



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2017 – 09:10 AM EDT

PDB ID : 2HOI
Title : Crystal structure of the tetrameric pre-cleavage synaptic complex in the cre-loxp site-specific recombination
Authors : Ghosh, K.; Van Duyne, G.D.
Deposited on : unknown
Resolution : 2.60 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

<http://wwpdb.org/validation/2016/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) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030345

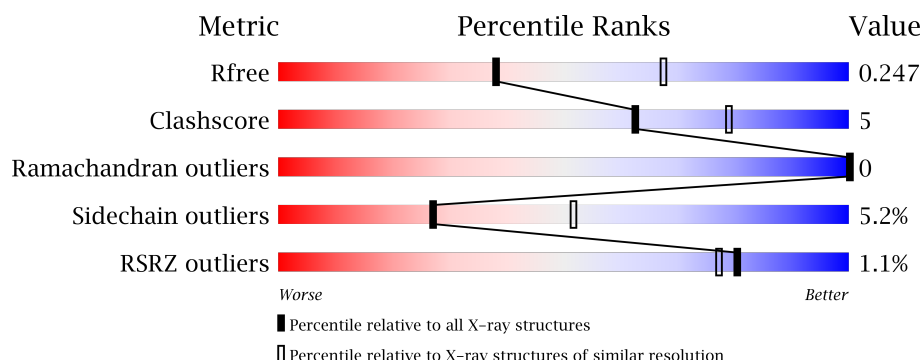
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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}	100719	2542 (2.60-2.60)
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)
RSRZ outliers	101464	2550 (2.60-2.60)

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. 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	C	35	<div> <div>3%</div> <div>57%</div> <div>43%</div> </div>
1	E	35	<div> <div>66%</div> <div>31%</div> <div>.</div> </div>
2	D	35	<div> <div>3%</div> <div>77%</div> <div>20%</div> <div>.</div> </div>
2	F	35	<div> <div>3%</div> <div>71%</div> <div>26%</div> <div>.</div> </div>
3	A	343	<div> <div>80%</div> <div>12%</div> <div>6%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	B	343	<div><div></div><div>83%</div><div>11%</div><div>6%</div></div>
3	G	343	<div>%<div><div></div><div>80%</div><div>12%</div><div>• 6%</div></div></div>
3	H	343	<div>%<div><div></div><div>82%</div><div>10%</div><div>• 6%</div></div></div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13811 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 DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	35	Total	C	N	O	P	0	0	0
			710	344	124	208	34			
1	E	34	Total	C	N	O	P	0	0	0
			693	334	122	203	34			

- Molecule 2 is a DNA chain called LoxP DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	34	Total	C	N	O	P	0	0	0
			701	337	125	205	34			
2	F	34	Total	C	N	O	P	0	0	0
			701	337	125	205	34			

- Molecule 3 is a protein called Recombinase Cre.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	322	Total	C	N	O	S	0	0	0
			2546	1581	485	465	15			
3	B	322	Total	C	N	O	S	0	0	0
			2546	1581	485	465	15			
3	G	322	Total	C	N	O	S	0	0	0
			2546	1581	485	465	15			
3	H	322	Total	C	N	O	S	0	0	0
			2546	1581	485	465	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	ALA	LYS	ENGINEERED	UNP P06956
B	201	ALA	LYS	ENGINEERED	UNP P06956
G	201	ALA	LYS	ENGINEERED	UNP P06956
H	201	ALA	LYS	ENGINEERED	UNP P06956

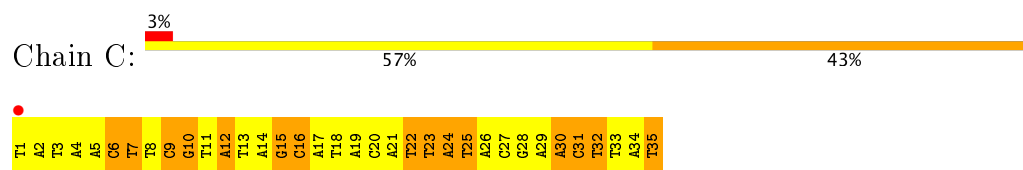
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	50	Total 50	O 50	0	0
4	D	42	Total 42	O 42	0	0
4	E	36	Total 36	O 36	0	0
4	F	50	Total 50	O 50	0	0
4	A	168	Total 168	O 168	0	0
4	B	177	Total 177	O 177	0	0
4	G	144	Total 144	O 144	0	0
4	H	155	Total 155	O 155	0	0

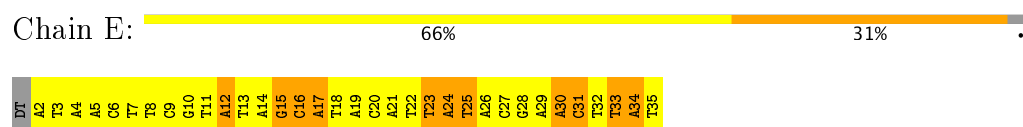
3 Residue-property plots [i](#)

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.

