



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2017 – 02:38 PM EDT

PDB ID : 5W0M  
Title : Structure of human TUT7 catalytic module (CM) in complex with U5 RNA  
Authors : Faehnle, C.R.; Walleshauser, J.; Joshua-Tor, L.  
Deposited on : unknown  
Resolution : 2.30 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:

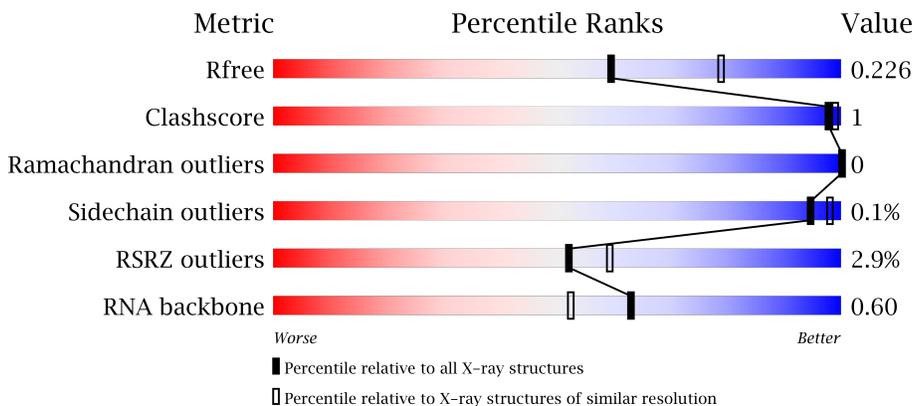
MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20029824

# 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 2.30 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	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)
RNA backbone	2435	1106 (2.80-1.80)

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	389	 3% 94% 2% 1% 1%
1	B	389	 0% 88% 10% 1% 1%
1	C	389	 5% 92% 5% 1% 1%
2	H	5	 80% 20%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	I	5	 60% 20% 20%
2	J	5	 60% 20% 20%

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
3	SO4	B	1401	-	-	-	X
3	SO4	C	1401	-	-	-	X
4	ZN	A	1403	-	-	-	X
5	IOD	A	1404	-	-	-	X
5	IOD	C	1404	-	-	-	X

## 2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 18313 atoms, of which 8988 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 Terminal uridylyltransferase 7.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	372	6037	1953	3019	504	541	20	0	0	0
1	B	350	5674	1835	2837	472	514	16	0	0	0
1	C	371	6017	1944	3009	502	542	20	0	0	0

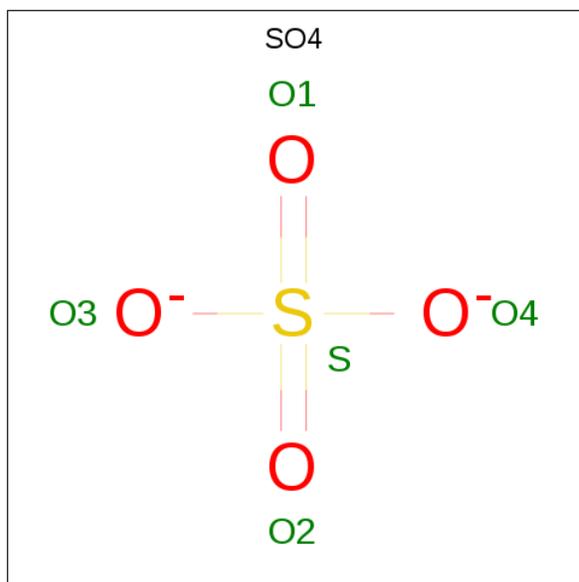
There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	977	GLY	-	expression tag	UNP Q5VYS8
A	978	ALA	-	expression tag	UNP Q5VYS8
A	979	GLY	-	expression tag	UNP Q5VYS8
A	980	ALA	-	expression tag	UNP Q5VYS8
A	981	GLY	-	expression tag	UNP Q5VYS8
A	982	SER	-	expression tag	UNP Q5VYS8
A	1060	ALA	ASP	engineered mutation	UNP Q5VYS8
B	977	GLY	-	expression tag	UNP Q5VYS8
B	978	ALA	-	expression tag	UNP Q5VYS8
B	979	GLY	-	expression tag	UNP Q5VYS8
B	980	ALA	-	expression tag	UNP Q5VYS8
B	981	GLY	-	expression tag	UNP Q5VYS8
B	982	SER	-	expression tag	UNP Q5VYS8
B	1060	ALA	ASP	engineered mutation	UNP Q5VYS8
C	977	GLY	-	expression tag	UNP Q5VYS8
C	978	ALA	-	expression tag	UNP Q5VYS8
C	979	GLY	-	expression tag	UNP Q5VYS8
C	980	ALA	-	expression tag	UNP Q5VYS8
C	981	GLY	-	expression tag	UNP Q5VYS8
C	982	SER	-	expression tag	UNP Q5VYS8
C	1060	ALA	ASP	engineered mutation	UNP Q5VYS8

