



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 2, 2021 – 10:36 PM EDT

PDB ID : 3L8B  
Title : Crystal structure of a replicative DNA polymerase bound to the oxidized guanine lesion guanidinohydantoin  
Authors : Aller, P.; Ye, Y.; Wallace, S.S.; Burrows, C.J.; Doublié, S.  
Deposited on : 2009-12-30  
Resolution : 2.15 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
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)	:	2.23.2

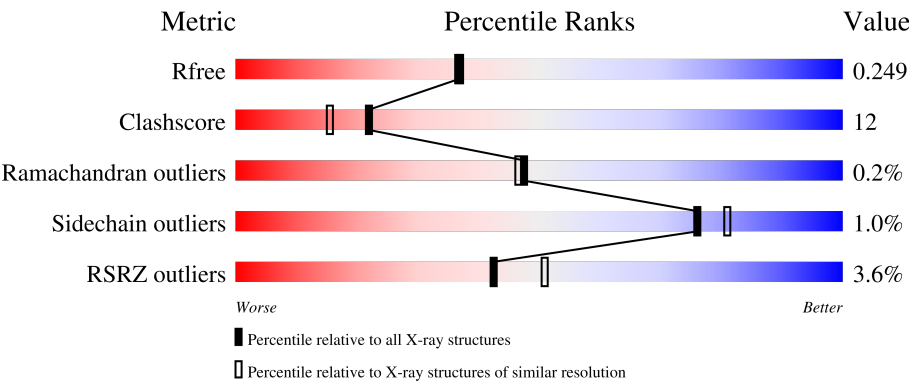
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.15 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 <sub>free</sub>	130704	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	A	906	<div><div>3%</div><div><div></div><div>76%</div><div>23%</div><div>.</div></div></div>
1	B	906	<div><div>4%</div><div><div></div><div>78%</div><div>20%</div><div>.</div></div></div>
2	C	18	<div><div></div><div><div>28%</div><div>44%</div><div>17%</div><div>11%</div></div></div>
2	E	18	<div><div></div><div><div>17%</div><div>61%</div><div>11%</div><div>11%</div></div></div>
3	D	13	<div><div></div><div><div>23%</div><div>77%</div></div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	F	13	 31% 69%

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 16717 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 protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	896	Total	C	N	O	S	0	0	0
			7270	4671	1205	1362	32			
1	B	896	Total	C	N	O	S	0	0	0
			7257	4660	1203	1362	32			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	222	ALA	ASP	engineered mutation	UNP Q38087
A	327	ALA	ASP	engineered mutation	UNP Q38087
A	904	HIS	-	expression tag	UNP Q38087
A	905	HIS	-	expression tag	UNP Q38087
A	906	HIS	-	expression tag	UNP Q38087
B	222	ALA	ASP	engineered mutation	UNP Q38087
B	327	ALA	ASP	engineered mutation	UNP Q38087
B	904	HIS	-	expression tag	UNP Q38087
B	905	HIS	-	expression tag	UNP Q38087
B	906	HIS	-	expression tag	UNP Q38087

- Molecule 2 is a DNA chain called DNA (5'-D(\*AP\*C\*TP\*(G35)P\*TP\*TP\*AP\*AP\*GP\*C P\*AP\*GP\*TP\*CP\*CP\*GP\*CP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	16	Total	C	N	O	P	0	0	0
			326	155	60	96	15			
2	E	16	Total	C	N	O	P	0	0	0
			326	155	60	96	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	13	Total	C	N	O	P	0	0	0
			265	127	50	76	12			
3	F	13	Total	C	N	O	P	0	0	0
			265	127	50	76	12			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		

