wwPDB X-ray Structure Validation Summary Report (i)

May 25, 2020 – 07:18 am BST

PDB ID	:	5A39
Title	:	Structure of Rad14 in complex with cisplatin containing DNA
Authors	:	Koch, S.C.; Kuper, J.; Gasteiger, K.L.; Wichlein, N.; Strasser, R.; Eisen, D.; Geiger, S.; Schneider, S.; Kisker, C.; Carell, T.
Deposited on	:	2015-05-28
Resolution	:	2.80 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 following versions of software and data (see references (1)) were used in the production of this report:

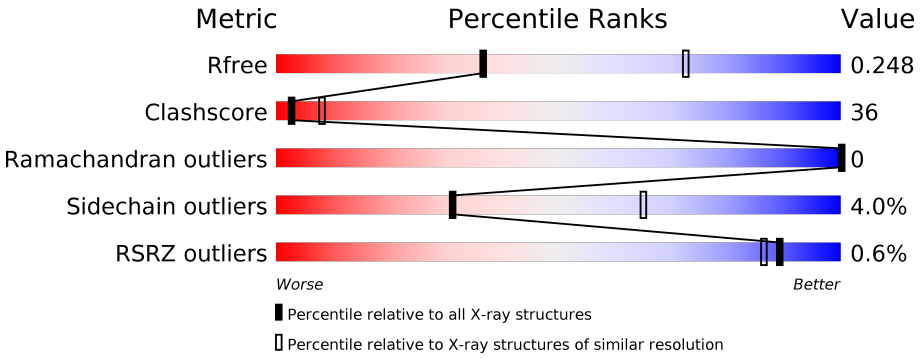
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (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 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 <=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	115	% 88% 11% .
1	B	115	90% 9% .
2	C	15	80% 20%
2	D	15	7% 80% 20%
3	E	13	23% 77%
3	F	13	23% 77%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	G	13	8% 92%
4	H	13	100%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 4088 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 REPAIR PROTEIN RAD14.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	114	Total	C	N	O	S	0	0	0
			940	594	161	177	8			
1	B	115	Total	C	N	O	S	0	0	0
			936	591	159	178	8			

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	15	Total	C	N	O	P	0	15	0
			557	270	93	168	26			
2	D	15	Total	C	N	O	P	0	15	0
			557	270	93	168	26			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	13	Total	C	N	O	P	0	13	0
			270	128	52	77	13			
3	F	13	Total	C	N	O	P	0	13	0
			270	128	52	77	13			

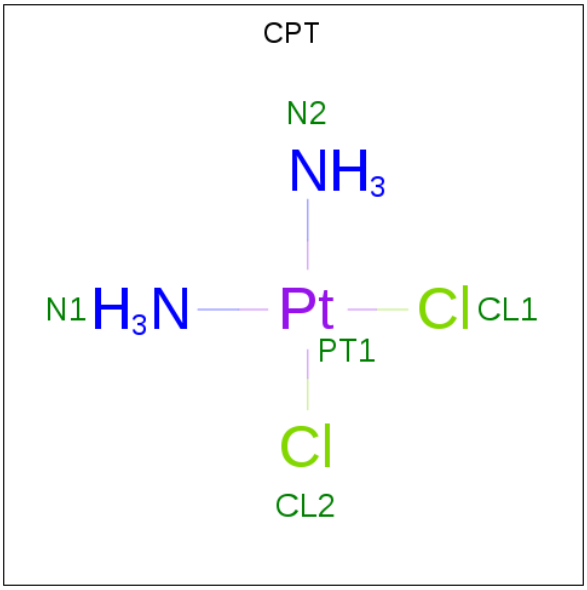
- Molecule 4 is a DNA chain called 5'-D(*DGP*AP*TP*GP*AP*CP*CP*GP*TP*AP*GP*AP*GP)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	G	13	Total	C	N	O	P	0	13	0
			272	128	55	76	13			
4	H	13	Total	C	N	O	P	0	13	0
			272	128	55	76	13			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total Zn 1 1	0	0
5	A	1	Total Zn 1 1	0	0

- Molecule 6 is Cisplatin (three-letter code: CPT) (formula: $Cl_2H_6N_2Pt$).



