



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 6, 2023 – 01:52 AM EDT

PDB ID : 4CRX  
Title : ASYMMETRIC DNA-BENDING IN THE CRE-LOXP SITE-SPECIFIC RE-COMBINATION SYNAPSE  
Authors : Guo, F.; Gopaul, D.N.; Van Duyne, G.D.  
Deposited on : 1999-04-20  
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

<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 : 2.35  
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) : 2.35

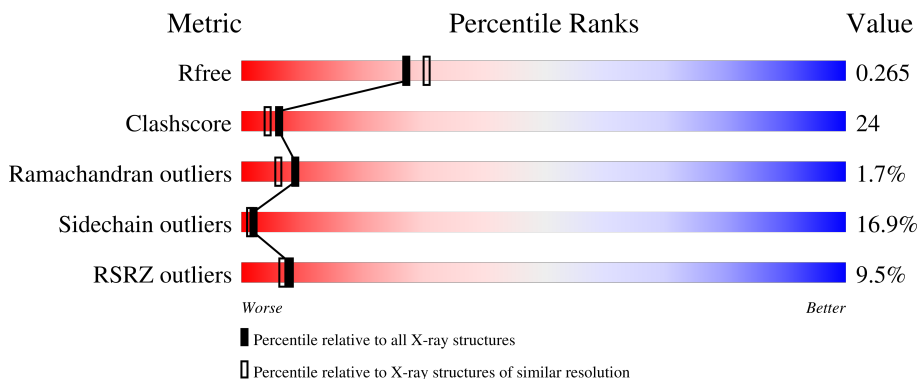
# 1 Overall quality at a glance

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))
$R_{free}$	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	C	35	 11% 69% 31%
1	D	35	 3% 54% 37% 9%
2	A	322	 11% 58% 32% 9% .
2	B	322	 9% 59% 32% 8% .

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6999 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 DNA (35 NUCLEOTIDE CRE RECOGNITION SITE).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	35	714	345	126	209	34	0	0	0
1	D	35	714	345	126	209	34	0	0	0

- Molecule 2 is a protein called PROTEIN (CRE RECOMBINASE).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	322	2549	1584	484	466	15	0	0	0
2	B	322	2549	1584	484	466	15	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	173	LYS	ARG	engineered mutation	UNP P06956
B	173	LYS	ARG	engineered mutation	UNP P06956

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	71	Total 71	O 71	0	0
3	D	87	Total 87	O 87	0	0
3	A	149	Total 149	O 149	0	0
3	B	166	Total 166	O 166	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.

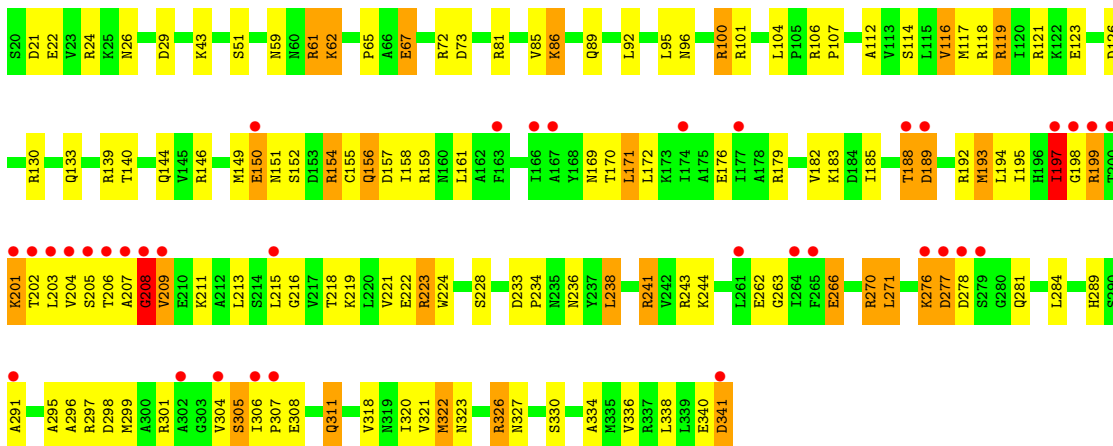
- Molecule 1: DNA (35 NUCLEOTIDE CRE RECOGNITION SITE)



- Molecule 1: DNA (35 NUCLEOTIDE CRE RECOGNITION SITE)

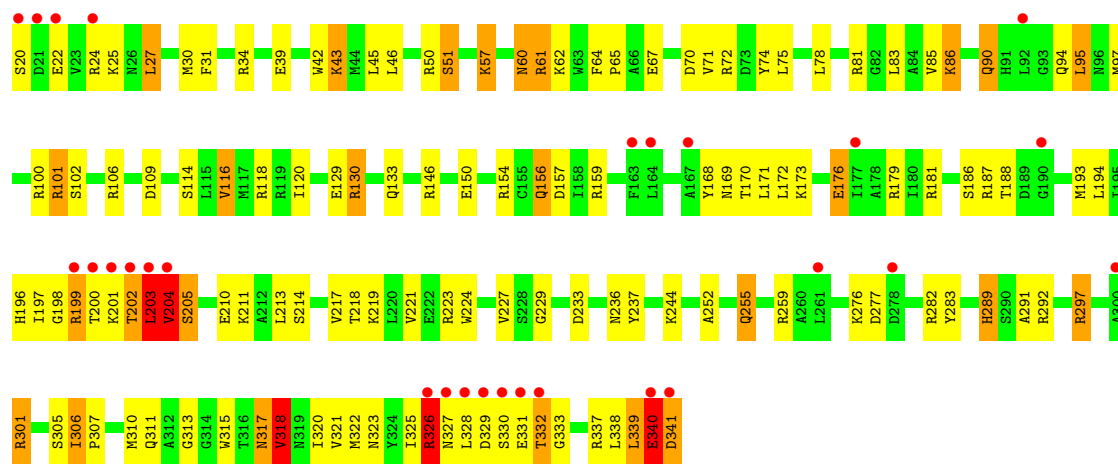


- Molecule 2: PROTEIN (CRE RECOMBINASE)