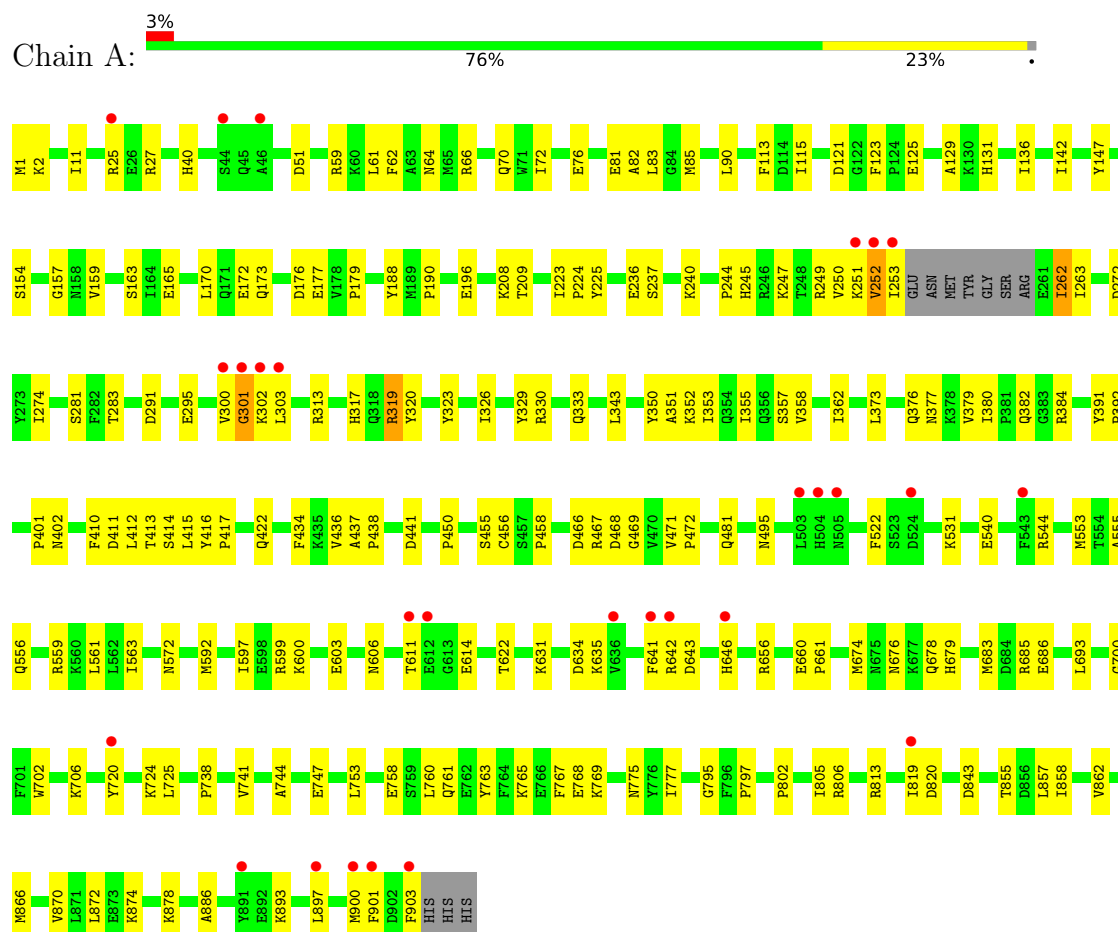
- Molecule 5 is water.

