W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2022 – 03:19 pm GMT

PDB ID	:	6ZX9
Title	:	Crystal structure of SIV Vpr,fused to T4 lysozyme, isolated from moustached monkey, bound to human DDB1 and human DCAF1 (amino acid residues 1046-1396)
Authors	:	Schwefel, D.; Banchenko, S.
Deposited on	:	2020-07-29
Resolution	:	2.52 Å(reported)

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

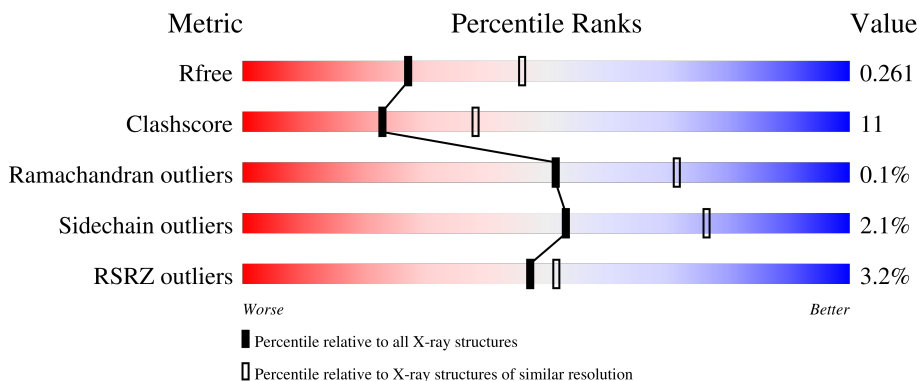|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.26
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

1 Overall quality at a glance ⓘ

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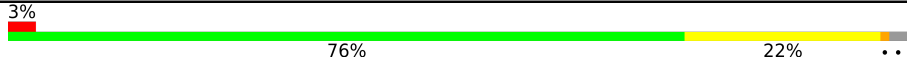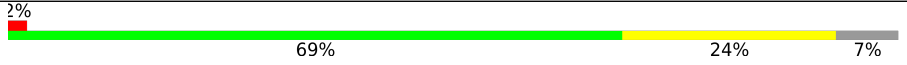*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.52 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		0.261
Clashscore		11
Ramachandran outliers		0.1%
Sidechain outliers		2.1%
RSRZ outliers		3.2%

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	5743 (2.54-2.50)
Clashscore	141614	6463 (2.54-2.50)
Ramachandran outliers	138981	6335 (2.54-2.50)
Sidechain outliers	138945	6337 (2.54-2.50)
RSRZ outliers	127900	5630 (2.54-2.50)

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	1142	3% / 76% 22% ··
2	B	360	2% / 69% 24% 7%
3	C	258	5% / 73% 25% ·

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 13858 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 damage-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1124	Total	C	N	O	S	0	1	0
			8808	5581	1485	1695	47			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP Q16531
A	0	SER	-	expression tag	UNP Q16531

- Molecule 2 is a protein called DDB1- and CUL4-associated factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	335	Total	C	N	O	S	0	0	0
			2670	1684	467	501	18			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1045	MET	-	initiating methionine	UNP Q9Y4B6
B	1397	GLU	-	expression tag	UNP Q9Y4B6
B	1398	LEU	-	expression tag	UNP Q9Y4B6
B	1399	ALA	-	expression tag	UNP Q9Y4B6
B	1400	LEU	-	expression tag	UNP Q9Y4B6
B	1401	VAL	-	expression tag	UNP Q9Y4B6
B	1402	PRO	-	expression tag	UNP Q9Y4B6
B	1403	ARG	-	expression tag	UNP Q9Y4B6
B	1404	GLY	-	expression tag	UNP Q9Y4B6

- Molecule 3 is a protein called Vpr protein fused to T4 lysozyme.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	257	Total	C	N	O	S	0	0	0
			2078	1306	384	379	9			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-156	HIS	GLU	conflict	UNP P00720
C	-155	GLY	ARG	conflict	UNP P00720
C	-113	THR	CYS	conflict	UNP P00720
C	-70	ALA	CYS	conflict	UNP P00720
C	-30	ARG	ILE	conflict	UNP P00720
C	-2	ALA	-	linker	UNP P00720
C	-1	ALA	-	linker	UNP P00720
C	0	ALA	-	linker	UNP P00720

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

C3

OH O3

C1 C2

HO OH

O1 O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		
4	A	1	Total	C	O	0	0
			6	3	3		

Continued on next page...

