



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 07:57 am BST

PDB ID : 5O6I
Title : Structures and dynamics of mesophilic variants from the homing endonuclease I-DmoI
Authors : Molina, R.; Marcaida, M.J.
Deposited on : 2017-06-06
Resolution : 2.25 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
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

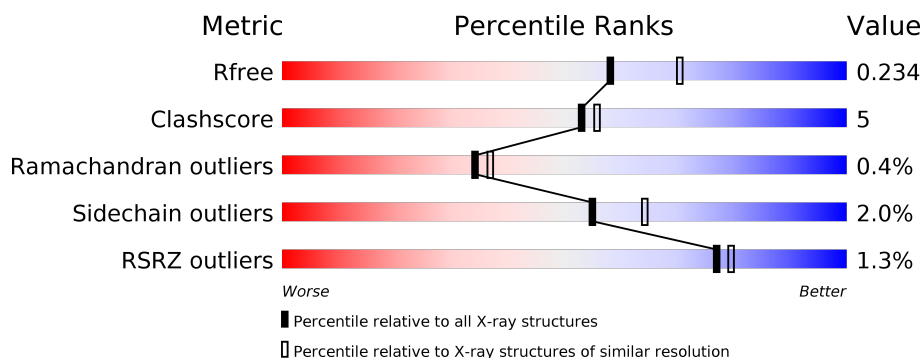
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.25 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	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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 $\leq 5\%$. 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	200	<div> <div>2%</div> <div> <div></div> <div>79%</div> <div>16%</div> <div>• 5%</div> </div> </div>
1	F	200	<div> <div>%</div> <div> <div></div> <div>74%</div> <div>15%</div> <div>• 11%</div> </div> </div>
1	K	200	<div> <div>2%</div> <div> <div></div> <div>77%</div> <div>12%</div> <div>• 11%</div> </div> </div>
2	C	25	<div> <div></div> <div> <div>80%</div> <div>20%</div> </div> </div>
2	G	25	<div> <div></div> <div> <div>88%</div> <div>12%</div> </div> </div>
2	L	25	<div> <div></div> <div> <div>92%</div> <div>8%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	D	25	 88% 12%
3	I	25	 84% 16%
3	N	25	 88% 12%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7973 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 Homing endonuclease I-Dmol.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	191	Total	C	N	O	S	8	5	0
			1607	1036	294	273	4			
1	F	179	Total	C	N	O	S	0	5	0
			1517	978	278	257	4			
1	K	179	Total	C	N	O	S	0	3	0
			1501	968	272	257	4			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P21505
A	1	ALA	-	expression tag	UNP P21505
A	52	PHE	ILE	conflict	UNP P21505
A	92	THR	ALA	conflict	UNP P21505
A	101	CYS	PHE	conflict	UNP P21505
A	189	ALA	-	expression tag	UNP P21505
A	190	ALA	-	expression tag	UNP P21505
A	191	ALA	-	expression tag	UNP P21505
A	192	LEU	-	expression tag	UNP P21505
A	193	GLU	-	expression tag	UNP P21505
A	194	HIS	-	expression tag	UNP P21505
A	195	HIS	-	expression tag	UNP P21505
A	196	HIS	-	expression tag	UNP P21505
A	197	HIS	-	expression tag	UNP P21505
A	198	HIS	-	expression tag	UNP P21505
A	199	HIS	-	expression tag	UNP P21505
F	0	MET	-	initiating methionine	UNP P21505
F	1	ALA	-	expression tag	UNP P21505
F	52	PHE	ILE	conflict	UNP P21505
F	92	THR	ALA	conflict	UNP P21505
F	101	CYS	PHE	conflict	UNP P21505
F	189	ALA	-	expression tag	UNP P21505
F	190	ALA	-	expression tag	UNP P21505

