



# wwPDB X-ray Structure Validation Summary Report

Sep 17, 2023 – 03:41 PM EDT

PDB ID : 4YIR  
Title : Crystal structure of Rad4-Rad23 crosslinked to an undamaged DNA  
Authors : Min, J.-H.; Chen, X.; Kim, Y.  
Deposited on : 2015-03-02  
Resolution : 3.05 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 : 2.35.1  
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

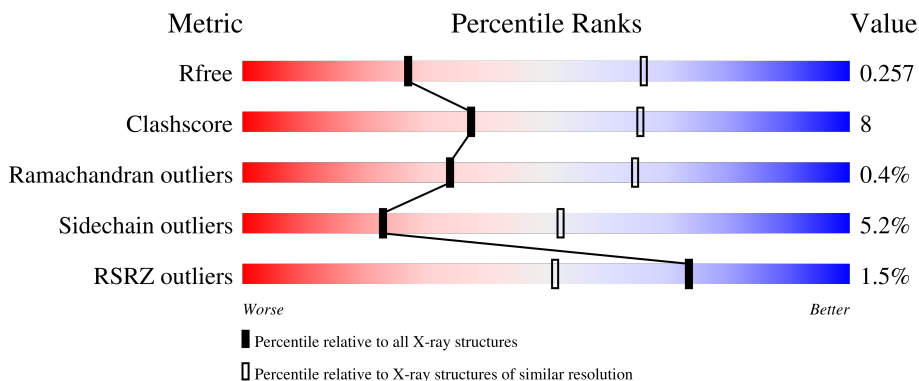
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.05 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	1754 (3.10-3.02)
Clashscore	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)
RSRZ outliers	127900	1713 (3.10-3.02)

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	538	 75% 17% • 7%
2	X	171	 20% 11% 69%
3	W	24	 4% 46% 21% 33%
4	Y	24	 58% 25% 12% •

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 10590 atoms, of which 5131 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 protein RAD4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	499	8300	2619	4189	740	725	27	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	95	GLY	-	expression tag	UNP P14736
A	96	SER	-	expression tag	UNP P14736
A	97	SER	-	expression tag	UNP P14736
A	98	ARG	-	expression tag	UNP P14736
A	99	ALA	-	expression tag	UNP P14736
A	100	MET	-	expression tag	UNP P14736
A	115	THR	LYS	conflict	UNP P14736
A	131	CYS	VAL	engineered mutation	UNP P14736
A	132	SER	CYS	engineered mutation	UNP P14736
A	223	GLU	VAL	conflict	UNP P14736
A	427	ARG	GLN	conflict	UNP P14736

- Molecule 2 is a protein called UV excision repair protein RAD23.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	X	53	827	260	418	69	78	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
X	228	GLY	-	expression tag	UNP P32628
X	229	SER	-	expression tag	UNP P32628

- Molecule 3 is a DNA chain called DNA (5'-D(\*TP\*TP\*GP\*AP\*CP\*TP\*CP\*(G47)P\*AP\*CP\*AP\*TP\*CP\*CP\*CP\*CP\*CP\*GP\*CP\*TP\*AP\*CP\*AP\*A)-3').

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P	S			
3	W	24	756	232	274	85	141	23	1	0	0	0

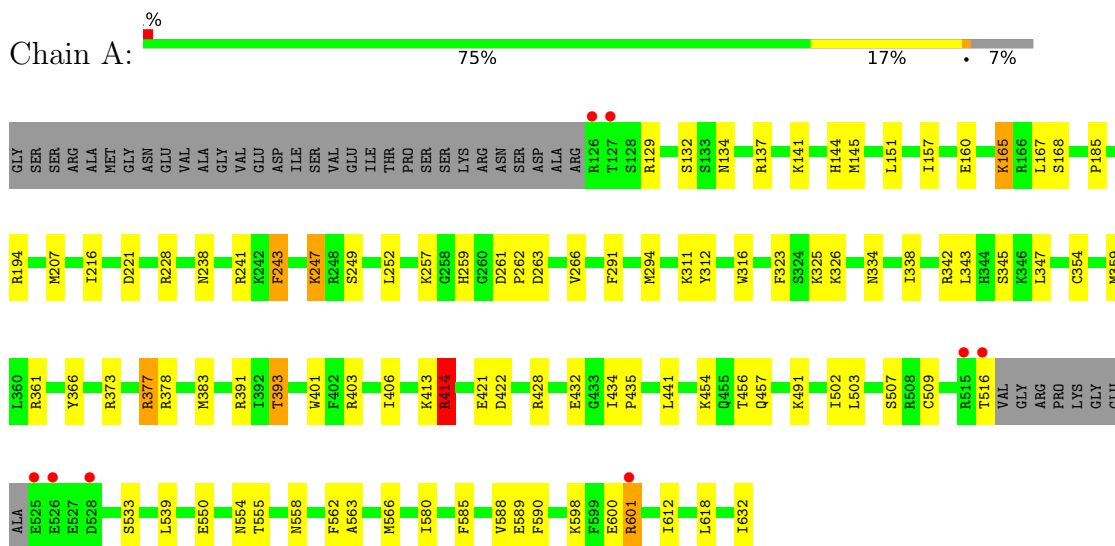
- Molecule 4 is a DNA chain called DNA (5'-D(\*AP\*TP\*TP\*GP\*TP\*AP\*GP\*CP\*G\*GP\*G  
P\*GP\*AP\*TP\*GP\*TP\*CP\*GP\*AP\*GP\*TP\*CP\*A)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
4	Y	23	707	218	250	86	132	21	0	0	0