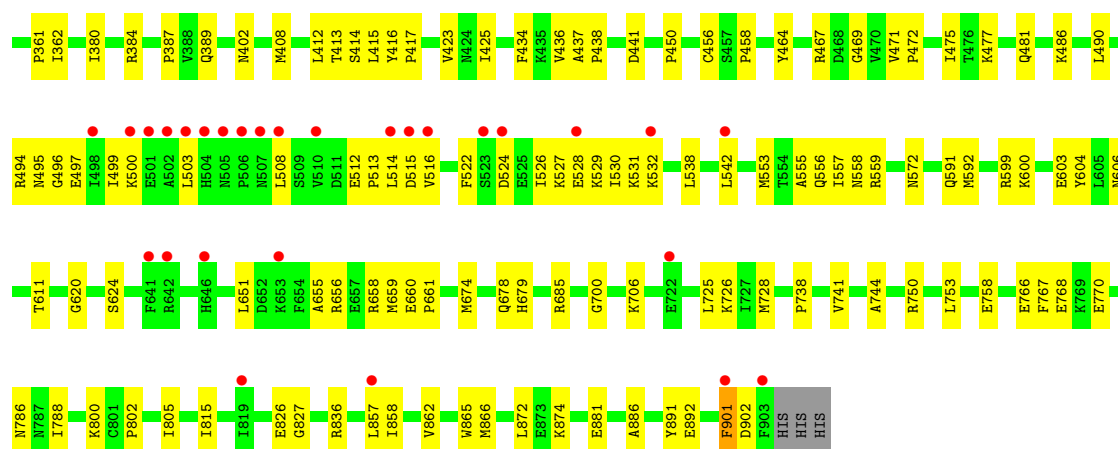
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	462	Total	O	0	0
			462	462		
5	B	428	Total	O	0	0
			428	428		
5	C	27	Total	O	0	0
			27	27		
5	D	19	Total	O	0	0
			19	19		
5	E	39	Total	O	0	0
			39	39		
5	F	23	Total	O	0	0
			23	23		

### 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 polymerase





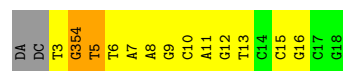
• Molecule 2: DNA (5'-D(\*AP\*C\*TP\*(G35)P\*TP\*TP\*AP\*AP\*GP\*CP\*AP\*GP\*TP\*CP\*CP\*GP\*CP\*G)-3')

Chain C: 28% 44% 17% 11%



• Molecule 2: DNA (5'-D(\*AP\*C\*TP\*(G35)P\*TP\*TP\*AP\*AP\*GP\*CP\*AP\*GP\*TP\*CP\*CP\*GP\*CP\*G)-3')

Chain E: 17% 61% 11% 11%



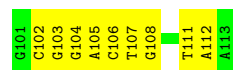
• Molecule 3: DNA (5'-D(\*GP\*CP\*GP\*GP\*AP\*CP\*TP\*GP\*CP\*TP\*TP\*AP\*A)-3')

Chain D: 23% 77%



• Molecule 3: DNA (5'-D(\*GP\*CP\*GP\*GP\*AP\*CP\*TP\*GP\*CP\*TP\*TP\*AP\*A)-3')

Chain F: 31% 69%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.28Å 120.25Å 136.96Å 90.00° 95.96° 90.00°	Depositor
Resolution (Å)	48.81 – 2.15 48.81 – 2.15	Depositor EDS
% Data completeness (in resolution range)	88.8 (48.81-2.15) 95.5 (48.81-2.15)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.36 (at 2.14Å)	Xtriage
Refinement program	CNS 1.21	Depositor
R, $R_{free}$	0.219 , 0.255 0.212 , 0.249	Depositor DCC
$R_{free}$ test set	25471 reflections (9.44%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	21.6	Xtriage
Anisotropy	0.054	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 46.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	16717	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	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 45.01 % 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 1.4061e-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: G35, SO4

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	A	0.36	0/7449	0.57	0/10077
1	B	0.35	0/7435	0.57	0/10062
2	C	0.47	0/339	0.76	0/519
2	E	0.53	0/339	0.86	0/519
3	D	0.40	0/297	0.80	0/457
3	F	0.39	0/297	0.78	0/457
All	All	0.36	0/16156	0.60	0/22091

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
2	C	0	2
2	E	0	1
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	5	DT	Sidechain
2	C	8	DA	Sidechain
2	E	5	DT	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	A	7270	0	7105	156	0
1	B	7257	0	7081	148	0
2	C	326	0	183	17	0
2	E	326	0	183	23	0
3	D	265	0	148	16	0
3	F	265	0	148	10	0
4	B	10	0	0	0	0
5	A	462	0	0	18	0
5	B	428	0	0	13	0
5	C	27	0	0	1	0
5	D	19	0	0	0	0
5	E	39	0	0	1	0
5	F	23	0	0	1	0
All	All	16717	0	14848	362	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 12.

