



# Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 05:01 am BST

PDB ID : 5QIV  
Title : Covalent fragment group deposition – Crystal Structure of OUTB2 in complex with PCM-0102998  
Authors : Sethi, R.; Douangamath, A.; Resnick, E.; Bradley, A.R.; Collins, P.; Brandao-Neto, J.; Talon, R.; Krojer, T.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.; London, N.; von Delft, F.  
Deposited on : 2018-08-10  
Resolution : 1.39 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 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 : 2.11  
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) : 2.11

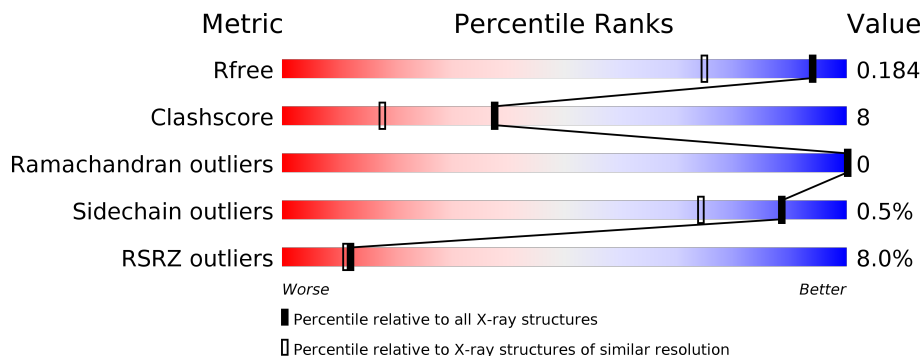
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.39 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	1714 (1.40-1.40)
Clashscore	141614	1812 (1.40-1.40)
Ramachandran outliers	138981	1763 (1.40-1.40)
Sidechain outliers	138945	1762 (1.40-1.40)
RSRZ outliers	127900	1674 (1.40-1.40)

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 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	225	

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
2	EDO	A	605	-	-	X	-
4	UNL	A	608	-	-	X	-

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Type</b>	<b>Chain</b>	<b>Res</b>	<b>Chirality</b>	<b>Geometry</b>	<b>Clashes</b>	<b>Electron density</b>
4	UNL	A	609	-	-	X	-

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 2133 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.

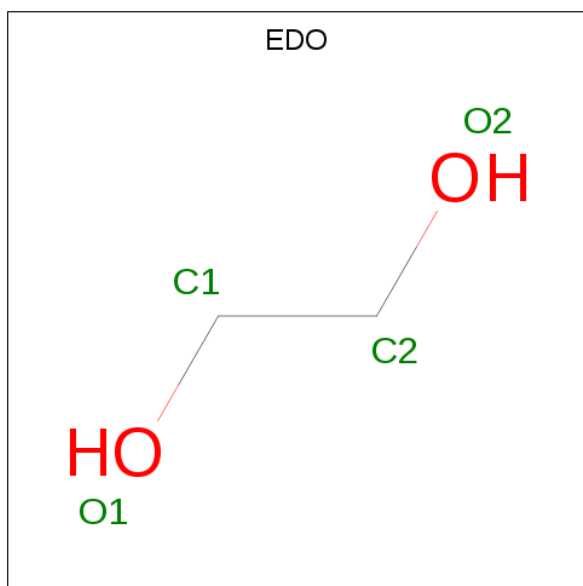
- Molecule 1 is a protein called Ubiquitin thioesterase OTUB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	225	1858	1191	313	345	9	0	5	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	49	ARG	GLY	conflict	UNP Q96DC9

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C<sub>2</sub>H<sub>6</sub>O<sub>2</sub>).



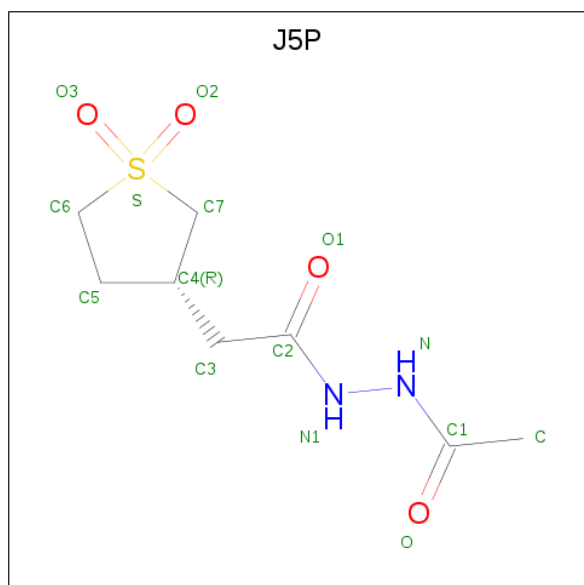
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	4	2	2	0	0
2	A	1	4	2	2	0	0
2	A	1	4	2	2	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 4 2 2	0	0
2	A	1	Total C O 4 2 2	0	0

- Molecule 3 is N'-acetyl-2-[(3R)-1,1-dioxo-1lambda 6 -thiolan-3-yl]acetohydrazide (three-letter code: J5P) (formula: C<sub>8</sub>H<sub>14</sub>N<sub>2</sub>O<sub>4</sub>S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C N O S 15 8 2 4 1	0	0

- Molecule 4 is UNKNOWN LIGAND (three-letter code: UNL) (formula: ).

