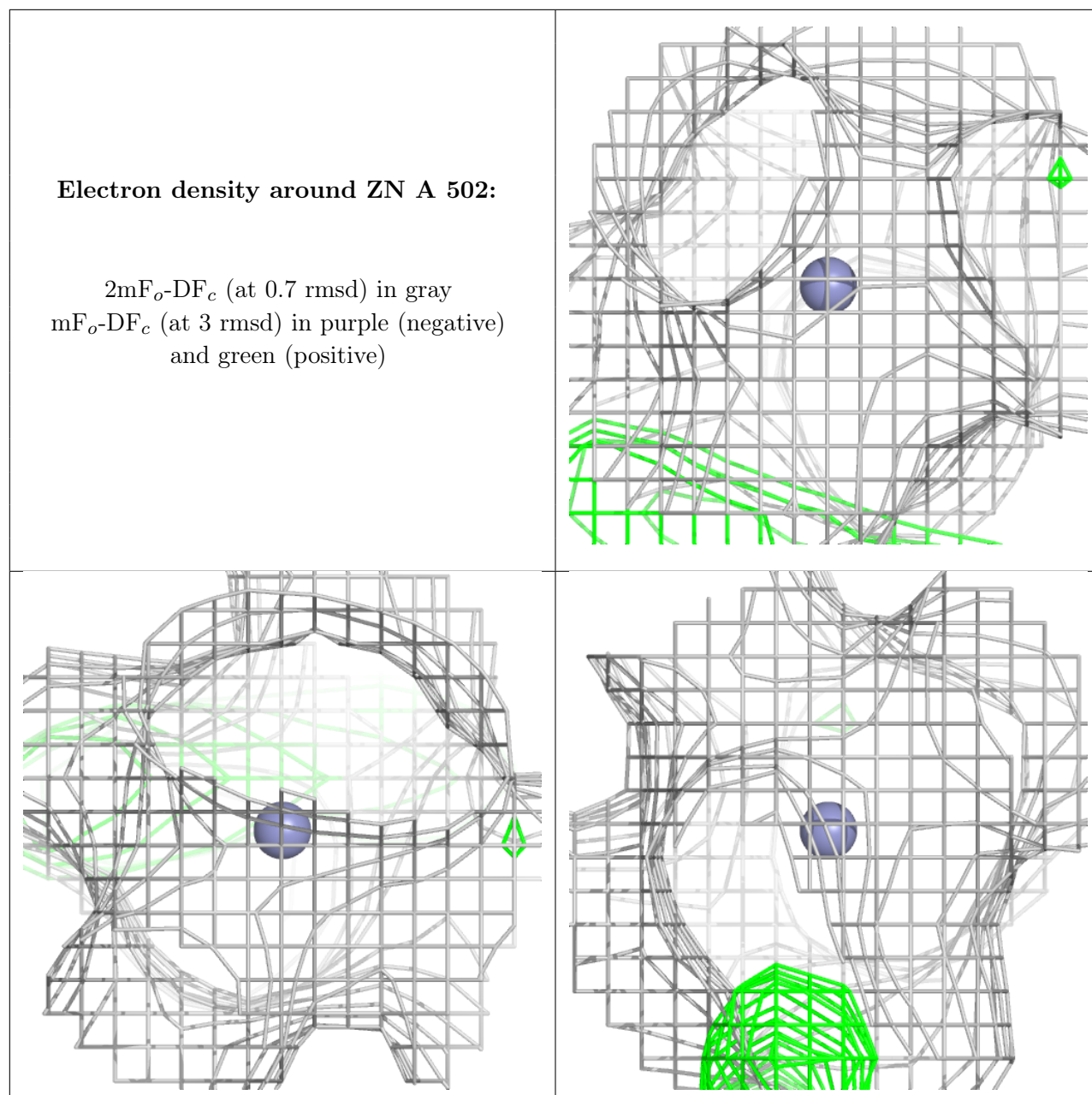## 6.3 Carbohydrates ⓘ

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
2	EDO	A	501	4/4	0.91	0.12	41,50,53,53	0
3	ZN	B	501	1/1	0.99	0.03	54,54,54,54	0
3	ZN	A	502	1/1	1.00	0.02	44,44,44,44	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 ZN B 501:

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

Electron density around ZN A 502:

$2mF_o$-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.