



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 7, 2020 – 06:05 PM EDT

PDB ID : 6UIN
Title : Role of Beta-hairpin motifs in the DNA duplex opening by the Rad4/XPC nucleotide excision repair complex
Authors : Paul, D.; Min, J.-H.
Deposited on : 2019-10-01
Resolution : 3.35 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.14.6
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)	:	2.14.6

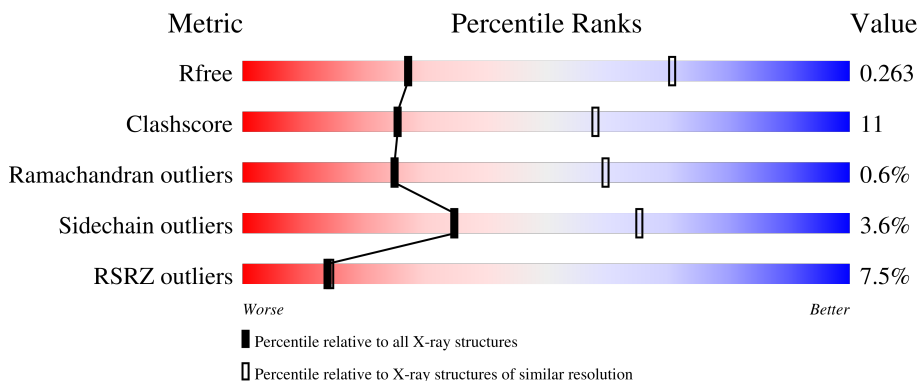
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.35 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	1060 (3.38-3.30)
Clashscore	141614	1111 (3.38-3.30)
Ramachandran outliers	138981	1090 (3.38-3.30)
Sidechain outliers	138945	1089 (3.38-3.30)
RSRZ outliers	127900	1028 (3.38-3.30)

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 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	538	<div> <div>7%</div> <div>65%</div> <div>23%</div> <div>11%</div> </div>
2	X	171	<div> <div>%</div> <div>24%</div> <div>7%</div> <div>68%</div> </div>
3	W	24	<div> <div>4%</div> <div>21%</div> <div>42%</div> <div>25%</div> <div>13%</div> </div>
4	Y	24	<div> <div>8%</div> <div>38%</div> <div>29%</div> <div>17%</div> <div>17%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10071 atoms, of which 4893 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 repair protein RAD4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	478	Total	C	H	N	O	S	0	0	0
			7930	2516	4002	697	688	27			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	GLY	-	expression tag	UNP P14736
A	96	SER	-	expression tag	UNP P14736
A	97	SER	-	expression tag	UNP P14736
A	98	ARG	-	expression tag	UNP P14736
A	99	ALA	-	expression tag	UNP P14736
A	100	MET	-	expression tag	UNP P14736
A	115	THR	LYS	conflict	UNP P14736
A	131	CYS	VAL	conflict	UNP P14736
A	132	SER	CYS	conflict	UNP P14736
A	223	GLU	VAL	conflict	UNP P14736

- Molecule 2 is a protein called UV excision repair protein RAD23.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	X	54	Total	C	H	N	O	S	0	0	0
			835	263	421	70	79	2			

There are 2 discrepancies between the modelled and reference sequences:

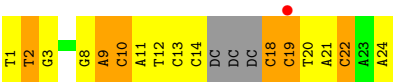
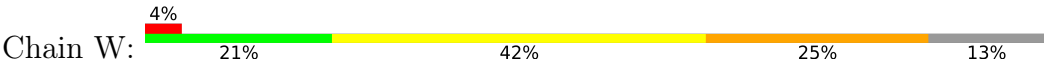
Chain	Residue	Modelled	Actual	Comment	Reference
X	228	GLY	-	expression tag	UNP P32628
X	229	SER	-	expression tag	UNP P32628

- Molecule 3 is a DNA chain called DNA (5'-D(*TP*TP*GP*AP*CP*TP*CP*(G47)P*AP*CP*AP*TP*CP*CP*CP*CP*CP*CP*CP*TP*AP*CP*AP*A)-3').

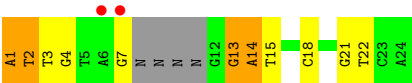
Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
3	W	21	Total	C	H	N	O	P	S	0	0	0
			664	204	242	74	123	20	1			

- Molecule 4 is a DNA chain called DNA (5'-D(*AP*TP*TP*GP*TP*AP*GP*NP*NP*NP*NP*GP*GP*AP*TP*GP*TP*CP*GP*AP*GP*TP*CP*A)-3').

