



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 01:04 am GMT

PDB ID : 1KBU  
Title : CRE RECOMBINASE BOUND TO A LOXP HOLLIDAY JUNCTION  
Authors : Martin, S.S.; Pulido, E.; Chu, V.C.; Lechner, T.; Baldwin, E.P.  
Deposited on : 2001-11-06  
Resolution : 2.20 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](mailto: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	:	trunk28620
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)	:	recalc28949

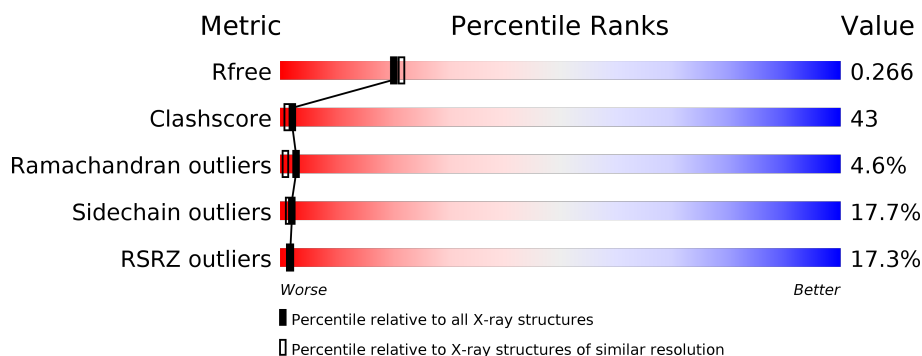
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4002 (2.20-2.20)
Clashscore	112137	4730 (2.20-2.20)
Ramachandran outliers	110173	4656 (2.20-2.20)
Sidechain outliers	110143	4657 (2.20-2.20)
RSRZ outliers	101464	4033 (2.20-2.20)

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  $\geq 3$ , 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	34	<div> <div>12%</div> <div>12% 53% 35%</div> </div>
2	D	34	<div> <div>9%</div> <div>21% 41% 38%</div> </div>
3	A	349	<div> <div>22%</div> <div>39% 38% 13% 7%</div> </div>
3	B	349	<div> <div>11%</div> <div>46% 35% 9% 9%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6837 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
1	C	34	Total	C	N	O	P	0	4	0
			740	357	132	216	35			

- Molecule 2 is a DNA chain called LOXP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	34	Total	C	N	O	P	0	4	0
			771	373	137	224	37			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	30	DC	G	see remark 999	GB 215626

- Molecule 3 is a protein called CRE RECOMBINASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	324	Total	C	N	O	S	54	0	0
			2562	1591	488	468	15			
3	B	319	Total	C	N	O	S	12	0	0
			2527	1572	483	457	15			

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

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
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

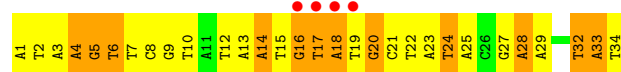
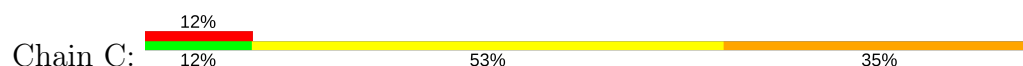
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	59	Total	O	0	0
			59	59		
4	B	124	Total	O	0	0
			124	124		
4	C	20	Total	O	0	0
			20	20		
4	D	34	Total	O	0	0
			34	34		

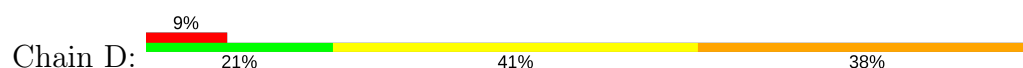
### 3 Residue-property plots

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.

