Full wwPDB X-ray Structure Validation Report ⓘ

Jan 8, 2024 – 02:13 am GMT

PDB ID	:	6FB2
Title	:	Crystal Structure of a Tailored I-CreI Homing Endonuclease Protein (3115 variant) in complex with its target DNA (Haemoglobin beta subunit gene) in the presence of Manganese
Authors	:	Molina, R.; Prieto, J.
Deposited on	:	2017-12-18
Resolution	:	2.95 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

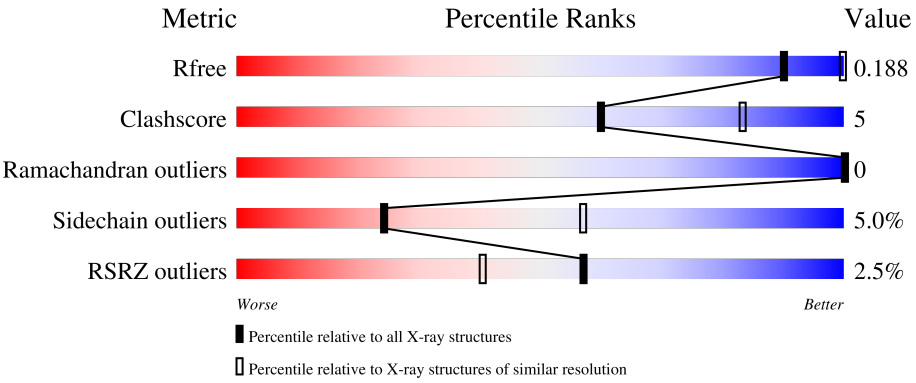
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	153	5% 84% 14%
2	B	154	% 92% 6%
3	D	14	64% 29% 7%
4	F	14	79% 21%
5	E	10	50% 50%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	G	10	

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 3480 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-CreI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	153	1236	794	211	230	1	0	0	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19	SER	GLY	conflict	UNP P05725
A	33	VAL	TYR	conflict	UNP P05725
A	38	ARG	GLN	conflict	UNP P05725
A	40	GLN	SER	conflict	UNP P05725
A	44	ASP	GLN	conflict	UNP P05725
A	68	ALA	ARG	conflict	UNP P05725
A	70	SER	ARG	conflict	UNP P05725
A	75	LYS	ASP	conflict	UNP P05725
A	77	ARG	ILE	conflict	UNP P05725

- Molecule 2 is a protein called DNA endonuclease I-CreI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	154	1250	810	207	232	1	0	1	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	33	GLY	TYR	conflict	UNP P05725
B	38	LYS	GLN	conflict	UNP P05725
B	44	LYS	GLN	conflict	UNP P05725
B	68	TYR	ARG	conflict	UNP P05725
B	70	SER	ARG	conflict	UNP P05725
B	75	ASN	ASP	conflict	UNP P05725
B	77	TYR	ILE	conflict	UNP P05725

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	132	VAL	ILE	conflict	UNP P05725

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	14	Total	C	N	O	P	0	0	0
			276	134	46	83	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	14	Total	C	N	O	P	0	0	0
			282	136	47	86	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	10	Total	C	N	O	P	0	0	0
			212	99	45	58	10			

- Molecule 6 is a DNA chain called DNA (5'-D(P*AP*GP*AP*AP*GP*TP*CP*TP*GP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	10	Total	C	N	O	P	0	0	0
			210	99	42	59	10			

- Molecule 7 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	3	Total	Mn	0	0
			3	3		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	6	Total	O	0	0
			6	6		

Continued on next page...

Continued from previous page...

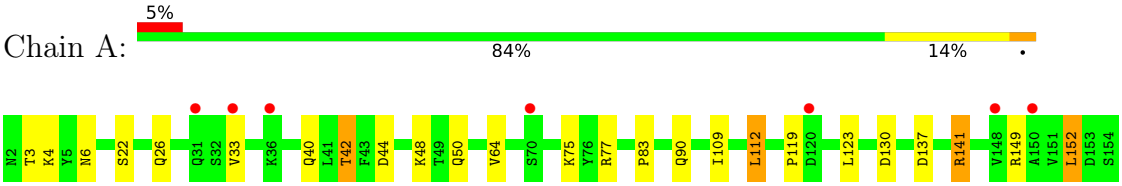
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	B	1	Total O 1 1	0	0
8	F	2	Total O 2 2	0	0
8	E	1	Total O 1 1	0	0
8	G	1	Total O 1 1	0	0

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.

