



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 5, 2023 – 07:29 PM EDT

PDB ID : 4E5Z  
Title : Damaged DNA induced UV-damaged DNA-binding protein (UV-DDB) dimerization and its roles in chromatinized DNA repair  
Authors : Yeh, J.I.; Du, S.  
Deposited on : 2012-03-15  
Resolution : 3.22 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 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

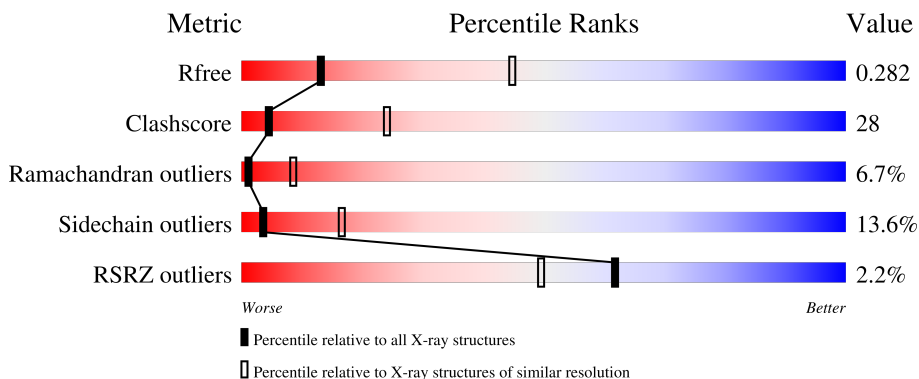
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.22 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	1335 (3.24-3.20)
Clashscore	141614	1460 (3.24-3.20)
Ramachandran outliers	138981	1437 (3.24-3.20)
Sidechain outliers	138945	1436 (3.24-3.20)
RSRZ outliers	127900	1291 (3.24-3.20)

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	1150	
2	B	436	
3	F	24	
4	G	24	

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	3DR	G	11	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 12985 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 damage-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1140	8851	5596	1497	1712	46	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-9	MET	-	expression tag	UNP Q16531
A	-8	HIS	-	expression tag	UNP Q16531
A	-7	HIS	-	expression tag	UNP Q16531
A	-6	HIS	-	expression tag	UNP Q16531
A	-5	HIS	-	expression tag	UNP Q16531
A	-4	HIS	-	expression tag	UNP Q16531
A	-3	HIS	-	expression tag	UNP Q16531
A	-2	HIS	-	expression tag	UNP Q16531
A	-1	HIS	-	expression tag	UNP Q16531
A	0	HIS	-	expression tag	UNP Q16531
A	1	HIS	-	expression tag	UNP Q16531

- Molecule 2 is a protein called DNA damage-binding protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	402	3157	2001	570	566	20	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-7	MET	-	expression tag	UNP Q92466
B	-6	ASP	-	expression tag	UNP Q92466
B	-5	TYR	-	expression tag	UNP Q92466
B	-4	LYS	-	expression tag	UNP Q92466
B	-3	ASP	-	expression tag	UNP Q92466

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	ASP	-	expression tag	UNP Q92466
B	-1	ASP	-	expression tag	UNP Q92466
B	0	ASP	-	expression tag	UNP Q92466
B	1	LYS	-	expression tag	UNP Q92466
B	420	GLU	-	expression tag	UNP Q92466

- Molecule 3 is a DNA chain called AP24 DNA strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	F	24	497	236	91	146	24	0	0	0