W O R L D W I D E
PDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 6	C 3	O 3	0	0
4	A	1	Total 6	C 3	O 3	0	0
4	B	1	Total 6	C 3	O 3	0	0
4	C	1	Total 6	C 3	O 3	0	0

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	1	Total 1	Zn 1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	202	Total 202	O 202	0	0
6	B	37	Total 37	O 37	0	0
6	C	14	Total 14	O 14	0	0

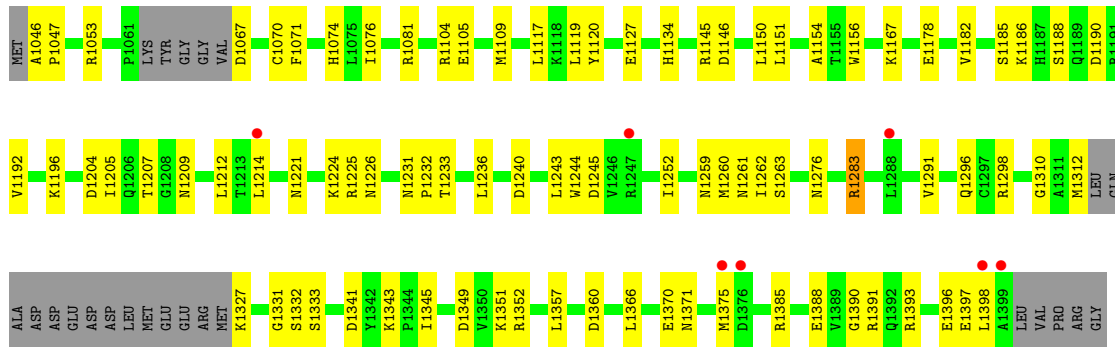# 3 Residue-property plots ⓘ

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.
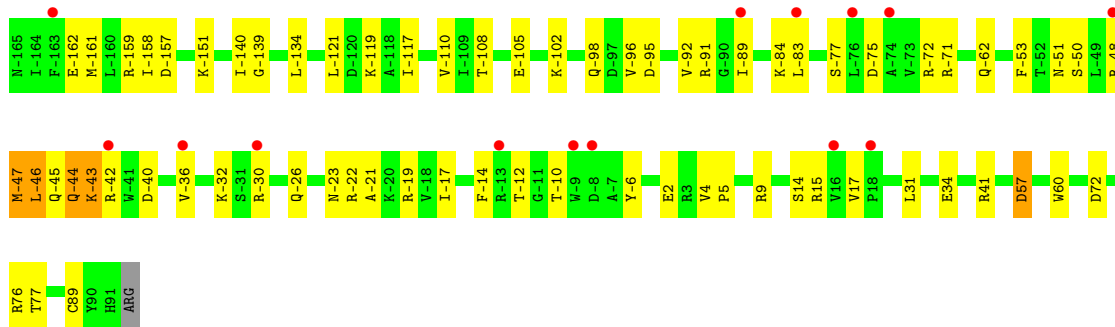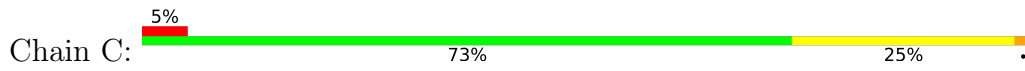
• Molecule 1: DNA damage-binding protein 1

• Molecule 2: DDB1- and CUL4-associated factor 1

Chain B:

