



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 4, 2026 – 07:19 PM UTC

PDB ID : 3MKY / pdb_00003mkky
Title : Structure of SopB(155-323)-18mer DNA complex, I23 form
Authors : Schumacher, M.A.; Piro, K.; Xu, W.
Deposited on : 2010-04-15
Resolution : 2.86 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

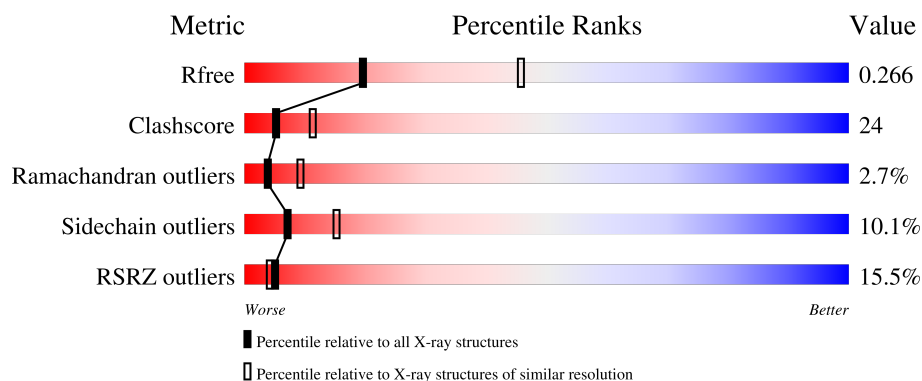
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.86 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}	180053	1407 (2.88-2.84)
Clashscore	190562	1446 (2.88-2.84)
Ramachandran outliers	187476	1406 (2.88-2.84)
Sidechain outliers	187428	1407 (2.88-2.84)
RSRZ outliers	180081	1408 (2.88-2.84)

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 $\leq 5\%$. 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	T	18	6% 33% 50% 17%
1	U	18	33% 67%
2	B	189	14% 33% 23% 5% 39%
2	P	189	7% 32% 23% 5% 40%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2531 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*TP*GP*GP*GP*AP*CP*CP*AP*TP*GP*GP*TP*CP*CP*CP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	U	18	Total	C	N	O	P	0	0	0
			366	174	69	106	17			
1	T	18	Total	C	N	O	P	0	0	0
			366	174	69	106	17			

- Molecule 2 is a protein called Protein sopB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	115	Total	C	N	O	S	0	0	0
			877	546	157	173	1			
2	P	114	Total	C	N	O	S	0	0	0
			872	543	156	172	1			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	135	MET	-	expression tag	UNP P62558
B	136	GLY	-	expression tag	UNP P62558
B	137	SER	-	expression tag	UNP P62558
B	138	SER	-	expression tag	UNP P62558
B	139	HIS	-	expression tag	UNP P62558
B	140	HIS	-	expression tag	UNP P62558
B	141	HIS	-	expression tag	UNP P62558
B	142	HIS	-	expression tag	UNP P62558
B	143	HIS	-	expression tag	UNP P62558
B	144	HIS	-	expression tag	UNP P62558
B	145	SER	-	expression tag	UNP P62558
B	146	SER	-	expression tag	UNP P62558
B	147	GLY	-	expression tag	UNP P62558
B	148	LEU	-	expression tag	UNP P62558
B	149	VAL	-	expression tag	UNP P62558

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	150	PRO	-	expression tag	UNP P62558
B	151	ARG	-	expression tag	UNP P62558
B	152	GLY	-	expression tag	UNP P62558
B	153	SER	-	expression tag	UNP P62558
B	154	HIS	-	expression tag	UNP P62558
B	255	ASP	GLU	conflict	UNP P62558
P	135	MET	-	expression tag	UNP P62558
P	136	GLY	-	expression tag	UNP P62558
P	137	SER	-	expression tag	UNP P62558
P	138	SER	-	expression tag	UNP P62558
P	139	HIS	-	expression tag	UNP P62558
P	140	HIS	-	expression tag	UNP P62558
P	141	HIS	-	expression tag	UNP P62558
P	142	HIS	-	expression tag	UNP P62558
P	143	HIS	-	expression tag	UNP P62558
P	144	HIS	-	expression tag	UNP P62558
P	145	SER	-	expression tag	UNP P62558
P	146	SER	-	expression tag	UNP P62558
P	147	GLY	-	expression tag	UNP P62558
P	148	LEU	-	expression tag	UNP P62558
P	149	VAL	-	expression tag	UNP P62558
P	150	PRO	-	expression tag	UNP P62558
P	151	ARG	-	expression tag	UNP P62558
P	152	GLY	-	expression tag	UNP P62558
P	153	SER	-	expression tag	UNP P62558
P	154	HIS	-	expression tag	UNP P62558
P	255	ASP	GLU	conflict	UNP P62558

- Molecule 3 is SULFATE ION (CCD ID: SO4) (formula: O₄S).

