Full wwPDB X-ray Structure Validation Report (i)

Feb 19, 2016 – 10:38 PM GMT

PDB ID	:	5DBB
Title	:	Structure of human DNA polymerase beta Host-Guest complex with the dG base paired with a dA
Authors	:	Koag, M.C.; Lee, S.
Deposited on	:	2015-08-21
Resolution	:	2.25 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

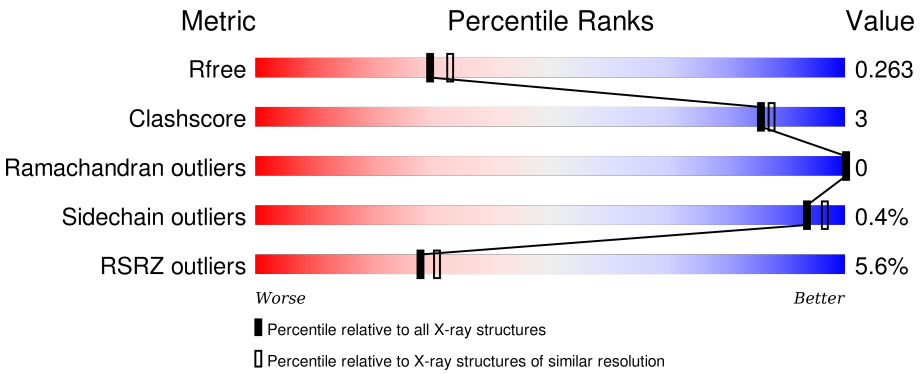
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 2.25 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}	91344	1640 (2.28-2.24)
Clashscore	102246	1095 (2.26-2.26)
Ramachandran outliers	100387	1063 (2.26-2.26)
Sidechain outliers	100360	1063 (2.26-2.26)
RSRZ outliers	91569	1647 (2.28-2.24)

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 >=3, 2, 1 and 0 types of geometric quality criteria. 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	335	6%, 90%, 8%, •
2	T	16	81%, 13%, 6%
3	P	10	80%, 20%
4	D	5	80%, 20%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3392 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 beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	327	2593	1642	454	488	9	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	16	323	154	62	92	15	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	P	10	205	98	40	58	9	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(P*GP*TP*CP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	5	106	49	20	32	5	0	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Na		
5	A	2	2	2	0	0

- Molecule 6 is water.

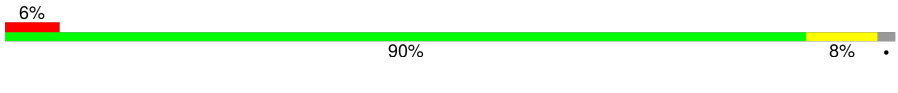
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	123	123	123	0	0
			Total	O		
6	T	16	16	16	0	0
			Total	O		
6	P	14	14	14	0	0
			Total	O		
6	D	10	10	10	0	0

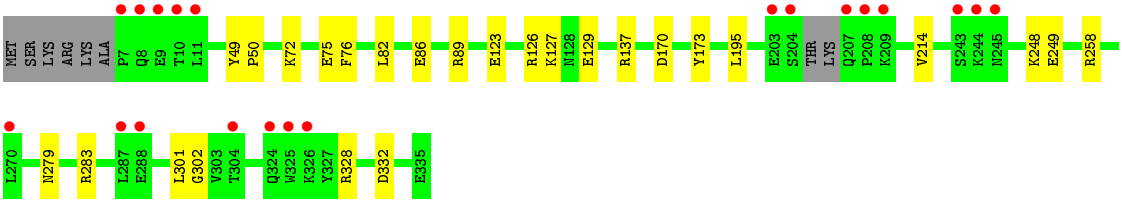
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: DNA polymerase beta

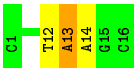
Chain A:





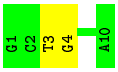
- Molecule 2: DNA (5'-D(*CP*CP*GP*AP*CP*GP*TP*CP*GP*CP*AP*TP*AP*AP*GP*C)-3')

Chain T:



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

Chain P:



- Molecule 4: DNA (5'-D(P*GP*TP*CP*GP*G)-3')

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	54.39Å 79.37Å 54.94Å 90.00° 105.45° 90.00°	Depositor
Resolution (Å)	19.84 – 2.25 19.84 – 2.25	Depositor EDS
% Data completeness (in resolution range)	96.5 (19.84-2.25) 96.2 (19.84-2.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.04 (at 2.26Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, R_{free}	0.206 , 0.265 0.205 , 0.263	Depositor DCC
R_{free} test set	1066 reflections (5.14%)	DCC
Wilson B-factor (Å^2)	28.8	Xtriage
Anisotropy	0.485	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 31.0	EDS
Estimated twinning fraction	0.023 for l,-k,h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 21446 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3392	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 6.71% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: NA

