



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 15, 2017 – 12:32 AM EST

PDB ID : 5UXZ  
Title : X-ray crystal structure of Halotag bound to the P9 benzothiadiazole fluorogenic ligand  
Authors : Dunham, N.P.; Boal, A.K.  
Deposited on : unknown  
Resolution : 1.92 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

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

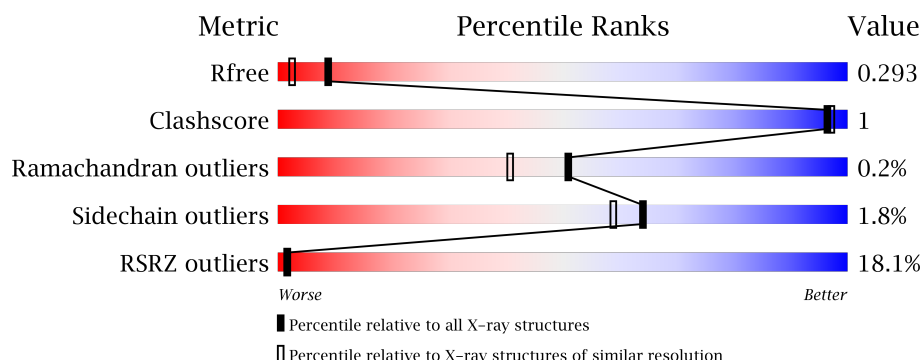
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.92 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))
$R_{free}$	100719	6276 (1.94-1.90)
Clashscore	112137	7025 (1.94-1.90)
Ramachandran outliers	110173	6947 (1.94-1.90)
Sidechain outliers	110143	6948 (1.94-1.90)
RSRZ outliers	101464	6332 (1.94-1.90)

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	307	<div> <div>10%</div> <div> <div></div> <div>92%</div> <div>6%</div> </div> </div>
1	B	307	<div> <div>24%</div> <div> <div></div> <div>89%</div> <div>7%</div> </div> </div>

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	EDO	A	404	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 4827 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 Haloalkane dehalogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	290	Total	C	N	O	S	0	0	0
			2328	1515	392	412	9			
1	B	295	Total	C	N	O	S	0	0	0
			2363	1537	397	420	9			

There are 78 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	47	VAL	LEU	conflict	UNP P0A3G2
A	58	THR	SER	conflict	UNP P0A3G2
A	78	GLY	ASP	conflict	UNP P0A3G2
A	87	PHE	TYR	conflict	UNP P0A3G2
A	88	MET	LEU	conflict	UNP P0A3G2
A	128	PHE	CYS	conflict	UNP P0A3G2
A	155	THR	ALA	conflict	UNP P0A3G2
A	160	LYS	GLU	conflict	UNP P0A3G2
A	167	VAL	ALA	conflict	UNP P0A3G2
A	172	THR	ALA	conflict	UNP P0A3G2
A	175	MET	LYS	conflict	UNP P0A3G2
A	176	GLY	CYS	conflict	UNP P0A3G2
A	195	ASN	LYS	conflict	UNP P0A3G2
A	224	GLU	ALA	conflict	UNP P0A3G2
A	227	ASP	ASN	conflict	UNP P0A3G2
A	257	LYS	GLU	conflict	UNP P0A3G2
A	264	ALA	THR	conflict	UNP P0A3G2
A	272	ASN	HIS	conflict	UNP P0A3G2
A	273	LEU	TYR	conflict	UNP P0A3G2
A	291	SER	-	expression tag	UNP P0A3G2
A	293	THR	-	expression tag	UNP P0A3G2
A	294	LEU	-	expression tag	UNP P0A3G2
A	295	GLU	-	expression tag	UNP P0A3G2
A	296	ILE	-	expression tag	UNP P0A3G2
A	297	SER	-	expression tag	UNP P0A3G2

