



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 08:28 am GMT

PDB ID : 3EPG  
Title : Structure of Human DNA Polymerase Iota complexed with N2-ethylguanine  
Authors : Pence, M.G.  
Deposited on : 2008-09-29  
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  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
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

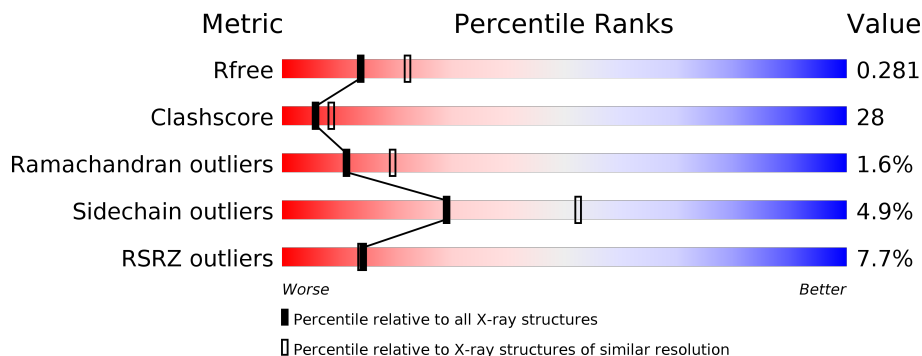
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	A	420	
2	B	18	
2	C	18	

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3307 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.

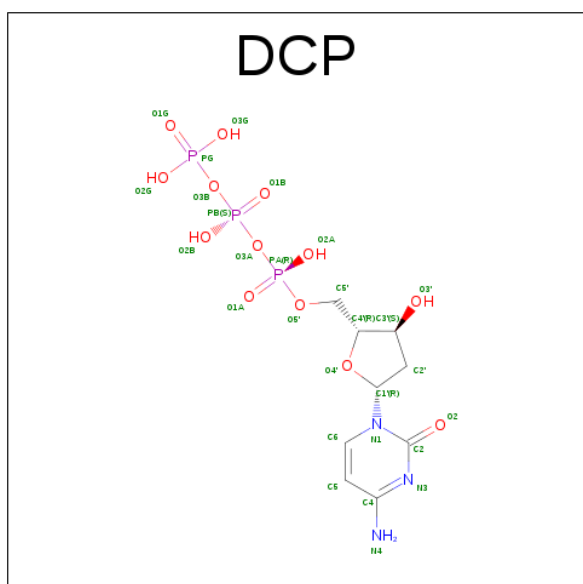
- Molecule 1 is a protein called DNA polymerase iota.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	375	Total	C	N	O	S	0	0	0
			2896	1825	504	546	21			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	7	Total	C	N	O	P	0	0	0
			142	67	29	39	7			
2	C	8	Total	C	N	O	P	0	0	0
			168	80	30	50	8			

- Molecule 3 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: C<sub>9</sub>H<sub>16</sub>N<sub>3</sub>O<sub>13</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			28	9	3	13	3		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	Na	0	0
			3	3		
4	C	1	Total	Na	0	0
			1	1		

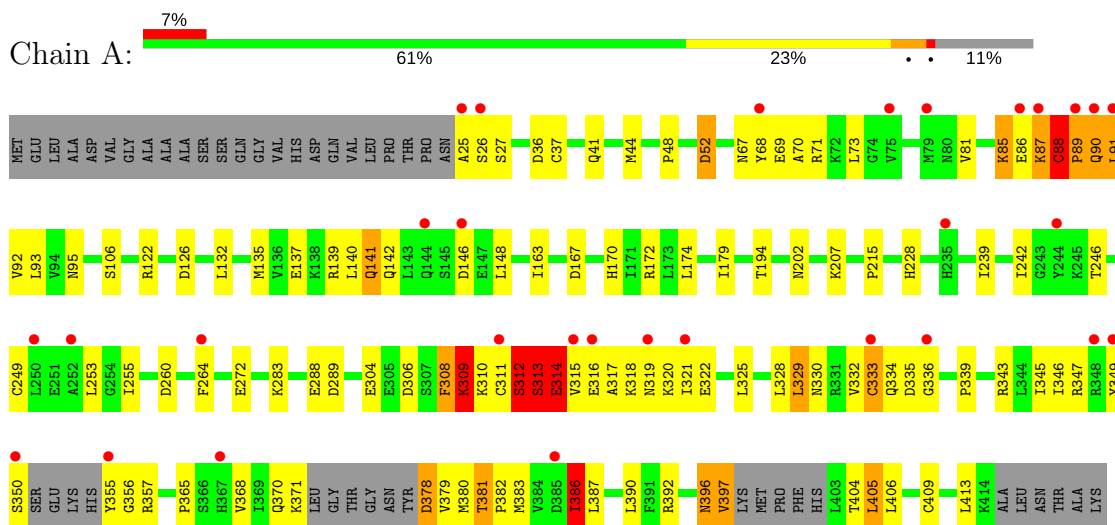
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	67	Total	O	0	0
			67	67		
5	C	2	Total	O	0	0
			2	2		

