



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 04:02 pm BST

PDB ID : 1U0D  
Title : Y33H Mutant of Homing endonuclease I-CreI  
Authors : Sussman, D.; Chadsey, M.; Fauce, S.; Engel, A.; Bruett, A.; Monnat, R.;  
Stoddard, B.L.; Seligman, L.M.  
Deposited on : 2004-07-13  
Resolution : 2.90 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 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	:	2.11
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)	:	2.11

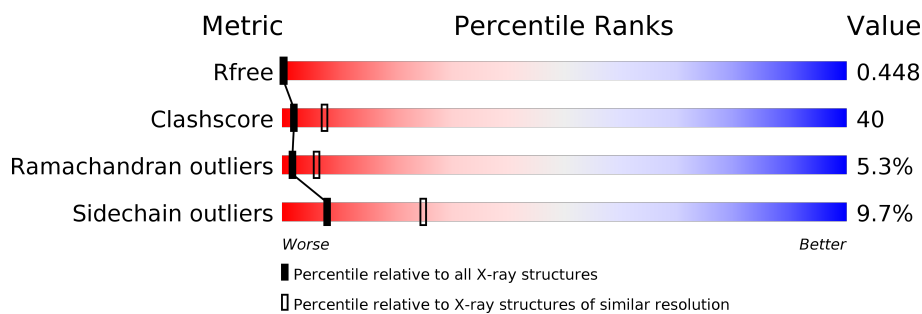
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)

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 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	1-C	24	<div> <div>25%</div> <div>71%</div> <div>.</div> </div>
1	2-C	24	<div> <div>46%</div> <div>50%</div> <div>.</div> </div>
2	1-D	24	<div> <div>13%</div> <div>83%</div> <div>.</div> </div>
2	2-D	24	<div> <div>25%</div> <div>75%</div> </div>
3	1-A	163	<div> <div>44%</div> <div>44%</div> <div>6%</div> <div>7%</div> </div>
3	1-B	163	<div> <div>35%</div> <div>46%</div> <div>11%</div> <div>7%</div> <div>.</div> </div>
3	2-A	163	<div> <div>44%</div> <div>44%</div> <div>6%</div> <div>7%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	2-B	163	<div><div></div><div>33%</div><div>48%</div><div>11%</div><div>• 7%</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 6892 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 5'-D(\*GP\*CP\*GP\*AP\*AP\*AP\*CP\*GP\*TP\*CP\*GP\*T  
P\*GP\*AP\*GP\*AP\*CP\*AP\*GP\*TP\*TP\*CP\*CP\*G)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	1-C	24	Total	C	N	O	P	0	0	0
			493	234	96	140	23			
1	2-C	24	Total	C	N	O	P	0	0	0
			493	234	96	140	23			

- Molecule 2 is a DNA chain called 5'-D(\*CP\*GP\*GP\*AP\*AP\*CP\*TP\*GP\*TP\*CP\*TP\*C  
P\*AP\*CP\*GP\*AP\*CP\*GP\*TP\*TP\*TP\*CP\*GP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	1-D	24	Total	C	N	O	P	0	0	0
			485	232	86	144	23			
2	2-D	24	Total	C	N	O	P	0	0	0
			485	232	86	144	23			

- Molecule 3 is a protein called DNA endonuclease I-CreI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	1-A	152	Total	C	N	O	S	0	0	0
			1234	793	211	229	1			
3	2-A	152	Total	C	N	O	S	0	0	0
			1234	793	211	229	1			
3	1-B	152	Total	C	N	O	S	0	0	0
			1234	793	211	229	1			
3	2-B	152	Total	C	N	O	S	0	0	0
			1234	793	211	229	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	33	HIS	TYR	ENGINEERED MUTATION	UNP P05725
A	42	THR	ALA	ENGINEERED MUTATION	UNP P05725

*Continued on next page...*

*Continued from previous page...*

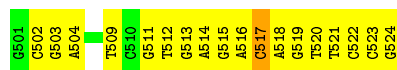
Chain	Residue	Modelled	Actual	Comment	Reference
A	47	GLU	GLN	ENGINEERED MUTATION	UNP P05725
A	110	GLU	TRP	ENGINEERED MUTATION	UNP P05725
A	111	GLN	ARG	ENGINEERED MUTATION	UNP P05725
B	333	HIS	TYR	ENGINEERED MUTATION	UNP P05725
B	342	THR	ALA	ENGINEERED MUTATION	UNP P05725
B	347	GLU	GLN	ENGINEERED MUTATION	UNP P05725
B	410	GLU	TRP	ENGINEERED MUTATION	UNP P05725
B	411	GLN	ARG	ENGINEERED MUTATION	UNP P05725

### 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. 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. 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.