2%

69% 24% 7%

MET A1046 P1047 R1053 P1061 LYS TYR GLY GLY VAL D1067 C1070 F1071 H1074 L1075 I1076 R1081 R1104 E1105 M1109 L1117 L1118 L1119 Y1120 E1127 H1134 R1145 D1146 L1150 L1151 A1154 T1155 W1156 K1167 E1178 V1182 S1185 K1186 H1187 S1188 Q1189 D1190 R1191

V1192 K1196 D1204 I1205 Q1206 T1207 G1208 N1209 L1212 T1213 L1214 N1221 K1224 R1225 N1226 N1231 P1232 T1233 L1236 D1240 L1243 W1244 D1245 V1246 R1247 I1252 N1259 M1260 N1261 I1262 S1263 N1276 R1283 L1288 V1291 Q1296 C1297 R1298 G1310 A1311 M1312 LEU GLN

ALA ASP ASP GLU ASP ASP LEU MET GLU GLU ARG MET K1327 G1331 S1332 S1333 D1341 Y1342 K1343 P1344 I1345 D1349 V1350 K1351 R1352 L1357 D1360 L1366 E1370 N1371 M1375 D1376 R1385 E1388 V1389 G1390 R1391 Q1392 R1393 E1396 E1397 L1398 A1399 LEU VAL PRO ARG GLY

• Molecule 3: Vpr protein fused to T4 lysozyme

Chain C:

5%

73% 25% •

N-165 I-164 F-163 E-162 M-161 L-160 R-159 I-158 D-157 K-151 I-140 G-139 L-134 L-121 D-120 K-119 A-118 I-117 V-110 I-109 T-108 E-105 K-102 Q-98 D-97 V-96 D-95 V-92 R-91 G-90 I-89 K-84 L-83 S-77 L-76 D-75 A-74 V-73 R-72 R-71 Q-62 F-53 T-52 N-51 S-50 L-49 R-48

M-47 L-46 Q-45 Q-44 K-43 R-42 W-41 D-40 V-36 K-32 S-31 R-30 Q-26 N-23 R-22 A-21 K-20 R-19 V-18 I-17 F-14 R-13 T-12 G-11 T-10 W-9 D-8 A-7 Y-6 E2 R3 V4 P5 R9 S14 R15 V16 V17 P18 L31 E34 R41 D57 W60 D72

R76 T77 C89 Y90 H91 ARG

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	265.90Å 95.54Å 98.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	79.07 – 2.52 79.07 – 2.52	Depositor EDS
% Data completeness (in resolution range)	99.2 (79.07-2.52) 89.0 (79.07-2.52)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.81 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.12_2829, PHENIX 1.12_2829	Depositor
R, R_{free}	0.216 , 0.260 0.216 , 0.261	Depositor DCC
R_{free} test set	3948 reflections (4.62%)	wwPDB-VP
Wilson B-factor (Å^2)	46.3	Xtriage
Anisotropy	0.133	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.018 for -h,l,k	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	13858	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.63% 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, GOL

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	0/8973	0.58	3/12154 (0.0%)
2	B	0.28	0/2731	0.50	0/3696
3	C	0.31	0/2124	0.48	0/2878
All	All	0.29	0/13828	0.55	3/18728 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1030	PHE	CB-CG-CD1	5.67	124.77	120.80
1	A	491	LYS	CD-CE-NZ	5.34	123.97	111.70
1	A	155	PHE	CB-CG-CD1	5.03	124.32	120.80

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	928	ARG	Peptide

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	8808	0	8782	195	0
2	B	2670	0	2568	71	0
3	C	2078	0	2059	51	0
4	A	36	0	48	2	0
4	B	6	0	8	0	0
4	C	6	0	8	0	0
5	C	1	0	0	0	0
6	A	202	0	0	8	0
6	B	37	0	0	1	0
6	C	14	0	0	2	0
All	All	13858	0	13473	309	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 11.