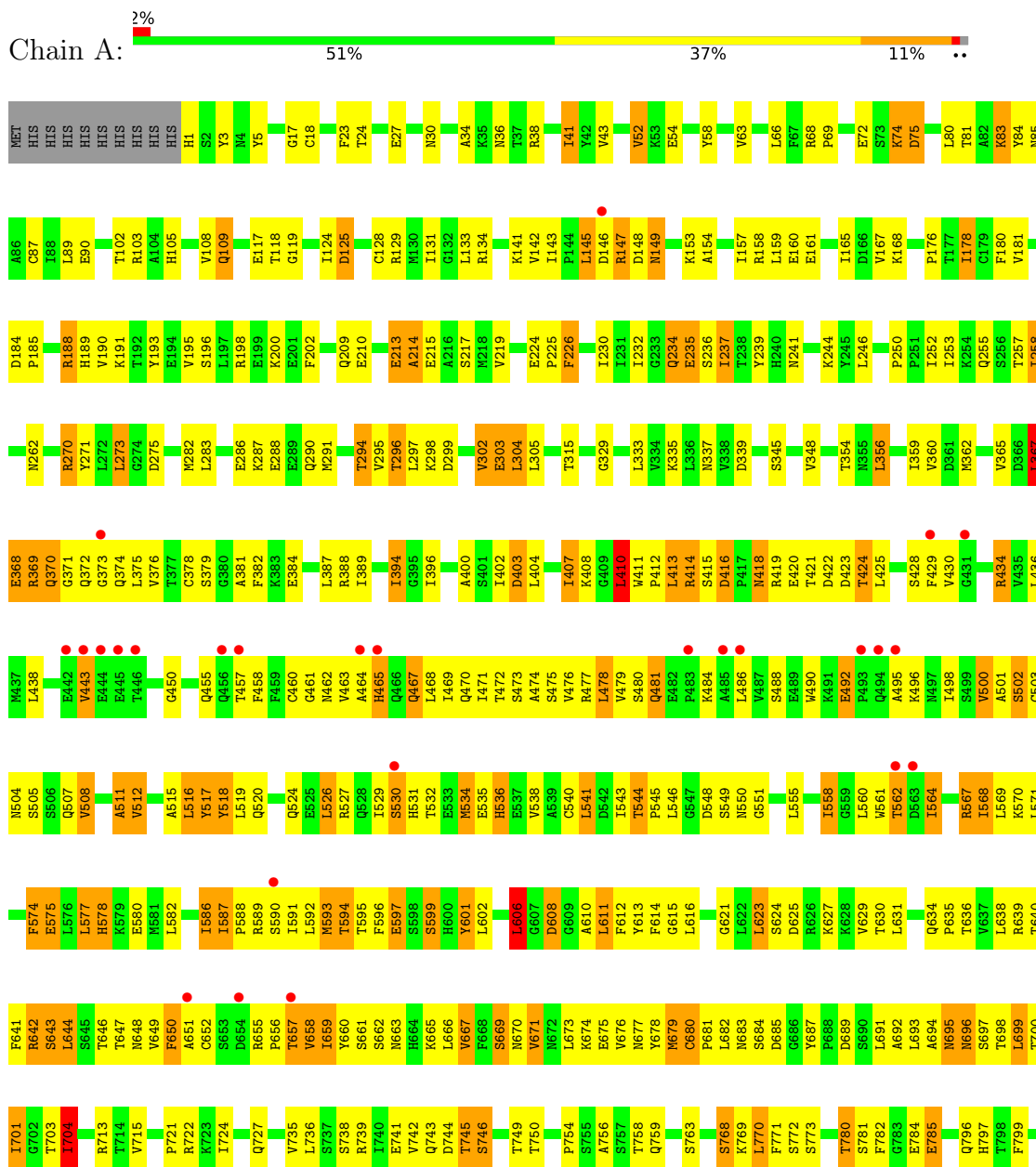
- Molecule 4 is a DNA chain called AP24 DNA complementary strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	G	24	480	228	86	142	24	0	0	0

### 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 damage-binding protein 1









## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.42Å 76.50Å 389.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.09 – 3.22 41.09 – 3.22	Depositor EDS
% Data completeness (in resolution range)	93.6 (41.09-3.22) 93.7 (41.09-3.22)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.73 (at 3.25Å)	Xtrriage
Refinement program	PHENIX 1.8_1069, REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.228 , 0.284 0.231 , 0.282	Depositor DCC
$R_{free}$ test set	1690 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	109.3	Xtrriage
Anisotropy	0.074	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.05 , 19.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	12985	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	87.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.43% 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

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	0.34	0/9013	0.62	6/12220 (0.0%)
2	B	0.42	2/3244 (0.1%)	0.75	9/4404 (0.2%)
3	F	0.66	1/557 (0.2%)	1.22	4/857 (0.5%)
4	G	0.88	3/524 (0.6%)	1.24	3/801 (0.4%)
All	All	0.41	6/13338 (0.0%)	0.73	22/18282 (0.1%)