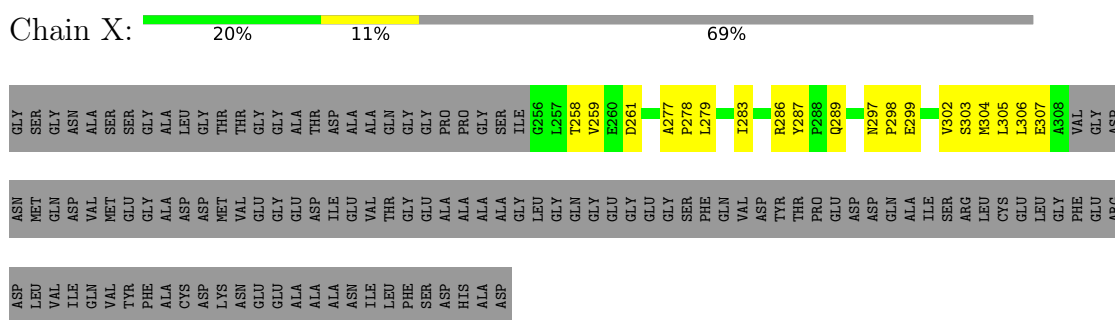
### 3 Residue-property plots [i](#)

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.

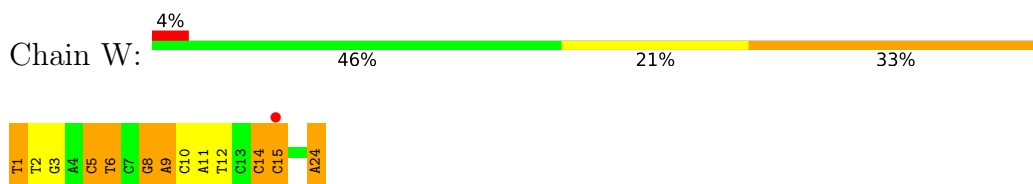
- Molecule 1: DNA repair protein RAD4



- Molecule 2: UV excision repair protein RAD23



- Molecule 3: DNA (5'-D(\*TP\*TP\*GP\*AP\*CP\*TP\*CP\*(G47)P\*AP\*CP\*AP\*TP\*CP\*CP\*CP\*CP\*CP\*GP\*CP\*TP\*AP\*CP\*AP\*A)-3')



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

Chain Y:  58% 25% 12% .



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	79.41Å 79.41Å 404.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.70 – 3.05 49.08 – 3.05	Depositor EDS
% Data completeness (in resolution range)	97.7 (39.70-3.05) 97.5 (49.08-3.05)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.82 (at 3.07Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.192 , 0.253 0.198 , 0.257	Depositor DCC
$R_{free}$ test set	1951 reflections (7.73%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	63.6	Xtrriage
Anisotropy	0.096	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 47.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	10590	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.45% 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: G47

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	0.86	0/4200	0.87	4/5643 (0.1%)
2	X	0.76	0/415	0.78	0/565
3	W	1.80	9/509 (1.8%)	1.35	4/778 (0.5%)
4	Y	1.45	5/512 (1.0%)	1.39	4/789 (0.5%)
All	All	1.04	14/5636 (0.2%)	0.98	12/7775 (0.2%)