*Continued on next page...*

*Continued from previous page...*

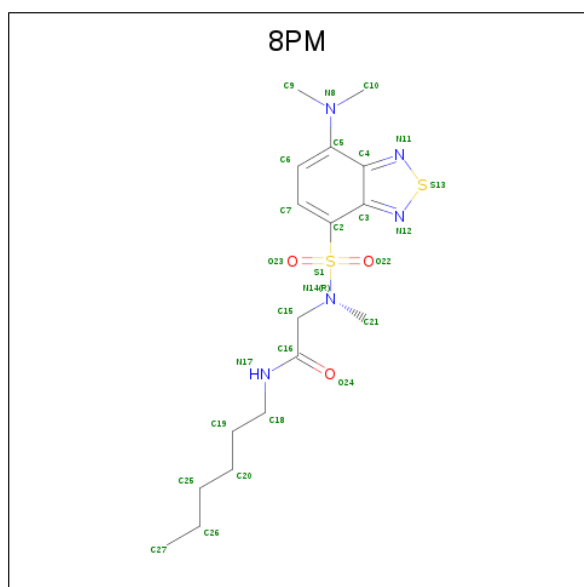
Chain	Residue	Modelled	Actual	Comment	Reference
A	298	GLY	-	expression tag	UNP P0A3G2
A	299	LEU	-	expression tag	UNP P0A3G2
A	300	GLU	-	expression tag	UNP P0A3G2
A	301	ILE	-	expression tag	UNP P0A3G2
A	302	SER	-	expression tag	UNP P0A3G2
A	303	GLY	-	expression tag	UNP P0A3G2
A	304	LEU	-	expression tag	UNP P0A3G2
A	305	GLU	-	expression tag	UNP P0A3G2
A	306	HIS	-	expression tag	UNP P0A3G2
A	307	HIS	-	expression tag	UNP P0A3G2
A	308	HIS	-	expression tag	UNP P0A3G2
A	309	HIS	-	expression tag	UNP P0A3G2
A	310	HIS	-	expression tag	UNP P0A3G2
A	311	HIS	-	expression tag	UNP P0A3G2
B	47	VAL	LEU	conflict	UNP P0A3G2
B	58	THR	SER	conflict	UNP P0A3G2
B	78	GLY	ASP	conflict	UNP P0A3G2
B	87	PHE	TYR	conflict	UNP P0A3G2
B	88	MET	LEU	conflict	UNP P0A3G2
B	128	PHE	CYS	conflict	UNP P0A3G2
B	155	THR	ALA	conflict	UNP P0A3G2
B	160	LYS	GLU	conflict	UNP P0A3G2
B	167	VAL	ALA	conflict	UNP P0A3G2
B	172	THR	ALA	conflict	UNP P0A3G2
B	175	MET	LYS	conflict	UNP P0A3G2
B	176	GLY	CYS	conflict	UNP P0A3G2
B	195	ASN	LYS	conflict	UNP P0A3G2
B	224	GLU	ALA	conflict	UNP P0A3G2
B	227	ASP	ASN	conflict	UNP P0A3G2
B	257	LYS	GLU	conflict	UNP P0A3G2
B	264	ALA	THR	conflict	UNP P0A3G2
B	272	ASN	HIS	conflict	UNP P0A3G2
B	273	LEU	TYR	conflict	UNP P0A3G2
B	291	SER	-	expression tag	UNP P0A3G2
B	292	THR	-	expression tag	UNP P0A3G2
B	293	LEU	-	expression tag	UNP P0A3G2
B	294	GLU	-	expression tag	UNP P0A3G2
B	295	ILE	-	expression tag	UNP P0A3G2
B	296	SER	-	expression tag	UNP P0A3G2
B	297	GLY	-	expression tag	UNP P0A3G2
B	298	LEU	-	expression tag	UNP P0A3G2
B	299	GLU	-	expression tag	UNP P0A3G2

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	300	ILE	-	expression tag	UNP P0A3G2
B	301	SER	-	expression tag	UNP P0A3G2
B	302	GLY	-	expression tag	UNP P0A3G2
B	303	LEU	-	expression tag	UNP P0A3G2
B	304	GLU	-	expression tag	UNP P0A3G2
B	305	HIS	-	expression tag	UNP P0A3G2
B	306	HIS	-	expression tag	UNP P0A3G2
B	307	HIS	-	expression tag	UNP P0A3G2
B	308	HIS	-	expression tag	UNP P0A3G2
B	309	HIS	-	expression tag	UNP P0A3G2
B	310	HIS	-	expression tag	UNP P0A3G2

- Molecule 2 is N 2 -{[7-(dimethylamino)-2,1,3-benzothiadiazol-4-yl]sulfonyl}-N-hexyl-N 2 -methylglycinamide (three-letter code: 8PM) (formula: C<sub>17</sub>H<sub>27</sub>N<sub>5</sub>O<sub>3</sub>S<sub>2</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	S	0	0
			27	17	5	3	2		
2	B	1	Total	C	N	O	S	0	0
			27	17	5	3	2		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

