wwPDB X-ray Structure Validation Summary Report (i)

Sep 8, 2025 – 02:25 pm BST

PDB ID	:	8S8T / pdb_00008s8t
Title	:	KOD-H4 DNA polymerase mutant in a binary complex with DNA/DNA
Authors	:	Gutfreund, C.; Betz, K.
Deposited on	:	2024-03-07
Resolution	:	3.00 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 (1)) were used in the production of this report:

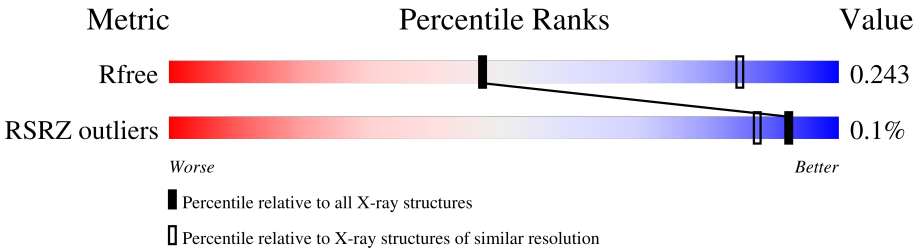
MolProbity	:	FAILED
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.45.1

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 3.00 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}	164625	2511 (3.00-3.00)
RSRZ outliers	164620	2523 (3.00-3.00)

MolProbity failed to run properly - the sequence quality summary graphics cannot be shown.

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 6818 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 DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	756	Total	C	N	O	S	0	0	0
			6200	3988	1054	1141	17			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	93	GLN	VAL	conflict	UNP D0VWU9
A	114	THR	ILE	conflict	UNP D0VWU9
A	141	ALA	ASP	conflict	UNP D0VWU9
A	143	ALA	GLU	conflict	UNP D0VWU9
A	147	HIS	GLU	conflict	UNP D0VWU9
A	383	LYS	SER	conflict	UNP D0VWU9
A	429	GLY	LYS	conflict	UNP D0VWU9
A	445	LEU	PHE	conflict	UNP D0VWU9
A	485	LEU	ALA	conflict	UNP D0VWU9
A	493	VAL	TYR	conflict	UNP D0VWU9
A	496	HIS	TYR	conflict	UNP D0VWU9
A	497	MET	TYR	conflict	UNP D0VWU9
A	499	PHE	TYR	conflict	UNP D0VWU9
A	500	GLU	ALA	conflict	UNP D0VWU9
A	501	ASN	ARG	conflict	UNP D0VWU9
A	521	LEU	ILE	conflict	UNP D0VWU9
A	584	LYS	GLU	conflict	UNP D0VWU9
A	664	LYS	GLU	conflict	UNP D0VWU9
A	726	ARG	LYS	conflict	UNP D0VWU9
A	735	LYS	ASN	conflict	UNP D0VWU9

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	13	Total	C	N	O	P	0	0	0
			261	124	53	72	12			

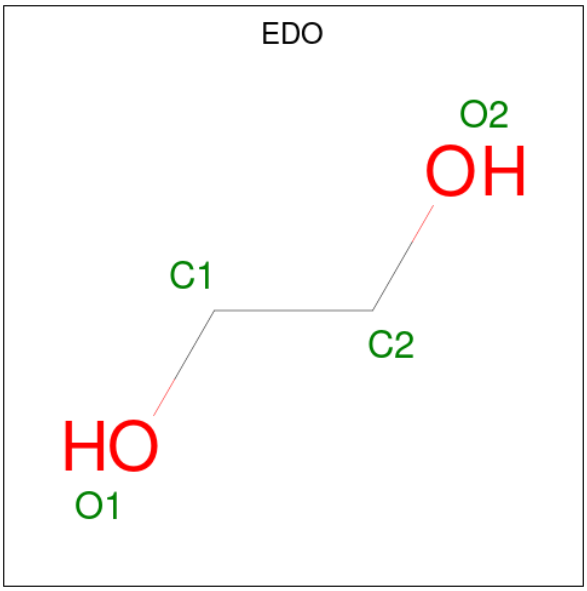
- Molecule 3 is a DNA chain called DNA (5'-D(*AP*CP*TP*GP*TP*GP*GP*CP*CP*GP*TP*GP*GP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	T	15	Total	C	N	O	P	0	0	0
			306	146	55	91	14			