The worst 5 of 309 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)
2:B:1393:ARG:HD2	2:B:1398:LEU:CD2	1.28	1.56
2:B:1393:ARG:CD	2:B:1398:LEU:HD21	1.30	1.55
1:A:111:ARG:NH1	1:A:112:ILE:CG2	1.68	1.54
1:A:111:ARG:NH1	1:A:112:ILE:HG22	0.83	1.15
1:A:906:TYR:OH	2:B:1396:GLU:OE1	1.71	1.07

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	1117/1142 (98%)	1061 (95%)	56 (5%)	0	100	100
2	B	329/360 (91%)	309 (94%)	20 (6%)	0	100	100
3	C	255/258 (99%)	247 (97%)	7 (3%)	1 (0%)	34	53
All	All	1701/1760 (97%)	1617 (95%)	83 (5%)	1 (0%)	51	71

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	-44	GLN

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	987/1000 (99%)	970 (98%)	17 (2%)	60	81
2	B	294/315 (93%)	289 (98%)	5 (2%)	60	81
3	C	217/218 (100%)	207 (95%)	10 (5%)	27	47
All	All	1498/1533 (98%)	1466 (98%)	32 (2%)	53	76

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	9	ARG
3	C	41	ARG
1	A	713	ARG
1	A	699	LEU
3	C	57	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	467	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	1296	GLN
3	C	-45	GLN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 9 ligands modelled in this entry, 1 is monoatomic - leaving 8 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	$\#|Z| > 2$
4	GOL	A	1202	-	5,5,5	0.95	0	5,5,5	0.90	0
4	GOL	A	1206	-	5,5,5	0.92	0	5,5,5	0.97	0
4	GOL	B	1501	-	5,5,5	0.90	0	5,5,5	1.00	0
4	GOL	A	1205	-	5,5,5	0.93	0	5,5,5	0.97	0
4	GOL	C	102	-	5,5,5	0.92	0	5,5,5	1.00	0
4	GOL	A	1201	-	5,5,5	0.90	0	5,5,5	1.01	0
4	GOL	A	1203	-	5,5,5	0.89	0	5,5,5	1.03	0
4	GOL	A	1204	-	5,5,5	1.00	0	5,5,5	0.82	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
4	GOL	A	1202	-	-	4/4/4/4	-
4	GOL	A	1206	-	-	3/4/4/4	-
4	GOL	B	1501	-	-	2/4/4/4	-
4	GOL	A	1205	-	-	2/4/4/4	-
4	GOL	C	102	-	-	0/4/4/4	-
4	GOL	A	1201	-	-	0/4/4/4	-
4	GOL	A	1203	-	-	4/4/4/4	-
4	GOL	A	1204	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 18 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1202	GOL	O1-C1-C2-C3
4	A	1202	GOL	C1-C2-C3-O3
4	A	1203	GOL	C1-C2-C3-O3
4	A	1203	GOL	O2-C2-C3-O3
4	B	1501	GOL	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1203	GOL	1	0
4	A	1204	GOL	1	0

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	1124/1142 (98%)	0.28	34 (3%)	50	54	29, 61, 121, 175	0
2	B	335/360 (93%)	0.24	7 (2%)	63	67	39, 67, 99, 132	0
3	C	257/258 (99%)	0.45	14 (5%)	25	28	47, 86, 126, 149	0
All	All	1716/1760 (97%)	0.30	55 (3%)	47	52	29, 67, 121, 175	0

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	16	VAL	7.7
1	A	111	ARG	5.7
1	A	448	LEU	5.3
2	B	1398	LEU	5.0
1	A	644	LEU	4.9