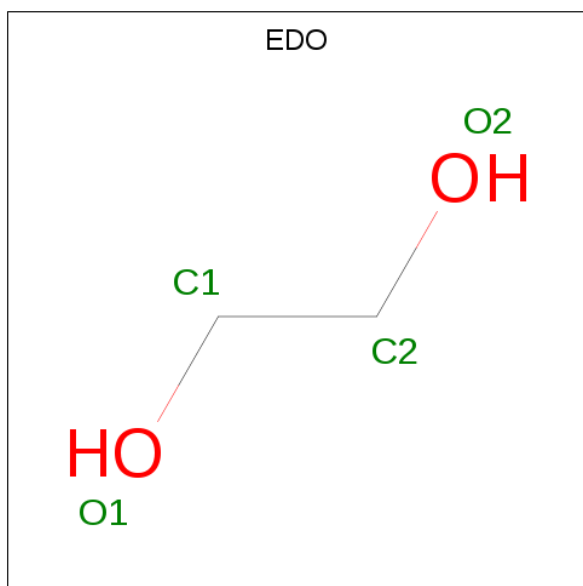
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cl	0	0
			1	1		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Cl	0	0
			1	1		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

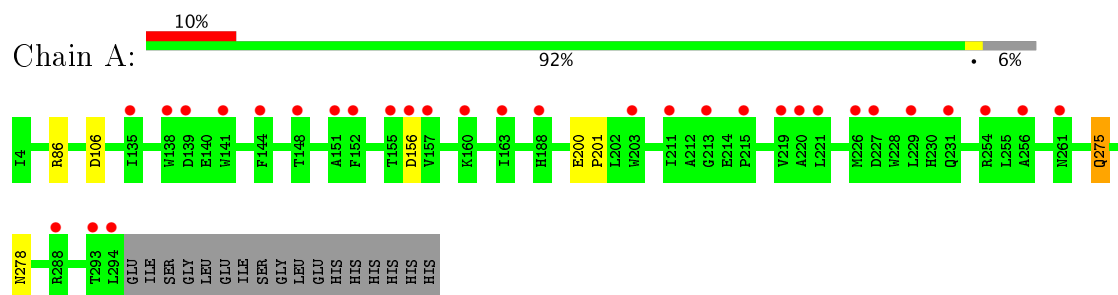
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	53	Total	O	0	0
			53	53		
5	B	19	Total	O	0	0
			19	19		

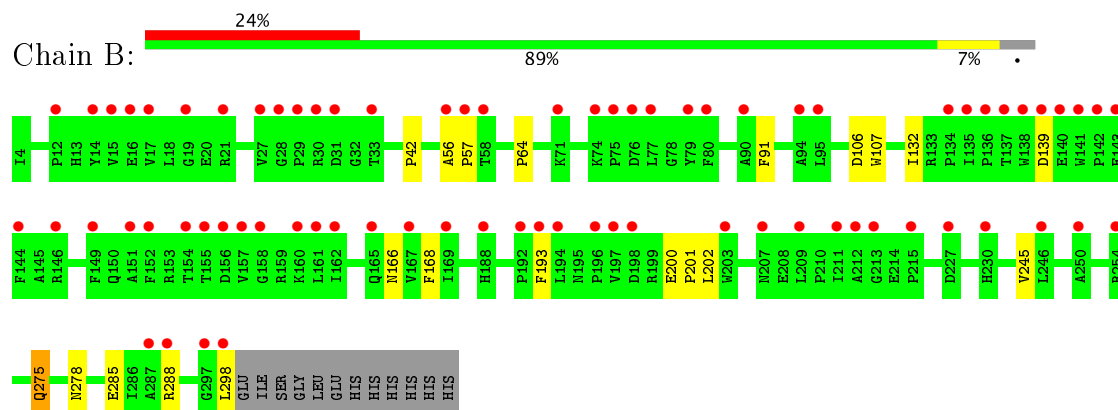
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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: Haloalkane dehalogenase



#### • Molecule 1: Haloalkane dehalogenase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.94Å 69.12Å 83.90Å 90.00° 94.67° 90.00°	Depositor
Resolution (Å)	50.00 – 1.92 49.77 – 1.92	Depositor EDS
% Data completeness (in resolution range)	81.9 (50.00-1.92) 81.9 (49.77-1.92)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.44 (at 1.92Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.269 , 0.287 0.274 , 0.293	Depositor DCC
$R_{free}$ test set	1806 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	19.7	Xtriage
Anisotropy	0.091	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 44.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	4827	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.78% 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: EDO, 8PM, CL

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	0.33	1/2407 (0.0%)	0.46	0/3292
1	B	0.33	1/2442 (0.0%)	0.45	0/3339
All	All	0.33	2/4849 (0.0%)	0.45	0/6631

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	106	ASP	CG-OD1	8.96	1.46	1.25
1	A	106	ASP	CG-OD1	8.70	1.45	1.25

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	2328	0	2271	2	0
1	B	2363	0	2307	9	0
2	A	27	0	0	0	0
2	B	27	0	0	1	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	8	0	12	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	53	0	0	0	0
5	B	19	0	0	0	0
All	All	4827	0	4590	11	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 1.