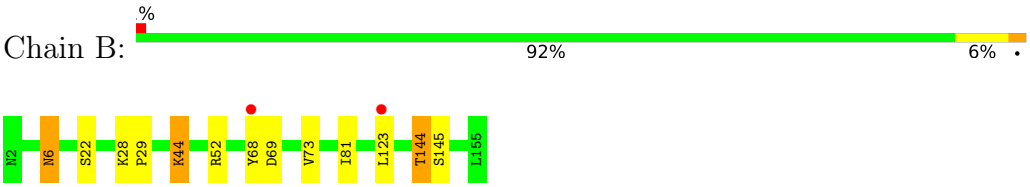
• Molecule 1: DNA endonuclease I-CreI

Chain A:



• Molecule 2: DNA endonuclease I-CreI

Chain B:

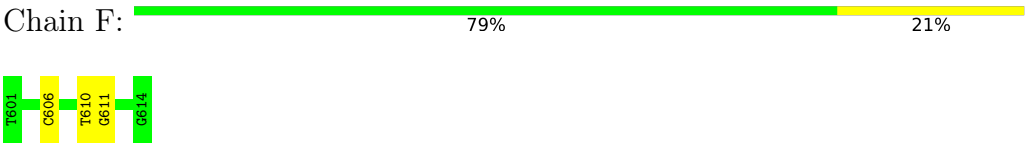


• Molecule 3: DNA (5'-D(*TP*CP*AP*GP*AP*CP*TP*TP*CP*TP*CP*CP*AP*C)-3')

Chain D:

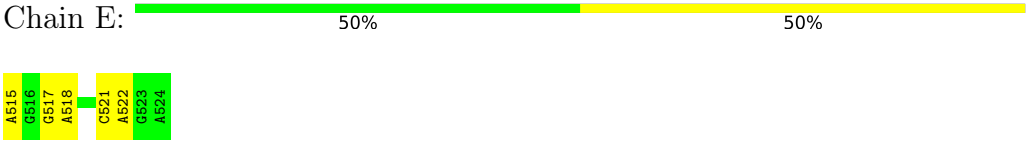
• Molecule 4: DNA (5'-D(*TP*CP*TP*GP*AP*CP*TP*CP*CP*TP*GP*TP*GP*G)-3')

Chain F:



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

Chain E:



• Molecule 6: DNA (5'-D(P*AP*GP*AP*AP*GP*TP*CP*TP*GP*A)-3')

Chain G:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	44.89Å 67.36Å 90.87Å 90.00° 96.80° 90.00°	Depositor
Resolution (Å)	41.99 – 2.95 41.99 – 2.95	Depositor EDS
% Data completeness (in resolution range)	97.1 (41.99-2.95) 97.4 (41.99-2.95)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.19 (at 2.95Å)	Xtriage
Refinement program	PHENIX 1.8.1_1168	Depositor
R, R_{free}	0.187 , 0.235 0.192 , 0.188	Depositor DCC
R_{free} test set	532 reflections (4.75%)	wwPDB-VP
Wilson B-factor (Å^2)	68.2	Xtriage
Anisotropy	0.636	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 43.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	3480	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.32% 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: MN

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	0/1258	0.39	0/1698
2	B	0.24	0/1278	0.38	0/1726
3	D	0.49	0/307	1.25	1/470 (0.2%)
4	F	0.41	0/314	1.18	1/483 (0.2%)
5	E	0.82	1/239 (0.4%)	1.06	0/366
6	G	0.81	1/236 (0.4%)	1.07	1/361 (0.3%)
All	All	0.40	2/3632 (0.1%)	0.73	3/5104 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	515	DA	OP3-P	-10.78	1.48	1.61
6	G	615	DA	OP3-P	-10.69	1.48	1.61

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	506	DC	O4'-C1'-N1	6.56	112.59	108.00
6	G	616	DG	O4'-C1'-N9	5.67	111.97	108.00
4	F	606	DC	C1'-O4'-C4'	-5.20	104.90	110.10

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	1236	0	1276	13	0
2	B	1250	0	1285	7	0
3	D	276	0	160	6	0
4	F	282	0	160	1	0
5	E	212	0	112	3	0
6	G	210	0	113	1	0
7	A	3	0	0	0	0
8	A	6	0	0	1	0
8	B	1	0	0	0	0
8	E	1	0	0	1	0
8	F	2	0	0	0	0
8	G	1	0	0	0	0
All	All	3480	0	3106	28	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.

