wwPDB X-ray Structure Validation Summary Report (i)

Jan 3, 2024 – 02:09 am GMT

PDB ID	:	5A78
Title	:	Crystal structure of the homing endonuclease I-CvuI in complex with I- CreI target (C1221) in the presence of 2 mM Mg revealing DNA not cleaved
Authors	:	Molina, R.; Redondo, P.; LopezMendez, B.; Villate, M.; Merino, N.; Blanco, F.J.; Valton, J.; Grizot, S.; Duchateau, P.; Prieto, J.; Montoya, G.
Deposited on	:	2015-07-03
Resolution	:	2.50 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:

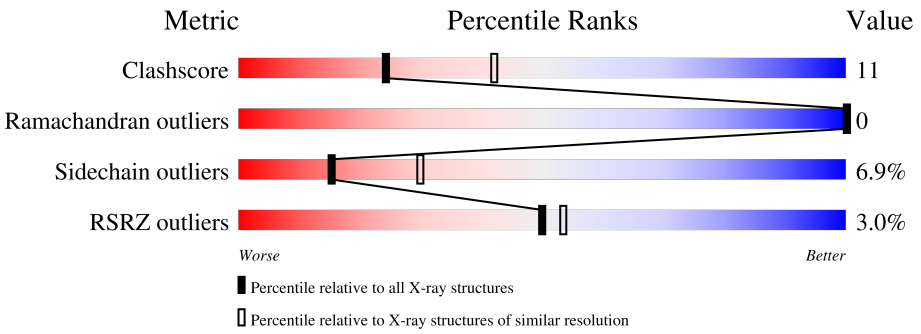
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.50 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	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 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	A	172	3% 64% 24% • 9%
1	B	172	3% 70% 20% • • 8%
2	C	24	58% 42%
2	D	24	63% 38%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3474 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 ENDONUCLEASE I-CVUI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	157	1233	791	215	221	6	0	0	0
			Total	C	N	O	S			
1	B	158	1239	794	216	223	6	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2	ALA	-	expression tag	UNP P56347
A	163	SER	-	expression tag	UNP P56347
A	164	ALA	-	expression tag	UNP P56347
A	165	ALA	-	expression tag	UNP P56347
A	166	LEU	-	expression tag	UNP P56347
A	167	GLU	-	expression tag	UNP P56347
A	168	HIS	-	expression tag	UNP P56347
A	169	HIS	-	expression tag	UNP P56347
A	170	HIS	-	expression tag	UNP P56347
A	171	HIS	-	expression tag	UNP P56347
A	172	HIS	-	expression tag	UNP P56347
A	173	HIS	-	expression tag	UNP P56347
A	54	GLN	ARG	conflict	UNP P56347
B	2	ALA	-	expression tag	UNP P56347
B	163	SER	-	expression tag	UNP P56347
B	164	ALA	-	expression tag	UNP P56347
B	165	ALA	-	expression tag	UNP P56347
B	166	LEU	-	expression tag	UNP P56347
B	167	GLU	-	expression tag	UNP P56347
B	168	HIS	-	expression tag	UNP P56347
B	169	HIS	-	expression tag	UNP P56347
B	170	HIS	-	expression tag	UNP P56347
B	171	HIS	-	expression tag	UNP P56347
B	172	HIS	-	expression tag	UNP P56347
B	173	HIS	-	expression tag	UNP P56347

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	54	GLN	ARG	conflict	UNP P56347

- Molecule 2 is a DNA chain called 24MER DNA, 5'-D(*TP*CP*AP*AP*AP*AP*CP*GP*TP*CP*GP*TP*AP *CP*GP*AP*CP*GP*TP*TP*TP*TP*GP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	24	489	235	89	142	23	0	0	0
			Total	C	N	O	P			
2	D	24	489	235	89	142	23	0	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
3	B	1	1	1	0	0
			Total	Mg		
3	D	1	1	1	0	0

- Molecule 4 is water.

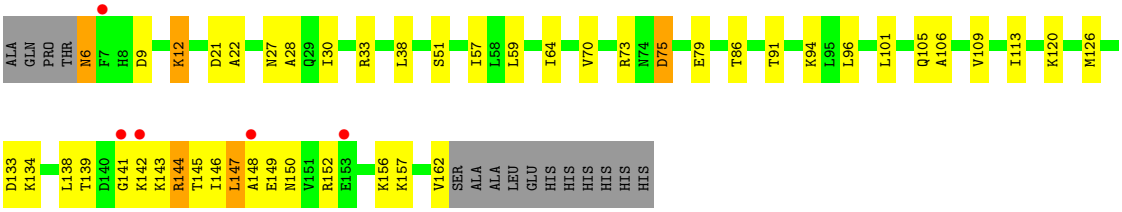
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	12	12	12	0	0
			Total	O		
4	B	6	6	6	0	0
			Total	O		
4	C	4	4	4	0	0

