



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 16, 2017 – 05:41 AM EDT

PDB ID : 3C25  
Title : Crystal Structure of NotI Restriction Endonuclease Bound to Cognate DNA  
Authors : Lambert, A.R.; Sussman, D.; Shen, B.; Stoddard, B.L.  
Deposited on : unknown  
Resolution : 2.50 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	:	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

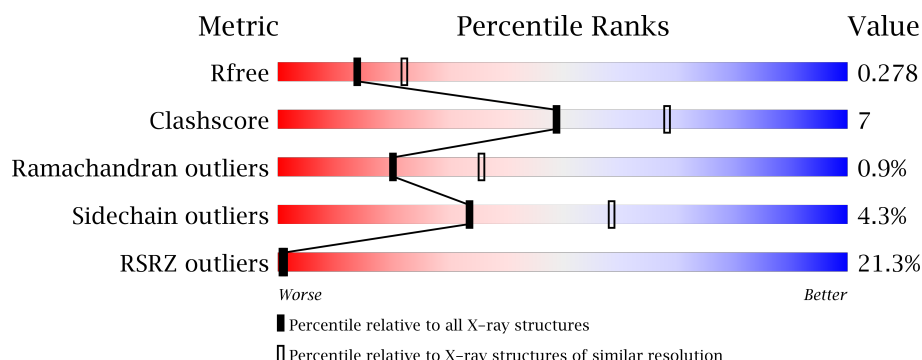
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	22	<div> <div>18%</div> <div> <div>64%</div> <div>36%</div> </div> </div>
2	D	22	<div> <div>27%</div> <div> <div>77%</div> <div>23%</div> </div> </div>
3	A	383	<div> <div>19%</div> <div> <div>77%</div> <div>13%</div> <div>8%</div> </div> </div>
3	B	383	<div> <div>20%</div> <div> <div>74%</div> <div>17%</div> <div>8%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	CA	A	802	-	-	-	X



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 (5'-D(\*DCP\*DGP\*DGP\*DAP\*DGP\*DGP\*DCP\*DGP\*DCP\*DGP\*DGP\*DCP\*DCP\*DGP\*DCP\*DCP\*DCP\*DGP\*DCP\*DCP\*DG)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	22	Total	C	N	O	P	0	0	0
			450	210	90	129	21			

- Molecule 2 is a DNA chain called DNA (5'-D(\*DCP\*DGP\*DGP\*DCP\*DGP\*DGP\*DCP\*DGP\*DCP\*DGP\*DGP\*DCP\*DCP\*DGP\*DCP\*DGP\*DCP\*DCP\*DTP\*DCP\*DCP\*DG)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	22	Total	C	N	O	P	0	0	0
			446	209	85	131	21			

- Molecule 3 is a protein called NotI restriction endonuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	354	Total 2720	C 1743	N 469	O 498	S 10	0	0	0
3	B	353	Total 2718	C 1740	N 467	O 501	S 10	0	0	0

- Molecule 4 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Fe 1 1	0	0
4	A	1	Total Fe 1 1	0	0

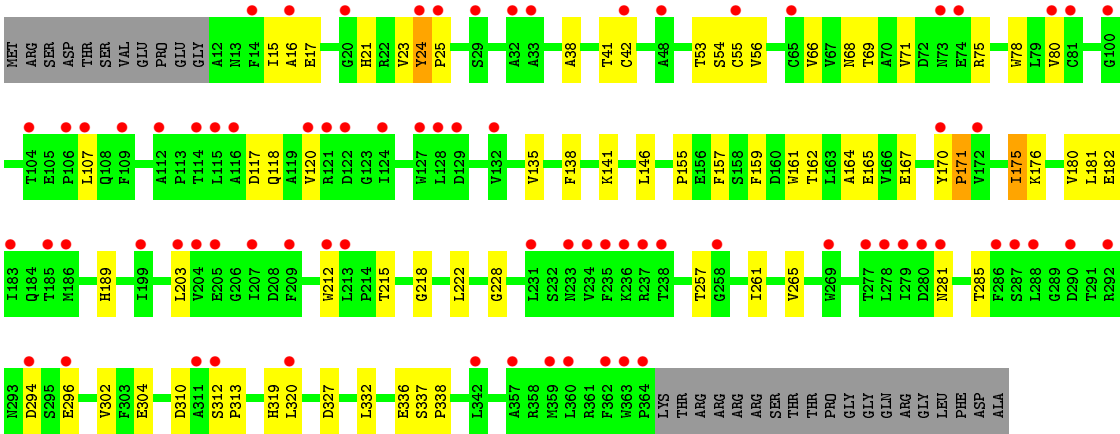
- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	2	Total	Ca	0	0
			2	2		
5	A	2	Total	Ca	0	0
			2	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	C	25	Total	O	0	0
			25	25		
6	D	27	Total	O	0	0
			27	27		
6	A	56	Total	O	0	0
			56	56		
6	B	60	Total	O	0	0
			60	60		