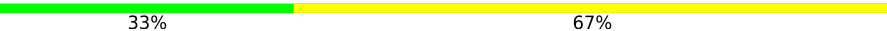


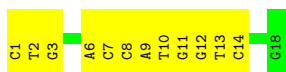
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	P	1	Total	O	S	0	0
			5	4	1		
3	P	1	Total	O	S	0	0
			5	4	1		
3	P	1	Total	O	S	0	0
			5	4	1		
3	P	1	Total	O	S	0	0
			5	4	1		
3	P	1	Total	O	S	0	0
			5	4	1		

3 Residue-property plots

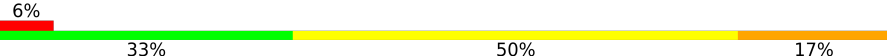
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

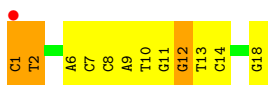
- Molecule 1: DNA (5'-D(*CP*TP*GP*GP*GP*AP*CP*CP*AP*TP*GP*GP*TP*CP*CP*CP*AP*G)-3')

Chain U: 

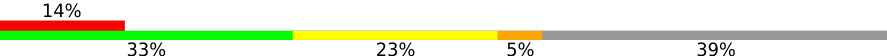


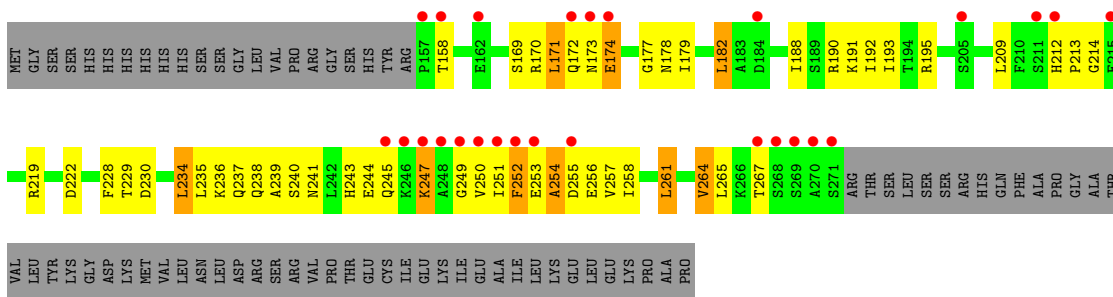
- Molecule 1: DNA (5'-D(*CP*TP*GP*GP*GP*AP*CP*CP*AP*TP*GP*GP*TP*CP*CP*CP*AP*G)-3')

Chain T: 



- Molecule 2: Protein sopB

Chain B: 



- Molecule 2: Protein sopB

Chain P: 



L209	F210	S211	H212	P213	G214	E215	G221	D222	Q225	F228	E233	L234	L235	Q238	A239	S240	N241	L242	H243	E244	K247	A248	G249	V250	I251	F252	E253	A254	D255	E256	V257	L261	T262	S263	V264	L265	S268	S269	A270	SER	ARG	THR	SER	SER	LEU	SER	SER	ARG	HIS	GLN	PHE	ALA
PRO	GLY	ALA	THR	VAL	LEU	TYR	LYS	GLY	ASP	LYS	MET	VAL	LEU	ASN	LEU	ASP	ARG	SER	ARG	VAL	PRO	THR	GLU	CYS	ILE	GLU	LYS	ILE	ALA	LEU	LYS	GLU	LEU	LYS	PRO	ALA	PRO	SER	ARG	THR	SER	SER	LEU	SER	SER	ARG	HIS	GLN	PHE	ALA		

4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, α , β , γ	200.78Å 200.78Å 200.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	63.49 – 2.86 63.49 – 2.86	Depositor EDS
% Data completeness (in resolution range)	98.0 (63.49-2.86) 98.1 (63.49-2.86)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.04	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.86 (at 2.86Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.242 , 0.266 0.242 , 0.266	Depositor DCC
R_{free} test set	3067 reflections (10.03%)	wwPDB-VP
Wilson B-factor (Å ²)	72.8	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 80.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.030 for -l,-k,-h	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	2531	wwPDB-VP
Average B, all atoms (Å ²)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.37% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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	T	0.34	0/410	0.83	0/631
1	U	0.34	0/410	0.82	0/631
2	B	0.42	0/887	0.89	2/1194 (0.2%)
2	P	0.42	0/882	0.98	4/1187 (0.3%)
All	All	0.40	0/2589	0.90	6/3643 (0.2%)