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: DNA ENDONUCLEASE I-CVUI

Chain A:

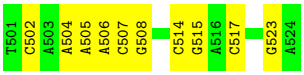


• Molecule 1: DNA ENDONUCLEASE I-CVUI

Chain B:

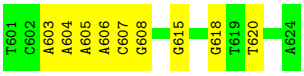
• Molecule 2: 24MER DNA, 5'-D(*TP*CP*AP*AP*AP*AP*CP*GP*TP*CP*GP*TP*AP *CP *GP*AP*CP*GP*TP*TP*TP*TP*GP*A)-3'

Chain C:



• Molecule 2: 24MER DNA, 5'-D(*TP*CP*AP*AP*AP*AP*CP*GP*TP*CP*GP*TP*AP *CP *GP*AP*CP*GP*TP*TP*TP*TP*GP*A)-3'

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	48.57Å 95.08Å 52.15Å 90.00° 103.71° 90.00°	Depositor
Resolution (Å)	47.54 – 2.50 47.54 – 2.50	Depositor EDS
% Data completeness (in resolution range)	97.9 (47.54-2.50) 98.7 (47.54-2.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.86 (at 2.51Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.215 , 0.280 0.225 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	36.9	Xtriage
Anisotropy	0.970	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 26.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3474	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.52% 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)

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

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	0.59	0/1249	0.75	1/1683 (0.1%)
1	B	0.54	0/1255	0.88	5/1691 (0.3%)
2	C	0.81	0/548	1.12	2/844 (0.2%)
2	D	0.84	0/548	1.05	0/844
All	All	0.66	0/3600	0.92	8/5062 (0.2%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	141	GLY	N-CA-C	-13.19	80.11	113.10
1	B	142	LYS	N-CA-C	-10.99	81.33	111.00
1	B	140	ASP	CB-CA-C	-9.17	92.07	110.40
1	A	142	LYS	N-CA-CB	-6.48	98.93	110.60
1	B	57	ILE	C-N-CA	-6.20	106.19	121.70

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	A	1233	0	1306	41	1
1	B	1239	0	1309	26	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	489	0	272	11	0
2	D	489	0	272	7	1
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	A	12	0	0	5	0
4	B	6	0	0	1	0
4	C	4	0	0	1	0
All	All	3474	0	3159	73	1

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 73 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:143:LYS:NZ	2:C:515:DG:N2	1.64	1.40
4:A:2009:HOH:O	1:B:22:ALA:HA	1.68	0.94
4:A:2003:HOH:O	2:C:515:DG:H5'	1.68	0.93
1:A:143:LYS:HZ2	2:C:515:DG:N2	1.34	0.92
4:B:2001:HOH:O	2:D:615:DG:H5'	1.76	0.86

All (1) 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)
1:A:157:LYS:NZ	2:D:620:DT:OP1[2_445]	2.15	0.05

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	155/172 (90%)	146 (94%)	9 (6%)	0	100	100
1	B	156/172 (91%)	153 (98%)	3 (2%)	0	100	100
All	All	311/344 (90%)	299 (96%)	12 (4%)	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	137/149 (92%)	129 (94%)	8 (6%)	20	38
1	B	138/149 (93%)	127 (92%)	11 (8%)	12	23
All	All	275/298 (92%)	256 (93%)	19 (7%)	15	30

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	144	ARG
1	B	152	ARG
1	B	153	GLU
1	B	147	LEU
1	B	35	ASP

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

Mol	Chain	Res	Type
1	A	6	ASN
1	A	29	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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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.

No monomer is involved in short contacts.

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	157/172 (91%)	0.18	5 (3%) 47 51	24, 41, 59, 78	0
1	B	158/172 (91%)	0.24	6 (3%) 40 43	24, 43, 62, 75	0
2	C	24/24 (100%)	-0.65	0 100 100	34, 44, 54, 56	0
2	D	24/24 (100%)	-0.56	0 100 100	36, 43, 51, 53	0
All	All	363/392 (92%)	0.10	11 (3%) 50 53	24, 42, 60, 78	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	142	LYS	6.7
1	A	153	GLU	5.1
1	A	142	LYS	3.8
1	B	141	GLY	3.2
1	A	141	GLY	2.7

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)

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	MG	B	1164	1/1	0.75	0.11	51,51,51,51	0
3	MG	D	1625	1/1	0.94	0.14	53,53,53,53	0

6.5 Other polymers (i)

There are no such residues in this entry.