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.89Å 81.71Å 73.58Å 90.00° 99.51° 90.00°	Depositor
Resolution (Å)	50.00 – 2.50 47.64 – 2.49	Depositor EDS
% Data completeness (in resolution range)	97.7 (50.00-2.50) 97.3 (47.64-2.49)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.68 (at 2.48Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.223 , 0.277 0.224 , 0.278	Depositor DCC
$R_{free}$ test set	1484 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	55.3	Xtriage
Anisotropy	0.349	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 53.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.017 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6508	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 45.46 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.3178e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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

Bond lengths and bond angles in the following residue types are not validated in this section: CA, FE

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	0.62	0/505	1.27	5/778 (0.6%)
2	D	0.63	0/499	1.16	3/768 (0.4%)
3	A	0.39	2/2795 (0.1%)	0.51	0/3812
3	B	0.37	0/2793	0.51	0/3809
All	All	0.43	2/6592 (0.0%)	0.68	8/9167 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	42	CYS	CB-SG	5.73	1.92	1.82
3	A	65	CYS	CB-SG	5.56	1.91	1.82

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1	DC	O3'-P-O5'	-11.24	82.64	104.00
1	C	1	DC	OP2-P-O3'	-8.04	87.51	105.20
1	C	1	DC	OP1-P-O3'	-7.88	87.86	105.20
1	C	10	DG	O4'-C1'-N9	7.45	113.22	108.00
2	D	10	DG	O4'-C1'-N9	6.64	112.65	108.00

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	450	0	244	3	0
2	D	446	0	245	3	0
3	A	2720	0	2615	42	0
3	B	2718	0	2614	38	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0
6	A	56	0	0	0	0
6	B	60	0	0	1	0
6	C	25	0	0	0	0
6	D	27	0	0	0	0
All	All	6508	0	5718	82	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 7.

The worst 5 of 82 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:29:SER:CB	3:A:210:HIS:CE1	1.91	1.53
3:A:29:SER:CB	3:A:210:HIS:ND1	1.97	1.24
3:A:283:ASP:OD1	3:A:285:THR:HG22	1.40	1.16
3:A:283:ASP:CG	3:A:285:THR:HG22	1.76	1.05
3:A:29:SER:CB	3:A:210:HIS:HE1	1.61	0.91

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	352/383 (92%)	329 (94%)	20 (6%)	3 (1%)	20	36
3	B	351/383 (92%)	328 (93%)	20 (6%)	3 (1%)	20	36
All	All	703/766 (92%)	657 (94%)	40 (6%)	6 (1%)	20	36

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	28	VAL
3	B	294	ASP
3	A	171	PRO
3	A	24	TYR
3	B	171	PRO

### 5.3.2 Protein sidechains ⓘ

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	278/317 (88%)	264 (95%)	14 (5%)	28	51
3	B	280/317 (88%)	270 (96%)	10 (4%)	40	67
All	All	558/634 (88%)	534 (96%)	24 (4%)	33	58

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	177	ARG
3	A	287	SER
3	B	310	ASP
3	A	193	LYS
3	A	234	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	281	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	B	281	ASN
3	B	118	GLN
3	A	189	HIS
3	B	189	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 carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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	22/22 (100%)	1.16	4 (18%) <b>1</b> <b>1</b>	56, 61, 72, 76	0
2	D	22/22 (100%)	1.26	6 (27%) <b>1</b> <b>0</b>	57, 63, 76, 78	0
3	A	354/383 (92%)	1.23	73 (20%) <b>1</b> <b>1</b>	26, 67, 72, 77	0
3	B	353/383 (92%)	1.31	77 (21%) <b>1</b> <b>1</b>	58, 67, 72, 76	0
All	All	751/810 (92%)	1.26	160 (21%) <b>1</b> <b>1</b>	26, 67, 72, 78	0

The worst 5 of 160 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	B	204	VAL	9.3
3	A	166	VAL	8.8
3	B	124	ILE	8.3
3	A	207	ILE	7.8
3	B	32	ALA	7.1

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	CA	A	802	1/1	0.86	0.37	2.13	97,97,97,97	0
5	CA	B	804	1/1	0.96	0.32	0.65	89,89,89,89	0
5	CA	A	801	1/1	0.65	0.26	0.19	103,103,103,103	0
4	FE	A	901	1/1	0.92	0.21	-1.81	64,64,64,64	0
4	FE	B	902	1/1	0.88	0.20	-2.04	64,64,64,64	0
5	CA	B	803	1/1	0.63	0.24	-	115,115,115,115	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.