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	B	0	3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	G	9	DC	O3'-P	12.42	1.76	1.61
3	F	1	DT	OP3-P	-10.76	1.48	1.61
4	G	1	DG	OP3-P	-10.69	1.48	1.61
2	B	256	TRP	CG-CD1	7.70	1.47	1.36
2	B	256	TRP	CG-CD2	-6.40	1.32	1.43

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	9	DC	P-O3'-C3'	-13.46	103.54	119.70
2	B	273	ARG	CG-CD-NE	-9.08	92.73	111.80
4	G	1	DG	OP1-P-OP2	-8.30	107.15	119.60
3	F	1	DT	OP1-P-OP2	-8.11	107.43	119.60

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	B	256	TRP	CA-CB-CG	7.59	128.12	113.70

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	28	LEU	Peptide
2	B	29	GLU	Peptide
2	B	63	LEU	Peptide

## 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	8851	0	8754	487	1
2	B	3157	0	3126	210	1
3	F	497	0	272	12	0
4	G	480	0	267	33	0
All	All	12985	0	12419	722	2

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 722 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:G:9:DC:H2''	4:G:10:DG:C8	1.73	1.23
1:A:508:VAL:HG13	1:A:519:LEU:HD22	1.44	0.99
1:A:224:GLU:HA	1:A:226:PHE:H	1.34	0.93
2:B:67:ARG:HA	2:B:67:ARG:HE	1.29	0.93
1:A:372:GLN:HB2	1:A:374:GLN:HE22	1.32	0.93

All (2) 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:A:339:ASP:OD2	1:A:1046:SER:OG[1_455]	2.16	0.04
2:B:42:SER:OG	2:B:187:LYS:O[1_455]	2.19	0.01

### 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	1138/1150 (99%)	932 (82%)	142 (12%)	64 (6%)	2	13
2	B	400/436 (92%)	301 (75%)	60 (15%)	39 (10%)	0	3
All	All	1538/1586 (97%)	1233 (80%)	202 (13%)	103 (7%)	1	8

5 of 103 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	235	GLU
1	A	295	VAL
1	A	303	GLU
1	A	370	GLN
1	A	502	SER

#### 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	980/1009 (97%)	849 (87%)	131 (13%)	4	17
2	B	347/380 (91%)	297 (86%)	50 (14%)	3	14

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1327/1389 (96%)	1146 (86%)	181 (14%)	<b>3</b> <b>16</b>

5 of 181 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1019	GLU
2	B	141	PHE
1	A	1044	SER
1	A	1131	LYS
2	B	213	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	168	ASN
2	B	221	ASN
2	B	287	HIS
1	A	803	HIS
1	A	797	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](#)

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	$\# Z  > 2$
4	3DR	G	11	4	8,11,12	<b>3.10</b>	<b>3 (37%)</b>	9,14,17	1.52	<b>2 (22%)</b>

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
4	3DR	G	11	4	-	2/3/15/16	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	G	11	3DR	C2'-C3'	-7.31	1.39	1.52
4	G	11	3DR	O4'-C4'	-3.21	1.39	1.44
4	G	11	3DR	O5'-C5'	-2.05	1.39	1.44

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	11	3DR	O5'-C5'-C4'	2.56	117.71	108.99
4	G	11	3DR	C1'-C2'-C3'	2.40	105.92	103.20

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	G	11	3DR	C4'-C5'-O5'-P
4	G	11	3DR	O4'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	G	11	3DR	8	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides 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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
4	G	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	G	11:3DR	O3'	12:DC	P	2.90
1	G	9:DC	O3'	10:DG	P	1.76

## 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	1140/1150 (99%)	-0.66	26 (2%) 60 48	29, 73, 182, 360	0
2	B	402/436 (92%)	-0.63	9 (2%) 62 49	36, 69, 176, 276	0
3	F	24/24 (100%)	-0.86	0 100 100	60, 93, 120, 138	0
4	G	23/24 (95%)	-0.74	0 100 100	64, 95, 133, 165	0
All	All	1589/1634 (97%)	-0.66	35 (2%) 62 49	29, 73, 180, 360	0

The worst 5 of 35 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	26	SER	9.7
2	B	27	PRO	7.2
1	A	651	ALA	5.6
1	A	464	ALA	4.9
1	A	445	GLU	4.7

### 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
4	3DR	G	11	11/12	0.78	0.21	71,109,126,187	0

### 6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.



## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.