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

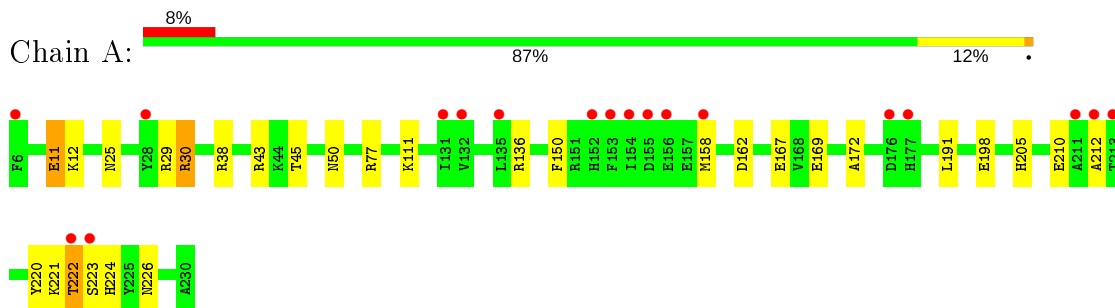
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	235	Total O 235 235	0	0

### 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: Ubiquitin thioesterase OTUB2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.24Å 58.63Å 49.79Å 90.00° 116.18° 90.00°	Depositor
Resolution (Å)	29.33 – 1.39 29.32 – 1.39	Depositor EDS
% Data completeness (in resolution range)	97.3 (29.33-1.39) 97.3 (29.32-1.39)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.30 (at 1.39Å)	Xtrriage
Refinement program	REFMAC 5.8.0230	Depositor
R, $R_{free}$	0.137 , 0.184 0.137 , 0.184	Depositor DCC
$R_{free}$ test set	2152 reflections (4.51%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	13.0	Xtrriage
Anisotropy	0.789	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 54.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.019 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	2133	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.76% 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: J5P, UNL, EDO

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	1.01	4/1914 (0.2%)	0.96	5/2583 (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	4

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	11	GLU	CD-OE1	11.31	1.38	1.25
1	A	11	GLU	CD-OE2	8.39	1.34	1.25
1	A	158	MET	C-O	5.38	1.33	1.23
1	A	167	GLU	CD-OE2	-5.17	1.20	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	221	LYS	C-N-CA	-7.17	103.79	121.70
1	A	150	PHE	CB-CA-C	6.09	122.57	110.40
1	A	30	ARG	NE-CZ-NH2	-5.87	117.36	120.30
1	A	221	LYS	O-C-N	-5.68	113.61	122.70
1	A	77	ARG	NE-CZ-NH2	-5.41	117.60	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:



Mol	Chain	Res	Type	Group
1	A	222	THR	Peptide
1	A	29	ARG	Sidechain
1	A	30	ARG	Sidechain
1	A	38	ARG	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	1858	0	1822	23	1
2	A	20	0	30	10	2
3	A	15	0	0	0	3
4	A	5	0	0	3	0
5	A	235	0	0	8	5
All	All	2133	0	1852	29	6

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.

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:A:608:UNL:O	4:A:609:UNL:O	1.92	0.87
1:A:162:ASP:HB3	5:A:860:HOH:O	1.79	0.83
2:A:604:EDO:H22	5:A:868:HOH:O	1.79	0.81
4:A:607:UNL:O	4:A:608:UNL:O	2.02	0.78
1:A:210:GLU:O	5:A:701:HOH:O	2.04	0.74
1:A:50:ASN:HD21	1:A:172:ALA:H	1.44	0.66
1:A:43:ARG:HH12	2:A:605:EDO:H22	1.59	0.65
1:A:12:LYS:H	2:A:605:EDO:H11	1.66	0.59
2:A:604:EDO:C1	5:A:758:HOH:O	2.51	0.58
4:A:609:UNL:O	4:A:610:UNL:O	2.22	0.57
1:A:50:ASN:ND2	1:A:172:ALA:H	2.05	0.55
1:A:50:ASN:HD22	1:A:136:ARG:HH21	1.56	0.53
1:A:205:HIS:HD2	5:A:726:HOH:O	1.92	0.52
1:A:224:HIS:HE1	1:A:226:ASN:ND2	2.07	0.52
1:A:45:THR:H	1:A:226:ASN:ND2	2.08	0.52