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	191	ALA	-	expression tag	UNP P21505
F	192	LEU	-	expression tag	UNP P21505
F	193	GLU	-	expression tag	UNP P21505
F	194	HIS	-	expression tag	UNP P21505
F	195	HIS	-	expression tag	UNP P21505
F	196	HIS	-	expression tag	UNP P21505
F	197	HIS	-	expression tag	UNP P21505
F	198	HIS	-	expression tag	UNP P21505
F	199	HIS	-	expression tag	UNP P21505
K	0	MET	-	initiating methionine	UNP P21505
K	1	ALA	-	expression tag	UNP P21505
K	52	PHE	ILE	conflict	UNP P21505
K	92	THR	ALA	conflict	UNP P21505
K	101	CYS	PHE	conflict	UNP P21505
K	189	ALA	-	expression tag	UNP P21505
K	190	ALA	-	expression tag	UNP P21505
K	191	ALA	-	expression tag	UNP P21505
K	192	LEU	-	expression tag	UNP P21505
K	193	GLU	-	expression tag	UNP P21505
K	194	HIS	-	expression tag	UNP P21505
K	195	HIS	-	expression tag	UNP P21505
K	196	HIS	-	expression tag	UNP P21505
K	197	HIS	-	expression tag	UNP P21505
K	198	HIS	-	expression tag	UNP P21505
K	199	HIS	-	expression tag	UNP P21505

- Molecule 2 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	25	Total	C	N	O	P	0	0	0
			511	242	94	151	24			
2	G	25	Total	C	N	O	P	0	0	0
			511	242	94	151	24			
2	L	25	Total	C	N	O	P	0	0	0
			511	242	94	151	24			

- Molecule 3 is a DNA chain called DNA (25-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	25	Total	C	N	O	P	0	0	0
			508	240	99	145	24			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	25	Total	C	N	O	P	0	0	0
			508	240	99	145	24			
3	N	25	Total	C	N	O	P	0	0	0
			508	240	99	145	24			

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Mn	0	0
			2	2		
4	K	2	Total	Mn	0	0
			2	2		
4	F	2	Total	Mn	0	0
			2	2		

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	G	1	Total	Cl	1	0
			1	1		
5	A	1	Total	Cl	1	0
			1	1		

- Molecule 6 is water.


Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	56	Total	O	0	0
			56	56		
6	C	19	Total	O	0	0
			19	19		
6	D	24	Total	O	0	0
			24	24		
6	F	51	Total	O	0	0
			51	51		
6	G	18	Total	O	0	0
			18	18		
6	I	16	Total	O	0	0
			16	16		
6	K	49	Total	O	0	0
			49	49		
6	L	27	Total	O	0	0
			27	27		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	N	23	Total	O	0	0
			23	23		


● Molecule 2: DNA (25-MER)

Chain G:  88% 12%


● Molecule 2: DNA (25-MER)

Chain L:  92% 8%


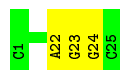
● Molecule 3: DNA (25-MER)

Chain D:  88% 12%

● Molecule 3: DNA (25-MER)

Chain I:  84% 16%

● Molecule 3: DNA (25-MER)

Chain N:  88% 12%

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	107.00 Å 70.48 Å 107.32 Å 90.00° 119.81° 90.00°	Depositor
Resolution (Å)	56.14 – 2.25 56.14 – 2.25	Depositor EDS
% Data completeness (in resolution range)	98.0 (56.14-2.25) 93.4 (56.14-2.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.86 (at 2.25 Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.196 , 0.228 0.201 , 0.234	Depositor DCC
R_{free} test set	3282 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	39.8	Xtriage
Anisotropy	0.476	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 42.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.022 for -h-l,k,h 0.022 for l,k,-h-l 0.018 for h,-k,-h-l 0.014 for -h-l,-k,l 0.016 for l,-k,h	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7973	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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

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	0.23	0/1650	0.42	0/2221
1	F	0.24	0/1555	0.44	0/2088
1	K	0.25	0/1533	0.46	2/2060 (0.1%)
2	C	0.58	0/572	0.92	0/882
2	G	0.52	0/572	0.93	0/882
2	L	0.54	0/572	0.93	0/882
3	D	0.56	0/570	0.88	0/877
3	I	0.53	0/570	0.84	0/877
3	N	0.61	0/570	0.88	1/877 (0.1%)
All	All	0.41	0/8164	0.69	3/11646 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
3	N	22	DA	O4'-C1'-N9	7.30	113.11	108.00
1	K	104	ARG	NE-CZ-NH2	-6.01	117.30	120.30
1	K	104	ARG	CG-CD-NE	-5.04	101.22	111.80

