wwPDB X-ray Structure Validation Summary Report (i)

Feb 10, 2024 – 07:05 PM EST

PDB ID	:	2QSG
Title	:	Crystal structure of Rad4-Rad23 bound to a UV-damaged DNA
Authors	:	Min, J.-H.; Pavletich, N.P.
Deposited on	:	2007-07-31
Resolution	:	3.10 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:

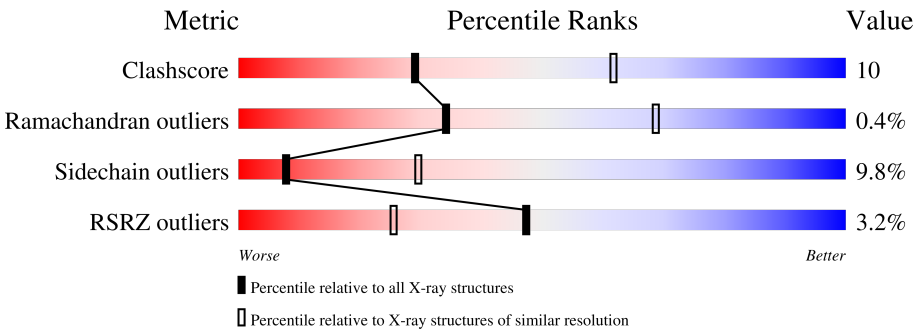
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.10 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))
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	W	24	54% 33% 12%
2	Y	24	4% 63% 21% 8% 8%
3	A	533	3% 69% 22% • 6%
4	X	171	% 22% 9% • 68%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5480 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 DNA chain called native strand of the CPD-mismatch DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	W	24	481	233	82	143	23	0	0	0

- Molecule 2 is a DNA chain called damaged strand of the CPD-mismatch DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	Y	22	451	218	82	131	20	0	0	0

- Molecule 3 is a protein called DNA repair protein RAD4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	502	4132	2633	744	728	27	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	100	MET	-	initiating methionine	UNP P14736
A	223	GLU	VAL	SEE REMARK 999	UNP P14736
A	225	LEU	ILE	SEE REMARK 999	UNP P14736

- Molecule 4 is a protein called UV excision repair protein RAD23.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	X	54	416	265	70	79	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

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

Continued on next page...

Continued from previous page...

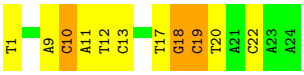
Chain	Residue	Modelled	Actual	Comment	Reference
X	229	SER	-	expression tag	UNP P32628

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.

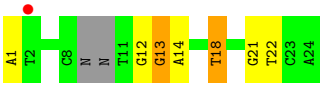
• Molecule 1: native strand of the CPD-mismatch DNA

Chain W:



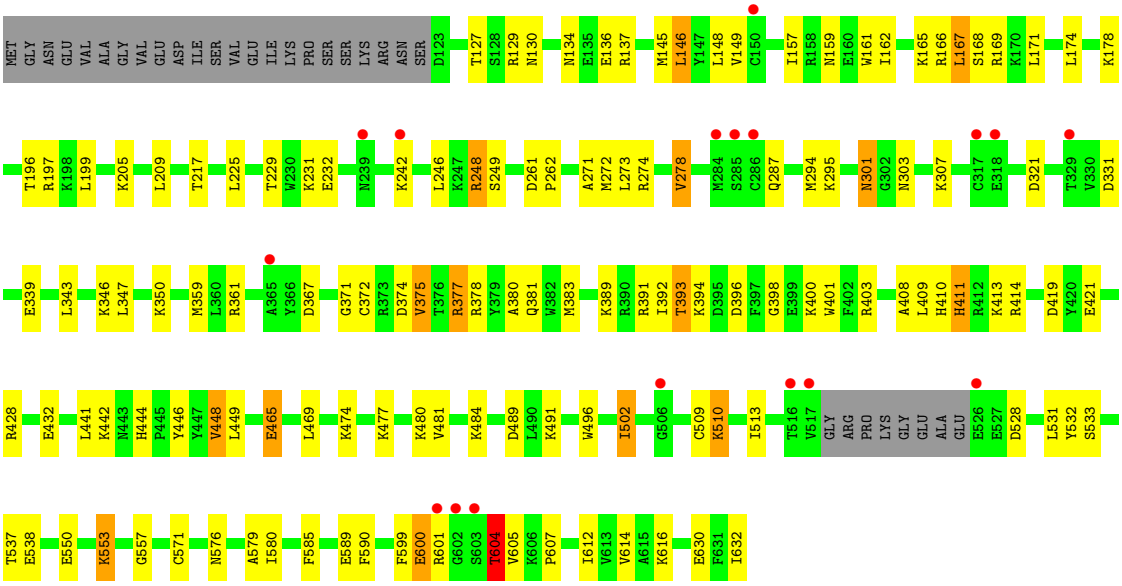
• Molecule 2: damaged strand of the CPD-mismatch DNA

Chain Y:



• Molecule 3: DNA repair protein RAD4

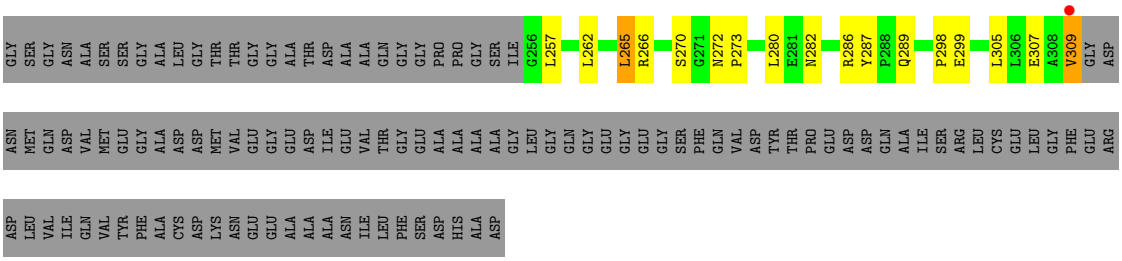
Chain A:



• Molecule 4: UV excision repair protein RAD23

Chain X:

%
22%
9%
68%


GLY SER GLY ASN ALA SER SER GLY ALA LEU GLY THR THR GLY GLY ALA THR ASP ALA ALA GLN GLY GLY PRO PRO GLY SER ILE G256 L257 L262 L265 R266 S270 G271 N272 P273 L280 E281 N282 R286 Y287 P288 Q289 P298 E299 L305 L306 E307 A308 V309 GLY ASP
ASN MET GLN ASP VAL MET GLU GLY ALA ASP ASP MET VAL GLU GLY GLU ASP ILE GLU VAL THR GLY GLU ALA ALA ALA ALA GLY LEU GLY GLN GLY GLU GLY GLU GLY SER PHE GLN VAL ASP TYR THR PRO GLU ASP ASP GLN ALA ILE SER ARG LEU CYS GLU LEU GLY PHE GLU ARG
ASP LEU VAL ILE GLN VAL TYR PHE ALA ALA CYS ASP ASP LYS ASN GLU GLU GLU ALA ALA ALA ALA ASN ILE LEU PHE SER ASP HIS ALA ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	79.64Å 79.64Å 403.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.10 29.11 – 3.05	Depositor EDS
% Data completeness (in resolution range)	93.4 (30.00-3.10) 92.6 (29.11-3.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.76 (at 3.05Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.246 , 0.277 0.266 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	83.0	Xtriage
Anisotropy	0.215	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 14.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	5480	wwPDB-VP
Average B, all atoms (Å^2)	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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 (i)

5.1 Standard geometry (i)

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	W	0.73	0/537	1.52	9/825 (1.1%)
2	Y	0.72	0/505	1.53	9/778 (1.2%)
3	A	0.38	0/4221	0.54	0/5673
4	X	0.37	0/422	0.55	0/575
All	All	0.46	0/5685	0.84	18/7851 (0.2%)

There are no bond length outliers.

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	Y	12	DG	O4'-C1'-N9	8.96	114.28	108.00
1	W	19	DC	O4'-C1'-N1	8.27	113.79	108.00
2	Y	13	DG	P-O3'-C3'	7.08	128.19	119.70
1	W	22	DC	O4'-C1'-N1	6.84	112.79	108.00
2	Y	21	DG	O4'-C1'-N9	6.83	112.78	108.00

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	W	481	0	274	8	0
2	Y	451	0	254	2	0
3	A	4132	0	4214	89	0
4	X	416	0	427	12	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	5480	0	5169	102	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 10.

The worst 5 of 102 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)
3:A:159:ASN:HD21	3:A:274:ARG:HH22	1.19	0.90
3:A:444:HIS:HD2	3:A:446:TYR:H	1.29	0.80
3:A:580:ILE:HA	3:A:612:ILE:HD11	1.66	0.78
3:A:599:PHE:O	3:A:600:GLU:HG2	1.85	0.76
1:W:18:DG:H4'	1:W:18:DG:OP1	1.89	0.72

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
3	A	498/533 (93%)	460 (92%)	36 (7%)	2 (0%)	34 69
4	X	52/171 (30%)	46 (88%)	6 (12%)	0	100 100
All	All	550/704 (78%)	506 (92%)	42 (8%)	2 (0%)	34 69

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	600	GLU
3	A	604	THR

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
3	A	452/477 (95%)	407 (90%)	45 (10%)	7 28
4	X	47/129 (36%)	43 (92%)	4 (8%)	10 37
All	All	499/606 (82%)	450 (90%)	49 (10%)	8 29

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	393	THR
3	A	465	GLU
3	A	403	ARG
3	A	441	LEU
3	A	489	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	303	ASN
3	A	410	HIS
4	X	272	ASN
3	A	444	HIS
4	X	267	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 (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	W	24/24 (100%)	-0.00	0 100 100	55, 70, 78, 99	0
2	Y	22/24 (91%)	0.09	1 (4%) 33 16	56, 72, 85, 107	0
3	A	502/533 (94%)	0.01	17 (3%) 45 24	58, 65, 80, 108	0
4	X	54/171 (31%)	-0.34	1 (1%) 66 46	58, 65, 77, 84	0
All	All	602/752 (80%)	-0.02	19 (3%) 47 25	55, 66, 81, 108	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	517	VAL	4.3
3	A	317	CYS	4.1
3	A	516	THR	4.0
3	A	526	GLU	3.2
3	A	602	GLY	2.8

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 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.