



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:18 PM GMT

PDB ID : 4REA
Title : A Nuclease DNA complex
Authors : Zhao, Q.; Xue, X.; Longerich, S.; Sung, P.; Xiong, Y.
Deposited on : 2014-09-22
Resolution : 3.81 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
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
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) : trunk26865

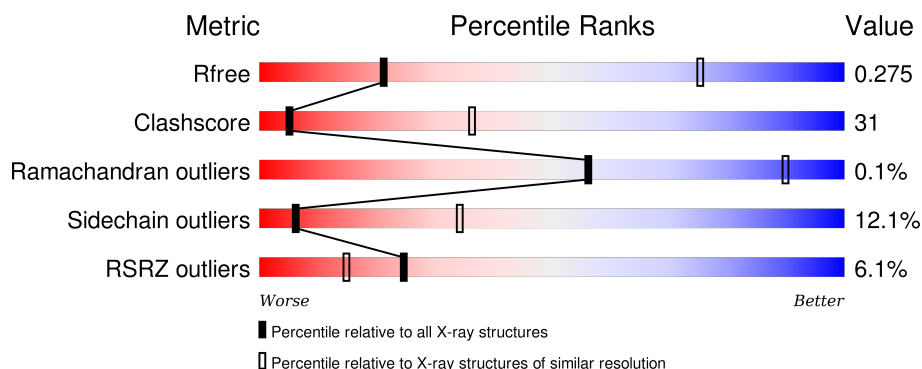
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.81 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1324 (4.14-3.50)
Clashscore	102246	1028 (4.12-3.52)
Ramachandran outliers	100387	1404 (4.14-3.50)
Sidechain outliers	100360	1399 (4.14-3.50)
RSRZ outliers	91569	1332 (4.14-3.50)

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 $\leq 5\%$. 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	647	<div> <div>7%</div> <div>52%</div> <div>35%</div> <div>•</div> <div>9%</div> </div>
1	B	647	<div> <div>4%</div> <div>51%</div> <div>34%</div> <div>6%</div> <div>9%</div> </div>
2	C	10	<div> <div>20%</div> <div>80%</div> </div>
3	D	10	<div> <div>10%</div> <div>90%</div> </div>
4	E	17	<div> <div>18%</div> <div>24%</div> <div>76%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10270 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 Fanconi-associated nuclease 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	588	Total	C	N	O	S	0	1	0
			4752	3031	841	856	24			
1	A	590	Total	C	N	O	S	0	1	0
			4764	3039	846	854	25			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	371	GLU	-	EXPRESSION TAG	UNP Q9Y2M0
B	372	PHE	-	EXPRESSION TAG	UNP Q9Y2M0
B	960	ALA	ASP	ENGINEERED MUTATION	UNP Q9Y2M0
A	371	GLU	-	EXPRESSION TAG	UNP Q9Y2M0
A	372	PHE	-	EXPRESSION TAG	UNP Q9Y2M0
A	960	ALA	ASP	ENGINEERED MUTATION	UNP Q9Y2M0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	10	Total	C	N	O	P	0	0	0
			197	95	34	59	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	10	Total	C	N	O	P	0	0	0
			208	97	41	60	10			