All (28) 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:B:22:SER:HB3	2:B:44:LYS:HG2	1.81	0.62
2:B:144:THR:OG1	2:B:145:SER:N	2.33	0.62
1:A:119:PRO:HB3	1:A:152:LEU:HD21	1.83	0.61
1:A:130:ASP:OD1	1:A:141:ARG:NH2	2.33	0.60
1:A:44:ASP:OD2	1:A:75:LYS:NZ	2.35	0.59
1:A:4:LYS:HA	1:A:90:GLN:HE22	1.67	0.59
2:B:6:ASN:OD1	2:B:6:ASN:N	2.31	0.58
6:G:622:DT:H2"	6:G:623:DG:C8	2.39	0.57
1:A:22:SER:HB3	1:A:44:ASP:HB3	1.89	0.55
1:A:42:THR:OG1	8:A:301:HOH:O	2.18	0.52
1:A:26:GLN:NE2	8:E:701:HOH:O	2.34	0.51
3:D:512:DC:H2"	3:D:513:DA:C8	2.46	0.50
5:E:521:DC:H2"	5:E:522:DA:C8	2.46	0.50
1:A:109:ILE:HA	1:A:112:LEU:HD22	1.95	0.49
2:B:52:ARG:NH1	2:B:69:ASP:OD1	2.47	0.48
1:A:48:LYS:HE2	1:A:50:GLN:HB3	1.95	0.47
2:B:68:TYR:OH	3:D:506:DC:O5'	2.27	0.45
1:A:64:VAL:HG11	1:A:83:PRO:HB3	1.98	0.44
1:A:123:LEU:HD13	1:A:149:ARG:HD3	1.98	0.44
1:A:137:ASP:OD1	1:A:137:ASP:N	2.51	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:512:DC:H4'	3:D:513:DA:OP1	2.17	0.44
5:E:517:DG:H2'	5:E:518:DA:C8	2.53	0.43
4:F:610:DT:H2"	4:F:611:DG:C8	2.53	0.43
3:D:506:DC:H2'	3:D:507:DT:H71	1.99	0.43
3:D:505:DA:H2"	3:D:506:DC:O5'	2.18	0.43
1:A:77:ARG:NH2	5:E:517:DG:O6	2.50	0.43
2:B:28:LYS:HA	2:B:29:PRO:HD3	1.90	0.43
2:B:68:TYR:HH	3:D:506:DC:P	2.42	0.40

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	151/153 (99%)	143 (95%)	8 (5%)	0	100	100
2	B	153/154 (99%)	143 (94%)	10 (6%)	0	100	100
All	All	304/307 (99%)	286 (94%)	18 (6%)	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	140/140 (100%)	132 (94%)	8 (6%)	20	52

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	141/140 (101%)	135 (96%)	6 (4%)	29 62
All	All	281/280 (100%)	267 (95%)	14 (5%)	24 57

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3	THR
1	A	6	ASN
1	A	33	VAL
1	A	40	GLN
1	A	42	THR
1	A	112	LEU
1	A	141	ARG
1	A	152	LEU
2	B	6	ASN
2	B	44	LYS
2	B	73	VAL
2	B	81	ILE
2	B	123	LEU
2	B	144	THR

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 3 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

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	153/153 (100%)	0.23	7 (4%) 32 20	26, 55, 99, 123	0
2	B	154/154 (100%)	0.18	2 (1%) 77 61	25, 49, 91, 109	0
3	D	14/14 (100%)	-0.32	0 100 100	35, 57, 71, 75	0
4	F	14/14 (100%)	-0.12	0 100 100	47, 58, 77, 88	0
5	E	10/10 (100%)	-0.14	0 100 100	30, 55, 75, 79	0
6	G	10/10 (100%)	-0.25	0 100 100	22, 45, 76, 76	0
All	All	355/355 (100%)	0.15	9 (2%) 57 40	22, 54, 96, 123	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	68	TYR	4.7
1	A	120	ASP	2.6
1	A	31	GLN	2.5
1	A	33	VAL	2.3
1	A	36	LYS	2.3
1	A	70	SER	2.2
1	A	148	VAL	2.1
1	A	150	ALA	2.0
2	B	123	LEU	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 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, 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
7	MN	A	203	1/1	0.95	0.13	52,52,52,52	0
7	MN	A	202	1/1	0.99	0.09	54,54,54,54	0
7	MN	A	201	1/1	0.99	0.13	52,52,52,52	0

6.5 Other polymers (i)

There are no such residues in this entry.