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	1607	0	1688	20	0
1	F	1517	0	1604	25	0
1	K	1501	0	1578	19	0
2	C	511	0	282	5	0
2	G	511	0	282	3	0
2	L	511	0	282	2	0
3	D	508	0	279	2	0
3	I	508	0	279	4	0
3	N	508	0	279	1	0
4	A	2	0	0	0	0
4	F	2	0	0	0	0
4	K	2	0	0	0	0
5	A	1	0	0	0	0
5	G	1	0	0	0	0
6	A	56	0	0	4	0
6	C	19	0	0	2	0
6	D	24	0	0	1	0
6	F	51	0	0	5	0
6	G	18	0	0	1	0
6	I	16	0	0	4	0
6	K	49	0	0	2	0
6	L	27	0	0	1	0
6	N	23	0	0	0	0
All	All	7973	0	6553	75	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 5.

The worst 5 of 75 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:F:68:LYS:O	6:F:301:HOH:O	1.83	0.94
1:A:68:LYS:O	6:A:301:HOH:O	1.87	0.93
1:K:144:ASN:OD1	6:K:301:HOH:O	1.89	0.87
3:I:1:DC:O5'	6:I:101:HOH:O	1.97	0.82
1:A:61:ASP:OD2	6:A:302:HOH:O	2.00	0.80

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	194/200 (97%)	189 (97%)	5 (3%)	0	100	100
1	F	182/200 (91%)	176 (97%)	5 (3%)	1 (0%)	29	29
1	K	180/200 (90%)	172 (96%)	7 (4%)	1 (1%)	25	25
All	All	556/600 (93%)	537 (97%)	17 (3%)	2 (0%)	34	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	182	LEU
1	K	6	ASN

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	177/180 (98%)	173 (98%)	4 (2%)	50	59
1	F	168/180 (93%)	165 (98%)	3 (2%)	59	68
1	K	166/180 (92%)	162 (98%)	4 (2%)	49	58
All	All	511/540 (95%)	500 (98%)	11 (2%)	55	61

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	77	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	95	LEU
1	K	42	GLN
1	A	120	LYS
1	K	32	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 8 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 ⓘ

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 [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	191/200 (95%)	0.02	3 (1%) 72 74	28, 44, 65, 84	0
1	F	179/200 (89%)	0.11	2 (1%) 80 82	31, 45, 67, 83	0
1	K	179/200 (89%)	0.10	4 (2%) 62 65	29, 41, 69, 126	0
2	C	25/25 (100%)	-0.40	0 100 100	34, 52, 66, 70	0
2	G	25/25 (100%)	-0.55	0 100 100	37, 48, 60, 63	0
2	L	25/25 (100%)	-0.50	0 100 100	36, 50, 56, 62	0
3	D	25/25 (100%)	-0.36	0 100 100	37, 48, 76, 78	0
3	I	25/25 (100%)	-0.37	0 100 100	36, 51, 64, 69	0
3	N	25/25 (100%)	-0.51	0 100 100	37, 49, 68, 71	0
All	All	699/750 (93%)	-0.04	9 (1%) 77 79	28, 45, 67, 126	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	182	LEU	12.2
1	K	183	ASN	10.5
1	K	181	HIS	3.9
1	F	181	HIS	3.5
1	F	75	ASP	2.9

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](#)

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(\AA^2)	Q<0.9
5	CL	G	101	1/1	-	-	48,48,48,48	1
4	MN	A	201	1/1	0.99	0.13	40,40,40,40	0
5	CL	A	203	1/1	-	-	47,47,47,47	1
4	MN	F	202	1/1	0.95	0.16	42,42,42,42	0
4	MN	A	202	1/1	0.96	0.18	40,40,40,40	0
4	MN	K	202	1/1	0.97	0.15	38,38,38,38	0
4	MN	F	201	1/1	0.98	0.15	44,44,44,44	0
4	MN	K	201	1/1	0.98	0.15	41,41,41,41	0

6.5 Other polymers [i](#)

There are no such residues in this entry.