- Molecule 4 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

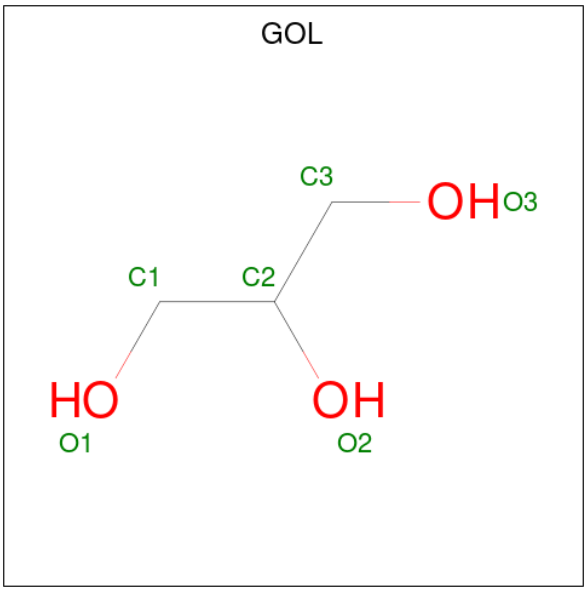
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	5	Total	Mg	0	0
			5	5		

- Molecule 5 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 6 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	P	1	Total C O 6 3 3	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	30	Total O 30 30	0	0
7	P	1	Total O 1 1	0	0
7	T	5	Total O 5 5	0	0

MolProbity failed to run properly - this section is therefore empty.

3 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	111.55Å 149.30Å 66.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.45 – 3.00 45.45 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.1 (45.45-3.00) 99.5 (45.45-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.08 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.21.2_5419	Depositor
R, R_{free}	0.185 , 0.242 0.186 , 0.243	Depositor DCC
R_{free} test set	859 reflections (1.62%)	wwPDB-VP
Wilson B-factor (Å^2)	55.3	Xtriage
Anisotropy	0.933	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 29.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6818	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.03% 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.

4 Model quality

4.1 Standard geometry

MolProbity failed to run properly - this section is therefore empty.

4.2 Too-close contacts

MolProbity failed to run properly - this section is therefore empty.

4.3 Torsion angles

4.3.1 Protein backbone

MolProbity failed to run properly - this section is therefore empty.

4.3.2 Protein sidechains

MolProbity failed to run properly - this section is therefore empty.

4.3.3 RNA

MolProbity failed to run properly - this section is therefore empty.

4.4 Non-standard residues in protein, DNA, RNA chains

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

4.5 Carbohydrates

There are no oligosaccharides in this entry.

4.6 Ligand geometry

Of 7 ligands modelled in this entry, 5 are monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
6	GOL	P	101	-	5,5,5	0.35	0	5,5,5	0.47	0
5	EDO	A	806	-	3,3,3	0.26	0	2,2,2	0.36	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	GOL	P	101	-	-	2/4/4/4	-
5	EDO	A	806	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	P	101	GOL	O1-C1-C2-O2
6	P	101	GOL	O1-C1-C2-C3

There are no ring outliers.

No monomer is involved in short contacts.

4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data

5.1 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	756/774 (97%)	-0.44	1 (0%) 92 88	35, 58, 90, 147	0
2	P	13/13 (100%)	-0.42	0 100 100	50, 57, 82, 82	0
3	T	15/16 (93%)	-0.59	0 100 100	45, 58, 81, 110	0
All	All	784/803 (97%)	-0.45	1 (0%) 92 88	35, 58, 89, 147	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	51	VAL	2.9

5.2 Non-standard residues in protein, DNA, RNA chains

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

5.3 Carbohydrates

There are no oligosaccharides in this entry.

5.4 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	A	804	1/1	0.80	0.17	54,54,54,54	0
4	MG	A	802	1/1	0.81	0.18	50,50,50,50	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	MG	A	805	1/1	0.83	0.25	62,62,62,62	0
6	GOL	P	101	6/6	0.84	0.18	45,58,70,79	0
5	EDO	A	806	4/4	0.85	0.13	46,51,54,54	0
4	MG	A	801	1/1	0.91	0.11	37,37,37,37	0
4	MG	A	803	1/1	0.94	0.12	50,50,50,50	0

5.5 Other polymers (i)

There are no such residues in this entry.