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:224:HIS:HE1	1:A:226:ASN:HD21	1.57	0.51
1:A:11:GLU:HA	2:A:605:EDO:H22	1.92	0.50
1:A:220:TYR:OH	1:A:223:SER:HA	2.13	0.48
1:A:111:LYS:CE	5:A:803:HOH:O	2.61	0.48
1:A:25:ASN:HD21	1:A:198:GLU:CD	2.17	0.47
1:A:162:ASP:CB	5:A:860:HOH:O	2.52	0.46
1:A:43:ARG:NH1	2:A:605:EDO:H22	2.30	0.46
1:A:169:GLU:HB3	2:A:603:EDO:H21	1.98	0.46
1:A:25:ASN:ND2	1:A:198:GLU:OE2	2.47	0.45
1:A:222:THR:O	1:A:223:SER:HB2	2.16	0.45
1:A:222:THR:N	1:A:224:HIS:H	2.15	0.44
2:A:604:EDO:H11	5:A:758:HOH:O	2.15	0.44
1:A:43:ARG:HH12	2:A:605:EDO:C2	2.28	0.43
1:A:12:LYS:H	2:A:605:EDO:C1	2.30	0.42

All (6) 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:A:606:J5P:C7	5:A:701:HOH:O[2_656]	1.76	0.44
1:A:212:ALA:O	3:A:606:J5P:O3[2_646]	1.97	0.23
2:A:603:EDO:O1	5:A:810:HOH:O[2_656]	2.16	0.04
2:A:603:EDO:C1	5:A:847:HOH:O[2_656]	2.17	0.03
5:A:711:HOH:O	5:A:767:HOH:O[1_554]	2.17	0.03
3:A:606:J5P:O2	5:A:738:HOH:O[2_656]	2.18	0.02

## 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	228/225 (101%)	223 (98%)	5 (2%)	0	<b>100</b> <b>100</b>

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	206/204 (101%)	205 (100%)	1 (0%)	88 74

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	191	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	50	ASN
1	A	83	ASN
1	A	97	ASN
1	A	133	GLN
1	A	205	HIS
1	A	224	HIS
1	A	226	ASN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 11 ligands modelled in this entry, 5 are unknown - leaving 6 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
2	EDO	A	602	-	3,3,3	0.17	0	2,2,2	0.82	0
2	EDO	A	605	-	3,3,3	1.15	0	2,2,2	1.48	0
2	EDO	A	604	-	3,3,3	0.47	0	2,2,2	0.54	0
2	EDO	A	601	-	3,3,3	0.60	0	2,2,2	0.08	0
3	J5P	A	606	1	15,15,15	0.69	1 (6%)	16,21,21	1.39	3 (18%)
2	EDO	A	603	-	3,3,3	0.30	0	2,2,2	0.44	0

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	EDO	A	602	-	-	1/1/1/1	-
2	EDO	A	605	-	-	0/1/1/1	-
2	EDO	A	604	-	-	0/1/1/1	-
2	EDO	A	601	-	-	1/1/1/1	-
3	J5P	A	606	1	-	1/9/20/20	0/1/1/1
2	EDO	A	603	-	-	0/1/1/1	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	606	J5P	C-C1	2.01	1.54	1.50

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	606	J5P	C4-C7-S	-4.15	98.15	104.62
3	A	606	J5P	C4-C3-C2	-2.29	107.68	112.87
3	A	606	J5P	C1-N-N1	2.07	123.07	119.22

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	601	EDO	O1-C1-C2-O2
2	A	602	EDO	O1-C1-C2-O2
3	A	606	J5P	C2-C3-C4-C7

There are no ring outliers.

4 monomers are involved in 15 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	605	EDO	6	0
2	A	604	EDO	3	0
3	A	606	J5P	0	3
2	A	603	EDO	1	2

## 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	225/225 (100%)	0.50	18 (8%) <b>12</b> <b>11</b>	10, 17, 35, 52	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	153	PHE	7.5
1	A	154	ILE	6.3
1	A	222	THR	4.0
1	A	132	VAL	3.5
1	A	158	MET	3.4
1	A	156	GLU	3.2
1	A	131	ILE	3.2
1	A	6	PHE	3.1
1	A	152	HIS	3.1
1	A	177	HIS	3.0
1	A	211	ALA	2.9
1	A	223	SER	2.8
1	A	176	ASP	2.8
1	A	28	TYR	2.6
1	A	213	THR	2.5
1	A	155	ASP	2.4
1	A	135	LEU	2.4
1	A	212	ALA	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 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. 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	J5P	A	606	15/15	0.69	0.26	15,23,45,49	0
2	EDO	A	605	4/4	0.83	0.34	33,39,39,62	0
2	EDO	A	603	4/4	0.86	0.20	27,35,47,56	0
2	EDO	A	602	4/4	0.94	0.11	30,30,40,46	0
2	EDO	A	604	4/4	0.94	0.32	36,45,47,49	0
2	EDO	A	601	4/4	0.95	0.08	32,36,44,54	0
4	UNL	A	610	1/-	0.97	0.25	21,21,21,21	0
4	UNL	A	607	1/-	0.97	0.24	12,12,12,12	0
4	UNL	A	609	1/-	0.98	0.25	30,30,30,30	0
4	UNL	A	611	1/-	0.99	0.18	19,19,19,19	0
4	UNL	A	608	1/-	0.99	0.17	13,13,13,13	0

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

There are no such residues in this entry.