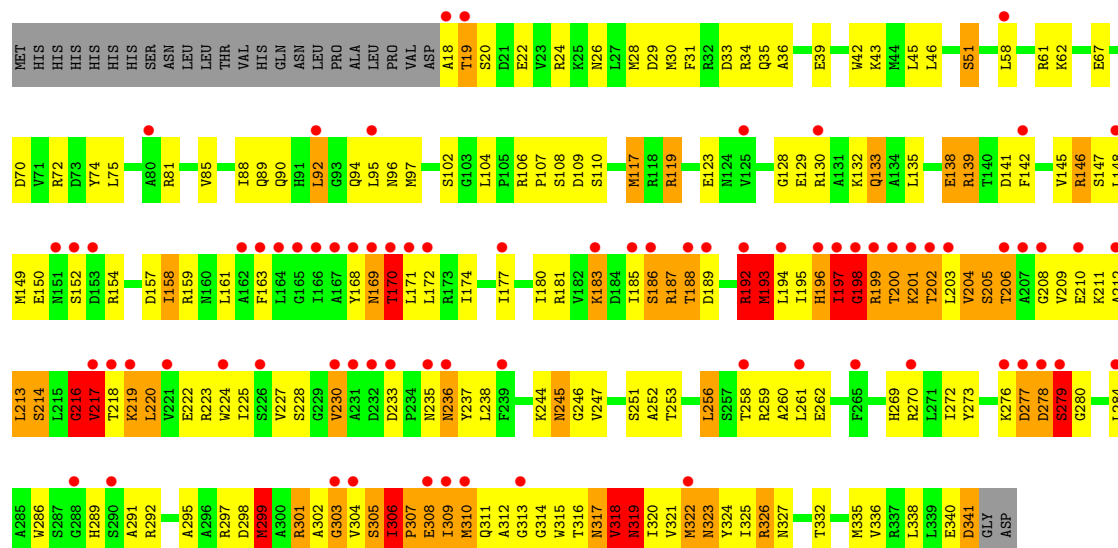
#### • Molecule 1: LOXP



#### • Molecule 2: LOXP

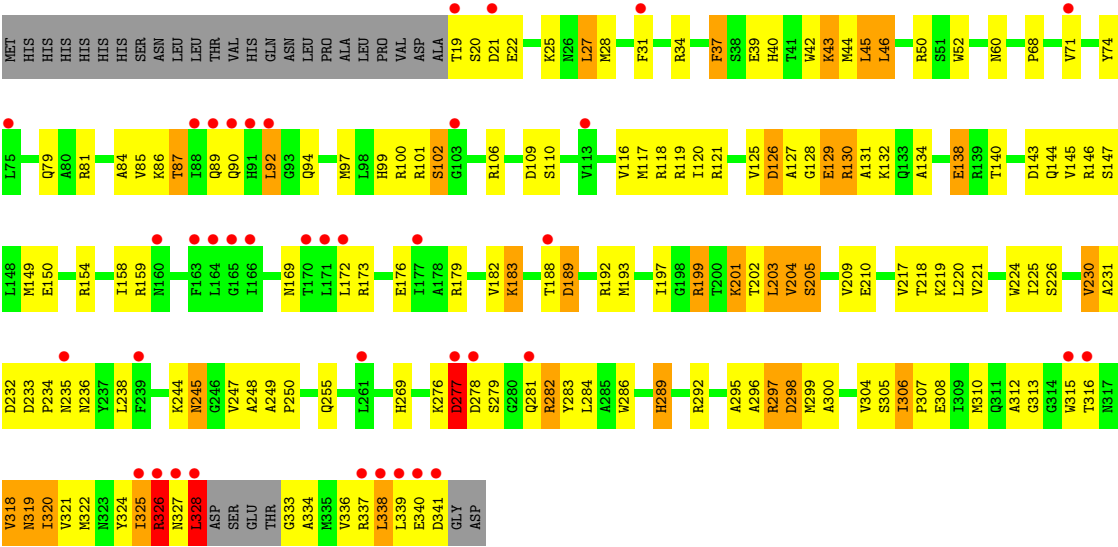


#### • Molecule 3: CRE RECOMBINASE



#### • Molecule 3: CRE RECOMBINASE





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.17Å 121.60Å 179.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	5.00 – 2.20 24.52 – 2.20	Depositor EDS
% Data completeness (in resolution range)	89.0 (5.00-2.20) 89.3 (24.52-2.20)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	0.04	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.47 (at 2.19Å)	Xtriage
Refinement program	TNT	Depositor
R, $R_{free}$	0.231 , 0.279 0.238 , 0.266	Depositor DCC
$R_{free}$ test set	2682 reflections (5.31%)	DCC
Wilson B-factor (Å <sup>2</sup> )	48.0	Xtriage
Anisotropy	0.289	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 83.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6837	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.07% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.95	3/830 (0.4%)	2.60	27/1281 (2.1%)
2	D	0.54	0/864	1.67	24/1330 (1.8%)
3	A	0.42	0/2603	1.05	16/3510 (0.5%)
3	B	0.42	0/2567	0.89	5/3459 (0.1%)
All	All	0.53	3/6864 (0.0%)	1.41	72/9580 (0.8%)

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
3	A	1	0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	16[A]	DG	O3'-P	13.54	1.77	1.61
1	C	16[B]	DG	O3'-P	13.54	1.77	1.61
1	C	18[B]	DA	O3'-P	9.57	1.72	1.61

The worst 5 of 72 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	16[A]	DG	P-O3'-C3'	40.18	167.91	119.70
1	C	16[B]	DG	P-O3'-C3'	40.18	167.91	119.70
1	C	18[B]	DA	OP1-P-O3'	-36.31	25.32	105.20
3	A	301	ARG	NE-CZ-NH2	-21.76	109.42	120.30
1	C	18[B]	DA	OP2-P-O3'	16.78	142.12	105.20

All (1) chirality outliers are listed below:



Mol	Chain	Res	Type	Atom
3	A	217	VAL	CA

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	740	0	413	45	0