- Molecule 2 is a RNA chain called U5 single-stranded RNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
2	H	4	Total	C	H	N	O	P	0	0	0
			118	36	41	8	30	3			
2	I	4	Total	C	H	N	O	P	0	0	0
			118	36	41	8	30	3			
2	J	4	Total	C	H	N	O	P	0	0	0
			118	36	41	8	30	3			

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 5 is IODIDE ION (three-letter code: IOD) (formula: I).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total I 1 1	0	0
5	A	1	Total I 1 1	0	0
5	C	2	Total I 2 2	0	0

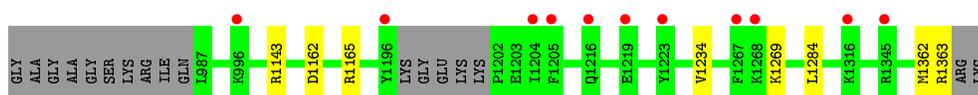
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	61	Total O 61 61	0	0
6	B	48	Total O 48 48	0	0
6	C	76	Total O 76 76	0	0
6	H	6	Total O 6 6	0	0
6	I	4	Total O 4 4	0	0
6	J	10	Total O 10 10	0	0

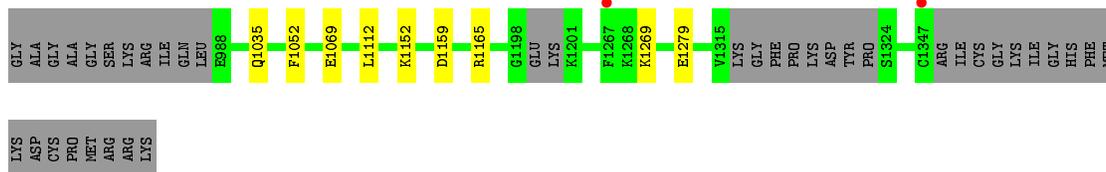
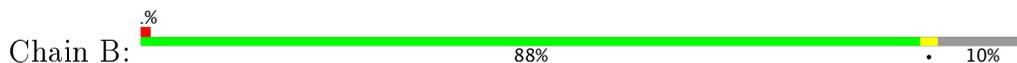
### 3 Residue-property plots [i](#)

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.

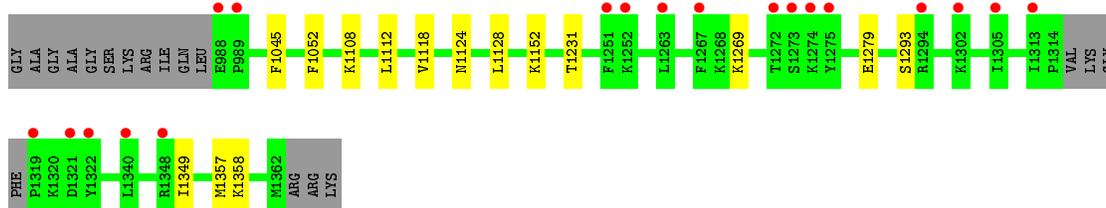
- Molecule 1: Terminal uridylyltransferase 7



- Molecule 1: Terminal uridylyltransferase 7



- Molecule 1: Terminal uridylyltransferase 7



- Molecule 2: U5 single-stranded RNA



- Molecule 2: U5 single-stranded RNA





- Molecule 2: U5 single-stranded RNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	135.47Å 135.47Å 179.27Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	98.17 – 2.30 98.17 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.8 (98.17-2.30) 99.9 (98.17-2.30)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.20 (at 2.29Å)	Xtrriage
Refinement program	PHENIX 1.10_2155	Depositor
R, $R_{free}$	0.188 , 0.229 0.184 , 0.226	Depositor DCC
$R_{free}$ test set	4138 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	43.7	Xtrriage
Anisotropy	0.334	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 52.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.037 for h,-h-k,-l	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	18313	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.36% 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: ZN, IOD, SO4

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	0.30	0/3093	0.49	1/4181 (0.0%)
1	B	0.29	0/2904	0.45	1/3927 (0.0%)
1	C	0.30	0/3082	0.46	0/4164
2	H	0.35	0/84	0.81	0/128
2	I	0.36	0/84	0.72	0/128
2	J	0.32	0/84	0.77	0/128
All	All	0.30	0/9331	0.48	2/12656 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	1362	MET	C-N-CA	9.39	145.18	121.70
1	B	1112	LEU	C-N-CA	5.47	135.37	121.70

There are no chirality outliers.

