Full wwPDB X-ray Structure Validation Report ⓘ

Oct 9, 2023 – 10:38 PM EDT

PDB ID	:	7SOZ
Title	:	Replication Initiator Protein REPE54 and cognate DNA sequence with terminal three prime phosphates chemically crosslinked (5 mg/mL EDC, 12 hours).
Authors	:	Ward, A.R.; Snow, C.D.
Deposited on	:	2021-11-01
Resolution	:	3.14 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:

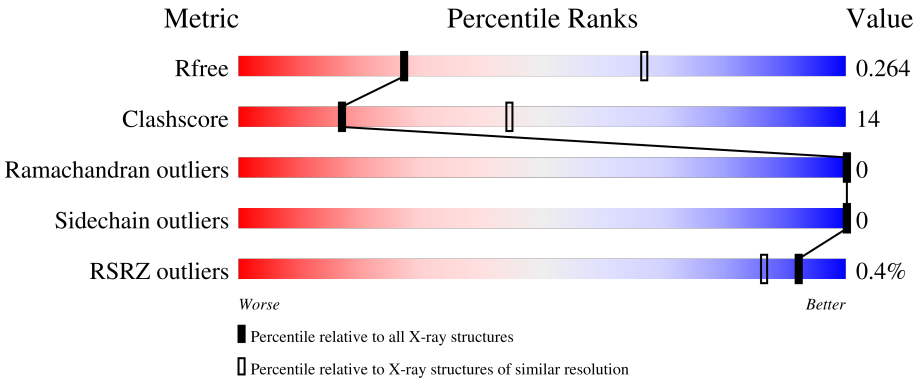
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.14 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	1626 (3.18-3.10)
Clashscore	141614	1735 (3.18-3.10)
Ramachandran outliers	138981	1677 (3.18-3.10)
Sidechain outliers	138945	1677 (3.18-3.10)
RSRZ outliers	127900	1588 (3.18-3.10)

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	22	64% 27% 9%
2	B	22	55% 36% 9%
3	C	263	59% 26% 15%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 2691 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 DNA chain called DNA (5'-D(*CP*CP*TP*GP*TP*GP*AP*CP*AP*AP*AP*TP*TP*GP*CP*CP*CP*TP*CP*AP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	A	22	427	203	76	127	21	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	22	436	206	82	127	21	0	0	0

- Molecule 3 is a protein called RepB family plasmid replication initiator protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	224	1824	1173	317	327	7	0	1	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-11	MET	-	initiating methionine	UNP Q0E856
C	-10	ARG	-	expression tag	UNP Q0E856
C	-9	GLY	-	expression tag	UNP Q0E856
C	-8	SER	-	expression tag	UNP Q0E856
C	-7	HIS	-	expression tag	UNP Q0E856
C	-6	HIS	-	expression tag	UNP Q0E856
C	-5	HIS	-	expression tag	UNP Q0E856
C	-4	HIS	-	expression tag	UNP Q0E856
C	-3	HIS	-	expression tag	UNP Q0E856
C	-2	HIS	-	expression tag	UNP Q0E856
C	-1	GLY	-	expression tag	UNP Q0E856
C	0	SER	-	expression tag	UNP Q0E856

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	118	PRO	ARG	engineered mutation	UNP Q0E856

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	1	Total Mg 1 1	0	0

- Molecule 5 is water.

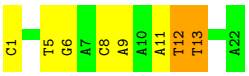
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total O 1 1	0	0
5	C	2	Total O 2 2	0	0

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

Chain A:

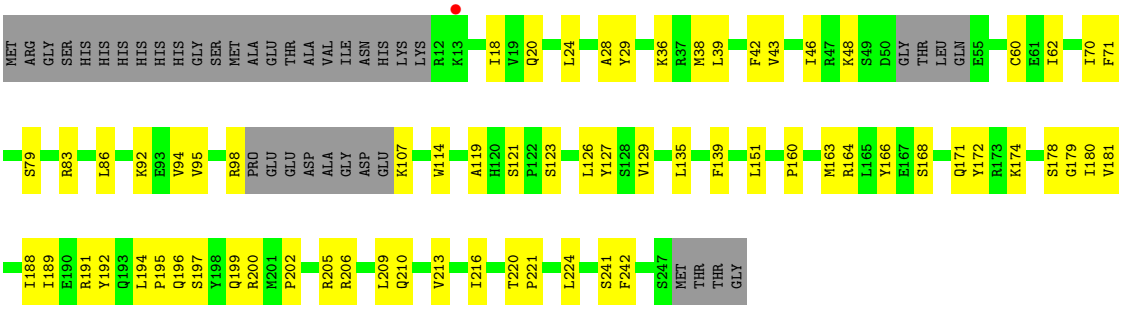


• Molecule 2: DNA (5'-D(*CP*TP*GP*AP*GP*GP*GP*CP*AP*AP*TP*TP*TP*GP*TP*CP *AP*CP*AP*GP*GP*A)-3')

