Full wwPDB X-ray Structure Validation Report ⓘ

Apr 27, 2023 – 11:06 am BST

PDB ID	:	8AAS
Title	:	Crystal structure of the Pyrococcus abyssi RPA trimerization core bound to poly-dT20 ssDNA
Authors	:	Madru, C.; Legrand, P.; Sauguet, L.
Deposited on	:	2022-07-01
Resolution	:	3.20 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

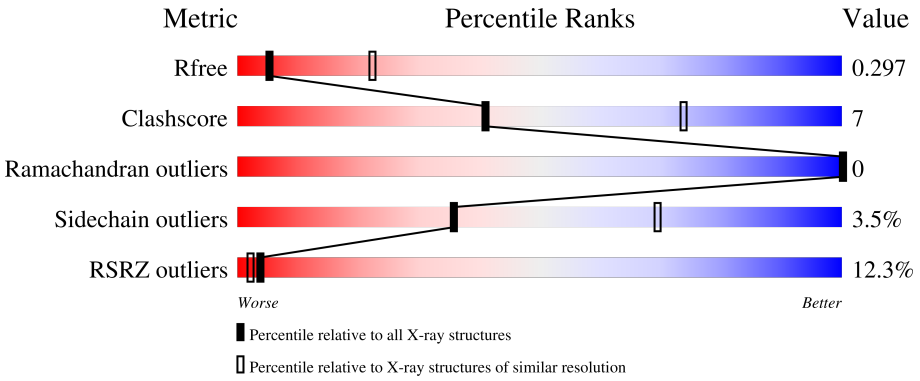
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.32.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.32.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.20 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	295	13% 48% 10% · 41%
2	B	183	5% 84% 14% ·
3	C	122	9% 84% 9% 7%
4	F	20	5% 25% 40% 5% 30%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	CA	A	402	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 4057 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 Replication factor A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	173	Total	C	N	O	S	0	0	0
			1389	879	233	272	5			

- Molecule 2 is a protein called RPA32 subunit of the hetero-oligomeric complex involved in homologous recombination.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	178	Total	C	N	O	S	0	0	0
			1457	947	244	263	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP Q9V1Z1
B	1	SER	-	expression tag	UNP Q9V1Z1

- Molecule 3 is a protein called RPA14 subunit of the hetero-oligomeric complex involved in homologous recombination.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	113	Total	C	N	O	S	0	0	0
			925	583	165	175	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-4	GLY	-	expression tag	UNP Q9V1Z0
C	-3	THR	-	expression tag	UNP Q9V1Z0
C	-2	GLY	-	expression tag	UNP Q9V1Z0
C	-1	ASP	-	expression tag	UNP Q9V1Z0
C	0	GLY	-	expression tag	UNP Q9V1Z0
C	1	SER	-	expression tag	UNP Q9V1Z0

- Molecule 4 is a DNA chain called DNA (5'-D(P*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	14	Total	C	N	O	P	0	0	0
			280	140	28	98	14			

- 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	A	1	Total	Zn	0	0
			1	1		

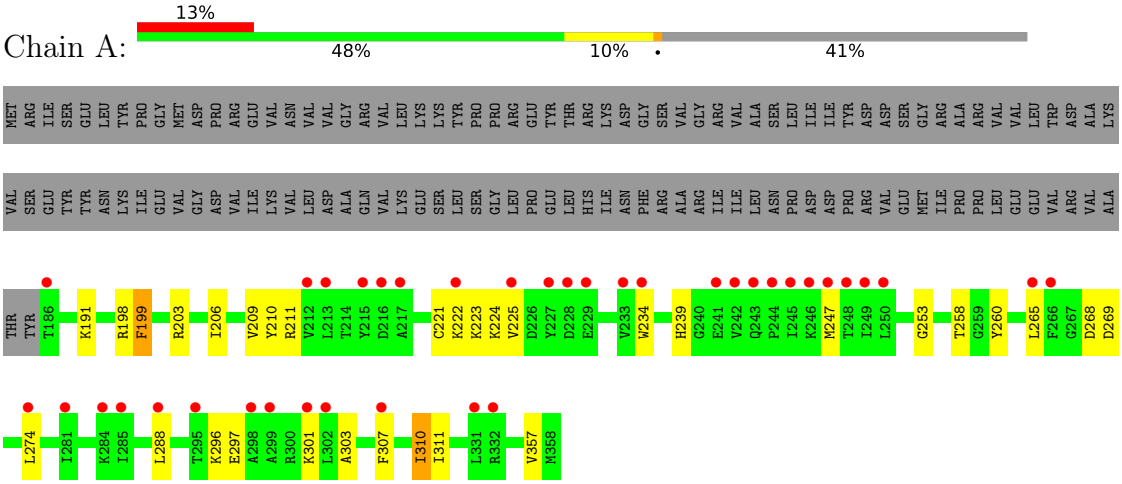
- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

