



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 10, 2018 – 02:57 am GMT

PDB ID : 4PHU  
Title : Crystal structure of Human GPR40 bound to allosteric agonist TAK-875  
Authors : Srivastava, A.; Yano, J.K.; Hirozane, Y.; Kefala, G.; Snell, G.; Lane, W.;  
Gruswitz, F.; Ivetac, A.; Aertgeerts, K.; Nguyen, J.; Jennings, A.; Okada, K.  
Deposited on : 2014-05-07  
Resolution : 2.33 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 : 1.7.3 (157068), CSD as539be (2018)  
Xtrriage (Phenix) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

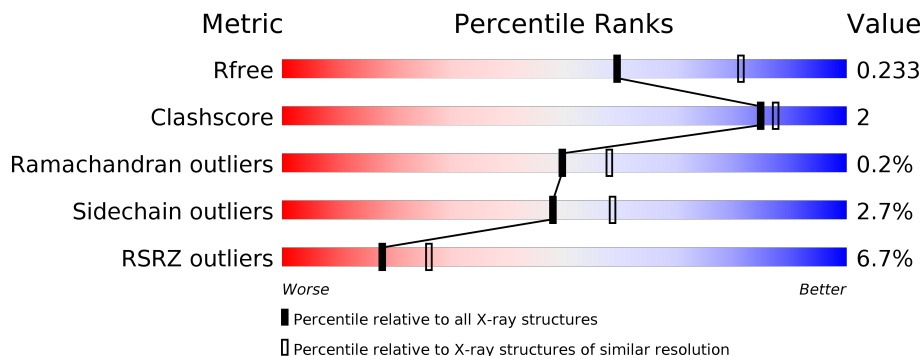
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.33 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))
$R_{free}$	111664	1763 (2.36-2.32)
Clashscore	122126	1858 (2.36-2.32)
Ramachandran outliers	120053	1834 (2.36-2.32)
Sidechain outliers	120020	1835 (2.36-2.32)
RSRZ outliers	108989	1737 (2.36-2.32)

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	491	

## 2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 3503 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 Free fatty acid receptor 1, Lysozyme.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	434	3286	2131	567	574	14	0	3	0

There are 39 discrepancies between the modelled and reference sequences:

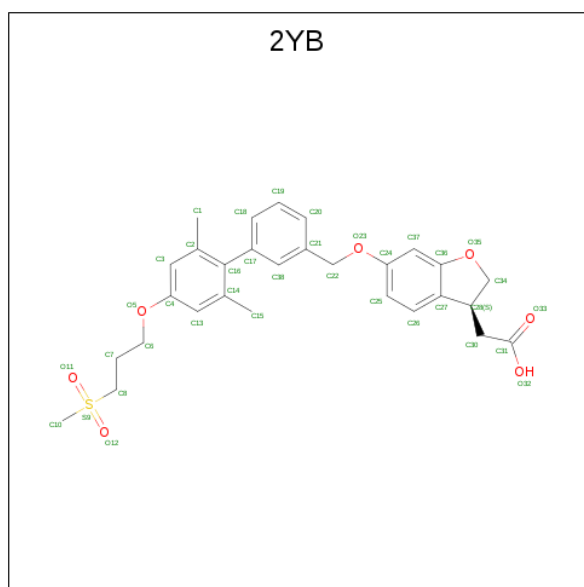
Chain	Residue	Modelled	Actual	Comment	Reference
A	-12	MET	-	expression tag	UNP O14842
A	-11	ASP	-	expression tag	UNP O14842
A	-10	TYR	-	expression tag	UNP O14842
A	-9	LYS	-	expression tag	UNP O14842
A	-8	ASP	-	expression tag	UNP O14842
A	-7	ASP	-	expression tag	UNP O14842
A	-6	ASP	-	expression tag	UNP O14842
A	-5	ASP	-	expression tag	UNP O14842
A	-4	LYS	-	expression tag	UNP O14842
A	-3	GLY	-	expression tag	UNP O14842
A	-2	SER	-	expression tag	UNP O14842
A	-1	ALA	-	expression tag	UNP O14842
A	0	THR	-	expression tag	UNP O14842
A	42	ALA	LEU	engineered mutation	UNP O14842
A	88	ALA	PHE	engineered mutation	UNP O14842
A	103	ALA	GLY	engineered mutation	UNP O14842
A	202	PHE	TYR	engineered mutation	UNP O14842
A	900	GLY	-	linker	UNP O14842
A	901	SER	-	linker	UNP O14842
A	1054	THR	CYS	engineered mutation	UNP P00720
A	1097	ALA	CYS	engineered mutation	UNP P00720
A	1902	GLY	-	linker	UNP P00720
A	1903	SER	-	linker	UNP P00720
A	2301	ALA	-	expression tag	UNP O14842
A	2302	GLU	-	expression tag	UNP O14842
A	2303	ASN	-	expression tag	UNP O14842
A	2304	LEU	-	expression tag	UNP O14842

*Continued on next page...*

Continued from previous page...