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	2

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	W	9	DA	N3-C4	9.79	1.40	1.34
3	W	9	DA	C6-N1	8.95	1.41	1.35
4	Y	20	DA	C6-N1	7.77	1.41	1.35
3	W	15	DC	C1'-N1	7.17	1.58	1.49
3	W	6	DT	C2-N3	-6.63	1.32	1.37

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	Y	18	DC	O5'-P-OP2	-11.33	95.50	105.70
3	W	14	DC	O4'-C1'-N1	7.06	112.94	108.00
4	Y	12	DG	O4'-C1'-N9	6.65	112.66	108.00
3	W	9	DA	O5'-P-OP2	-6.33	100.00	105.70

*Continued on next page...*



Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
4	Y	18	DC	OP1-P-OP2	6.31	129.07	119.60

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	247	LYS	Peptide
1	A	601	ARG	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	4111	4189	4189	55	0
2	X	409	418	418	14	0
3	W	482	274	273	9	2
4	Y	457	250	251	4	1
All	All	5459	5131	5131	77	3

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 8.

The worst 5 of 77 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance ( $\text{\AA}$ )	Clash overlap ( $\text{\AA}$ )
1:A:137:ARG:NE	1:A:294:MET:SD	2.43	0.91
1:A:598:LYS:NZ	1:A:600:GLU:OE1	2.18	0.76
2:X:289:GLN:N	2:X:289:GLN:OE1	2.18	0.75
1:A:562:PHE:HB2	1:A:566:MET:HE1	1.73	0.68
3:W:15:DC:O2	4:Y:11:DG:N2	2.29	0.66

All (3) 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)
3:W:1:DT:O5'	3:W:24:DA:O3'[1_455]	2.01	0.19
4:Y:1:DA:O5'	4:Y:24:DA:O3'[1_655]	2.04	0.16
3:W:1:DT:O5'	3:W:24:DA:HO3'[1_455]	1.52	0.08

### 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	495/538 (92%)	470 (95%)	23 (5%)	2 (0%)	34	64
2	X	51/171 (30%)	47 (92%)	4 (8%)	0	100	100
All	All	546/709 (77%)	517 (95%)	27 (5%)	2 (0%)	34	64

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	247	LYS
1	A	238	ASN

#### 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	450/480 (94%)	425 (94%)	25 (6%)	21	49
2	X	46/129 (36%)	45 (98%)	1 (2%)	52	76
All	All	496/609 (81%)	470 (95%)	26 (5%)	23	52

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	413	LYS
1	A	456	THR
1	A	555	THR
1	A	441	LEU
1	A	491	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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
3	G47	W	8	3,1,4	19,27,28	4.77	11 (57%)	17,38,41	2.33	7 (41%)

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	G47	W	8	3,1,4	-	4/7/25/26	0/3/3/3

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	W	8	G47	C2'-C3'	-11.29	1.23	1.52
3	W	8	G47	O4'-C4'	-8.57	1.25	1.45
3	W	8	G47	C1'-N9	-6.61	1.29	1.49
3	W	8	G47	C2-N1	6.54	1.47	1.36
3	W	8	G47	C2-N2	5.23	1.45	1.34

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	W	8	G47	O6-C6-C5	-4.73	115.13	124.37
3	W	8	G47	O6-C6-N1	4.70	126.20	120.65
3	W	8	G47	C2'-C1'-N9	2.84	120.82	114.27
3	W	8	G47	C4'-O4'-C1'	-2.83	102.61	109.45
3	W	8	G47	C5-C6-N1	2.67	118.67	113.95

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	W	8	G47	N2-C6A-C7A-SG
3	W	8	G47	C7A-C6A-N2-C2
3	W	8	G47	C3'-C4'-C5'-O5'
3	W	8	G47	O4'-C4'-C5'-O5'

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	W	8	G47	1	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](#)

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	499/538 (92%)	-0.07	8 (1%) 72 49	29, 59, 112, 172	0
2	X	53/171 (30%)	-0.22	0 100 100	41, 74, 119, 122	0
3	W	23/24 (95%)	-0.11	1 (4%) 35 16	52, 90, 120, 124	0
4	Y	23/24 (95%)	0.14	0 100 100	52, 106, 120, 165	0
All	All	598/757 (78%)	-0.08	9 (1%) 73 51	29, 62, 120, 172	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	516	THR	3.1
1	A	526	GLU	3.1
1	A	126	ARG	3.1
1	A	525	GLU	2.4
1	A	515	ARG	2.3

### 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
3	G47	W	8	25/26	0.91	0.22	75,93,100,104	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.