wwPDB X-ray Structure Validation Summary Report (i)

Feb 3, 2024 – 01:43 PM EST

PDB ID	:	1MA7
Title	:	Crystal structure of Cre site-specific recombinase complexed with a mutant DNA substrate, LoxP-A8/T27
Authors	:	Martin, S.S.; Chu, V.C.; Baldwin, E.P.
Deposited on	:	2002-08-01
Resolution	:	2.30 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:

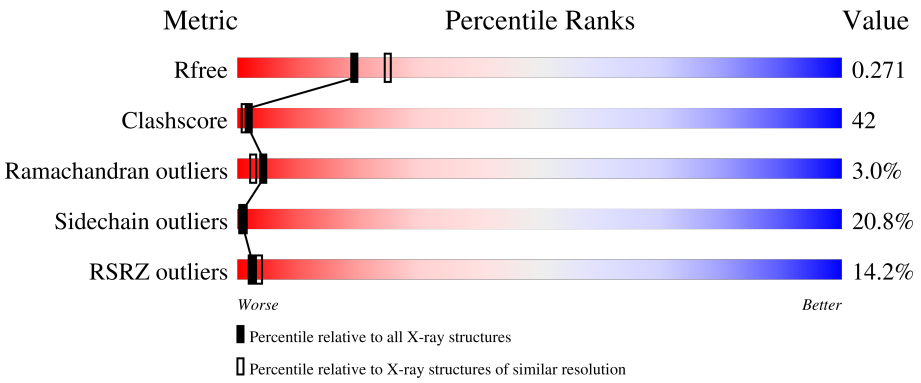
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	C	34	6%; 18%, 56%, 21%, 6%
2	D	34	6%; 32%, 53%, 15%
3	A	349	17%; 32%, 43%, 16%, • 7%
3	B	349	10%; 48%, 36%, 6%, • 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6699 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 LOXP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	32	654	317	115	191	31	0	0	0

- Molecule 2 is a DNA chain called LOXP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	34	693	336	123	201	33	0	0	0

- Molecule 3 is a protein called CRE RECOMBINASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	323	2557	1588	487	467	15	45	0	0
			Total	C	N	O	S			
3	B	317	2512	1562	481	454	15	11	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	HIS	-	expression tag	UNP P06956
A	-3	HIS	-	expression tag	UNP P06956
A	-2	HIS	-	expression tag	UNP P06956
A	-1	HIS	-	expression tag	UNP P06956
A	0	HIS	-	expression tag	UNP P06956
A	1	HIS	-	expression tag	UNP P06956
B	-4	HIS	-	expression tag	UNP P06956
B	-3	HIS	-	expression tag	UNP P06956
B	-2	HIS	-	expression tag	UNP P06956
B	-1	HIS	-	expression tag	UNP P06956
B	0	HIS	-	expression tag	UNP P06956
B	1	HIS	-	expression tag	UNP P06956

- Molecule 4 is water.

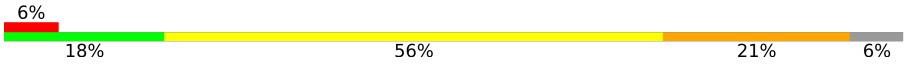
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	38	Total O 38 38	0	0
4	D	41	Total O 41 41	0	0
4	A	60	Total O 60 60	0	0
4	B	144	Total O 144 144	0	0

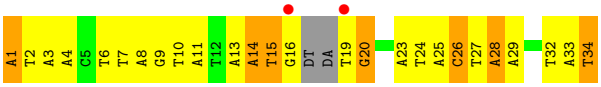
3 Residue-property plots (i)

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: LOXP

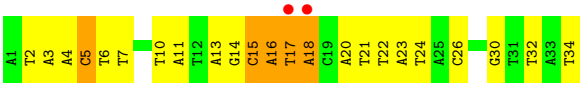
Chain C:





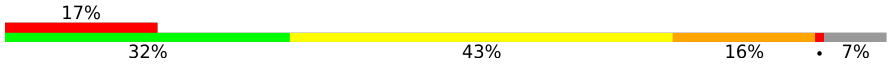
• Molecule 2: LOXP

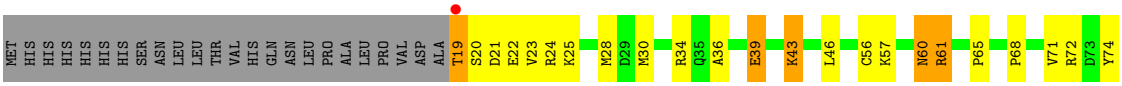
Chain D:

