



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 30, 2024 – 10:22 AM EDT

PDB ID : 9DP3  
Title : APE1 N174D Product Complex with Abasic DNA  
Authors : Hoitsma, N.M.; DeHart, K.D.; Freudenthal, B.D.  
Deposited on : 2024-09-20  
Resolution : 2.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.39

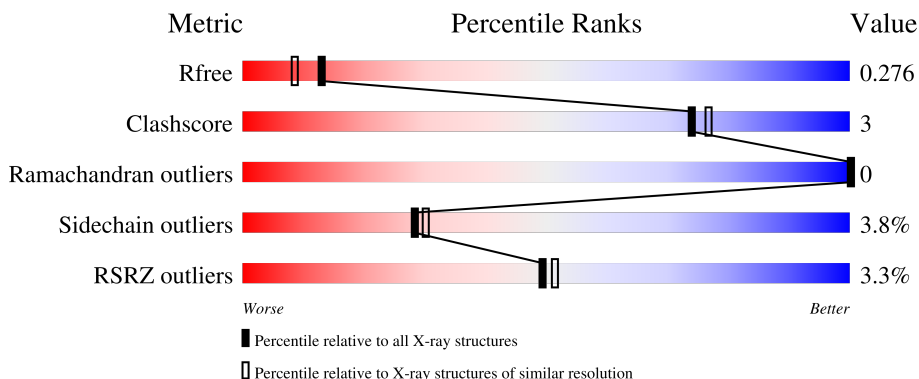
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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))
$R_{free}$	164625	6234 (2.10-2.10)
Clashscore	180529	6893 (2.10-2.10)
Ramachandran outliers	177936	6839 (2.10-2.10)
Sidechain outliers	177891	6840 (2.10-2.10)
RSRZ outliers	164620	6234 (2.10-2.10)

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	276	 2% 90% 9%
1	B	276	 4% 85% 11%
2	D	11	 9% 91% 9%
3	E	10	 100%
4	F	21	 71% 29%

## 2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 5445 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 repair nuclease/redox regulator APEX1, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	276	Total	C	N	O	S	0	1	0
			2187	1398	374	407	8			
1	B	266	Total	C	N	O	S	0	1	0
			2107	1350	365	384	8			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	138	ALA	CYS	engineered mutation	UNP P27695
A	174	ASP	ASN	engineered mutation	UNP P27695
B	138	ALA	CYS	engineered mutation	UNP P27695
B	174	ASP	ASN	engineered mutation	UNP P27695

- Molecule 2 is a DNA chain called DNA (5'-D(P\*(3DR)P\*CP\*GP\*AP\*CP\*GP\*GP\*AP\*T P\*CP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	11	Total	C	N	O	P	0	0	0
			216	101	39	65	11			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	E	10	Total	C	N	O	P	0	0	0
			203	97	38	59	9			

- Molecule 4 is a DNA chain called DNA (5'-D(\*GP\*GP\*AP\*TP\*CP\*CP\*GP\*TP\*CP\*GP\*GP\*GP\*CP\*GP\*CP\*AP\*TP\*CP\*AP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	F	21	429	203	82	124	20	0	0	0

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

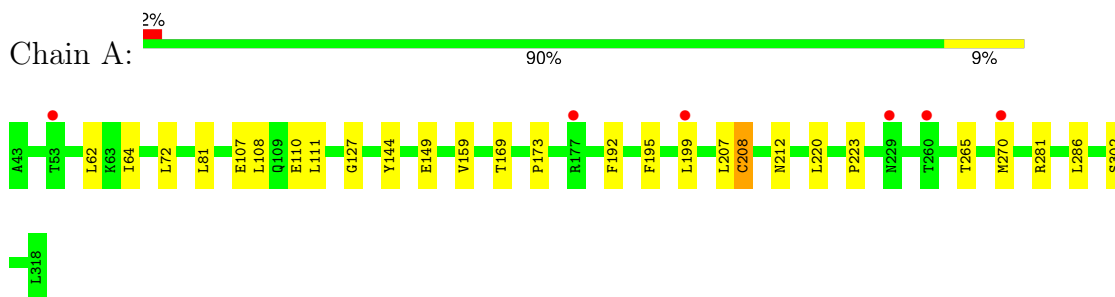
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	148	Total	O	0	0
			148	148		
6	B	104	Total	O	0	0
			104	104		
6	D	7	Total	O	0	0
			7	7		
6	E	14	Total	O	0	0
			14	14		
6	F	25	Total	O	0	0
			25	25		

