



wwPDB X-ray Structure Validation Summary Report ⓘ

May 28, 2020 – 08:02 pm BST

PDB ID : 1LWS
Title : Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence
Authors : Moure, C.M.; Gimble, F.S.; Quiocho, F.A.
Deposited on : 2002-06-03
Resolution : 3.50 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

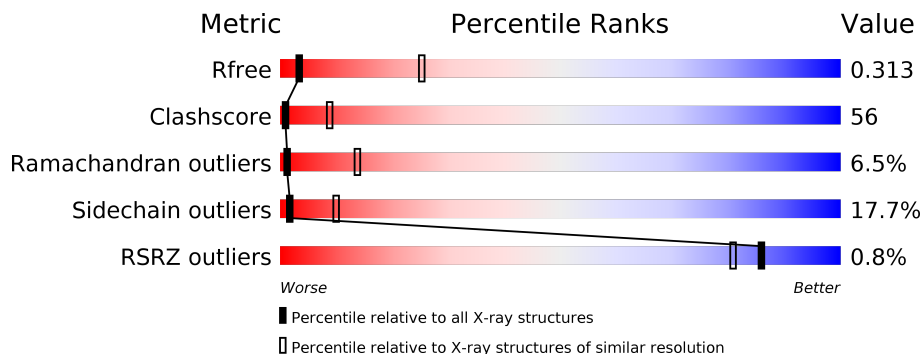
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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))
R_{free}	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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	B	37	
2	C	37	
3	A	454	

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4756 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 PI-SceI DNA recognition region top strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	34	712	337	140	201	34	0	0	0

- Molecule 2 is a DNA chain called PI-SceI DNA recognition region bottom strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	34	682	328	113	207	34	0	0	0

- Molecule 3 is a protein called ENDONUCLEASE PI-SCEI.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
3	A	427	3360	2125	583	638	6	8	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	28	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	47	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	109	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	193	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	236	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	372	MSE	MET	MODIFIED RESIDUE	UNP P17255
A	385	MSE	MET	MODIFIED RESIDUE	UNP P17255

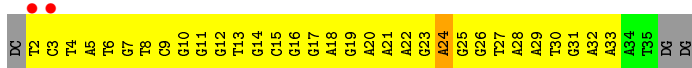
- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Ca	0	0
			2	2		

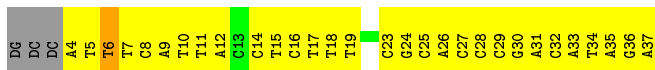
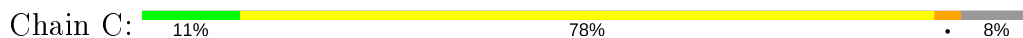
3 Residue-property plots i

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.

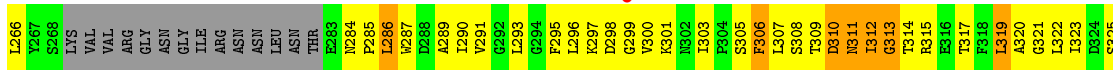
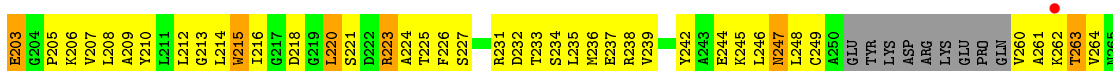
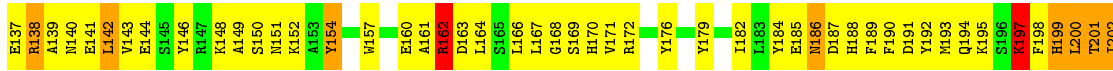
- Molecule 1: PI-SceI DNA recognition region top strand



- Molecule 2: PI-SceI DNA recognition region bottom strand



- Molecule 3: ENDONUCLEASE PI-SCEI



C388	A399	G400	F404	R405	A409	F412	A413	R414	E415	C416	R417	G418	F419	Y420	F421	E422	L423	Q424	E425	L426	K427	E428	D429	D430	Y431	Y432	G433	L434	I435	L436	S437	D438	D439	Q443	F444	L445	L446	A447	W448	Q449	V450	V451	V452	H453	M454
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α , β , γ	123.00Å 123.00Å 211.70Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	24.92 – 3.50 35.02 – 3.49	Depositor EDS
% Data completeness (in resolution range)	97.6 (24.92-3.50) 98.9 (35.02-3.49)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	11.01 (at 3.47Å)	Xtrriage
Refinement program	CNS	Depositor
R, R_{free}	0.287 , 0.310 0.280 , 0.313	Depositor DCC
R_{free} test set	622 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å ²)	95.8	Xtrriage
Anisotropy	0.144	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 70.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.86	EDS
Total number of atoms	4756	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% 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 [i](#)

5.1 Standard geometry [i](#)

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

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	B	0.81	0/802	0.91	0/1239
2	C	0.75	0/760	0.95	1/1167 (0.1%)
3	A	0.65	0/3412	0.82	1/4586 (0.0%)
All	All	0.70	0/4974	0.86	2/6992 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	65	ARG	N-CA-C	-5.58	95.92	111.00
2	C	6	DT	N1-C1'-C2'	5.37	122.80	112.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	24	DA	Sidechain

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	B	712	0	383	64	0
2	C	682	0	386	42	0
3	A	3360	0	3367	402	0
4	A	2	0	0	0	0
All	All	4756	0	4136	498	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 56.

The worst 5 of 498 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:C:36:DG:H2''	2:C:37:DA:H5'	1.27	1.15
1:B:16:DG:H1'	1:B:17:DG:H5'	1.28	1.13
3:A:199:HIS:NE2	3:A:201:THR:HG23	1.64	1.11
3:A:448:ASN:HD21	3:A:450:VAL:HB	1.12	1.09
3:A:218:ASP:O	3:A:226:PHE:HB2	1.60	1.00

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	A	417/454 (92%)	314 (75%)	76 (18%)	27 (6%)	1 14

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	14	SER
3	A	97	LYS
3	A	197	LYS
3	A	202	ILE
3	A	341	THR

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
3	A	368/384 (96%)	303 (82%)	65 (18%)	2 10

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	187	ASP
3	A	203	GLU
3	A	414	ARG
3	A	194	GLN
3	A	199	HIS

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

Mol	Chain	Res	Type
3	A	170	HIS
3	A	454	ASN
3	A	448	ASN
3	A	56	HIS
3	A	186	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 carbohydrates 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	B	34/37 (91%)	0.39	2 (5%) 22 20	96, 96, 96, 96	0
2	C	34/37 (91%)	0.08	0 100 100	96, 96, 96, 96	0
3	A	419/454 (92%)	-0.21	2 (0%) 91 88	23, 58, 84, 106	0
All	All	487/528 (92%)	-0.15	4 (0%) 86 81	23, 61, 96, 106	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	299	GLY	3.0
3	A	262	LYS	2.8
1	B	3	DC	2.7
1	B	2	DT	2.4

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
4	CA	A	502	1/1	0.84	0.16	29,29,29,29	0
4	CA	A	501	1/1	0.94	0.27	30,30,30,30	0

6.5 Other polymers [i](#)

There are no such residues in this entry.