wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 11:39 AM EDT

PDB ID	:	2GPY
Title	:	Crystal structure of putative O-methyltransferase from Bacillus halodurans
Authors	:	Ramagopal, U.A.; Patskovsky, Y.V.; Almo, S.C.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2006-04-18
Resolution	:	1.90 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 (i) 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 (i)) were used in the production of this report:

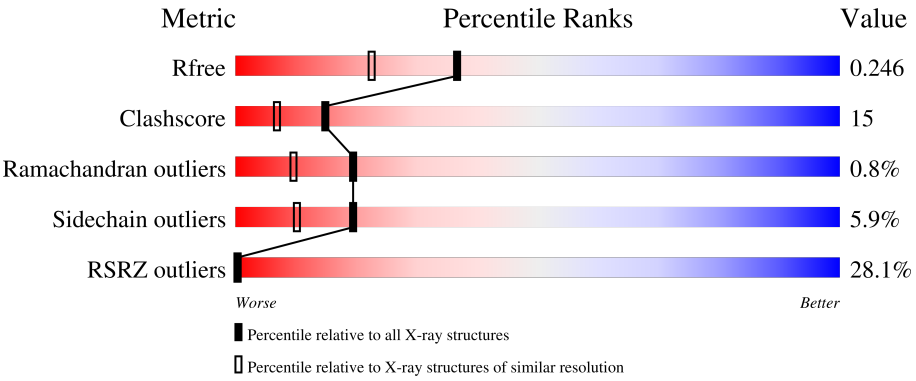
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	233	24% 59% 15% 5% 21%
1	B	233	20% 63% 18% • 18%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3365 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 O-methyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	185	Total	C	N	O	Se	0	10	0
			1593	1021	277	287	8			
1	B	192	Total	C	N	O	Se	0	2	0
			1583	1012	281	283	7			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	modified residue	GB 10173888
A	2	SER	-	cloning artifact	GB 10173888
A	3	LEU	-	cloning artifact	GB 10173888
A	26	MSE	MET	modified residue	GB 10173888
A	38	MSE	MET	modified residue	GB 10173888
A	43	MSE	MET	modified residue	GB 10173888
A	52	MSE	MET	modified residue	GB 10173888
A	72	MSE	MET	modified residue	GB 10173888
A	145	MSE	MET	modified residue	GB 10173888
A	149	MSE	MET	modified residue	GB 10173888
A	226	GLU	-	cloning artifact	GB 10173888
A	227	GLY	-	cloning artifact	GB 10173888
A	228	HIS	-	expression tag	GB 10173888
A	229	HIS	-	expression tag	GB 10173888
A	230	HIS	-	expression tag	GB 10173888
A	231	HIS	-	expression tag	GB 10173888
A	232	HIS	-	expression tag	GB 10173888
A	233	HIS	-	expression tag	GB 10173888
B	1	MSE	-	modified residue	GB 10173888
B	2	SER	-	cloning artifact	GB 10173888
B	3	LEU	-	cloning artifact	GB 10173888
B	26	MSE	MET	modified residue	GB 10173888
B	38	MSE	MET	modified residue	GB 10173888
B	43	MSE	MET	modified residue	GB 10173888
B	52	MSE	MET	modified residue	GB 10173888

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	72	MSE	MET	modified residue	GB 10173888
B	145	MSE	MET	modified residue	GB 10173888
B	149	MSE	MET	modified residue	GB 10173888
B	226	GLU	-	cloning artifact	GB 10173888
B	227	GLY	-	cloning artifact	GB 10173888
B	228	HIS	-	expression tag	GB 10173888
B	229	HIS	-	expression tag	GB 10173888
B	230	HIS	-	expression tag	GB 10173888
B	231	HIS	-	expression tag	GB 10173888
B	232	HIS	-	expression tag	GB 10173888
B	233	HIS	-	expression tag	GB 10173888

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	4	Total Zn 4 4	0	0
2	B	5	Total Zn 5 5	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Mg 1 1	0	0

- Molecule 4 is water.