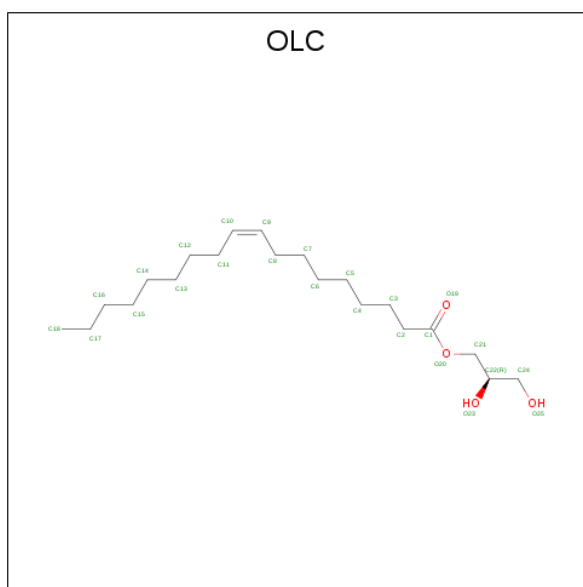
Chain	Residue	Modelled	Actual	Comment	Reference
A	2305	TYR	-	expression tag	UNP O14842
A	2306	PHE	-	expression tag	UNP O14842
A	2307	GLN	-	expression tag	UNP O14842
A	2308	GLY	-	expression tag	UNP O14842
A	2309	HIS	-	expression tag	UNP O14842
A	2310	HIS	-	expression tag	UNP O14842
A	2311	HIS	-	expression tag	UNP O14842
A	2312	HIS	-	expression tag	UNP O14842
A	2313	HIS	-	expression tag	UNP O14842
A	2314	HIS	-	expression tag	UNP O14842
A	2315	HIS	-	expression tag	UNP O14842
A	2316	HIS	-	expression tag	UNP O14842

- Molecule 2 is [(3S)-6-({2',6'-dimethyl-4'-[3-(methylsulfonyl)propoxy]biphenyl-3-yl}methoxy)-2,3-dihydro-1-benzofuran-3-yl]acetic acid (three-letter code: 2YB) (formula: C<sub>29</sub>H<sub>32</sub>O<sub>7</sub>S).



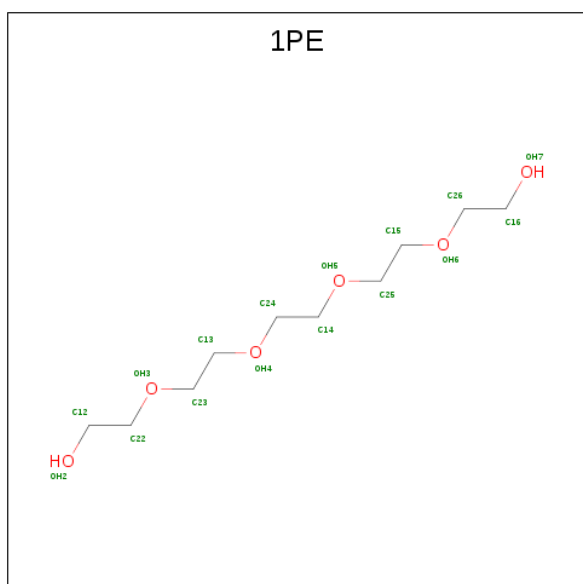
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	S		
2	A	1	37	29	7	1	0	0

- Molecule 3 is (2R)-2,3-dihydroxypropyl (9Z)-octadec-9-enoate (three-letter code: OLC) (formula: C<sub>21</sub>H<sub>40</sub>O<sub>4</sub>).



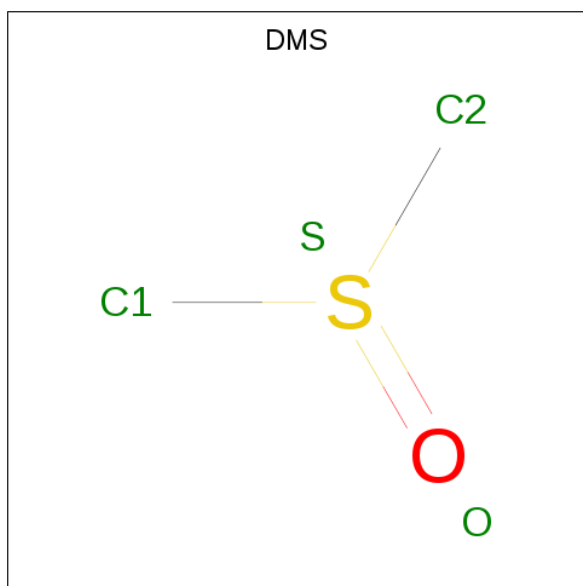
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C 16 16	0	0
3	A	1	Total C 8 8	0	0
3	A	1	Total C O 25 21 4	0	0
3	A	1	Total C 12 12	0	0
3	A	1	Total C 9 9	0	0

- Molecule 4 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula:  $C_{10}H_{22}O_6$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			16	10	6		

- Molecule 5 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C<sub>2</sub>H<sub>6</sub>OS).



