



# Preliminary Full wwPDB X-ray Structure Validation Report ⓘ

Apr 27, 2020 – 10:58 PM EDT

Deposition ID : D\_1000029360

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB Deposition System during initial deposition but before annotation of the structure.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

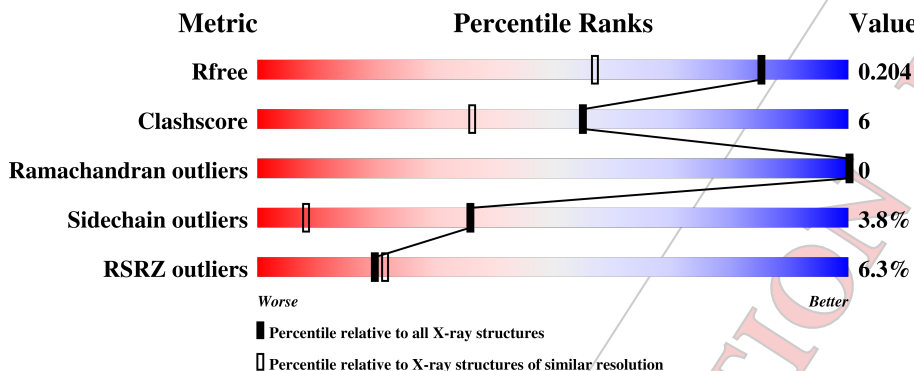
MolProbity	:	4.02b-467
Mogul	:	<b>FAILED</b>
Xtrriage (Phenix)	:	1.13
EDS	:	2.10.1
buster-report	:	<b>FAILED</b>
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.10.1

# 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 1.50 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))
$R_{free}$	111664	2534 (1.50-1.50)
Clashscore	122126	2727 (1.50-1.50)
Ramachandran outliers	120053	2661 (1.50-1.50)
Sidechain outliers	120020	2659 (1.50-1.50)
RSRZ outliers	108989	2481 (1.50-1.50)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	208	

## 2 Entry composition i

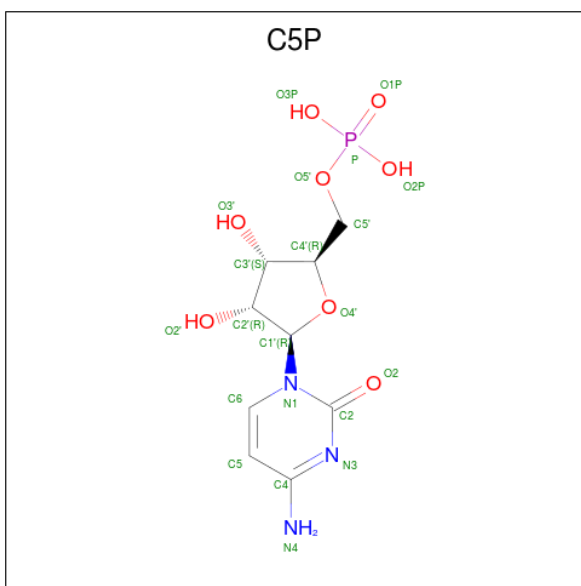
There are 3 unique types of molecules in this entry. The entry contains 1733 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Cytidylate kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	206	1577	990	295	289	3	0	0	0

- Molecule 2 is CYTIDINE-5'-MONOPHOSPHATE (three-letter code: C5P) (formula:  $C_9H_{14}N_3O_8P$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	21	9	3	8	1	0	0

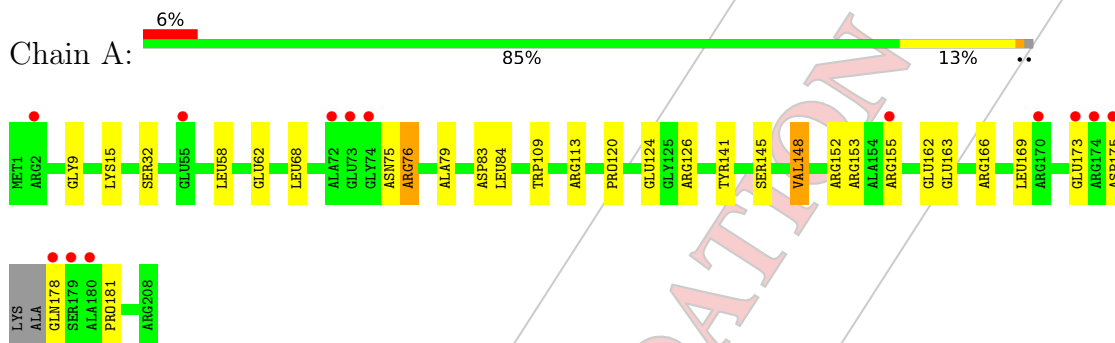
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	135	Total	O	0	0
			135	135		