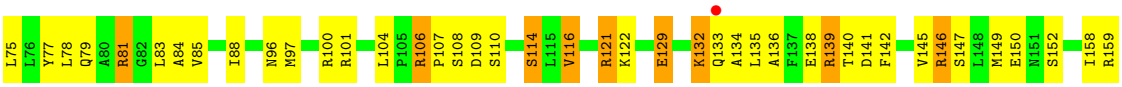


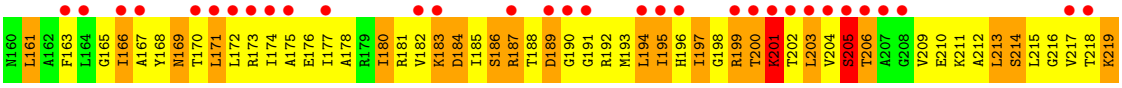
• Molecule 3: CRE RECOMBINASE

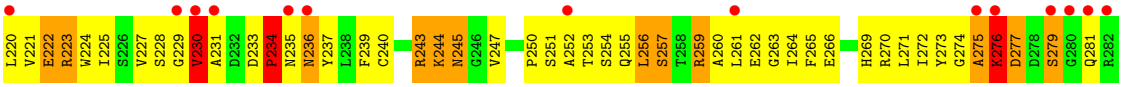
Chain A:

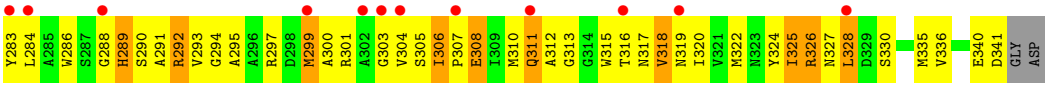






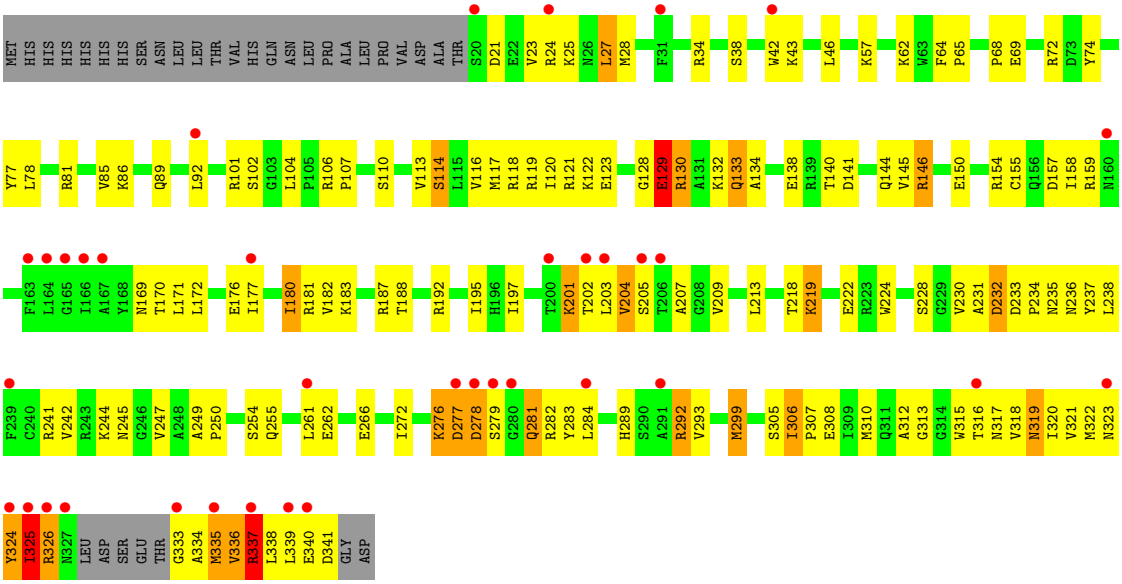






• Molecule 3: CRE RECOMBINASE

Chain B:


MET HIS HIS HIS HIS HIS HIS SER ASN LEU LEU THR VAL HIS GLN ASN LEU PRO ALA LEU PRO VAL ASP ALA THR S20 D21 E22 V23 R24 K25 N26 L27 M28 F31 R34 S38 W42 K43 L46 K57 K62 W63 F64 P65 P68 E69 R72 D73 Y74
Y77 L78 R81 V85 K86 Q89 L92 R101 S102 G103 L104 P105 R106 P107 S110 V113 S114 L115 V116 M117 R118 R119 I120 R121 K122 E123 G128 E129 R130 A131 K132 Q133 A134 E138 R139 T140 D141 Q144 V145 R146 E150 R154 C155 Q156 D157 I158 R159 N160
F163 L164 G165 I166 A167 Y168 N169 T170 L171 L172 E176 I177 I180 R181 V182 K183 R187 T188 R192 I195 H196 I197 T200 K201 T202 L203 V204 S205 T206 A207 G208 V209 L213 T218 K219 E222 R223 W224 S228 G229 V230 A231 D232 D233 P234 N235 N236 Y237 L238
F239 C240 R241 V242 R243 K244 N245 G246 V247 A248 A249 P250 S254 Q255 L261 E262 E266 I272 K276 D277 D278 S279 G280 Q281 R282 Y283 L284 H289 S290 A291 R292 V293 M299 S305 I306 P307 E308 I309 M310 Q311 A312 G313 G314 W315 T316 N317 V318 N319 I320 V321 M322 N323
Y324 I325 R326 N327 LEU ASP SER GLU THR G333 A334 M335 V336 R337 L338 L339 E340 D341 GLY ASP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	107.54Å 122.24Å 180.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	5.00 – 2.30 29.84 – 2.25	Depositor EDS
% Data completeness (in resolution range)	97.0 (5.00-2.30) 96.9 (29.84-2.25)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	10.55 (at 2.24Å)	Xtriage
Refinement program	TNT	Depositor
R, R_{free}	0.239 , 0.290 0.241 , 0.271	Depositor DCC
R_{free} test set	2787 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	45.1	Xtriage
Anisotropy	0.167	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 72.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6699	wwPDB-VP
Average B, all atoms (Å^2)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.71% 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)

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	C	0.58	0/732	1.35	9/1126 (0.8%)
2	D	0.65	0/777	1.42	11/1197 (0.9%)
3	A	0.47	0/2598	0.74	1/3503 (0.0%)
3	B	0.44	1/2552 (0.0%)	0.65	0/3438
All	All	0.50	1/6659 (0.0%)	0.92	21/9264 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	114	SER	CB-OG	-5.57	1.35	1.42

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	15	DC	O4'-C1'-N1	11.32	115.92	108.00
2	D	18	DA	O4'-C1'-N9	-10.29	100.79	108.00
1	C	4	DA	O4'-C1'-N9	7.97	113.58	108.00
1	C	34	DT	O4'-C1'-N1	6.97	112.88	108.00
2	D	32	DT	O4'-C1'-N1	6.76	112.73	108.00

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	C	654	0	368	49	0
2	D	693	0	389	38	0
3	A	2557	0	2578	290	0
3	B	2512	0	2537	177	0
4	A	60	0	0	8	0
4	B	144	0	0	2	0
4	C	38	0	0	2	0
4	D	41	0	0	1	0
All	All	6699	0	5872	505	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 42.

The worst 5 of 505 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)
3:A:235:ASN:HB2	3:A:252:ALA:HB1	1.15	1.14
2:D:23:DA:H5'	3:B:201:LYS:HB2	1.31	1.08
3:A:169:ASN:HB2	3:A:217:VAL:HG21	1.40	1.03
3:A:182:VAL:HB	3:A:231:ALA:HB2	1.40	1.03
3:A:202:THR:HB	3:A:205:SER:HB2	1.33	1.03

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	321/349 (92%)	273 (85%)	37 (12%)	11 (3%)	3 2
3	B	313/349 (90%)	283 (90%)	22 (7%)	8 (3%)	5 4
All	All	634/698 (91%)	556 (88%)	59 (9%)	19 (3%)	4 2

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	188	THR
3	A	201	LYS
3	A	205	SER
3	A	230	VAL
3	B	129	GLU

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	270/293 (92%)	197 (73%)	73 (27%)	0 0
3	B	264/293 (90%)	226 (86%)	38 (14%)	3 3
All	All	534/586 (91%)	423 (79%)	111 (21%)	1 1

5 of 111 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	271	LEU
3	B	338	LEU
3	A	330	SER
3	B	337	ARG
3	B	281	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	89	GLN
3	B	133	GLN
3	B	323	ASN
3	B	269	HIS
3	A	269	HIS

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)

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	C	32/34 (94%)	0.00	2 (6%)	20	25	33, 51, 69, 93	0
2	D	34/34 (100%)	-0.19	2 (5%)	22	28	33, 48, 83, 98	0
3	A	320/349 (91%)	1.11	60 (18%)	1	1	29, 65, 93, 99	7 (2%)
3	B	317/349 (90%)	0.55	36 (11%)	5	7	24, 44, 85, 100	3 (0%)
All	All	703/766 (91%)	0.74	100 (14%)	2	3	24, 54, 90, 100	10 (1%)

The worst 5 of 100 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	204	VAL	13.1
3	A	199	ARG	9.5
3	A	207	ALA	9.0
3	A	283	TYR	8.5
3	A	276	LYS	8.4

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 monosaccharides 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.