Chain B:

• Molecule 3: RepB family plasmid replication initiator protein

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	111.99Å 79.18Å 74.90Å 90.00° 123.02° 90.00°	Depositor
Resolution (Å)	37.05 – 3.14 37.05 – 3.14	Depositor EDS
% Data completeness (in resolution range)	95.9 (37.05-3.14) 95.9 (37.05-3.14)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.31 (at 3.12Å)	Xtriage
Refinement program	PHENIX 1.19_4092	Depositor
R, R_{free}	0.193 , 0.263 0.194 , 0.264	Depositor DCC
R_{free} test set	498 reflections (5.36%)	wwPDB-VP
Wilson B-factor (Å^2)	83.4	Xtriage
Anisotropy	0.703	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 72.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2691	wwPDB-VP
Average B, all atoms (Å^2)	109.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.74% 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: MG

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	1.20	1/477 (0.2%)	1.13	2/734 (0.3%)
2	B	1.19	0/489	1.24	4/755 (0.5%)
3	C	0.62	0/1871	0.81	0/2525
All	All	0.86	1/2837 (0.0%)	0.96	6/4014 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	13	DT	C3'-O3'	-5.46	1.36	1.44

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	23	DC	O4'-C4'-C3'	-8.83	100.70	106.00
2	B	43	DG	C1'-O4'-C4'	-7.00	103.10	110.10
1	A	11	DA	OP1-P-O3'	6.21	118.85	105.20
2	B	38	DC	O4'-C1'-N1	6.07	112.25	108.00
1	A	12	DT	O4'-C4'-C3'	-5.06	102.48	104.50
2	B	29	DG	O4'-C4'-C3'	-5.00	102.50	104.50

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	427	0	237	7	0
2	B	436	0	237	10	0
3	C	1824	0	1784	57	0
4	C	1	0	0	0	0
5	A	1	0	0	0	0
5	C	2	0	0	0	0
All	All	2691	0	2258	67	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 14.