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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: Cytidylate kinase



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.75Å 61.75Å 111.33Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 1.50 19.89 – 1.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (50.00-1.50) 99.9 (19.89-1.50)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.61 (at 1.50Å)	Xtrriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.206 , 0.229 0.208 , 0.204	Depositor DCC
$R_{free}$ test set	1920 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	14.2	Xtrriage
Anisotropy	0.040	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 43.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.053 for h,-h-k,-l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1733	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	19.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.80% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: C5P

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	0.80	0/1606	0.96	2/2180 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	175	ASP	N-CA-C	5.63	126.20	111.00
1	A	153	ARG	NE-CZ-NH1	5.32	122.96	120.30

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	1577	0	1597	18	0
2	A	21	0	12	0	0
3	A	135	0	0	0	0
All	All	1733	0	1609	18	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 6.

All (18) 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:58:LEU:O	1:A:62:GLU:HG3	1.86	0.74
1:A:163:GLU:HA	1:A:166:ARG:NH1	2.02	0.73
1:A:163:GLU:HA	1:A:166:ARG:HH12	1.54	0.71
1:A:145:SER:OG	1:A:148:VAL:HG13	1.98	0.63
1:A:76:ARG:HD3	1:A:83:ASP:OD2	1.99	0.62
1:A:109:TRP:HE1	1:A:113:ARG:HH21	1.47	0.62
1:A:166:ARG:HG3	1:A:166:ARG:HH11	1.63	0.62
1:A:32:SER:H	1:A:75:ASN:ND2	1.98	0.61
1:A:162:GLU:O	1:A:166:ARG:NH1	2.36	0.59
1:A:152:ARG:HA	1:A:155:ARG:HD3	1.90	0.54
1:A:141:TYR:CD1	1:A:181:PRO:HG3	2.48	0.49
1:A:169:LEU:O	1:A:173:GLU:HG3	2.13	0.49
1:A:68:LEU:HD11	1:A:75:ASN:HD22	1.79	0.47
1:A:162:GLU:CD	1:A:166:ARG:NH1	2.69	0.46
1:A:141:TYR:CG	1:A:181:PRO:HG3	2.52	0.43
1:A:163:GLU:CA	1:A:166:ARG:HH12	2.26	0.43
1:A:9:GLY:O	1:A:126:ARG:HD3	2.21	0.41
1:A:79:ALA:HB2	1:A:84:LEU:HD11	2.04	0.40

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	202/208 (97%)	200 (99%)	2 (1%)	0	<a href="#">100</a> <a href="#">100</a>

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	159/160 (99%)	153 (96%)	6 (4%)	36 8

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	15	LYS
1	A	76	ARG
1	A	120	PRO
1	A	124	GLU
1	A	148	VAL
1	A	178	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	75	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

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



## 5.6 Ligand geometry [i](#)

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

## 5.7 Other polymers [i](#)

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

## 5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

PRELIMINARY VALIDATION REPORT

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	206/208 (99%)	0.42	13 (6%) 20 21	10, 16, 31, 50	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	175	ASP	5.3
1	A	174	ARG	5.2
1	A	178	GLN	5.1
1	A	72	ALA	4.4
1	A	179	SER	4.0
1	A	73	GLU	4.0
1	A	180	ALA	3.2
1	A	2	ARG	3.1
1	A	173	GLU	3.1
1	A	155	ARG	2.6
1	A	170	ARG	2.4
1	A	74	GLY	2.3
1	A	55	GLU	2.1

### 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 carbohydrates 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	C5P	A	209	21/?	0.93	0.10	11,14,23,25	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.

PRELIMINARY VALIDATION REPORT