

wwPDB X-ray Structure Validation Summary Report (i)

Feb 10, 2024 – 10:14 AM EST

PDB ID	:	2QSH
Title	:	Crystal structure of Rad4-Rad23 bound to a mismatch DNA
Authors	:	Min, JH.; Pavletich, N.P.
Deposited on		
Resolution	:	2.81 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:

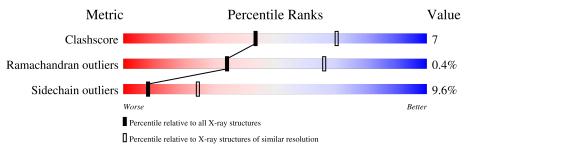
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: $X\text{-}RAY \, DIFFRACTION$

The reported resolution of this entry is 2.81 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	$egin{array}{llllllllllllllllllllllllllllllllllll$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)

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%

Mol	Chain	Length	Quality of chain				
1	W	24	67%		25%	8%	
2	Y	24	58%	21%	17%	•	
3	А	538	75%		17%	• 7%	
4	Х	171	22% 9% ·	68%			



2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5485 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 top strand of the mismatch DNA.

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

• Molecule 2 is a DNA chain called bottom strand of the mismatch DNA.

Ι	Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
	2	Y	23	Total 456	C 219	N 82	0 134	Р 21	0	0	0

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

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

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
А	95	GLY	-	expression tag	UNP P14736
А	96	SER	-	expression tag	UNP P14736
А	97	SER	-	expression tag	UNP P14736
А	98	ARG	-	expression tag	UNP P14736
А	99	ALA	-	expression tag	UNP P14736
А	100	MET	-	expression tag	UNP P14736
А	223	GLU	VAL	SEE REMARK 999	UNP P14736
А	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		
4	Х	54	Total 416	C 265	N 70	O 79	${ m S} { m 2}$	0	0	0



There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
Х	228	GLY	-	expression tag	UNP P32628
Х	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. 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. 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: top strand of the mismatch DNA

Chain W:	67%		25%	8%
T1 T2 G3 G3 A8 A11 C10 C10 C10 C22 C22 A23				
• Molecule 2: bott	om strand of the mism	atch DNA		
Chain Y:	58%	21%	17%	·
A1 75 711 711 712 612 613 714 715 718	021 122 038 1 24 1 24			
• Molecule 3: DNA	A repair protein RAD4			
Chain A:	75%		17%	• 7%
GLY SER SER ARG ALA MET MET MET ALA ALA ALA VAL CLU VAL	ASP ILE SER VAL VAL CU CU CU SER SER ASN ASN SER SER SER SER	M.30 1146 1146 1148 1148 1148 1148 1148 1148	S164 K165 R166 L167 L167 L174	K178 L182 P185 R197 K198
L199 1217 1217 1217 8231 8235 8238 8249 8249 8249 8249	K257 K266 Q267 Q267 Q267 Q267 K273 K274 K278 K278 K278 K278 Q287	P290 P290 P291 P293 P294 P295 P295 P293 P303 P303 P303 P303 P303 P303 P303 P3	u327 u327 L343 K350	M359 L360 R361 R372 R372 R377 R377 R377
Y 379 A 380 A 380 A 383 A 383 R 383 R 383 R 383 R 383 R 383 R 401 R 403 R 403	L409 H411 H411 R413 R413 R413 R414 P419 P419 P419 P429 C433 C433 C433 C433 C433	P435 P436 D436 P445 P445 Y446 Y446 Y446 Y446 Y446 Y446 Y446 Y	L469 V478 L479 K480	R485 K491 R494 Q495 W496 K511
W512 1513 V517 V517 V517 CLY CLY CLY CLY CLY CLY CLU CLU CLU	A645 A645 E550 K653 K653 K653 F559 F559 F559 F550 F550 F550 F550 F550	1612 V613 V614 F631 F631 1632		
• Molecule 4: UV e	excision repair protein	RAD23		
Chain X: 22%	9% •	68%		
GLY SER GLY ASN ASN ALA SER SER ALA ALA LEU CLY THR THR THR CLY	GLY ALA THR ASP ALA ALA ALA ALA GLY GLY GLY GLY GLY GLY GLZ GLG GLZ GLG	1262 1266 1266 1266 1266 1266 1266 1280 1283 1283 1283	R286 R291 M295	P298 E299 L305 L306 E306 A307 A307 GLY
ASP ASP ASN ASP CLN ASP GLU GLU ASP ASP ASP ASP ASP ASP ASP	GLU GLU GLU GLU GLU GLU GLU GLU ALA ALA ALA ALA CLU CLU	GLY GLY GLY GLY GLY GLY GLY VAL VAL TYR TYR TYR	PRU GLU ASP GLN ALA ILE	SER ARG LEU CYS CVS CVS CLU LEU CLY PHE FHE CLU