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

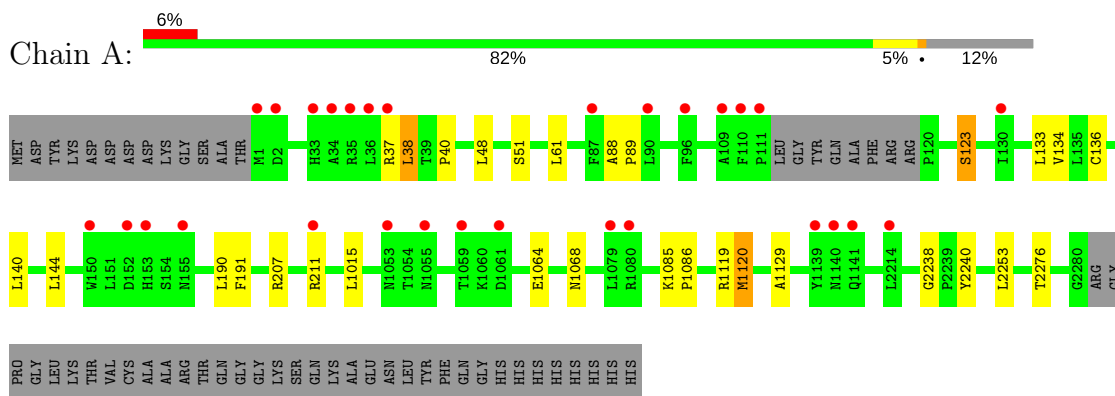
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	90	Total	O	0	0
			90	90		

### 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: Free fatty acid receptor 1, Lysozyme



## 4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	101.00Å 61.74Å 105.70Å 90.00° 108.90° 90.00°	Depositor
Resolution (Å)	30.00 – 2.33 30.03 – 2.33	Depositor EDS
% Data completeness (in resolution range)	97.8 (30.00-2.33) 97.9 (30.03-2.33)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.88 (at 2.34Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.198 , 0.233 0.203 , 0.233	Depositor DCC
$R_{free}$ test set	1312 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.2	Xtrriage
Anisotropy	0.267	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 54.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.022 for $1/2^*h+3/2^*k, 1/2^*h-1/2^*k, -1/2^*h-1/2^*k-l$ 0.015 for $1/2^*h-3/2^*k, -1/2^*h-1/2^*k, -1/2^*h+1/2^*k-l$	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3503	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.14% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: OLC, 2YB, DMS, 1PE

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	0.41	0/3377	0.55	0/4604

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

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	3286	0	3342	17	0
2	A	37	0	31	1	0
3	A	70	0	112	0	0
4	A	16	0	22	0	0
5	A	4	0	6	0	0
6	A	90	0	0	0	0
All	All	3503	0	3513	17	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 2.

All (17) 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:2240:TYR:OH	2:A:2401:2YB:O33	2.12	0.67
1:A:1120:MET:HB3	1:A:1129:ALA:HB2	1.83	0.60
1:A:48:LEU:HD23	1:A:2276:THR:HG21	1.84	0.59
1:A:37:ARG:O	1:A:38:LEU:CB	2.56	0.54
1:A:1120:MET:HG2	1:A:1129:ALA:N	2.27	0.50
1:A:134:VAL:HG22	1:A:190:LEU:CD2	2.41	0.50
1:A:88:ALA:HB3	1:A:89:PRO:HD3	1.93	0.50
1:A:191:PHE:CZ	1:A:2238:GLY:HA2	2.48	0.48
1:A:140:LEU:O	1:A:144:LEU:HB2	2.14	0.47
1:A:136:CYS:O	1:A:140:LEU:HB2	2.15	0.46
1:A:140:LEU:CD2	1:A:144:LEU:HD22	2.46	0.46
1:A:40:PRO:HA	1:A:123[A]:SER:OG	2.16	0.45
1:A:133:LEU:C	1:A:133:LEU:HD13	2.37	0.45
1:A:134:VAL:CG2	1:A:190:LEU:CD2	2.95	0.45
1:A:134:VAL:CG2	1:A:190:LEU:HD21	2.48	0.44
1:A:207:ARG:O	1:A:211:ARG:HG2	2.20	0.41
1:A:1085:LYS:HB3	1:A:1086:PRO:HD3	2.02	0.41

