wwPDB X-ray Structure Validation Summary Report (i)

Oct 14, 2023 – 08:58 PM EDT

PDB ID	:	8DW1
Title	:	Crystal structure of a host-guest complex with 5'-CTTAGTTATAACTAAG -3'
Authors	:	Georgiadis, M.M.
Deposited on	:	2022-07-30
Resolution	:	1.85 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:

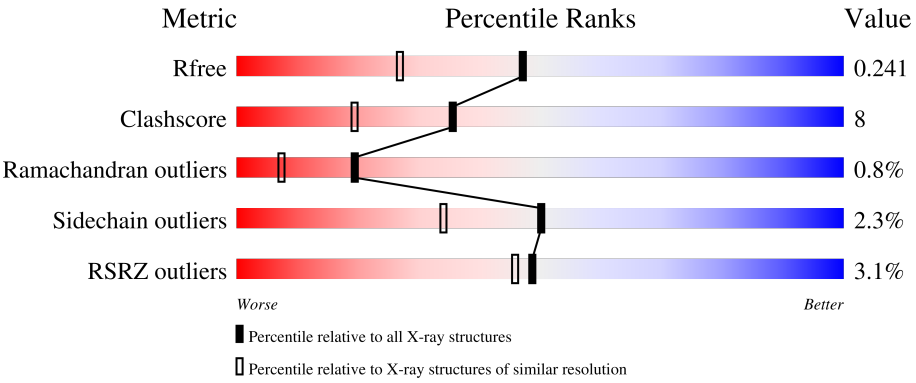
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.85 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	4003 (1.86-1.82)
Clashscore	141614	4233 (1.86-1.82)
Ramachandran outliers	138981	4185 (1.86-1.82)
Sidechain outliers	138945	4186 (1.86-1.82)
RSRZ outliers	127900	3957 (1.86-1.82)

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	266	3% 77% 13% 9%
2	D	8	38% 62%
3	G	8	12% 75% 12%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2502 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 reverse transcriptase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	242	Total	C	N	O	S	0	4	0
			1966	1269	344	347	6			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	GLY	-	expression tag	UNP Q8UN00
A	21	SER	-	expression tag	UNP Q8UN00
A	22	HIS	-	expression tag	UNP Q8UN00
A	23	MET	-	expression tag	UNP Q8UN00
A	249	ASN	ASP	conflict	UNP Q8UN00
A	279	LEU	-	expression tag	UNP Q8UN00
A	280	THR	-	expression tag	UNP Q8UN00
A	281	ARG	-	expression tag	UNP Q8UN00
A	282	GLY	-	expression tag	UNP Q8UN00
A	283	SER	-	expression tag	UNP Q8UN00
A	284	GLY	-	expression tag	UNP Q8UN00
A	285	CYS	-	expression tag	UNP Q8UN00

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	8	Total	C	N	O	P	0	0	0
			160	79	26	48	7			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	8	Total	C	N	O	P	0	0	0
			165	79	32	46	8			

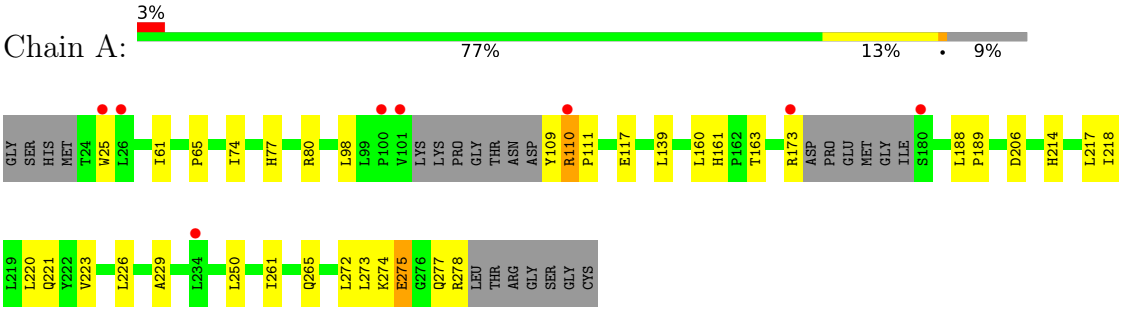
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	199	Total O 201 201	0	2
4	D	4	Total O 4 4	0	0
4	G	6	Total O 6 6	0	0

3 Residue-property plots (i)

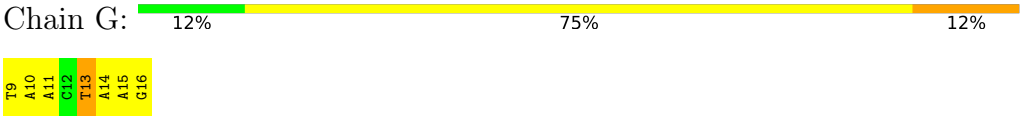
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density.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



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

• Molecule 3: DNA (5'-D(P*TP*AP*AP*CP*TP*AP*AP*G)-3')



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	55.13Å 145.79Å 46.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.97 – 1.85 43.97 – 1.85	Depositor EDS
% Data completeness (in resolution range)	95.0 (43.97-1.85) 98.7 (43.97-1.85)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.55 (at 1.84Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, R_{free}	0.200 , 0.238 0.199 , 0.241	Depositor DCC
R_{free} test set	1658 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	24.0	Xtriage
Anisotropy	0.232	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 55.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2502	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.01% 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

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	0.32	0/2030	0.53	0/2766
2	D	0.55	0/178	1.28	0/273
3	G	0.55	0/185	1.30	1/283 (0.4%)
All	All	0.37	0/2393	0.71	1/3322 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	G	13	DT	N3-C4-O4	5.30	123.08	119.90

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	1966	0	1998	22	0
2	D	160	0	94	6	1
3	G	165	0	91	9	1
4	A	201	0	0	1	0
4	D	4	0	0	0	0
4	G	6	0	0	1	0
All	All	2502	0	2183	36	1

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 8.

The worst 5 of 36 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:161:HIS:HD2	1:A:163[B]:THR:H	1.18	0.92
1:A:161:HIS:HD2	1:A:163[A]:THR:H	1.18	0.92
3:G:10:DA:H1'	3:G:11:DA:H5'	1.65	0.78
1:A:61:ILE:HD11	1:A:117:GLU:HG3	1.78	0.66
3:G:15:DA:H1'	4:G:103:HOH:O	1.95	0.66

All (1) 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)
2:D:8:DA:O3'	3:G:9:DT:P[2_655]	1.60	0.60

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	240/266 (90%)	236 (98%)	2 (1%)	2 (1%)	19 7

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	110	ARG
1	A	223	VAL

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	217/232 (94%)	212 (98%)	5 (2%)	50 34

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	98	LEU
1	A	139	LEU
1	A	220	LEU
1	A	273	LEU
1	A	275	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	263	GLN
1	A	277	GLN
1	A	144	GLN
1	A	161	HIS
1	A	213	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 monosaccharides 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

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	242/266 (90%)	0.09	8 (3%) 46 43	10, 26, 69, 87	0
2	D	8/8 (100%)	0.12	0 100 100	36, 50, 59, 59	0
3	G	8/8 (100%)	0.51	0 100 100	25, 61, 70, 73	0
All	All	258/282 (91%)	0.10	8 (3%) 49 46	10, 28, 69, 87	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	234	LEU	4.4
1	A	101	VAL	3.3
1	A	173	ARG	2.4
1	A	180	SER	2.3
1	A	25	TRP	2.0

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.