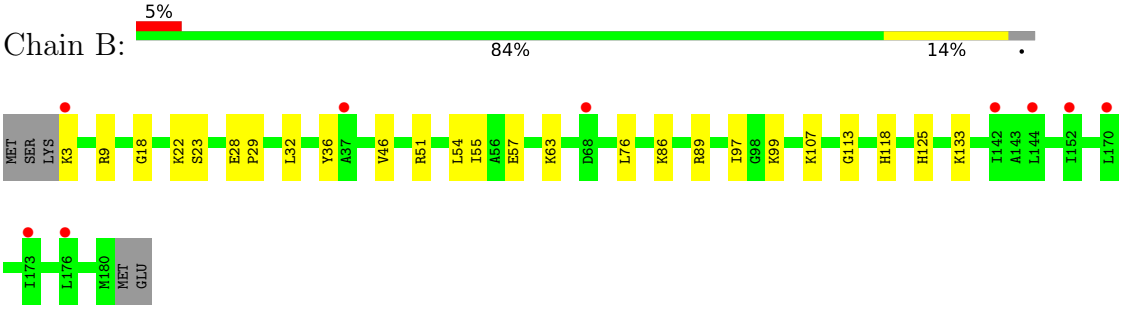
3 Residue-property plots (i)

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.

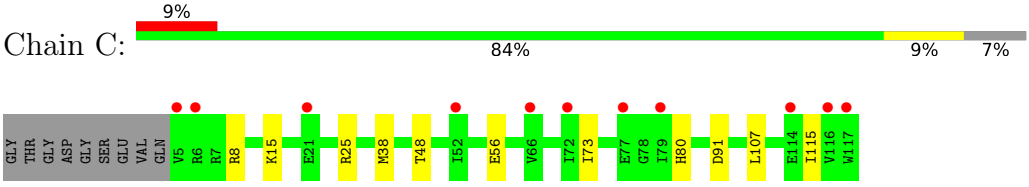
• Molecule 1: Replication factor A



• Molecule 2: RPA32 subunit of the hetero-oligomeric complex involved in homologous recombination

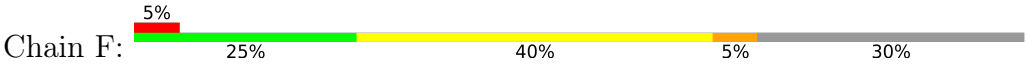


• Molecule 3: RPA14 subunit of the hetero-oligomeric complex involved in homologous recombination



• Molecule 4: DNA (5'-D(P*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*TP*T)-3')

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	75.23Å 75.23Å 215.49Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.26 – 3.20 48.26 – 3.20	Depositor EDS
% Data completeness (in resolution range)	81.5 (48.26-3.20) 81.5 (48.26-3.20)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.37 (at 3.19Å)	Xtriage
Refinement program	BUSTER 2.10.4 (3-FEB-2022)	Depositor
R, R_{free}	0.264 , 0.293 0.269 , 0.297	Depositor DCC
R_{free} test set	494 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	146.9	Xtriage
Anisotropy	0.060	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 112.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.056 for -h,-k,l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	4057	wwPDB-VP
Average B, all atoms (Å^2)	173.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% 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, CA

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	0.34	0/1408	0.49	0/1893
2	B	0.40	0/1484	0.52	0/1996
3	C	0.45	0/936	0.55	0/1259
4	F	0.93	0/307	1.08	1/472 (0.2%)
All	All	0.45	0/4135	0.59	1/5620 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	F	13	DT	C6-C5-C7	-5.04	119.88	122.90

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	1389	0	1393	26	0
2	B	1457	0	1512	19	0
3	C	925	0	948	16	0
4	F	280	0	169	19	0
5	A	1	0	0	0	0
6	A	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	1	0	0	0	0
6	C	1	0	0	0	0
All	All	4057	0	4022	57	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 7.