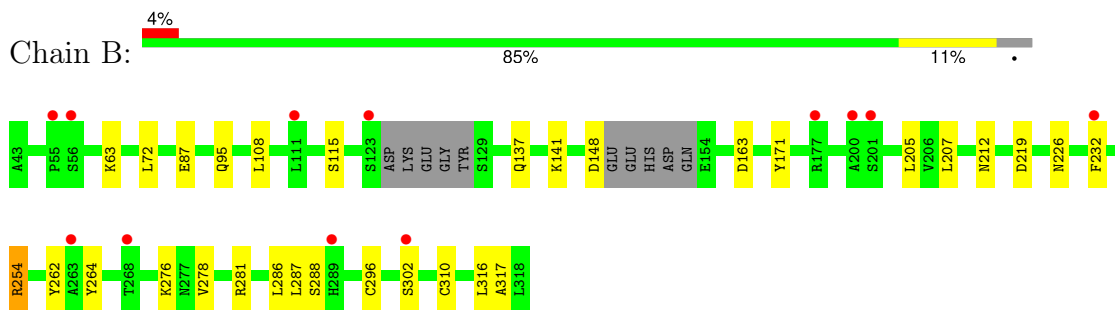
### 3 Residue-property plots [i](#)

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.

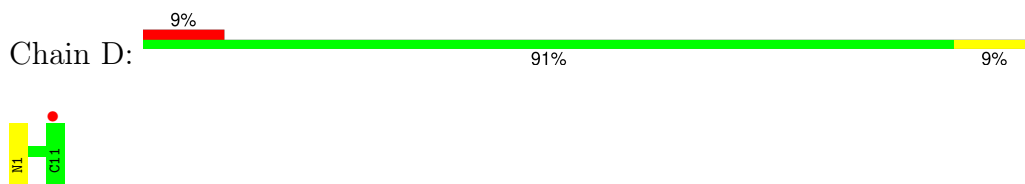
- Molecule 1: DNA repair nuclease/redox regulator APEX1, mitochondrial



- Molecule 1: DNA repair nuclease/redox regulator APEX1, mitochondrial



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



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



There are no outlier residues recorded for this chain.

- Molecule 4: DNA (5'-D(\*GP\*GP\*AP\*TP\*CP\*CP\*GP\*TP\*CP\*GP\*GP\*GP\*CP\*GP\*CP\*AP\*TP\*CP\*AP\*GP\*C)-3')

Chain F:  71% 29%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.20Å 60.83Å 73.62Å 82.55° 77.17° 85.85°	Depositor
Resolution (Å)	24.95 – 2.10 24.95 – 2.10	Depositor EDS
% Data completeness (in resolution range)	80.5 (24.95-2.10) 67.2 (24.95-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.13 (at 2.10Å)	Xtrriage
Refinement program	PHENIX 1.19.2-4158-000	Depositor
R, $R_{free}$	0.253 , 0.274 0.255 , 0.276	Depositor DCC
$R_{free}$ test set	33749 reflections (5.65%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	27.2	Xtrriage
Anisotropy	0.410	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 29.0	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	0.92	EDS
Total number of atoms	5445	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.97% 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: 3DR, 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	0.62	0/2245	0.87	0/3048
1	B	0.62	0/2162	0.88	0/2934
2	D	0.44	0/228	0.71	0/349
3	E	0.48	0/227	0.71	0/349
4	F	0.53	0/481	0.76	0/741
All	All	0.60	0/5343	0.85	0/7421

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	2187	0	2135	14	0
1	B	2107	0	2074	13	0
2	D	216	0	120	0	0
3	E	203	0	114	0	0
4	F	429	0	236	4	0
5	A	5	0	0	0	0
6	A	148	0	0	2	0
6	B	104	0	0	0	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	D	7	0	0	0	0
6	E	14	0	0	0	0
6	F	25	0	0	1	0
All	All	5445	0	4679	29	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 3.