- Molecule 2: PROTEIN (CRE RECOMBINASE)





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.80Å 122.10Å 181.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.30 – 2.20 29.15 – 2.20	Depositor EDS
% Data completeness (in resolution range)	91.7 (32.30-2.20) 91.0 (29.15-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.88 (at 2.20Å)	Xtrriage
Refinement program	X-PLOR 3.851	Depositor
R, $R_{free}$	0.213 , 0.251 0.232 , 0.265	Depositor DCC
$R_{free}$ test set	2822 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.3	Xtrriage
Anisotropy	0.148	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 77.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6999	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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 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	0.52	0/800	0.91	0/1233
1	D	0.72	3/800 (0.4%)	1.22	8/1233 (0.6%)
2	A	0.49	1/2590 (0.0%)	0.88	7/3490 (0.2%)
2	B	4.77	5/2589 (0.2%)	1.33	10/3487 (0.3%)
All	All	2.98	9/6779 (0.1%)	1.11	25/9443 (0.3%)

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	D	0	1
2	A	0	3
2	B	0	2
All	All	0	6

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	341	ASP	C-OXT	238.34	5.76	1.23
2	B	340	GLU	C-N	-27.33	0.71	1.34
2	B	204	VAL	C-N	21.66	1.83	1.34
2	B	326	ARG	C-N	18.94	1.77	1.34
2	A	208	GLY	C-N	-16.92	0.95	1.34

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	340	GLU	O-C-N	-58.24	29.51	122.70
2	A	208	GLY	O-C-N	-25.14	82.47	122.70
2	B	204	VAL	O-C-N	-21.72	87.95	122.70
2	A	197	ILE	O-C-N	-18.18	92.30	123.20

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	1	DT	C1'-O4'-C4'	-16.57	93.53	110.10

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	A	197	ILE	Mainchain
2	A	208	GLY	Peptide,Mainchain
2	B	340	GLU	Peptide,Mainchain
1	D	17	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	C	714	0	400	16	0
1	D	714	0	400	28	2
2	A	2549	0	2568	141	3
2	B	2549	0	2566	153	7
3	A	149	0	0	13	0
3	B	166	0	0	12	7
3	C	71	0	0	6	0
3	D	87	0	0	1	0
All	All	6999	0	5934	304	10

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 24.

The worst 5 of 304 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:340:GLU:CA	2:B:341:ASP:N	1.76	1.45
2:B:326:ARG:C	2:B:327:ASN:N	1.77	1.36
2:A:306:ILE:CG2	2:A:307:PRO:HD3	1.58	1.33
2:A:207:ALA:CB	2:B:130:ARG:HG2	1.56	1.33
2:B:204:VAL:C	2:B:205:SER:N	1.83	1.30



The worst 5 of 10 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:329:ASP:OD2	3:B:440:HOH:O[8_556]	0.72	1.48
1:D:1:DT:O5'	3:B:449:HOH:O[6_655]	0.86	1.34
2:A:126:ASP:OD1	2:B:199:ARG:NE[4_566]	1.47	0.73
2:B:329:ASP:CG	3:B:440:HOH:O[8_556]	1.48	0.72
2:A:126:ASP:OD1	2:B:199:ARG:CD[4_566]	1.94	0.26

## 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	
2	A	320/322 (99%)	302 (94%)	14 (4%)	4 (1%)	12	9
2	B	318/322 (99%)	297 (93%)	14 (4%)	7 (2%)	6	4
All	All	638/644 (99%)	599 (94%)	28 (4%)	11 (2%)	9	6

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	150	GLU
2	A	209	VAL
2	B	203	LEU
2	B	205	SER
2	B	318	VAL

### 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	
2	A	269/269 (100%)	222 (82%)	47 (18%)	2	1
2	B	269/269 (100%)	225 (84%)	44 (16%)	2	2
All	All	538/538 (100%)	447 (83%)	91 (17%)	2	1

5 of 91 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	62	LYS
2	B	187	ARG
2	B	90	GLN
2	B	118	ARG
2	B	255	GLN

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

Mol	Chain	Res	Type
2	B	133	GLN
2	B	196	HIS
2	A	156	GLN
2	A	269	HIS
2	A	323	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 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	5
2	A	1

The worst 5 of 6 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	332:THR	C	333:GLY	N	2.17
1	B	204:VAL	C	205:SER	N	1.83
1	B	326:ARG	C	327:ASN	N	1.77
1	B	198:GLY	C	199:ARG	N	1.04
1	A	208:GLY	C	209:VAL	N	0.95

## 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	35/35 (100%)	0.06	4 (11%) 5 4	22, 36, 68, 92	0
1	D	35/35 (100%)	-0.20	1 (2%) 51 49	22, 31, 48, 63	0
2	A	322/322 (100%)	0.71	35 (10%) 5 5	19, 40, 77, 99	0
2	B	322/322 (100%)	0.60	28 (8%) 10 8	16, 33, 70, 97	0
All	All	714/714 (100%)	0.59	68 (9%) 8 7	16, 36, 75, 99	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	200	THR	19.2
2	A	198	GLY	16.0
2	A	200	THR	14.6
2	A	204	VAL	14.1
2	A	207	ALA	13.9

### 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.