2	D	771	0	433	69	0
3	A	2562	0	2583	270	0
3	B	2527	0	2555	212	0
4	A	59	0	0	5	0
4	B	124	0	0	11	0
4	C	20	0	0	3	0
4	D	34	0	0	0	0
All	All	6837	0	5984	533	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 43.

The worst 5 of 533 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:C:18[B]:DA:H1'	1:C:19[B]:DT:H5''	1.19	1.16
2:D:16[B]:DA:H1'	2:D:17[B]:DT:H5''	1.23	1.15
3:A:188:THR:HG21	3:A:194:LEU:HG	1.22	1.07
2:D:33:DA:H2''	2:D:34:DT:H5''	1.33	1.06
3:A:309:ILE:HG22	3:A:321:VAL:HG11	1.35	1.04

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	322/349 (92%)	269 (84%)	32 (10%)	21 (6%)	1	0
3	B	315/349 (90%)	288 (91%)	19 (6%)	8 (2%)	6	3
All	All	637/698 (91%)	557 (87%)	51 (8%)	29 (5%)	3	1

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	188	THR
3	A	189	ASP
3	A	192	ARG
3	A	193	MET
3	A	236	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	
3	A	270/293 (92%)	208 (77%)	62 (23%)	1	0
3	B	266/293 (91%)	233 (88%)	33 (12%)	5	4
All	All	536/586 (92%)	441 (82%)	95 (18%)	2	1

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	245	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	A	316	THR
3	B	289	HIS
3	A	253	THR
3	A	279	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	245	ASN
3	A	269	HIS
3	B	255	GLN
3	A	235	ASN
3	B	90	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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	C	34/34 (100%)	0.52	4 (11%) 5 4	32, 50, 86, 90	2 (5%)
2	D	34/34 (100%)	0.03	3 (8%) 11 9	33, 51, 80, 82	0
3	A	320/349 (91%)	1.29	76 (23%) 1 1	30, 65, 96, 99	9 (2%)
3	B	319/349 (91%)	0.60	39 (12%) 5 4	27, 46, 84, 98	3 (0%)
All	All	707/766 (92%)	0.88	122 (17%) 2 2	27, 55, 94, 99	14 (1%)

The worst 5 of 122 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	304	VAL	15.3
1	C	17[B]	DT	11.2
3	A	207	ALA	8.8
3	B	328	LEU	8.2
3	A	18	ALA	8.1

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.