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	433/491 (88%)	419 (97%)	13 (3%)	1 (0%)	49 58

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	38	LEU

### 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	335/380 (88%)	325 (97%)	10 (3%)	44 54

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	51	SER
1	A	61	LEU
1	A	123[A]	SER
1	A	123[B]	SER
1	A	1015	LEU
1	A	1064	GLU
1	A	1068	ASN
1	A	1119	ARG
1	A	1120	MET
1	A	2253	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1053	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

8 ligands 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	2YB	A	2401	-	37,40,40	0.86	1 (2%)	52,57,57	1.44	6 (11%)
3	OLC	A	2402	-	15,15,24	0.28	0	14,14,25	0.41	0
3	OLC	A	2403	-	7,7,24	0.40	0	6,6,25	0.64	0
3	OLC	A	2404	-	24,24,24	0.98	1 (4%)	25,25,25	0.87	1 (4%)
3	OLC	A	2405	-	11,11,24	0.36	0	10,10,25	0.55	0
3	OLC	A	2406	-	8,8,24	0.38	0	7,7,25	0.73	0
4	1PE	A	2407	-	15,15,15	0.44	0	14,14,14	0.31	0
5	DMS	A	2408	-	3,3,3	0.42	0	3,3,3	0.80	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	2YB	A	2401	-	-	0/19/30/30	0/4/4/4
3	OLC	A	2402	-	-	0/13/13/24	0/0/0/0
3	OLC	A	2403	-	-	0/5/5/24	0/0/0/0
3	OLC	A	2404	-	-	0/24/24/24	0/0/0/0
3	OLC	A	2405	-	-	0/9/9/24	0/0/0/0
3	OLC	A	2406	-	-	0/6/6/24	0/0/0/0
4	1PE	A	2407	-	-	0/13/13/13	0/0/0/0
5	DMS	A	2408	-	-	0/0/0/0	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2401	2YB	C8-S9	3.03	1.82	1.78
3	A	2404	OLC	O20-C1	4.49	1.46	1.33

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	2401	2YB	C37-C36-C27	-4.90	118.49	123.60
2	A	2401	2YB	C30-C28-C27	-3.53	105.29	114.17
2	A	2401	2YB	O12-S9-C8	-2.91	106.34	108.35
2	A	2401	2YB	C26-C27-C36	2.69	121.29	119.58
2	A	2401	2YB	C34-C28-C27	2.81	103.74	100.84
3	A	2404	OLC	O20-C1-C2	2.94	120.42	111.92
2	A	2401	2YB	C34-O35-C36	3.72	111.03	106.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	2401	2YB	1	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

### 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	434/491 (88%)	0.34	29 (6%) <span style="border: 1px solid red; padding: 2px;">18</span> <span style="border: 1px solid red; padding: 2px;">26</span>	29, 44, 80, 92	0

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	36	LEU	6.8
1	A	153	HIS	6.0
1	A	155	ASN	5.9
1	A	150	TRP	4.0
1	A	2214	LEU	3.9
1	A	211	ARG	3.7
1	A	111	PRO	3.6
1	A	37	ARG	3.5
1	A	110	PHE	3.4
1	A	1053	ASN	3.3
1	A	1079	LEU	3.2
1	A	1080	ARG	3.2
1	A	34	ALA	2.9
1	A	1	MET	2.8
1	A	1055	ASN	2.7
1	A	1061	ASP	2.6
1	A	33	HIS	2.6
1	A	152	ASP	2.4
1	A	35	ARG	2.4
1	A	96	PHE	2.3
1	A	2	ASP	2.3
1	A	1141	GLN	2.3
1	A	130	ILE	2.3
1	A	1140	ASN	2.2
1	A	90	LEU	2.2
1	A	109	ALA	2.2
1	A	1139	TYR	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	87	PHE	2.1
1	A	1059	THR	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( $\text{\AA}^2$ )	Q<0.9
3	OLC	A	2404	25/25	0.72	0.35	52,59,91,92	0
3	OLC	A	2406	9/25	0.82	0.23	62,65,67,67	0
3	OLC	A	2405	12/25	0.84	0.20	55,58,66,68	0
3	OLC	A	2403	8/25	0.85	0.23	56,60,63,63	0
2	2YB	A	2401	37/37	0.85	0.17	35,38,79,87	0
5	DMS	A	2408	4/4	0.86	0.20	61,67,68,74	0
4	1PE	A	2407	16/16	0.89	0.19	64,67,77,77	0
3	OLC	A	2402	16/25	0.95	0.15	48,51,58,60	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.