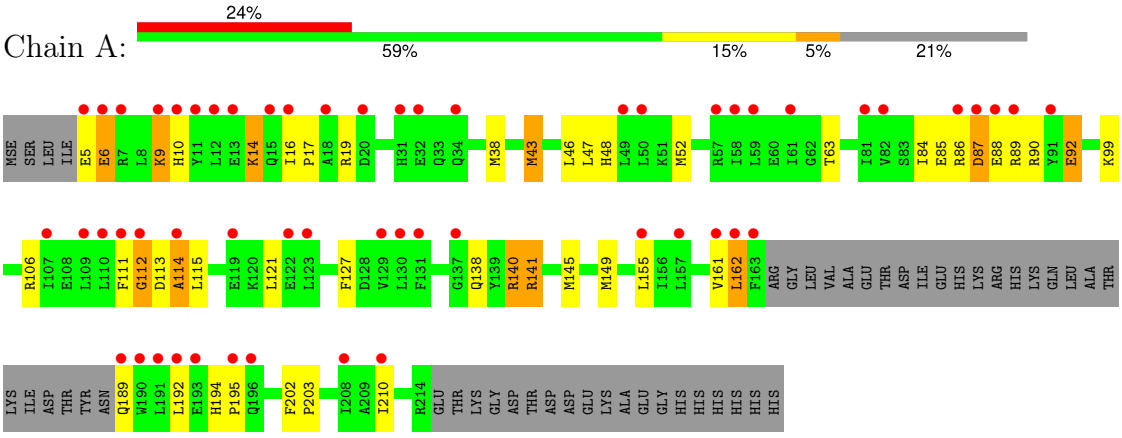
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	71	Total O 71 71	0	0
4	B	108	Total O 108 108	0	0

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.

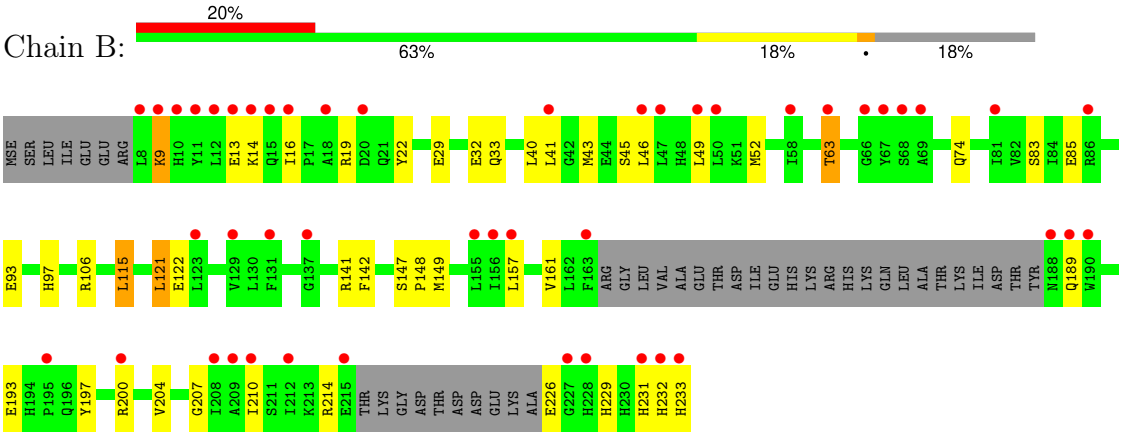
• Molecule 1: O-methyltransferase

Chain A:



• Molecule 1: O-methyltransferase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	50.57Å 62.81Å 137.75Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.44 – 1.90 34.44 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.7 (34.44-1.90) 99.7 (34.44-1.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.04	Depositor
$< I/\sigma(I) >$ [1]	3.41 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.206 , 0.248 0.205 , 0.246	Depositor DCC
R_{free} test set	1763 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	32.5	Xtriage
Anisotropy	0.400	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 55.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3365	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.65% 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, 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	0.72	1/1616 (0.1%)	0.75	2/2161 (0.1%)
1	B	0.69	1/1611 (0.1%)	0.73	2/2158 (0.1%)
All	All	0.70	2/3227 (0.1%)	0.74	4/4319 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	43	MSE	SE-CE	-13.14	1.18	1.95
1	B	43	MSE	SE-CE	-9.71	1.38	1.95

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	106	ARG	NE-CZ-NH2	-7.62	116.49	120.30
1	B	106	ARG	NE-CZ-NH1	-6.80	116.90	120.30
1	A	43	MSE	CG-SE-CE	-6.69	84.19	98.90
1	B	121	LEU	CA-CB-CG	-6.14	101.17	115.30

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	1593	0	1610	62	0
1	B	1583	0	1573	43	0
2	A	4	0	0	0	0
2	B	5	0	0	0	0
3	B	1	0	0	0	0
4	A	71	0	0	2	0
4	B	108	0	0	3	0
All	All	3365	0	3183	97	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 15.

The worst 5 of 97 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:43:MSE:CE	1:A:43:MSE:SE	1.17	1.37
1:A:43:MSE:CE	1:A:43:MSE:CG	2.19	1.18
1:A:43:MSE:SE	1:A:43:MSE:HE1	1.76	1.04
1:A:43:MSE:SE	1:A:43:MSE:HE2	1.76	1.04
1:A:43:MSE:SE	1:A:43:MSE:HE3	1.76	1.03

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	191/233 (82%)	186 (97%)	3 (2%)	2 (1%)	15 6
1	B	188/233 (81%)	183 (97%)	4 (2%)	1 (0%)	29 18
All	All	379/466 (81%)	369 (97%)	7 (2%)	3 (1%)	19 9

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	112	GLY
1	B	232	HIS
1	A	114	ALA

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	168/192 (88%)	154 (92%)	14 (8%)	11 4
1	B	166/192 (86%)	160 (96%)	6 (4%)	35 26
All	All	334/384 (87%)	314 (94%)	20 (6%)	19 9

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	9	LYS
1	B	63	THR
1	B	121	LEU
1	B	115	LEU
1	A	87	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	25	GLN
1	A	138	GLN
1	B	34	GLN
1	B	229	HIS

5.3.3 RNA (i)

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 10 ligands modelled in this entry, 10 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	178/233 (76%)	1.66	55 (30%)	0	0	33, 42, 57, 64	0
1	B	185/233 (79%)	1.51	47 (25%)	0	0	31, 41, 52, 74	0
All	All	363/466 (77%)	1.58	102 (28%)	0	0	31, 41, 56, 74	0

The worst 5 of 102 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	12	LEU	11.6
1	B	163	PHE	8.3
1	B	8	LEU	8.0
1	B	11	TYR	7.1
1	A	111	PHE	7.1

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
2	ZN	B	310	1/1	0.82	0.14	55,55,55,55	1
2	ZN	B	305	1/1	0.95	0.11	38,38,38,38	1
2	ZN	B	309	1/1	0.95	0.12	45,45,45,45	1
2	ZN	A	307	1/1	0.95	0.19	46,46,46,46	1
3	MG	B	301	1/1	0.95	0.24	35,35,35,35	0
2	ZN	A	306	1/1	0.96	0.13	34,34,34,34	1
2	ZN	B	303	1/1	0.96	0.13	42,42,42,42	1
2	ZN	A	304	1/1	0.97	0.17	41,41,41,41	1
2	ZN	A	308	1/1	0.98	0.17	44,44,44,44	1
2	ZN	B	302	1/1	0.99	0.11	41,41,41,41	1

6.5 Other polymers (i)

There are no such residues in this entry.