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
4	Y	20	Total	C	H	N	O	P		0	0	0
			642	198	228	78	119	19				



● Molecule 4: DNA (5'-D(*AP*TP*TP*GP*TP*AP*GP*NP*NP*NP*NP*GP*GP*AP*TP*GP*TP*CP*GP*AP*GP*TP*CP*A)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α , β , γ	77.67Å 77.67Å 264.05Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	38.83 – 3.35 38.83 – 3.35	Depositor EDS
% Data completeness (in resolution range)	99.8 (38.83-3.35) 94.6 (38.83-3.35)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.04	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.63 (at 3.32Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.209 , 0.262 0.209 , 0.263	Depositor DCC
R_{free} test set	1287 reflections (9.95%)	wwPDB-VP
Wilson B-factor (Å ²)	119.1	Xtriage
Anisotropy	0.229	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 79.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.087 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	10071	wwPDB-VP
Average B, all atoms (Å ²)	148.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.17% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: G47

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	0.37	0/4015	0.82	4/5395 (0.1%)
2	X	0.40	0/420	0.78	0/572
3	W	1.24	2/441 (0.5%)	1.55	11/671 (1.6%)
4	Y	1.32	2/464 (0.4%)	1.47	5/714 (0.7%)
All	All	0.63	4/5340 (0.1%)	0.98	20/7352 (0.3%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	W	12	DT	C1'-N1	6.72	1.57	1.49
4	Y	1	DA	C3'-O3'	-6.50	1.35	1.44
4	Y	15	DT	P-O5'	-5.53	1.54	1.59
3	W	19	DC	P-O5'	5.33	1.65	1.59

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	Y	7	DG	O4'-C1'-N9	8.77	114.14	108.00
4	Y	15	DT	O5'-P-OP2	-8.27	98.26	105.70
1	A	584	ARG	NE-CZ-NH2	7.53	124.06	120.30
3	W	18	DC	P-O3'-C3'	7.43	128.62	119.70
1	A	619	ARG	NE-CZ-NH2	-7.29	116.66	120.30

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	3928	4002	3994	85	1
2	X	414	421	420	10	0
3	W	422	242	242	11	1
4	Y	414	228	229	7	1
All	All	5178	4893	4885	104	2

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 11.

The worst 5 of 104 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:414:ARG:NH1	1:A:419:ASP:OD1	1.95	0.99
2:X:261:ASP:OD2	2:X:286:ARG:NH2	2.16	0.79
2:X:266:ARG:NH2	2:X:307:GLU:O	2.16	0.78
1:A:378:ARG:NH2	1:A:421:GLU:OE2	2.21	0.73
3:W:9:DA:H2''	3:W:10:DC:H5''	1.73	0.70

All (2) 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:W:1:DT:O5'	3:W:24:DA:O3'[1_445]	1.89	0.31
1:A:170:LYS:NZ	4:Y:13:DG:OP1[1_455]	2.14	0.06

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	470/538 (87%)	449 (96%)	19 (4%)	2 (0%)	34	68
2	X	52/171 (30%)	47 (90%)	4 (8%)	1 (2%)	8	36
All	All	522/709 (74%)	496 (95%)	23 (4%)	3 (1%)	25	60

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	222	ASN
2	X	307	GLU
1	A	238	ASN

5.3.2 Protein sidechains ⓘ

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	428/480 (89%)	411 (96%)	17 (4%)	31	63
2	X	46/129 (36%)	46 (100%)	0	100	100
All	All	474/609 (78%)	457 (96%)	17 (4%)	35	65

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	375	VAL
1	A	377	ARG
1	A	507	SER
1	A	345	SER
1	A	510	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	184	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	340	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	$\# Z > 2$
3	G47	W	8	1,3,4	20,27,28	4.69	10 (50%)	21,38,41	1.84	5 (23%)

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
3	G47	W	8	1,3,4	-	4/7/25/26	0/3/3/3

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	W	8	G47	C4-N3	9.55	1.50	1.35
3	W	8	G47	O4'-C1'	8.08	1.60	1.42
3	W	8	G47	C2-N2	8.08	1.46	1.34
3	W	8	G47	O4'-C4'	-6.58	1.30	1.45
3	W	8	G47	C6-C5	6.47	1.52	1.41

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	W	8	G47	C2-N3-C4	4.18	120.02	115.28
3	W	8	G47	N3-C2-N1	-3.72	120.35	126.23
3	W	8	G47	C5-C6-N1	-2.59	119.89	123.43
3	W	8	G47	C6-N1-C2	2.42	119.51	115.18
3	W	8	G47	N2-C2-N3	2.09	120.81	117.19

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	W	8	G47	C7A-C6A-N2-C2
3	W	8	G47	N2-C6A-C7A-SG
3	W	8	G47	O4'-C4'-C5'-O5'
3	W	8	G47	C3'-C4'-C5'-O5'

There are no ring outliers.

No monomer is involved in short contacts.

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 [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	478/538 (88%)	0.52	39 (8%) 11 12	85, 120, 175, 219	0
2	X	54/171 (31%)	0.10	1 (1%) 66 66	110, 134, 166, 183	0
3	W	20/24 (83%)	0.43	1 (5%) 28 29	155, 184, 265, 266	0
4	Y	20/24 (83%)	0.69	2 (10%) 7 7	146, 181, 259, 268	0
All	All	572/757 (75%)	0.48	43 (7%) 14 14	85, 124, 186, 268	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	472	HIS	8.3
1	A	473	GLY	6.9
1	A	469	LEU	6.0
1	A	354	CYS	5.6
1	A	474	LYS	5.6

6.2 Non-standard residues in protein, DNA, RNA chains [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, 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	B-factors(Å ²)	Q<0.9
3	G47	W	8	25/26	0.83	0.31	171,186,224,233	0

6.3 Carbohydrates [i](#)

There are no monosaccharides 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.