All (57) 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:211:ARG:NH2	4:F:3:DT:H3	1.51	1.06
1:A:223:LYS:HG2	4:F:2:DT:H4'	1.52	0.91
1:A:224:LYS:HB2	4:F:3:DT:O5'	1.86	0.75
1:A:223:LYS:HZ2	4:F:2:DT:H5'	1.52	0.74
2:B:9:ARG:NH1	2:B:125:HIS:ND1	2.36	0.73
2:B:29:PRO:HA	4:F:10:DT:OP1	1.89	0.73
2:B:51:ARG:CZ	3:C:8:ARG:HH11	2.03	0.71
3:C:8:ARG:NH2	3:C:25:ARG:HH22	1.89	0.69
2:B:51:ARG:NH2	3:C:8:ARG:NH1	2.43	0.67
1:A:311:ILE:HG22	2:B:113:GLY:HA3	1.76	0.66
1:A:206:ILE:HD13	1:A:310:ILE:HD12	1.79	0.65
2:B:51:ARG:CZ	3:C:8:ARG:NH1	2.60	0.64
2:B:51:ARG:NH2	3:C:8:ARG:HH11	1.94	0.64
1:A:209:VAL:HG12	1:A:303:ALA:HB1	1.79	0.63
2:B:99:LYS:NZ	4:F:10:DT:OP2	2.30	0.62
3:C:15:LYS:HD3	3:C:48:THR:HG21	1.81	0.62
1:A:224:LYS:NZ	4:F:4:DT:OP1	2.33	0.59
1:A:357:VAL:HG13	3:C:107:LEU:HD11	1.84	0.58
3:C:8:ARG:CZ	3:C:25:ARG:HH22	2.16	0.58
1:A:211:ARG:HH22	4:F:3:DT:H3	0.71	0.55
1:A:265:LEU:HD22	1:A:269:ASP:HB3	1.89	0.55
2:B:118:HIS:HE2	3:C:91:ASP:HB2	1.72	0.55
3:C:8:ARG:HH21	3:C:25:ARG:HH22	1.53	0.55
1:A:224:LYS:N	4:F:3:DT:OP1	2.34	0.54
2:B:36:TYR:HE2	2:B:133:LYS:HZ1	1.47	0.53
3:C:8:ARG:HE	3:C:25:ARG:NH2	2.08	0.52
1:A:206:ILE:HG21	1:A:310:ILE:HD12	1.92	0.51
3:C:8:ARG:NE	3:C:25:ARG:NH2	2.58	0.51
1:A:288:LEU:HD22	1:A:301:LYS:HD3	1.93	0.50
2:B:46:VAL:HG23	2:B:125:HIS:ND1	2.27	0.50
1:A:210:TYR:HE1	4:F:8:DT:OP1	1.95	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:F:4:DT:H2"	4:F:5:DT:C4'	2.41	0.49
3:C:73:ILE:CG2	3:C:80:HIS:CE1	2.95	0.49
3:C:73:ILE:HG21	3:C:80:HIS:CE1	2.48	0.48
1:A:223:LYS:HZ2	4:F:2:DT:C5'	2.22	0.47
1:A:224:LYS:H	4:F:3:DT:P	2.36	0.47
1:A:223:LYS:NZ	4:F:2:DT:H5'	2.27	0.47
3:C:8:ARG:NE	3:C:25:ARG:HH22	2.13	0.47
1:A:198:ARG:NH1	1:A:199:PHE:CZ	2.83	0.46
4:F:8:DT:H2"	4:F:9:DT:H5"	1.97	0.46
2:B:23:SER:HB3	2:B:28:GLU:HB2	1.98	0.45
1:A:221:CYS:HB3	1:A:239:HIS:HE1	1.81	0.45
4:F:4:DT:H2"	4:F:5:DT:H4'	2.01	0.43
2:B:55:ILE:HD11	2:B:86:LYS:HG2	1.99	0.43
2:B:54:LEU:HD23	2:B:63:LYS:HE2	2.00	0.43
4:F:12:DT:H2"	4:F:13:DT:OP1	2.19	0.43
1:A:223:LYS:NZ	4:F:2:DT:C5'	2.82	0.43
1:A:225:VAL:HG23	1:A:234:TRP:CE3	2.54	0.43
2:B:36:TYR:CE1	3:C:115:ILE:HD11	2.54	0.43
1:A:191:LYS:HD3	1:A:258:THR:HB	2.02	0.42
1:A:274:LEU:HD22	1:A:307:PHE:CD1	2.55	0.42
2:B:18:GLY:HA3	2:B:32:LEU:HD21	2.02	0.42
1:A:253:GLY:HA3	1:A:260:TYR:OH	2.21	0.41
2:B:76:LEU:CD2	4:F:13:DT:H72	2.50	0.41
2:B:55:ILE:CD1	2:B:86:LYS:HG2	2.51	0.41
2:B:36:TYR:CE2	2:B:133:LYS:NZ	2.73	0.41
1:A:222:LYS:HD2	1:A:247:MET:HE1	2.03	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