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	0.22	0/2642	0.41	1/3552 (0.0%)
2	T	0.62	0/362	0.95	2/556 (0.4%)
3	P	0.92	2/230 (0.9%)	0.87	0/354
4	D	1.09	1/118 (0.8%)	0.86	0/179
All	All	0.43	3/3352 (0.1%)	0.57	3/4641 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1	DG	OP3-P	-10.58	1.48	1.61
3	P	4	DG	O3'-P	-6.60	1.53	1.61
3	P	3	DT	O3'-P	-6.04	1.53	1.61

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	T	13	DA	P-O3'-C3'	6.96	128.05	119.70
2	T	12	DT	P-O3'-C3'	6.10	127.02	119.70
1	A	302	GLY	N-CA-C	-5.43	99.52	113.10

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	301	LEU	Peptide

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	2593	0	2595	15	0
2	T	323	0	180	2	0
3	P	205	0	114	0	0
4	D	106	0	57	0	0
5	A	2	0	0	0	0
6	A	123	0	0	3	0
6	D	10	0	0	0	0
6	P	14	0	0	0	0
6	T	16	0	0	1	0
All	All	3392	0	2946	17	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 3.

All (17) 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:72:LYS:O	6:A:501:HOH:O	1.95	0.84
1:A:76:PHE:N	6:A:501:HOH:O	2.30	0.64
2:T:13:DA:H1'	2:T:14:DA:H5'	1.81	0.62
2:T:14:DA:H2'	6:T:103:HOH:O	2.00	0.61
1:A:129:GLU:HG3	1:A:137:ARG:HE	1.69	0.56
1:A:123:GLU:OE1	1:A:123:GLU:N	2.31	0.53
1:A:49:TYR:HA	1:A:50:PRO:HD3	1.76	0.47
1:A:195:LEU:HD21	1:A:214:VAL:HG21	1.98	0.46
1:A:75:GLU:HB3	6:A:501:HOH:O	2.15	0.45
1:A:123:GLU:HG3	1:A:126:ARG:NH2	2.31	0.45
1:A:328:ARG:HD2	1:A:332:ASP:O	2.16	0.45
1:A:170:ASP:HB3	1:A:173:TYR:CD2	2.53	0.44
1:A:279:ASN:O	1:A:283:ARG:HG2	2.19	0.42
1:A:248:LYS:HD2	1:A:249:GLU:O	2.19	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:86:GLU:OE1	1:A:89:ARG:NH1	2.54	0.40
1:A:123:GLU:O	1:A:127:LYS:HG3	2.21	0.40
1:A:82:LEU:O	1:A:86:GLU:HG2	2.21	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	323/335 (96%)	312 (97%)	11 (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	280/295 (95%)	279 (100%)	1 (0%)	93 96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	258	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no

such sidechains identified.

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 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 (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	327/335 (97%)	0.30	20 (6%) 25 27	19, 31, 58, 89	0
2	T	16/16 (100%)	0.02	0 100 100	25, 34, 54, 56	0
3	P	10/10 (100%)	0.12	0 100 100	24, 34, 51, 55	0
4	D	5/5 (100%)	-0.35	0 100 100	25, 26, 32, 37	0
All	All	358/366 (97%)	0.27	20 (5%) 28 31	19, 31, 58, 89	0

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	7	PRO	9.3
1	A	10	THR	8.9
1	A	208	PRO	6.9
1	A	245	ASN	4.2
1	A	244	LYS	3.4
1	A	287	LEU	3.1
1	A	207	GLN	3.1
1	A	209	LYS	3.0
1	A	325	TRP	2.9
1	A	9	GLU	2.9
1	A	324	GLN	2.7
1	A	270	LEU	2.6
1	A	8	GLN	2.5
1	A	11	LEU	2.4
1	A	203	GLU	2.2
1	A	304	THR	2.1
1	A	326	LYS	2.1
1	A	243	SER	2.1
1	A	204	SER	2.1
1	A	288	GLU	2.0

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(Å^2)	Q<0.9
5	NA	A	402	1/1	0.88	0.12	-0.66	38,38,38,38	0
5	NA	A	401	1/1	0.93	0.12	-1.07	22,22,22,22	0

6.5 Other polymers (i)

There are no such residues in this entry.