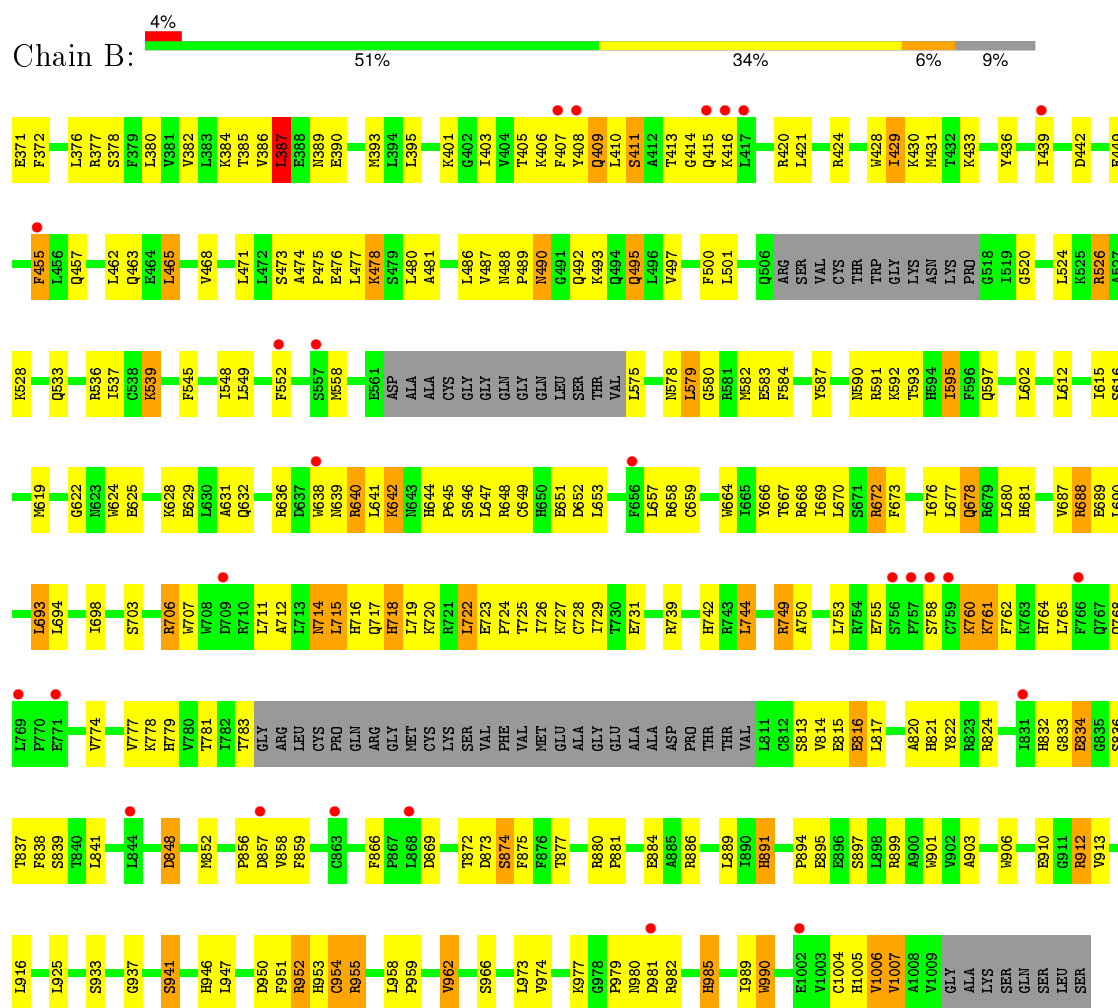
- Molecule 4 is a DNA chain called DNA (5'-D(P*GP*GP*CP*GP*AP*GP*CP*GP*CP*TP*CP*GP*CP*CP*AP*CP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	17	Total	C	N	O	P	0	0	0
			349	163	68	101	17			

