



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 10, 2026 – 05:45 AM UTC

PDB ID : 5AYS / pdb_00005ays
Title : Crystal structure of SAUGI/HSV UDG complex
Authors : Wang, H.C.; Ko, T.P.; Huang, M.F.; Wang, A.H.J.
Deposited on : 2015-09-02
Resolution : 2.09 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
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

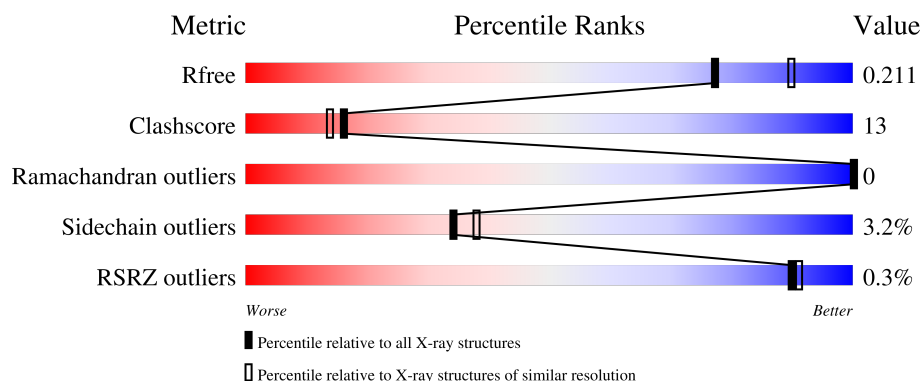
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.09 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	6658 (2.10-2.10)
Clashscore	190562	7164 (2.10-2.10)
Ramachandran outliers	187476	7099 (2.10-2.10)
Sidechain outliers	187428	7100 (2.10-2.10)
RSRZ outliers	180081	6662 (2.10-2.10)

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 $\leq 5\%$. 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	A	256	 70% 19% 10%
1	B	256	 70% 19% 10%
2	C	112	 68% 27% 5% 0%
2	D	112	 79% 13% 5% 3% 2%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6232 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 Uracil-DNA glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	230	Total	C	N	O	S	0	0	0
			1835	1170	344	312	9			
1	B	230	Total	C	N	O	S	0	0	0
			1835	1170	344	312	9			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	245	LEU	-	expression tag	UNP P10186
A	246	GLU	-	expression tag	UNP P10186
A	247	GLU	-	expression tag	UNP P10186
A	248	LEU	-	expression tag	UNP P10186
A	249	LEU	-	expression tag	UNP P10186
A	250	GLU	-	expression tag	UNP P10186
A	251	HIS	-	expression tag	UNP P10186
A	252	HIS	-	expression tag	UNP P10186
A	253	HIS	-	expression tag	UNP P10186
A	254	HIS	-	expression tag	UNP P10186
A	255	HIS	-	expression tag	UNP P10186
A	256	HIS	-	expression tag	UNP P10186
B	245	LEU	-	expression tag	UNP P10186
B	246	GLU	-	expression tag	UNP P10186
B	247	GLU	-	expression tag	UNP P10186
B	248	LEU	-	expression tag	UNP P10186
B	249	LEU	-	expression tag	UNP P10186
B	250	GLU	-	expression tag	UNP P10186
B	251	HIS	-	expression tag	UNP P10186
B	252	HIS	-	expression tag	UNP P10186
B	253	HIS	-	expression tag	UNP P10186
B	254	HIS	-	expression tag	UNP P10186
B	255	HIS	-	expression tag	UNP P10186
B	256	HIS	-	expression tag	UNP P10186

- Molecule 2 is a protein called Uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	109	Total	C	N	O	S	0	0	0
			925	603	141	177	4			
2	D	109	Total	C	N	O	S	0	0	0
			925	603	141	177	4			

- Molecule 3 is water.

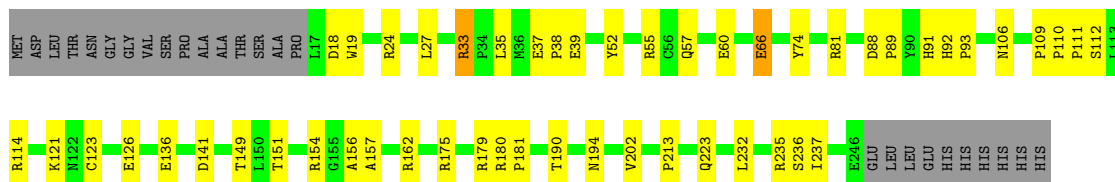
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	220	Total	O	0	0
			220	220		
3	C	129	Total	O	0	0
			129	129		
3	B	227	Total	O	0	0
			227	227		
3	D	136	Total	O	0	0
			136	136		

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.