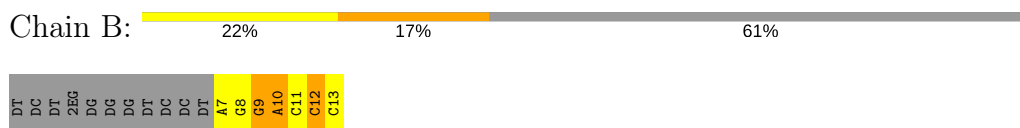
### 3 Residue-property plots [i](#)

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.

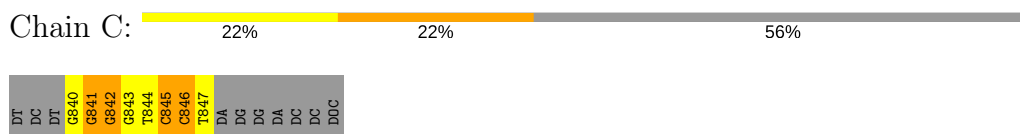
- Molecule 1: DNA polymerase iota



- Molecule 2: 5'-D(\*DTP\*DCP\*DTP\*(2EG)P\*DGP\*DGP\*DGP\*DTP\*DCP\*DCP\*DTP\*DAP\*DGP\*DGP\*DAP\*DCP\*DCP\*(DOC))-3'



- Molecule 2: 5'-D(\*DTP\*DCP\*DTP\*(2EG)P\*DGP\*DGP\*DGP\*DTP\*DCP\*DCP\*DTP\*DAP\*DGP\*DGP\*DAP\*DCP\*DCP\*(DOC))-3'



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	98.53Å 98.53Å 202.35Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	15.00 – 2.50 15.00 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.1 (15.00-2.50) 99.7 (15.00-2.60)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.18 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.234 , 0.283 0.240 , 0.281	Depositor DCC
$R_{free}$ test set	942 reflections (5.39%)	DCC
Wilson B-factor (Å <sup>2</sup> )	42.6	Xtriage
Anisotropy	0.004	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 45.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	3307	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.34% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, 2EG, DCP, DOC

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	0.72	0/2934	0.77	2/3964 (0.1%)
2	B	3.17	21/139 (15.1%)	3.35	22/212 (10.4%)
2	C	4.69	44/160 (27.5%)	4.57	42/245 (17.1%)
All	All	1.41	65/3233 (2.0%)	1.49	66/4421 (1.5%)

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	A	0	12

The worst 5 of 65 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	842	DG	P-O5'	-17.45	1.42	1.59
2	C	842	DG	N7-C5	-16.99	1.29	1.39
2	C	845	DC	C4'-O4'	-13.33	1.31	1.45
2	C	842	DG	N9-C8	-12.87	1.28	1.37
2	B	12	DC	N3-C4	12.57	1.42	1.33

The worst 5 of 66 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	842	DG	N3-C4-C5	-20.42	118.39	128.60
2	C	842	DG	C2-N3-C4	18.99	121.39	111.90
2	C	843	DG	C2-N3-C4	18.14	120.97	111.90
2	B	9	DG	C2-N3-C4	17.12	120.46	111.90
2	C	842	DG	C5-C6-N1	15.02	119.01	111.50

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	52	ASP	Peptide
1	A	87	LYS	Peptide
1	A	88	CYS	Peptide
1	A	89	PRO	Peptide
1	A	90	GLN	Peptide

## 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	A	2896	0	2929	168	0
2	B	142	0	78	4	0
2	C	168	0	90	4	0
3	A	28	0	12	8	0
4	A	3	0	0	0	0
4	C	1	0	0	0	0
5	A	67	0	0	5	0
5	C	2	0	0	0	0
All	All	3307	0	3109	176	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 28.