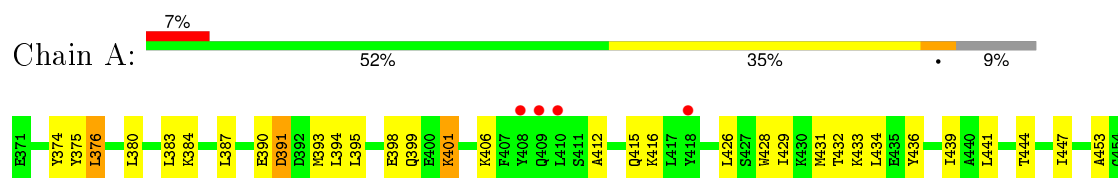
3 Residue-property plots

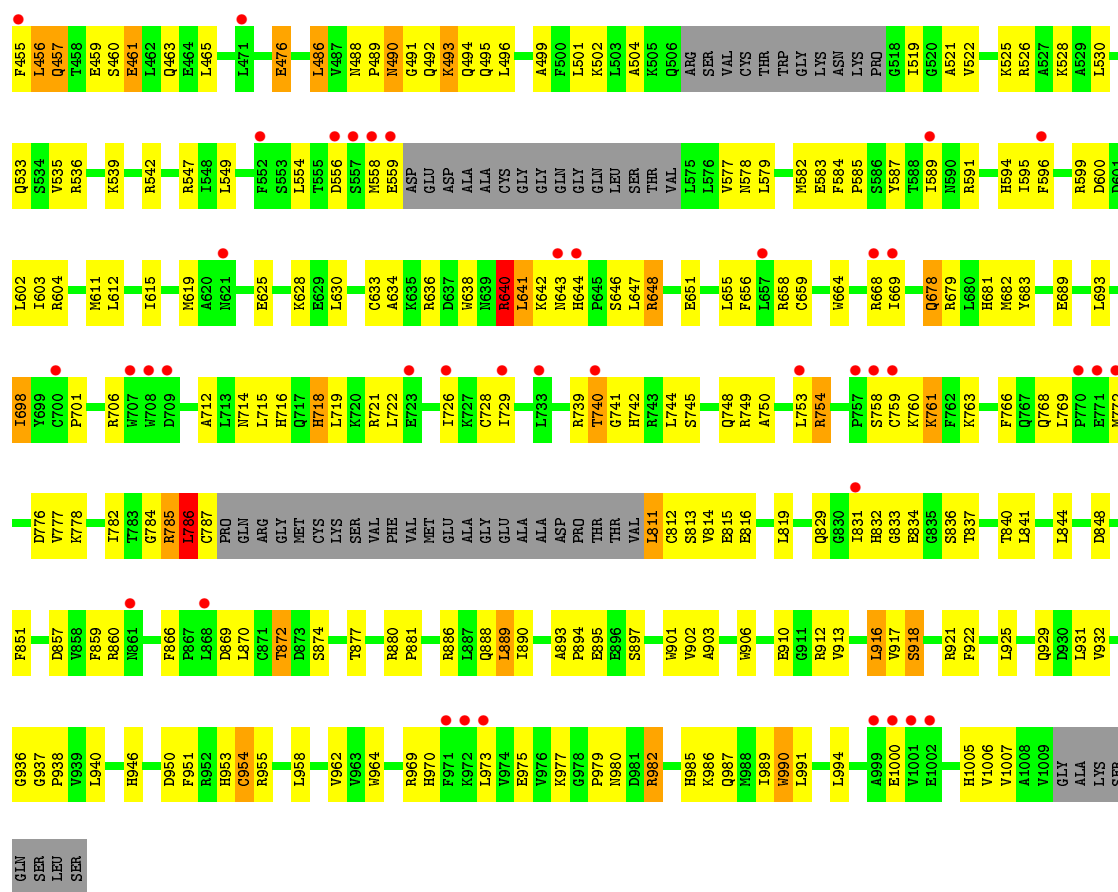
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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 ($\text{RSRZ} > 2$). 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: Fanconi-associated nuclease 1



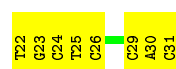
- Molecule 1: Fanconi-associated nuclease 1





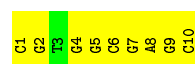
- Molecule 2: DNA (5'-D(*TP*GP*CP*TP*CP*GP*CP*CP*AP*C)-3')

Chain C: 20% 80%



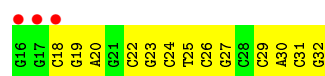
- Molecule 3: DNA (5'-D(P*CP*GP*TP*GP*GP*CP*GP*AP*GP*C)-3')