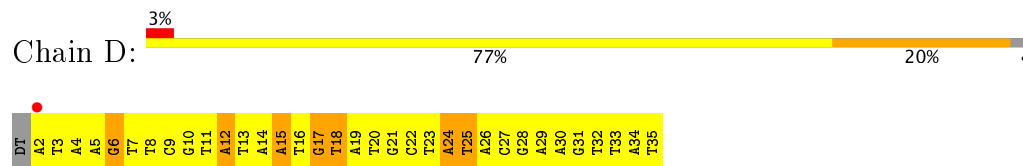
- Molecule 1: LoxP DNA



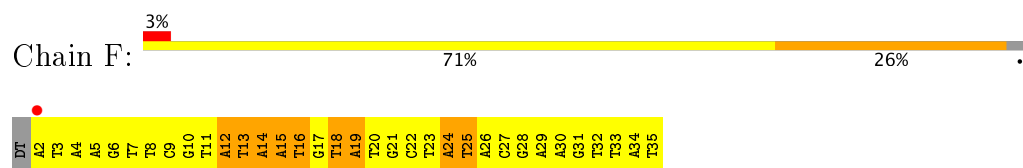
- Molecule 1: LoxP DNA



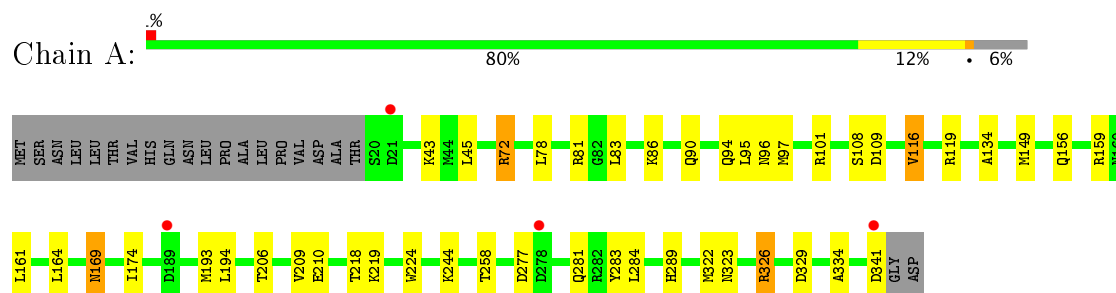
- Molecule 2: LoxP DNA




- Molecule 2: LoxP DNA

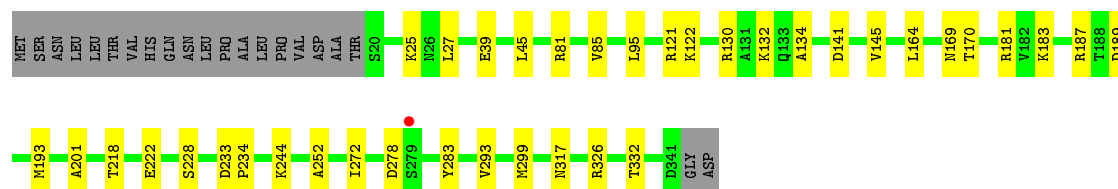


- Molecule 3: Recombinase Cre




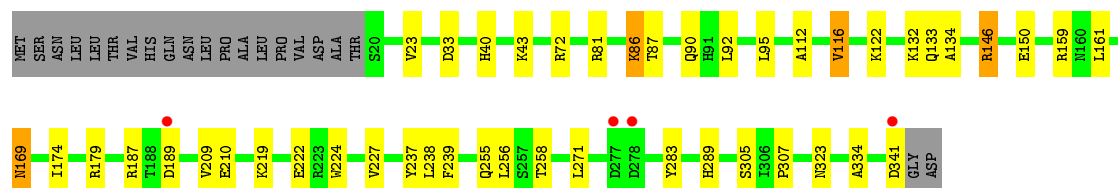
- Molecule 3: Recombinase Cre

Chain B:  83% 11% 6%




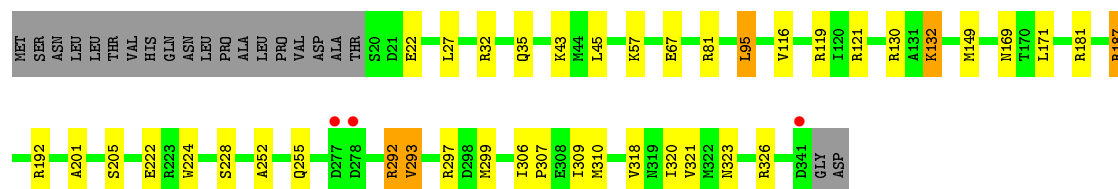
• Molecule 3: Recombinase Cre

Chain G:  80% 12% 6%



• Molecule 3: Recombinase Cre

Chain H:  82% 10% 6%



4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	136.78Å 136.78Å 218.28Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.92 – 2.60 19.92 – 2.60	Depositor EDS
% Data completeness (in resolution range)	81.5 (19.92-2.60) 81.5 (19.92-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	9.02 (at 2.59Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.184 , 0.246 0.187 , 0.247	Depositor DCC
R_{free} test set	3007 reflections (5.33%)	DCC
Wilson B-factor (Å ²)	32.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 48.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.018 for -h,-k,l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	13811	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.37% 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