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants	79.64Å 79.64Å 403.97Å	Depositor
a, b, c, α , β , γ	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 - 2.81	Depositor
Resolution (A)	29.79 - 2.80	EDS
% Data completeness	91.3 (30.00-2.81)	Depositor
(in resolution range)	90.4 (29.79-2.80)	EDS
R _{merge}	(Not available)	Depositor
R _{sym}	(Not available)	Depositor
$< I/\sigma(I) > 1$	$2.52 (at 2.80 \text{\AA})$	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
B B.	0.209 , 0.244	Depositor
R, R_{free}	0.269 , (Not available)	DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor $(Å^2)$	61.5	Xtriage
Anisotropy	0.188	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$, $B_{sol}(Å^2)$	0.31 , 25.2	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	0.90	EDS
Total number of atoms	5485	wwPDB-VP
Average B, all atoms $(Å^2)$	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 4.53% of the height of the origin peak. No significant pseudotranslation is detected.

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.



¹Intensities estimated from amplitudes.

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	Bo	ond angles
	Unam	RMSZ	# Z > 5	RMSZ	# Z > 5
1	W	0.71	0/537	1.44	10/825~(1.2%)
2	Y	0.69	0/510	1.54	11/786~(1.4%)
3	А	0.43	0/4221	0.56	0/5673
4	Х	0.40	0/422	0.54	0/575
All	All	0.49	0/5690	0.84	21/7859~(0.3%)

There are no bond length outliers.

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$\mathbf{Observed}(^{o})$	$Ideal(^{o})$
2	Y	21	DG	O4'-C1'-N9	8.21	113.75	108.00
2	Y	12	DG	O4'-C1'-N9	7.67	113.37	108.00
2	Y	6	DA	O4'-C1'-N9	-7.00	103.10	108.00
2	Y	11	DT	O4'-C4'-C3'	-6.93	101.73	104.50
1	W	18	DG	P-O3'-C3'	6.53	127.53	119.70

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	3	0
2	Y	456	0	253	6	0
3	А	4132	0	4214	59	0
4	Х	416	0	427	9	0

Continued on next page...



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	5485	0	5168	71	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 7.

The worst 5 of 71 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:372:CYS:H	3:A:410:HIS:HD2	1.07	0.99
3:A:159:ASN:HD21	3:A:274:ARG:HH22	1.26	0.83
3:A:287:GLN:HE22	3:A:359:MET:H	1.30	0.80
3:A:444:HIS:HD2	3:A:446:TYR:H	1.30	0.77
1:W:9:DA:H2"	1:W:10:DC:H5"	1.67	0.75

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	А	498/538~(93%)	474 (95%)	23~(5%)	1 (0%)	47 78
4	Х	52/171~(30%)	50 (96%)	1 (2%)	1 (2%)	8 26
All	All	550/709~(78%)	524 (95%)	24~(4%)	2~(0%)	34 66

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	А	603	SER
4	Х	308	ALA



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	Pe	erce	entiles
3	А	452/480~(94%)	410 (91%)	42 (9%)		9	26
4	Х	47/129~(36%)	41 (87%)	6 (13%)		4	13
All	All	499/609~(82%)	451 (90%)	48 (10%)		8	24

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	А	441	LEU
3	А	513	ILE
3	А	448	VAL
3	А	478	VAL
3	А	553	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
3	А	303	ASN
3	А	410	HIS
4	Х	272	ASN
3	А	444	HIS
3	А	570	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)

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