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	T	0	4
1	U	0	1
All	All	0	5

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	247	LYS	N-CA-C	-7.77	102.89	111.36
2	P	247	LYS	N-CA-C	-7.47	103.63	112.89
2	P	221	GLY	N-CA-C	-5.59	106.01	112.50
2	P	251	ILE	N-CA-C	5.30	116.04	106.61
2	B	172	GLN	N-CA-C	5.19	116.63	111.07

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	T	1	DC	Sidechain
1	T	12	DG	Sidechain
1	T	18	DG	Sidechain
1	T	2	DT	Sidechain
1	U	12	DG	Sidechain

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	T	366	0	203	25	0
1	U	366	0	203	16	0
2	B	877	0	892	43	0
2	P	872	0	890	42	0
3	B	20	0	0	0	0
3	P	30	0	0	2	0
All	All	2531	0	2188	114	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 24.

The worst 5 of 114 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:T:6:DA:H2''	1:T:7:DC:H5''	1.41	0.98
1:T:11:DG:H2''	1:T:12:DG:H5'	1.51	0.91
1:T:11:DG:H2''	1:T:12:DG:C5'	2.01	0.90
1:T:1:DC:H2'	1:T:2:DT:H71	1.58	0.85
1:T:6:DA:C2'	1:T:7:DC:H5''	2.05	0.85

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	
2	B	113/189 (60%)	98 (87%)	12 (11%)	3 (3%)	4	9
2	P	112/189 (59%)	97 (87%)	12 (11%)	3 (3%)	4	9
All	All	225/378 (60%)	195 (87%)	24 (11%)	6 (3%)	4	9

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	254	ALA
2	P	254	ALA
2	P	268	SER
2	B	174	GLU
2	P	174	GLU

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	
2	B	94/160 (59%)	86 (92%)	8 (8%)	10	22
2	P	94/160 (59%)	83 (88%)	11 (12%)	5	10
All	All	188/320 (59%)	169 (90%)	19 (10%)	7	15

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	P	234	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	P	261	LEU
2	P	265	LEU
2	P	252	PHE
2	P	174	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	237	GLN
2	B	238	GLN
2	P	187	ASN
2	B	173	ASN
2	B	172	GLN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

10 ligands are 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	$\# Z > 2$
3	SO4	B	324	-	4,4,4	0.40	0	6,6,6	0.07	0
3	SO4	P	945	-	4,4,4	0.39	0	6,6,6	0.10	0
3	SO4	B	936	-	4,4,4	0.40	0	6,6,6	0.12	0
3	SO4	P	846	-	4,4,4	0.36	0	6,6,6	0.21	0
3	SO4	B	835	-	4,4,4	0.38	0	6,6,6	0.09	0
3	SO4	P	844	-	4,4,4	0.40	0	6,6,6	0.07	0
3	SO4	P	849	-	4,4,4	0.40	0	6,6,6	0.11	0
3	SO4	P	845	-	4,4,4	0.37	0	6,6,6	0.17	0
3	SO4	P	344	-	4,4,4	0.37	0	6,6,6	0.09	0
3	SO4	B	836	-	4,4,4	0.38	0	6,6,6	0.13	0

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.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	P	844	SO4	1	0
3	P	344	SO4	1	0

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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	T	18/18 (100%)	0.53	1 (5%) 30 22	38, 49, 57, 59	0
1	U	18/18 (100%)	0.46	0 100 100	41, 53, 58, 58	0
2	B	115/189 (60%)	1.22	26 (22%) 2 2	45, 72, 99, 126	0
2	P	114/189 (60%)	1.02	14 (12%) 8 6	47, 71, 93, 112	0
All	All	265/414 (64%)	1.03	41 (15%) 5 4	38, 69, 96, 126	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	P	269	SER	8.0
2	B	271	SER	7.5
2	B	270	ALA	7.1
2	P	270	ALA	7.0
2	B	172	GLN	6.0

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 oligosaccharides 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, 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($\text{\AA}^2$)	Q<0.9
3	SO4	P	844	5/5	0.73	0.20	121,121,122,123	0
3	SO4	P	849	5/5	0.77	0.11	129,130,130,131	0
3	SO4	P	344	5/5	0.78	0.18	131,131,132,133	0
3	SO4	B	324	5/5	0.83	0.26	112,112,114,117	0
3	SO4	P	845	5/5	0.91	0.24	90,91,93,96	0
3	SO4	P	846	5/5	0.92	0.14	73,73,74,76	0
3	SO4	B	835	5/5	0.94	0.23	98,100,101,101	0
3	SO4	P	945	5/5	0.95	0.12	83,85,87,88	0
3	SO4	B	936	5/5	0.97	0.09	75,78,79,80	0
3	SO4	B	836	5/5	0.97	0.09	71,72,73,74	0

6.5 Other polymers [i](#)

There are no such residues in this entry.