The worst 5 of 11 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:B:285:GLU:OE1	1:B:288:ARG:NH1	2.37	0.57
1:A:275:GLN:H	1:A:275:GLN:HE21	1.53	0.56
1:B:275:GLN:HE21	1:B:275:GLN:H	1.59	0.51
1:B:200:GLU:N	1:B:201:PRO:HD2	2.29	0.46
1:B:245:VAL:HG11	2:B:401:8PM:O24	2.17	0.44

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	288/307 (94%)	275 (96%)	13 (4%)	0	100	100
1	B	293/307 (95%)	278 (95%)	14 (5%)	1 (0%)	44	33
All	All	581/614 (95%)	553 (95%)	27 (5%)	1 (0%)	51	41

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	166	ASN

### 5.3.2 Protein sidechains ⓘ

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	250/265 (94%)	246 (98%)	4 (2%)	68	63
1	B	254/265 (96%)	249 (98%)	5 (2%)	60	54
All	All	504/530 (95%)	495 (98%)	9 (2%)	64	59

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	139	ASP
1	B	298	LEU
1	B	275	GLN
1	A	275	GLN
1	B	168	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	217	ASN
1	B	278	ASN
1	B	272	ASN
1	A	275	GLN
1	B	275	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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 [i](#)

Of 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 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	8PM	A	401	1	26,28,28	3.13	5 (19%)	30,39,39	2.00	8 (26%)
4	EDO	A	403	-	3,3,3	0.50	0	2,2,2	0.15	0
4	EDO	A	404	-	3,3,3	0.44	0	2,2,2	0.28	0
2	8PM	B	401	1	26,28,28	3.09	5 (19%)	30,39,39	2.05	9 (30%)

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	8PM	A	401	1	-	0/27/27/27	0/2/2/2
4	EDO	A	403	-	-	0/1/1/1	0/0/0/0
4	EDO	A	404	-	-	0/1/1/1	0/0/0/0
2	8PM	B	401	1	-	0/27/27/27	0/2/2/2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	8PM	C2-S1	-11.66	1.60	1.77
2	B	401	8PM	C2-S1	-11.12	1.61	1.77
2	A	401	8PM	C7-C2	2.12	1.40	1.37
2	A	401	8PM	C4-C3	2.40	1.51	1.42
2	B	401	8PM	C7-C2	2.43	1.40	1.37

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	401	8PM	O22-S1-O23	-7.11	107.50	119.46
2	A	401	8PM	O22-S1-O23	-6.94	107.79	119.46
2	B	401	8PM	O24-C16-N17	-2.18	118.81	122.97
2	B	401	8PM	C18-N17-C16	2.01	126.71	122.84
2	B	401	8PM	C15-C16-N17	2.07	120.48	115.67

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	B	401	8PM	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 [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	290/307 (94%)	0.81	31 (10%) 7 8	5, 17, 37, 41	35 (12%)
1	B	295/307 (96%)	1.44	75 (25%) 1 1	15, 32, 51, 57	35 (11%)
All	All	585/614 (95%)	1.13	106 (18%) 1 2	5, 25, 48, 57	70 (11%)

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	193	PHE	7.8
1	B	194	LEU	6.6
1	B	203	TRP	6.1
1	B	140	GLU	5.3
1	A	157	VAL	5.1

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	EDO	A	404	4/4	0.92	0.25	2.10	16,16,16,16	0
2	8PM	A	401	27/27	0.86	0.18	1.59	14,23,25,25	0
4	EDO	A	403	4/4	0.89	0.14	0.49	17,18,18,18	0
2	8PM	B	401	27/27	0.84	0.18	0.27	20,31,33,33	0
3	CL	B	402	1/1	0.94	0.10	-1.95	26,26,26,26	0
3	CL	A	402	1/1	0.97	0.09	-2.20	16,16,16,16	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.