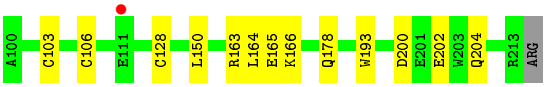
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	C	1	Total N Pt 3 2 1	0	0
6	C	1	Total N Pt 3 2 1	0	0
6	D	1	Total N Pt 3 2 1	0	0
6	D	1	Total N Pt 3 2 1	0	0

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 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.

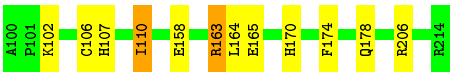
- Molecule 1: DNA REPAIR PROTEIN RAD14

Chain A:



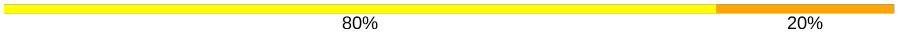
- Molecule 1: DNA REPAIR PROTEIN RAD14

Chain B:



- Molecule 2: DNA

Chain C:

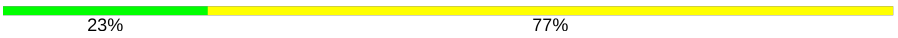


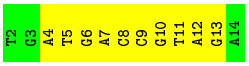
- Molecule 2: DNA

Chain D:

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

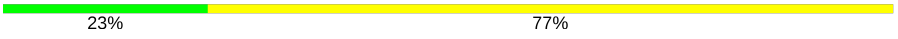
Chain E:

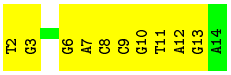




- Molecule 3: 5'-D(*DTP*GP*AP*TP*GP*AP*CP*CP*GP*TP*AP*GP*AP)-3'

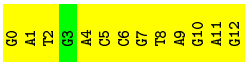
Chain F:





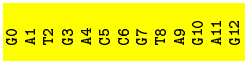
• Molecule 4: 5'-D(*DGP*AP*TP*GP*AP*CP*CP*GP*TP*AP*GP*AP*GP)-3'

Chain G:



• Molecule 4: 5'-D(*DGP*AP*TP*GP*AP*CP*CP*GP*TP*AP*GP*AP*GP)-3'

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	54.42Å 54.42Å 130.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	54.42 – 2.80 36.92 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.8 (54.42-2.80) 99.7 (36.92-2.80)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.38 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.213 , 0.256 0.209 , 0.248	Depositor DCC
R_{free} test set	430 reflections (4.57%)	wwPDB-VP
Wilson B-factor (Å^2)	58.0	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 25.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.458 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4088	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.19% 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

5.1 Standard geometry

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

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	0.81	0/964	0.83	0/1297
1	B	0.86	0/959	0.82	0/1292
2	C	1.79	22/622 (3.5%)	2.61	41/951 (4.3%)
2	D	1.76	24/622 (3.9%)	2.76	40/951 (4.2%)
3	E	0.24	0/303	0.69	0/466
3	F	0.33	0/303	0.73	0/466
4	G	0.27	0/306	0.75	0/471
4	H	0.35	0/306	0.73	0/471
All	All	1.11	46/4385 (1.0%)	1.61	81/6365 (1.3%)

The worst 5 of 46 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	8[A]	DG	C5-C4	14.25	1.48	1.38
2	C	8[B]	DG	C5-C4	14.25	1.48	1.38
2	D	8[C]	DG	C5-C4	13.81	1.48	1.38
2	D	8[D]	DG	C5-C4	13.81	1.48	1.38
2	D	9[C]	DG	C5-C4	11.79	1.46	1.38