The worst 5 of 362 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:D:103:DG:H2''	3:D:104:DG:O5'	1.65	0.96
2:C:11:DA:H2''	2:C:12:DG:H5'	1.50	0.94
1:A:330:ARG:HH11	1:A:333:GLN:HE22	1.20	0.90
1:A:495:ASN:HD21	1:A:522:PHE:H	1.17	0.88
1:A:196:GLU:HG3	5:A:1314:HOH:O	1.76	0.85

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	
1	A	892/906 (98%)	864 (97%)	25 (3%)	3 (0%)	41	37
1	B	892/906 (98%)	857 (96%)	35 (4%)	0	100	100
All	All	1784/1812 (98%)	1721 (96%)	60 (3%)	3 (0%)	47	46

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	252	VAL
1	A	301	GLY
1	A	622	THR

### 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	
1	A	782/803 (97%)	774 (99%)	8 (1%)	76	81
1	B	780/803 (97%)	773 (99%)	7 (1%)	78	83
All	All	1562/1606 (97%)	1547 (99%)	15 (1%)	76	81

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	843	ASP
1	B	678	GLN
1	B	25	ARG
1	B	901	PHE
1	B	202	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	495	ASN
1	B	761	GLN
1	B	507	ASN
1	B	558	ASN
1	B	823	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# $ Z  > 2$	Counts	RMSZ	# $ Z  > 2$
2	G35	E	4	2	18,23,24	1.28	4 (22%)	20,33,36	1.23	3 (15%)
2	G35	C	4	2	18,23,24	1.32	4 (22%)	20,33,36	1.00	2 (10%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	G35	E	4	2	-	1/10/41/42	0/2/2/2
2	G35	C	4	2	-	3/10/41/42	0/2/2/2

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	4	G35	C5-N7	-3.21	1.33	1.37
2	E	4	G35	C5-N7	-3.08	1.33	1.37
2	E	4	G35	C4-N3	2.69	1.47	1.44

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	4	G35	C8-N9	-2.65	1.33	1.37
2	C	4	G35	C4-N3	2.57	1.47	1.44

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	4	G35	C5-C4-N9	-2.89	98.43	102.28
2	E	4	G35	C4-C5-N7	2.72	110.21	106.88
2	C	4	G35	C5-C4-N9	-2.63	98.78	102.28
2	C	4	G35	C4-C5-N7	2.48	109.92	106.88
2	E	4	G35	C5-C4-N3	-2.25	107.98	112.76

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	4	G35	C5-C4-N3-C2
2	C	4	G35	N9-C4-N3-C2
2	E	4	G35	C2'-C1'-N9-C4
2	C	4	G35	C2'-C1'-N9-C4

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	4	G35	5	0
2	C	4	G35	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
4	SO4	B	907	-	4,4,4	0.25	0	6,6,6	0.16	0
4	SO4	B	908	-	4,4,4	0.25	0	6,6,6	0.10	0

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	A	896/906 (98%)	0.16	28 (3%) 49 58	10, 24, 48, 78	0
1	B	896/906 (98%)	0.26	38 (4%) 36 45	11, 26, 58, 82	0
2	C	15/18 (83%)	-0.10	0 100 100	20, 28, 42, 44	0
2	E	15/18 (83%)	0.02	0 100 100	21, 27, 44, 56	0
3	D	13/13 (100%)	0.24	0 100 100	21, 25, 79, 84	0
3	F	13/13 (100%)	0.17	0 100 100	22, 24, 85, 89	0
All	All	1848/1874 (98%)	0.21	66 (3%) 42 51	10, 25, 54, 89	0

The worst 5 of 66 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	252	VAL	7.2
1	A	253	ILE	6.4
1	B	508	LEU	5.4
1	B	46	ALA	5.3
1	B	260	ARG	4.8

### 6.2 Non-standard residues in protein, DNA, RNA chains [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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
2	G35	E	4	22/23	0.90	0.21	39,51,59,61	0
2	G35	C	4	22/23	0.93	0.17	35,46,57,58	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
4	SO4	B	908	5/5	0.79	0.22	78,79,80,80	0
4	SO4	B	907	5/5	0.95	0.20	70,70,70,71	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.