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

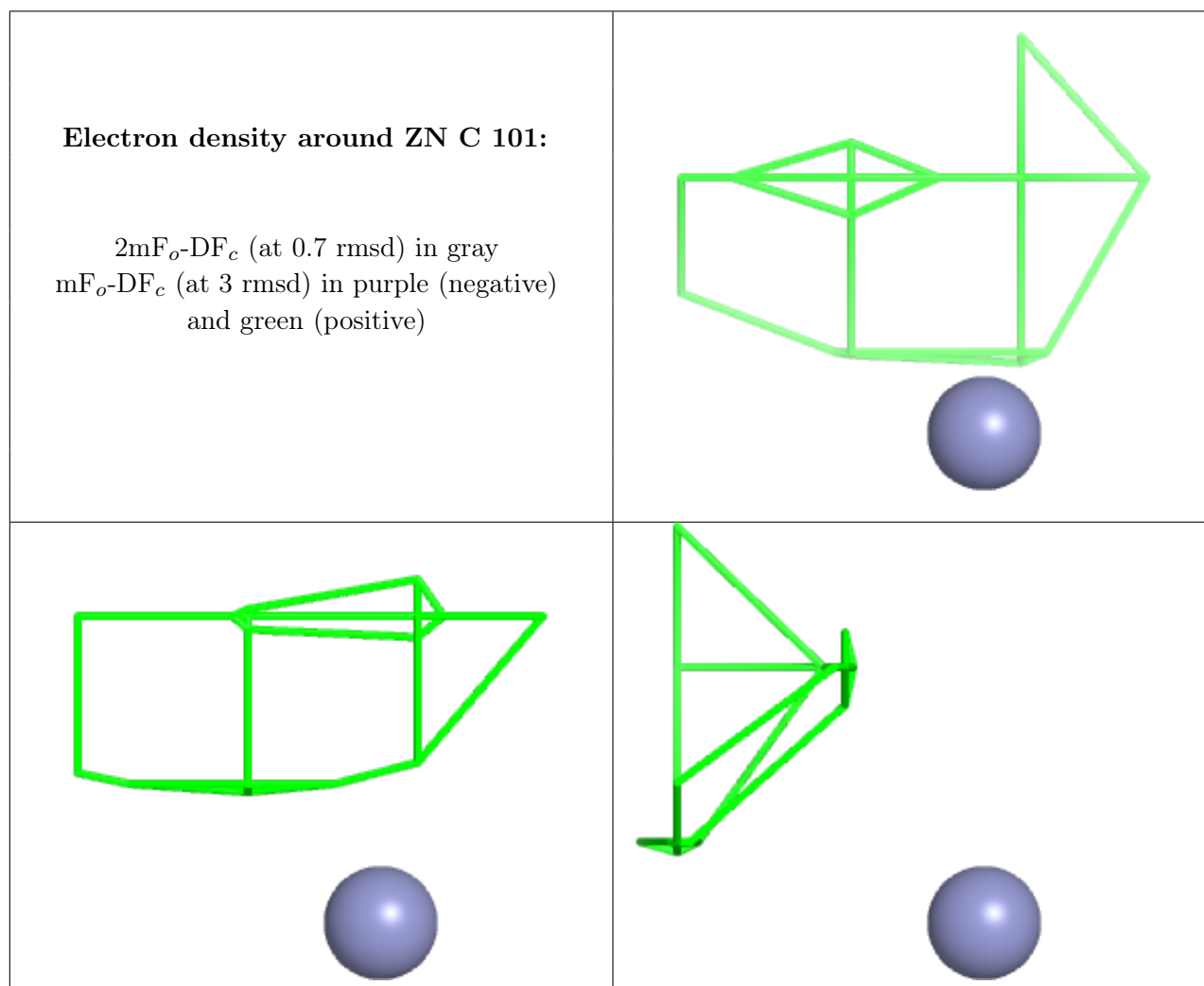There are no non-standard protein/DNA/RNA residues in this entry.

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
4	GOL	B	1501	6/6	0.75	0.18	73,89,90,90	0
4	GOL	A	1204	6/6	0.77	0.24	69,73,79,81	0
4	GOL	A	1206	6/6	0.87	0.20	80,89,91,92	0
4	GOL	A	1205	6/6	0.92	0.17	69,73,74,76	0
4	GOL	C	102	6/6	0.92	0.11	88,89,89,89	0
4	GOL	A	1203	6/6	0.93	0.18	61,65,74,75	0
4	GOL	A	1201	6/6	0.94	0.25	44,58,66,77	0
4	GOL	A	1202	6/6	0.96	0.11	51,57,58,62	0
5	ZN	C	101	1/1	0.97	0.17	79,79,79,79	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 C 101:

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

6.5 Other polymers (i)

There are no such residues in this entry.