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	171/295 (58%)	169 (99%)	2 (1%)	0	100	100
2	B	176/183 (96%)	172 (98%)	4 (2%)	0	100	100
3	C	111/122 (91%)	109 (98%)	2 (2%)	0	100	100
All	All	458/600 (76%)	450 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

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	149/258 (58%)	143 (96%)	6 (4%)	31	66
2	B	155/160 (97%)	149 (96%)	6 (4%)	32	67
3	C	101/107 (94%)	99 (98%)	2 (2%)	55	80
All	All	405/525 (77%)	391 (96%)	14 (4%)	36	69

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	199	PHE
1	A	203	ARG
1	A	268	ASP
1	A	296	LYS
1	A	297	GLU
1	A	310	ILE
2	B	3	LYS
2	B	22	LYS
2	B	57	GLU
2	B	89	ARG
2	B	97	ILE
2	B	107	LYS
3	C	38	MET
3	C	56	GLU

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

Mol	Chain	Res	Type
2	B	149	GLN
3	C	80	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 monosaccharides in this entry.

5.6 Ligand geometry

Of 6 ligands modelled in this entry, 6 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, 95^{th} 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	173/295 (58%)	0.98	38 (21%) 0 0	129, 182, 230, 238	0
2	B	178/183 (97%)	0.49	9 (5%) 28 16	105, 152, 244, 251	0
3	C	113/122 (92%)	0.48	11 (9%) 7 4	118, 143, 214, 233	0
4	F	14/20 (70%)	0.38	1 (7%) 16 9	183, 200, 234, 237	0
All	All	478/620 (77%)	0.66	59 (12%) 4 2	105, 168, 237, 251	0

All (59) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	227	TYR	5.9
1	A	215	TYR	5.8
3	C	116	VAL	5.6
1	A	216	ASP	5.6
1	A	244	PRO	5.4
1	A	246	LYS	5.0
1	A	234	TRP	5.0
1	A	245	ILE	4.6
1	A	241	GLU	4.5
2	B	152	ILE	4.4
1	A	281	ILE	4.2
1	A	229	GLU	4.2
1	A	249	ILE	4.1
1	A	225	VAL	4.0
3	C	117	TRP	3.8
1	A	228	ASP	3.7
1	A	299	ALA	3.7
1	A	248	THR	3.5
2	B	142	ILE	3.4
2	B	37	ALA	3.4
1	A	295	THR	3.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	247	MET	3.4
1	A	213	LEU	3.4
1	A	212	VAL	3.4
1	A	217	ALA	3.4
2	B	144	LEU	3.3
1	A	302	LEU	3.3
1	A	250	LEU	3.2
1	A	301	LYS	3.2
1	A	284	LYS	3.1
1	A	285	ILE	3.1
2	B	176	LEU	3.0
1	A	222	LYS	3.0
1	A	186	THR	3.0
1	A	332	ARG	3.0
3	C	79	ILE	2.9
1	A	298	ALA	2.8
1	A	243	GLN	2.7
3	C	66	VAL	2.7
1	A	266	PHE	2.6
1	A	288	LEU	2.6
3	C	21	GLU	2.6
1	A	307	PHE	2.5
2	B	173	ILE	2.5
3	C	72	ILE	2.5
1	A	265	LEU	2.5
3	C	5	VAL	2.5
3	C	114	GLU	2.4
1	A	331	LEU	2.4
1	A	242	VAL	2.3
2	B	3	LYS	2.3
1	A	274	LEU	2.3
1	A	233	VAL	2.2
3	C	77	GLU	2.1
4	F	3	DT	2.1
3	C	6	ARG	2.1
3	C	52	ILE	2.0
2	B	170	LEU	2.0
2	B	68	ASP	2.0