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	1.07	3/795 (0.4%)	1.93	60/1224 (4.9%)
1	E	1.09	4/776 (0.5%)	1.83	55/1194 (4.6%)
2	D	1.04	2/786 (0.3%)	1.90	51/1212 (4.2%)
2	F	1.01	1/786 (0.1%)	1.91	57/1212 (4.7%)
3	A	0.38	0/2587	0.58	1/3489 (0.0%)
3	B	0.36	0/2587	0.58	0/3489
3	G	0.35	0/2587	0.57	0/3489
3	H	0.37	0/2587	0.61	0/3489
All	All	0.60	10/13491 (0.1%)	1.09	224/18798 (1.2%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	1	DT	C5-C7	8.11	1.54	1.50
2	D	22	DC	N1-C2	-7.00	1.33	1.40
1	E	27	DC	N1-C2	-6.54	1.33	1.40
2	D	27	DC	N1-C2	-5.88	1.34	1.40
2	F	22	DC	N1-C2	-5.78	1.34	1.40

The worst 5 of 224 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	20	DC	O4'-C1'-N1	10.33	115.23	108.00
2	F	17	DG	O4'-C4'-C3'	-9.91	100.05	106.00
2	D	17	DG	O4'-C4'-C3'	-8.54	100.88	106.00
2	D	26	DA	O4'-C4'-C3'	-8.44	100.94	106.00
2	F	6	DG	O4'-C4'-C3'	-7.72	101.37	106.00

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	C	710	0	400	11	0
1	E	693	0	387	11	0
2	D	701	0	388	6	0
2	F	701	0	388	8	0
3	A	2546	0	2563	37	0
3	B	2546	0	2563	22	0
3	G	2546	0	2563	29	0
3	H	2546	0	2563	25	0
4	A	168	0	0	4	0
4	B	177	0	0	4	0
4	C	50	0	0	3	0
4	D	42	0	0	0	0
4	E	36	0	0	2	0
4	F	50	0	0	4	0
4	G	144	0	0	2	0
4	H	155	0	0	3	0
All	All	13811	0	11815	124	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 124 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)
4:F:428:HOH:O	3:H:121:ARG:HD2	1.69	0.92
1:C:22:DT:H72	4:A:430:HOH:O	1.70	0.91
3:A:72:ARG:HG3	3:A:116:VAL:HG11	1.56	0.86
3:A:72:ARG:HG3	3:A:116:VAL:CG1	2.05	0.85
4:C:60:HOH:O	3:A:97:MET:SD	2.36	0.82

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	320/343 (93%)	318 (99%)	2 (1%)	0	100	100
3	B	320/343 (93%)	319 (100%)	1 (0%)	0	100	100
3	G	320/343 (93%)	314 (98%)	6 (2%)	0	100	100
3	H	320/343 (93%)	317 (99%)	3 (1%)	0	100	100
All	All	1280/1372 (93%)	1268 (99%)	12 (1%)	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	
3	A	268/286 (94%)	253 (94%)	15 (6%)	25	48
3	B	268/286 (94%)	258 (96%)	10 (4%)	39	66
3	G	268/286 (94%)	253 (94%)	15 (6%)	25	48
3	H	268/286 (94%)	252 (94%)	16 (6%)	22	44
All	All	1072/1144 (94%)	1016 (95%)	56 (5%)	27	52

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	86	LYS
3	G	161	LEU
3	H	187	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	G	92	LEU
3	G	116	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	323	ASN
3	B	60	ASN
3	G	323	ASN
3	A	311	GLN
3	H	255	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 carbohydrates 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 ⓘ

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, 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	35/35 (100%)	-0.23	1 (2%) 52 45	16, 30, 60, 77	0
1	E	34/35 (97%)	-0.44	0 100 100	15, 31, 57, 62	0
2	D	34/35 (97%)	-0.55	1 (2%) 52 45	17, 31, 48, 74	0
2	F	34/35 (97%)	-0.55	1 (2%) 52 45	16, 33, 55, 79	0
3	A	322/343 (93%)	-0.49	4 (1%) 79 75	18, 28, 39, 50	0
3	B	322/343 (93%)	-0.54	1 (0%) 93 93	19, 28, 39, 42	0
3	G	322/343 (93%)	-0.42	4 (1%) 79 75	17, 32, 49, 54	0
3	H	322/343 (93%)	-0.55	3 (0%) 84 81	16, 26, 38, 46	0
All	All	1425/1512 (94%)	-0.49	15 (1%) 80 77	15, 29, 43, 79	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1	DT	5.5
3	H	341	ASP	5.0
3	B	279	SER	3.0
3	A	21	ASP	2.7
2	D	2	DA	2.6

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 carbohydrates 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.