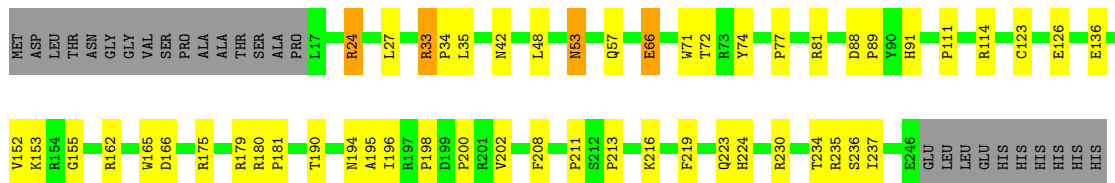
- Molecule 1: Uracil-DNA glycosylase

Chain A: 



- Molecule 1: Uracil-DNA glycosylase

Chain B: 



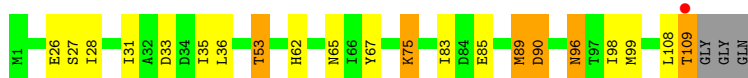
- Molecule 2: Uncharacterized protein

Chain C: 



- Molecule 2: Uncharacterized protein

Chain D: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	53.30Å 63.01Å 65.51Å 65.10° 89.58° 88.38°	Depositor
Resolution (Å)	20.00 – 2.09 20.00 – 2.09	Depositor EDS
% Data completeness (in resolution range)	94.4 (20.00-2.09) 94.9 (20.00-2.09)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.48 (at 2.09Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.195 , 0.223 (Not available) , 0.211	Depositor DCC
R_{free} test set	2191 reflections (4.77%)	wwPDB-VP
Wilson B-factor (Å ²)	22.6	Xtriage
Anisotropy	0.158	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 44.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.025 for h,-k,-l 0.000 for -h,l,k 0.000 for -h,-l,-k	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6232	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.78% 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](#)

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	0.61	0/1892	1.04	2/2584 (0.1%)
1	B	0.63	0/1892	1.04	4/2584 (0.2%)
2	C	0.55	0/948	0.99	2/1290 (0.2%)
2	D	0.61	0/948	1.03	4/1290 (0.3%)
All	All	0.61	0/5680	1.03	12/7748 (0.2%)

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	33	ARG	CA-C-N	-8.05	111.50	119.87
1	B	33	ARG	C-N-CA	-8.05	111.50	119.87
2	D	35	ILE	N-CA-C	7.55	117.59	111.62
2	C	35	ILE	N-CA-C	6.41	116.68	111.62
2	D	35	ILE	CB-CA-C	-5.94	105.77	111.05

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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	1835	0	1817	44	0
1	B	1835	0	1817	43	0
2	C	925	0	903	38	0
2	D	925	0	903	27	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	220	0	0	13	1
3	B	227	0	0	11	3
3	C	129	0	0	21	1
3	D	136	0	0	9	1
All	All	6232	0	5440	143	3

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 13.

The worst 5 of 143 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:C:13:ASN:HB3	3:C:290:HOH:O	1.34	1.25
2:D:85:GLU:HG3	3:D:209:HOH:O	1.61	0.97
1:B:230:ARG:NE	3:B:301:HOH:O	1.84	0.96
2:C:9:HIS:HB3	3:C:213:HOH:O	1.67	0.93
2:C:85:GLU:HG3	3:C:207:HOH:O	1.68	0.91

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:B:301:HOH:O	3:D:227:HOH:O[1_455]	1.84	0.36
3:A:375:HOH:O	3:B:356:HOH:O[1_565]	2.11	0.09
3:C:202:HOH:O	3:B:381:HOH:O[1_556]	2.18	0.02

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	228/256 (89%)	223 (98%)	5 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	228/256 (89%)	222 (97%)	6 (3%)	0	100	100
2	C	107/112 (96%)	101 (94%)	6 (6%)	0	100	100
2	D	107/112 (96%)	103 (96%)	4 (4%)	0	100	100
All	All	670/736 (91%)	649 (97%)	21 (3%)	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	196/219 (90%)	192 (98%)	4 (2%)	48	56
1	B	196/219 (90%)	192 (98%)	4 (2%)	48	56
2	C	104/105 (99%)	99 (95%)	5 (5%)	23	23
2	D	104/105 (99%)	98 (94%)	6 (6%)	18	16
All	All	600/648 (93%)	581 (97%)	19 (3%)	34	38

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

Mol	Chain	Res	Type
2	D	75	LYS
2	D	96	ASN
2	D	109	THR
2	D	90	ASP
2	C	109	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	6	GLN
2	D	9	HIS
2	D	96	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	62	HIS
1	B	91	HIS

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](#)

There are no ligands in this entry.

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	A	230/256 (89%)	-0.17	0	100 100	12, 23, 42, 54	0
1	B	230/256 (89%)	-0.15	0	100 100	13, 23, 41, 61	0
2	C	109/112 (97%)	-0.07	1 (0%)	81 83	15, 25, 40, 46	0
2	D	109/112 (97%)	-0.10	1 (0%)	81 83	14, 24, 39, 46	0
All	All	678/736 (92%)	-0.14	2 (0%)	90 91	12, 24, 41, 61	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	109	THR	2.7
2	C	109	THR	2.4

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.