The worst 5 of 176 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:A:405:LEU:HD22	1:A:406:LEU:N	1.04	1.37
1:A:44:MET:CE	1:A:67:ASN:HD22	1.37	1.35
1:A:405:LEU:CD2	1:A:406:LEU:N	1.94	1.29
1:A:316:GLU:HA	1:A:319:ASN:OD1	1.37	1.25
1:A:405:LEU:C	1:A:405:LEU:HD22	1.58	1.16

There are no symmetry-related clashes.



## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	367/420 (87%)	324 (88%)	37 (10%)	6 (2%)	11	19

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	312	SER
1	A	313	SER
1	A	386	ILE
1	A	141	GLN
1	A	88	CYS

### 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	327/376 (87%)	311 (95%)	16 (5%)	29	52

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	312	SER
1	A	314	GLU
1	A	381	THR
1	A	310	LYS
1	A	386	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such

sidechains are listed below:

Mol	Chain	Res	Type
1	A	217	GLN
1	A	412	ASN
1	A	279	GLN
1	A	189	ASN
1	A	262	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	DOC	B	13	2	13,19,20	2.34	3 (23%)	12,26,29	2.86	4 (33%)
2	2EG	C	840	2	18,26,27	1.13	1 (5%)	21,37,40	2.07	8 (38%)

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	DOC	B	13	2	-	0/3/18/19	0/2/2/2
2	2EG	C	840	2	-	0/6/24/25	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	13	DOC	O5'-C5'	-6.84	1.35	1.44
2	B	13	DOC	C2-N3	-2.26	1.33	1.38

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	13	DOC	C5-C4	-2.14	1.35	1.41
2	C	840	2EG	C6-N1	3.89	1.40	1.33

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	13	DOC	C2'-C1'-N1	-4.90	103.14	112.47
2	B	13	DOC	O4'-C1'-N1	-4.15	100.79	107.78
2	C	840	2EG	N3-C2-N1	-3.81	120.47	126.23
2	C	840	2EG	C5-C6-N1	-3.28	118.81	123.48
2	C	840	2EG	C21-N2-C2	-2.94	119.98	123.69

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	13	DOC	2	0
2	C	840	2EG	2	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 for Mogul analysis.

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
3	DCP	A	875	-	23,29,29	1.71	5 (21%)	25,45,45	2.12	4 (16%)

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
3	DCP	A	875	-	-	0/18/34/34	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	875	DCP	PG-O3G	-3.54	1.40	1.54
3	A	875	DCP	PG-O1G	-3.26	1.39	1.50
3	A	875	DCP	PG-O2G	-3.13	1.42	1.54
3	A	875	DCP	PA-O2A	-2.44	1.42	1.55
3	A	875	DCP	C5'-C4'	2.60	1.59	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	875	DCP	C6-N1-C2	-2.07	117.93	121.28
3	A	875	DCP	O5'-C5'-C4'	2.38	117.42	109.00
3	A	875	DCP	C5'-C4'-C3'	2.56	130.42	114.73
3	A	875	DCP	O4'-C1'-N1	8.79	122.60	107.78

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	875	DCP	8	0

## 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	375/420 (89%)	0.40	30 (8%) 13 13	21, 34, 44, 55	0
2	B	6/18 (33%)	-0.51	0 100 100	30, 32, 39, 40	0
2	C	7/18 (38%)	-0.82	0 100 100	32, 33, 34, 35	0
All	All	388/456 (85%)	0.36	30 (7%) 14 14	21, 34, 44, 55	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	355	TYR	6.0
1	A	86	GLU	5.4
1	A	25	ALA	4.9
1	A	244	TYR	4.0
1	A	333	CYS	3.9

### 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. 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(Å <sup>2</sup> )	Q<0.9
2	DOC	B	13	18/19	0.97	0.10	-	32,34,36,37	0
2	2EG	C	840	24/25	0.82	0.32	-	20,20,20,20	0

### 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(Å <sup>2</sup> )	Q<0.9
3	DCP	A	875	28/28	0.91	0.27	1.93	43,47,52,54	0
4	NA	A	421	1/1	0.95	0.16	0.65	39,39,39,39	0
4	NA	A	422	1/1	0.85	0.08	-	48,48,48,48	0
4	NA	C	1	1/1	0.96	0.42	-	20,20,20,20	0
4	NA	A	423	1/1	0.93	0.24	-	20,20,20,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.