All (67) 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)
3:C:166:TYR:HD2	3:C:216:ILE:HD11	1.48	0.77
3:C:98:ARG:HH12	3:C:191:ARG:HH22	1.36	0.74
3:C:98:ARG:HH22	3:C:191:ARG:HH21	1.36	0.73
1:A:12:DT:H2'	1:A:13:DT:C6	2.26	0.71
3:C:224:LEU:HD12	3:C:242:PHE:HB3	1.72	0.71
3:C:94:VAL:HB	3:C:114:TRP:CZ2	2.26	0.69
3:C:36:LYS:HZ1	3:C:164:ARG:HD2	1.57	0.69
2:B:25:DG:H2"	2:B:26:DA:C8	2.29	0.67
3:C:200:ARG:HG2	3:C:202:PRO:HD2	1.76	0.67
3:C:29:TYR:OH	3:C:36:LYS:NZ	2.31	0.64
3:C:166:TYR:CE2	3:C:220:THR:HG21	2.33	0.63
3:C:135:LEU:HB3	3:C:139:PHE:HE2	1.67	0.60
3:C:36:LYS:NZ	3:C:164:ARG:HD2	2.17	0.59
2:B:37:DT:OP2	3:C:197:SER:HB3	2.03	0.58
3:C:42:PHE:CE1	3:C:129:VAL:HG21	2.38	0.58
3:C:98:ARG:HH12	3:C:191:ARG:NH2	2.03	0.57
3:C:166:TYR:CD2	3:C:216:ILE:HD11	2.35	0.55
3:C:174:LYS:HG2	3:C:178:SER:O	2.07	0.54
1:A:6:DG:O6	3:C:206:ARG:NH1	2.39	0.54
3:C:92:LYS:O	3:C:114:TRP:HD1	1.92	0.53
1:A:5:DT:H3	2:B:39:DA:H61	1.56	0.52
3:C:188:ILE:HG23	3:C:192:TYR:CD2	2.45	0.51
3:C:119:ALA:HB2	3:C:129:VAL:HG12	1.93	0.51
3:C:28:ALA:HB3	3:C:95:VAL:HB	1.92	0.51
3:C:189:ILE:HD13	3:C:199[B]:GLN:HG2	1.92	0.51
3:C:121:SER:OG	3:C:123:SER:O	2.31	0.49
3:C:209:LEU:O	3:C:213:VAL:HG22	2.14	0.48
3:C:200:ARG:CG	3:C:202:PRO:HD2	2.44	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:20:GLN:HG3	3:C:24:LEU:HB3	1.97	0.47
3:C:42:PHE:O	3:C:46:ILE:HG12	2.14	0.47
3:C:79:SER:O	3:C:83:ARG:N	2.45	0.46
3:C:86:LEU:HD22	3:C:119:ALA:HB3	1.96	0.46
3:C:20:GLN:HG3	3:C:24:LEU:HD23	1.97	0.46
3:C:172:TYR:CE2	3:C:181:VAL:HG21	2.50	0.46
3:C:98:ARG:HH22	3:C:191:ARG:NH2	2.07	0.46
3:C:70:ILE:HG22	3:C:71:PHE:CD1	2.51	0.45
3:C:216:ILE:O	3:C:220:THR:HB	2.17	0.45
1:A:1:DC:C4	2:B:43:DG:O6	2.70	0.45
2:B:39:DA:H62	3:C:200:ARG:HH22	1.65	0.45
1:A:1:DC:N3	2:B:43:DG:C6	2.85	0.44
3:C:174:LYS:HE3	3:C:179:GLY:HA2	2.00	0.44
3:C:48:LYS:HB3	3:C:48:LYS:HE3	1.75	0.44
3:C:168:SER:O	3:C:171:GLN:N	2.50	0.44
2:B:30:DC:H2'	2:B:31:DA:C8	2.53	0.44
3:C:160:PRO:O	3:C:163:MET:HB2	2.18	0.44
3:C:220:THR:HG23	3:C:221:PRO:HD2	2.00	0.44
3:C:172:TYR:CD2	3:C:181:VAL:HG21	2.54	0.43
1:A:12:DT:H2'	1:A:13:DT:H6	1.78	0.43
3:C:98:ARG:HA	3:C:107:LYS:C	2.38	0.43
1:A:8:DC:H2"	1:A:9:DA:OP2	2.18	0.43
3:C:180:ILE:HD12	3:C:241:SER:HB3	2.00	0.43
2:B:39:DA:N6	3:C:200:ARG:HH22	2.17	0.42
2:B:29:DG:H2'	2:B:30:DC:C6	2.54	0.42
3:C:39:LEU:O	3:C:43:VAL:HG12	2.19	0.42
3:C:42:PHE:CE1	3:C:62:ILE:HD11	2.54	0.42
2:B:29:DG:H2'	2:B:30:DC:H6	1.84	0.42
3:C:194:LEU:HB3	3:C:195:PRO:HD2	2.01	0.42
3:C:135:LEU:HD23	3:C:135:LEU:HA	1.78	0.41
3:C:220:THR:CG2	3:C:221:PRO:HD2	2.50	0.41
3:C:42:PHE:HD1	3:C:60:CYS:HB2	1.86	0.41
3:C:151:LEU:HD23	3:C:151:LEU:O	2.20	0.41
3:C:38:MET:HB3	3:C:42:PHE:CE2	2.56	0.41
3:C:18:ILE:HD11	3:C:151:LEU:HD12	2.03	0.41
3:C:196:GLN:H	3:C:196:GLN:HG2	1.64	0.41
3:C:205:ARG:HB3	3:C:210:GLN:HE21	1.85	0.41
3:C:126:LEU:HG	3:C:127:TYR:N	2.35	0.40
3:C:202:PRO:HA	3:C:205:ARG:HG2	2.03	0.40

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
3	C	219/263 (83%)	200 (91%)	19 (9%)	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
3	C	195/236 (83%)	195 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

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

Mol	Chain	Res	Type
3	C	210	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 (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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 (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

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	22/22 (100%)	-1.07	0	100	100	89, 116, 141, 203	0
2	B	22/22 (100%)	-0.98	0	100	100	89, 127, 151, 200	0
3	C	224/263 (85%)	-0.37	1 (0%)	92	86	57, 97, 170, 264	0
All	All	268/307 (87%)	-0.47	1 (0%)	92	86	57, 101, 170, 264	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	13	LYS	2.5

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, 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
4	MG	C	301	1/1	0.98	0.09	30,30,30,30	0

6.5 Other polymers (i)

There are no such residues in this entry.