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

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

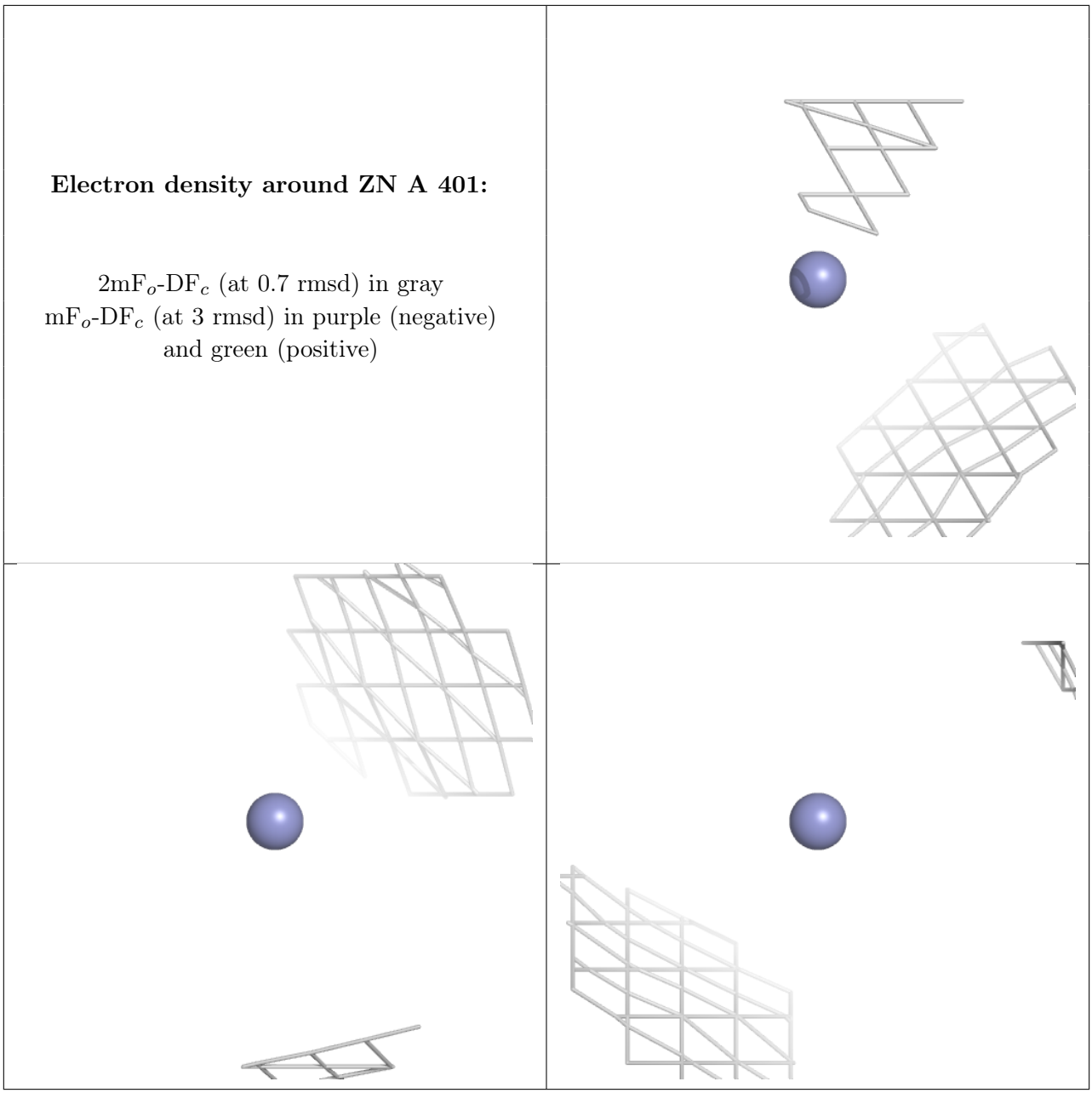
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
6	CA	C	201	1/1	0.77	0.15	151,151,151,151	0
6	CA	A	402	1/1	0.79	0.71	174,174,174,174	0
6	CA	B	201	1/1	0.88	0.58	121,121,121,121	0
6	CA	A	403	1/1	0.89	0.23	180,180,180,180	0
6	CA	A	404	1/1	0.89	0.06	169,169,169,169	0
5	ZN	A	401	1/1	0.92	0.07	255,255,255,255	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 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 (i)

There are no such residues in this entry.