All (29) 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)
4:F:12:DG:H5'	6:F:120:HOH:O	1.80	0.80
1:B:276:LYS:HB2	1:B:278:VAL:HG23	1.86	0.58
4:F:13:DC:H2'	4:F:14:DG:C8	2.40	0.57
1:B:72:LEU:HD21	1:B:108:LEU:HD11	1.86	0.56
1:A:169:THR:HA	1:A:208:CYS:O	2.07	0.54
1:A:62:LEU:HD21	1:A:64:ILE:HD11	1.95	0.48
1:B:226:ASN:O	1:B:232:PHE:HB3	2.15	0.47
1:B:254:ARG:HA	1:B:254:ARG:HD3	1.61	0.46
1:B:207:LEU:HD23	1:B:286:LEU:HD12	1.98	0.46
1:B:296:CYS:SG	1:B:317:ALA:HB2	2.56	0.46
1:A:159:VAL:HG21	1:A:195:PHE:HZ	1.80	0.46
1:A:207:LEU:HD23	1:A:286:LEU:HD12	1.98	0.46
1:B:262:TYR:HA	1:B:264:TYR:CZ	2.52	0.45
1:A:107[B]:GLU:O	1:A:110:GLU:HG2	2.18	0.44
1:A:107[A]:GLU:O	1:A:110:GLU:HG2	2.17	0.43
1:A:220:LEU:HD21	1:A:223:PRO:HB3	1.99	0.43
1:B:219:ASP:HB3	1:B:281:ARG:HB3	2.00	0.43
1:B:287:LEU:HD21	1:B:316:LEU:HD21	1.99	0.43
1:A:265:THR:HG23	1:A:281:ARG:HG3	2.01	0.42
1:A:111:LEU:N	6:A:514:HOH:O	2.53	0.42
4:F:9:DC:H2''	4:F:10:DG:C8	2.55	0.42
1:A:72:LEU:HD21	1:A:108:LEU:HD11	2.01	0.42
1:A:107[B]:GLU:H	1:A:107[B]:GLU:HG2	1.49	0.41
1:B:63:LYS:HE2	1:B:87:GLU:HG3	2.01	0.41
1:A:81:LEU:HD23	1:A:81:LEU:HA	1.85	0.41
1:A:127:GLY:HA3	4:F:16:DA:H5''	2.03	0.41
6:A:503:HOH:O	1:B:137:GLN:HG2	2.20	0.41
1:A:144:TYR:HE2	1:B:115:SER:HB2	1.85	0.41
1:B:95:GLN:HB3	1:B:171:TYR:HB2	2.03	0.40

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	275/276 (100%)	259 (94%)	16 (6%)	0	100	100
1	B	261/276 (95%)	247 (95%)	14 (5%)	0	100	100
All	All	536/552 (97%)	506 (94%)	30 (6%)	0	100	100

There are no Ramachandran outliers to report.

### 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	
1	A	231/235 (98%)	223 (96%)	8 (4%)	31	34
1	B	223/235 (95%)	214 (96%)	9 (4%)	27	28
All	All	454/470 (97%)	437 (96%)	17 (4%)	28	31

All (17) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	149	GLU
1	A	173	PRO
1	A	192	PHE
1	A	199	LEU
1	A	208	CYS

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type
1	A	212	ASN
1	A	270	MET
1	A	302	SER
1	B	141	LYS
1	B	148	ASP
1	B	163	ASP
1	B	205	LEU
1	B	212	ASN
1	B	254	ARG
1	B	288	SER
1	B	302	SER
1	B	310	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	153	GLN
1	B	245	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

1 non-standard protein/DNA/RNA residue is 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	3DR	D	1	2	12,12,12	2.87	4 (33%)	14,17,17	1.66	2 (14%)

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	3DR	D	1	2	-	0/6/16/16	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1	3DR	C3'-C4'	-7.17	1.34	1.53
2	D	1	3DR	O4'-C1'	-4.34	1.30	1.43
2	D	1	3DR	C2'-C1'	2.97	1.59	1.51
2	D	1	3DR	O3'-C3'	2.43	1.48	1.43

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	1	3DR	O4'-C4'-C3'	4.83	110.84	103.73
2	D	1	3DR	C1'-C2'-C3'	-3.00	100.05	103.26

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

1 ligand is 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
5	SO4	A	401	-	4,4,4	0.56	0	6,6,6	0.16	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

### 6.1 Protein, DNA and RNA chains

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	276/276 (100%)	0.22	6 (2%) 62 64	15, 29, 47, 62	1 (0%)
1	B	266/276 (96%)	0.57	12 (4%) 39 41	15, 37, 53, 63	1 (0%)
2	D	10/11 (90%)	0.49	1 (10%) 14 15	38, 62, 72, 81	0
3	E	10/10 (100%)	0.20	0 100 100	42, 47, 51, 52	0
4	F	21/21 (100%)	0.30	0 100 100	34, 52, 63, 74	0
All	All	583/594 (98%)	0.39	19 (3%) 49 51	15, 34, 54, 81	2 (0%)

All (19) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	289[A]	HIS	4.5
1	B	263	ALA	3.9
1	A	199	LEU	3.3
1	B	123	SER	3.2
1	B	200	ALA	3.0
1	A	260	THR	2.9
1	B	302	SER	2.6
1	B	177	ARG	2.5
1	A	270	MET	2.4
1	B	111	LEU	2.4
1	A	177	ARG	2.4
1	A	53	THR	2.4
1	B	55	PRO	2.3
1	B	56	SER	2.2
1	B	268	THR	2.1
1	B	232	PHE	2.1
1	B	201	SER	2.1
2	D	11	DC	2.1
1	A	229	ASN	2.0

## 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	3DR	D	1	12/12	0.94	0.08	25,30,36,39	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(Å <sup>2</sup> )	Q<0.9
5	SO4	A	401	5/5	0.95	0.13	30,30,30,30	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.