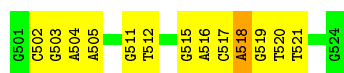
- Molecule 1: 5'-D(\*GP\*CP\*GP\*AP\*AP\*AP\*CP\*GP\*TP\*CP\*GP\*TP\*GP\*AP\*GP\*AP\*CP\*AP\*GP\*TP\*TP\*CP\*CP\*G)-3'

Chain 1-C: 



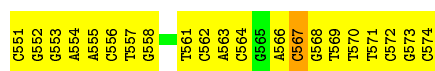
- Molecule 1: 5'-D(\*GP\*CP\*GP\*AP\*AP\*AP\*CP\*GP\*TP\*CP\*GP\*TP\*GP\*AP\*GP\*AP\*CP\*AP\*GP\*TP\*TP\*CP\*CP\*G)-3'

Chain 2-C: 



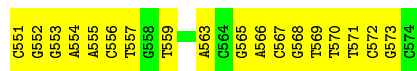
- Molecule 2: 5'-D(\*CP\*GP\*GP\*AP\*AP\*CP\*TP\*GP\*TP\*CP\*TP\*CP\*AP\*CP\*GP\*AP\*CP\*GP\*TP\*TP\*TP\*CP\*GP\*C)-3'

Chain 1-D: 



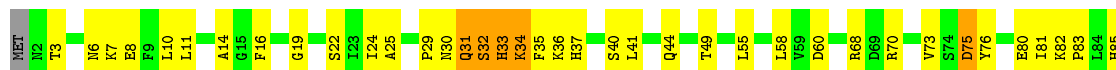
- Molecule 2: 5'-D(\*CP\*GP\*GP\*AP\*AP\*CP\*TP\*GP\*TP\*CP\*TP\*CP\*AP\*CP\*GP\*AP\*CP\*GP\*TP\*TP\*TP\*CP\*GP\*C)-3'

Chain 2-D: 



- Molecule 3: DNA endonuclease I-CreI

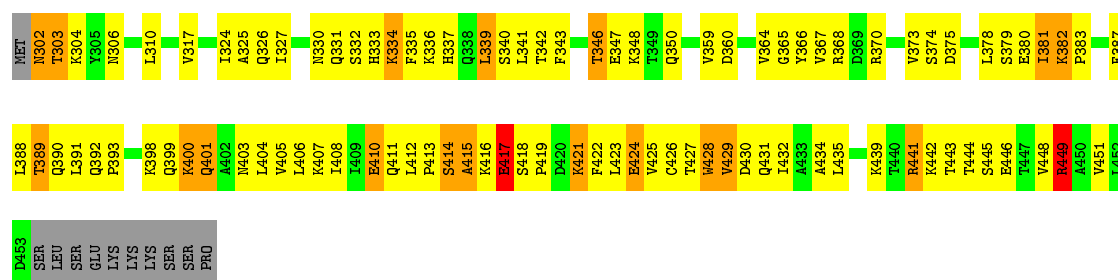
Chain 1-A: 