Chain D: 10% 90%



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

Chain E: 18% 24% 76%



4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α , β , γ	100.96Å 100.96Å 115.71Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.48 – 3.81 50.48 – 3.81	Depositor EDS
% Data completeness (in resolution range)	97.7 (50.48-3.81) 97.7 (50.48-3.81)	Depositor EDS
R_{merge}	0.21	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.21 (at 3.77Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.243 , 0.277 0.246 , 0.275	Depositor DCC
R_{free} test set	613 reflections (4.86%)	DCC
Wilson B-factor (Å ²)	177.6	Xtriage
Anisotropy	0.108	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 149.3	EDS
Estimated twinning fraction	0.490 for H, K, L 0.510 for K, H, -L 0.207 for -h,-k,l 0.349 for h,-h-k,-l 0.218 for -k,-h,-l	Xtriage
Reported twinning fraction	0.490 for H, K, L 0.510 for K, H, -L	Depositor
L-test for twinning ²	$\langle L \rangle = 0.35$, $\langle L^2 \rangle = 0.18$	Xtriage
Outliers	0 of 12619 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	10270	wwPDB-VP
Average B, all atoms (Å ²)	221.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.92% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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](#)

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	0.55	0/4864	0.80	4/6574 (0.1%)
1	B	0.56	0/4852	0.83	2/6559 (0.0%)
2	C	0.33	0/219	0.74	0/335
3	D	0.57	0/233	0.75	0/358
4	E	0.43	0/391	0.72	0/601
All	All	0.55	0/10559	0.81	6/14427 (0.0%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	387	LEU	CA-CB-CG	6.77	130.87	115.30
1	A	889	LEU	CA-CB-CG	6.58	130.43	115.30
1	B	693	LEU	CA-CB-CG	6.26	129.70	115.30
1	A	549	LEU	CA-CB-CG	6.04	129.20	115.30
1	A	630	LEU	CA-CB-CG	-5.48	102.69	115.30

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	A	4764	0	4787	296	0
1	B	4752	0	4761	308	0
2	C	197	0	114	10	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	D	208	0	112	28	0
4	E	349	0	189	29	0
All	All	10270	0	9963	613	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 31.

The worst 5 of 613 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:B:813:SER:CB	1:B:816:GLU:HB2	1.17	1.57
1:B:410:LEU:HD11	1:B:414:GLY:C	1.35	1.47
1:B:813:SER:HB3	1:B:816:GLU:CB	1.51	1.38
1:B:813:SER:CB	1:B:816:GLU:CB	2.07	1.29
1:B:781:THR:HG21	1:A:521:ALA:CB	1.66	1.25

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	583/647 (90%)	563 (97%)	19 (3%)	1 (0%)	52	86
1	B	581/647 (90%)	559 (96%)	22 (4%)	0	100	100
All	All	1164/1294 (90%)	1122 (96%)	41 (4%)	1 (0%)	56	90

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	786	LEU

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	516/562 (92%)	462 (90%)	54 (10%)	8	41
1	B	515/562 (92%)	444 (86%)	71 (14%)	4	30
All	All	1031/1124 (92%)	906 (88%)	125 (12%)	6	34

5 of 125 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	913	VAL
1	B	1006	VAL
1	A	811	LEU
1	B	941	SER
1	B	966	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	1005	HIS
1	A	953	HIS
1	A	678	GLN
1	B	907	HIS
1	A	452	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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](#)

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	590/647 (91%)	0.17	45 (7%) 17 10	143, 219, 293, 355	0
1	B	588/647 (90%)	0.02	26 (4%) 38 25	145, 210, 276, 432	0
2	C	10/10 (100%)	-0.04	0 100 100	232, 246, 258, 274	0
3	D	10/10 (100%)	-0.75	0 100 100	201, 212, 236, 238	0
4	E	17/17 (100%)	0.02	3 (17%) 2 2	190, 222, 263, 284	0
All	All	1215/1331 (91%)	0.08	74 (6%) 25 15	143, 215, 287, 432	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	758	SER	6.9
1	A	558	MET	6.7
1	A	1002	GLU	5.7
1	A	999	ALA	5.3
1	A	740	THR	4.8

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 carbohydrates 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.