The worst 5 of 81 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	9[C]	DG	C2-N3-C4	18.09	120.94	111.90
2	D	9[D]	DG	C2-N3-C4	18.09	120.94	111.90
2	D	8[C]	DG	N3-C4-C5	-17.88	119.66	128.60
2	D	8[D]	DG	N3-C4-C5	-17.88	119.66	128.60
2	D	8[C]	DG	C2-N3-C4	17.68	120.74	111.90

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	940	0	889	5	0
1	B	936	0	875	8	0
2	C	557	0	289	71	0
2	D	557	0	291	82	0
3	E	270	0	126	28	0
3	F	270	0	122	40	0
4	G	272	0	123	30	0
4	H	272	0	123	21	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
6	C	6	0	0	2	0
6	D	6	0	0	1	0
All	All	4088	0	2838	241	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 36.

The worst 5 of 241 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)
2:D:5[C]:DT:H2"	2:D:6[C]:DA:C8	1.37	1.57
2:D:8[D]:DG:N2	3:F:10[D]:DG:C2	1.88	1.37
3:F:8[D]:DC:H2"	3:F:9[D]:DC:C5'	1.65	1.26
2:C:5[A]:DT:H2"	2:C:6[A]:DA:C8	1.75	1.20
2:D:8[C]:DG:C2'	2:D:9[C]:DG:H5"	1.72	1.20

There are no symmetry-related clashes.

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	112/115 (97%)	109 (97%)	3 (3%)	0	100	100
1	B	113/115 (98%)	109 (96%)	4 (4%)	0	100	100
All	All	225/230 (98%)	218 (97%)	7 (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	102/104 (98%)	97 (95%)	5 (5%)	25	57
1	B	100/104 (96%)	97 (97%)	3 (3%)	41	75
All	All	202/208 (97%)	194 (96%)	8 (4%)	31	65

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	164	LEU
1	B	164	LEU
1	B	110	ILE
1	A	163	ARG
1	A	166	LYS

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 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 for Mogul analysis.

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	$\#\|Z\| > 2$
6	CPT	C	1016	6	0,2,4	0.00	-	-		
6	CPT	D	1016	6	0,2,4	0.00	-	-		
6	CPT	D	1015	6	0,2,4	0.00	-	-		
6	CPT	C	1015	6	0,2,4	0.00	-	-		

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.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	D	1015	CPT	1	0
6	C	1015	CPT	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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^{th} 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	114/115 (99%)	-0.43	1 (0%) 84 80	34, 49, 90, 109	0
1	B	115/115 (100%)	-0.53	0 100 100	32, 48, 87, 93	1 (0%)
2	C	15/15 (100%)	-0.59	0 100 100	24, 50, 75, 75	15 (100%)
2	D	15/15 (100%)	-0.20	1 (6%) 17 10	15, 61, 80, 86	15 (100%)
3	E	13/13 (100%)	-0.70	0 100 100	58, 63, 70, 74	13 (100%)
3	F	13/13 (100%)	-0.83	0 100 100	46, 54, 60, 68	13 (100%)
4	G	13/13 (100%)	-0.61	0 100 100	46, 54, 72, 74	13 (100%)
4	H	13/13 (100%)	-0.79	0 100 100	41, 45, 48, 53	13 (100%)
All	All	311/312 (99%)	-0.51	2 (0%) 89 86	15, 50, 86, 109	83 (26%)

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	1[C]	DT	5.2
1	A	111	GLU	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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^{th} 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
5	ZN	B	300	1/1	0.95	0.29	84,84,84,84	0
6	CPT	C	1016	3/5	0.97	0.12	50,50,50,54	3
6	CPT	D	1016	3/5	0.98	0.11	44,44,44,46	3
6	CPT	D	1015	3/5	0.99	0.08	40,40,41,44	3
6	CPT	C	1015	3/5	0.99	0.15	42,42,43,44	3
5	ZN	A	300	1/1	0.99	0.15	44,44,44,44	0

6.5 Other polymers

There are no such residues in this entry.