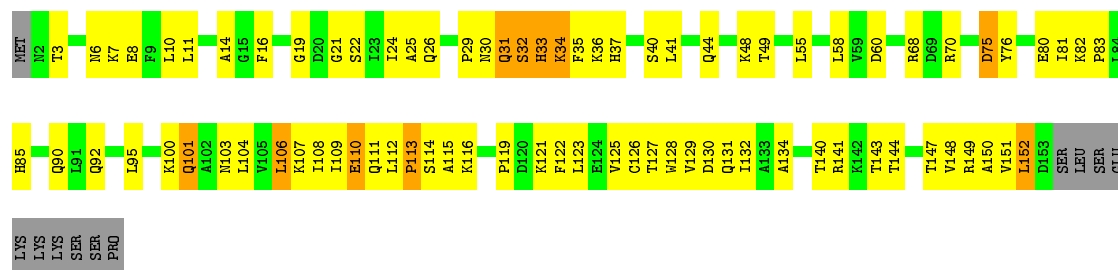
• Molecule 3: DNA endonuclease I-CreI

Chain 1-B: 35% 46% 11% 7%



• Molecule 3: DNA endonuclease I-CreI

Chain 2-A: 44% 44% 6% 7%



• Molecule 3: DNA endonuclease I-CreI

Chain 2-B: 33% 48% 11% 7%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	43.84Å 70.33Å 88.02Å 90.28° 89.94° 98.78°	Depositor
Resolution (Å)	29.50 – 2.90 29.47 – 2.91	Depositor EDS
% Data completeness (in resolution range)	(Not available) (29.50-2.90) 94.3 (29.47-2.91)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.07 (at 2.90Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.217 , 0.257 0.434 , 0.448	Depositor DCC
$R_{free}$ test set	2166 reflections (9.93%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.8	Xtriage
Anisotropy	0.270	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , -22.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.459 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.71	EDS
Total number of atoms	6892	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	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.*

<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	1-C	0.30	0/554	0.68	0/854
1	2-C	0.30	0/554	0.67	0/854
2	1-D	0.34	0/542	0.73	0/834
2	2-D	0.32	0/542	0.71	0/834
3	1-A	0.43	0/1257	0.71	0/1697
3	1-B	0.44	0/1257	0.72	1/1697 (0.1%)
3	2-A	0.43	0/1257	0.71	0/1697
3	2-B	0.44	0/1257	0.72	1/1697 (0.1%)
All	All	0.40	0/7220	0.71	2/10164 (0.0%)

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	1-C	0	1
1	2-C	0	1
2	1-D	0	1
All	All	0	3

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	1-B	449	ARG	N-CA-C	-5.23	96.87	111.00
3	2-B	449	ARG	N-CA-C	-5.23	96.87	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	1-C	517	DC	Sidechain
2	1-D	567	DC	Sidechain
1	2-C	518	DA	Sidechain

## 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	1-C	493	0	270	34	3
1	2-C	493	0	270	26	0
2	1-D	485	0	272	35	0
2	2-D	485	0	272	38	1
3	1-A	1234	0	1266	86	0
3	1-B	1234	0	1266	129	3
3	2-A	1234	0	1266	84	0
3	2-B	1234	0	1266	139	1
All	All	6892	0	6148	520	4

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 40.

The worst 5 of 520 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:B:401:GLN:HB2	3:B:432:ILE:HG23	1.47	0.96
1:C:521:DT:H2''	1:C:522:DC:H5'	1.47	0.96
3:B:401:GLN:HB2	3:B:432:ILE:HG23	1.47	0.96
3:A:81:ILE:HG23	3:A:109:ILE:HD12	1.52	0.91
3:A:81:ILE:HG23	3:A:109:ILE:HD12	1.52	0.91

All (4) 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)
1:C:509:DT:OP2	3:B:304:LYS:NZ[1_455]	0.95	1.25
1:C:509:DT:OP2	3:B:304:LYS:CE[1_455]	1.68	0.52

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:559:DT:OP2	3:B:304:LYS:NZ[1_455]	2.10	0.10
1:C:509:DT:P	3:B:304:LYS:NZ[1_455]	2.13	0.07

## 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	1-A	150/163 (92%)	116 (77%)	29 (19%)	5 (3%)	4	15
3	1-B	150/163 (92%)	110 (73%)	29 (19%)	11 (7%)	1	3
3	2-A	150/163 (92%)	116 (77%)	29 (19%)	5 (3%)	4	15
3	2-B	150/163 (92%)	110 (73%)	29 (19%)	11 (7%)	1	3
All	All	600/652 (92%)	452 (75%)	116 (19%)	32 (5%)	2	6

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	1-A	34	LYS
3	1-A	150	ALA
3	1-B	334	LYS
3	1-B	417	GLU
3	2-A	34	LYS

### 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	1-A	139/150 (93%)	128 (92%)	11 (8%)	12	34
3	1-B	139/150 (93%)	123 (88%)	16 (12%)	5	17
3	2-A	139/150 (93%)	128 (92%)	11 (8%)	12	34
3	2-B	139/150 (93%)	123 (88%)	16 (12%)	5	17
All	All	556/600 (93%)	502 (90%)	54 (10%)	8	25

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	1-B	439	LYS
3	2-A	68	ARG
3	2-B	421	LYS
3	1-B	441	ARG
3	2-A	3	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
3	1-B	411	GLN
3	2-A	37	HIS
3	2-B	350	GLN
3	2-A	6	ASN
3	1-A	92	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.