There are no planarity outliers.

### 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	3018	3019	3024	4	0
1	B	2837	2837	2844	5	0
1	C	3008	3009	3015	8	2

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	H	77	41	42	0	0
2	I	77	41	42	0	0
2	J	77	41	42	1	0
3	A	10	0	0	0	0
3	B	5	0	0	0	0
3	C	5	0	0	0	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
5	C	2	0	0	1	0
6	A	61	0	0	0	0
6	B	48	0	0	0	0
6	C	76	0	0	1	0
6	H	6	0	0	0	0
6	I	4	0	0	0	0
6	J	10	0	0	0	0
All	All	9325	8988	9009	16	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 1.

The worst 5 of 16 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:1162:ASP:OD2	1:A:1165:ARG:NH1	2.16	0.77
1:B:1269:LYS:NZ	1:B:1279:GLU:OE1	2.16	0.70
1:B:1069:GLU:N	1:B:1069:GLU:OE2	2.37	0.57
1:C:1269:LYS:NZ	1:C:1279:GLU:OE1	2.32	0.55
1:A:1269:LYS:NZ	1:A:1284:LEU:O	2.44	0.51

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:C:1231:THR:O	1:C:1358:LYS:NZ[6_555]	1.96	0.24
1:C:1231:THR:O	1:C:1358:LYS:HZ2[6_555]	1.56	0.04

## 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	368/389 (95%)	361 (98%)	7 (2%)	0	100	100
1	B	344/389 (88%)	335 (97%)	9 (3%)	0	100	100
1	C	367/389 (94%)	360 (98%)	7 (2%)	0	100	100
All	All	1079/1167 (92%)	1056 (98%)	23 (2%)	0	100	100

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	337/348 (97%)	337 (100%)	0	100	100
1	B	317/348 (91%)	317 (100%)	0	100	100
1	C	336/348 (97%)	335 (100%)	1 (0%)	94	97
All	All	990/1044 (95%)	989 (100%)	1 (0%)	94	98

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

Mol	Chain	Res	Type
1	C	1112	LEU

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

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	H	3/5 (60%)	0	0
2	I	3/5 (60%)	1 (33%)	0
2	J	3/5 (60%)	0	0
All	All	9/15 (60%)	1 (11%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	I	4	U

There are no RNA pucker outliers to report.

### 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 [i](#)

Of 10 ligands modelled in this entry, 6 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	$\# Z  > 2$
3	SO4	A	1401	-	4,4,4	0.19	0	6,6,6	0.15	0
3	SO4	A	1402	-	4,4,4	0.14	0	6,6,6	0.29	0
3	SO4	B	1401	-	4,4,4	0.11	0	6,6,6	0.09	0
3	SO4	C	1401	-	4,4,4	0.14	0	6,6,6	0.23	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
3	SO4	A	1401	-	-	0/0/0/0	0/0/0/0
3	SO4	A	1402	-	-	0/0/0/0	0/0/0/0
3	SO4	B	1401	-	-	0/0/0/0	0/0/0/0
3	SO4	C	1401	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	372/389 (95%)	0.29	11 (2%) 51 58	31, 55, 93, 113	0
1	B	350/389 (89%)	0.31	2 (0%) 89 92	32, 58, 91, 116	0
1	C	371/389 (95%)	0.34	19 (5%) 29 36	33, 54, 92, 115	0
2	H	4/5 (80%)	0.09	0 100 100	51, 54, 55, 98	0
2	I	4/5 (80%)	0.18	0 100 100	50, 51, 62, 107	0
2	J	4/5 (80%)	0.22	0 100 100	40, 48, 56, 96	0
All	All	1105/1182 (93%)	0.31	32 (2%) 52 59	31, 56, 93, 116	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1223	TYR	5.4
1	A	1196	TYR	4.3
1	C	1274	LYS	4.1
1	C	1275	TYR	3.9
1	C	1294	ARG	3.5

### 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(Å <sup>2</sup> )	Q<0.9
4	ZN	A	1403	1/1	0.77	0.44	9.86	115,115,115,115	0
3	SO4	C	1401	5/5	0.99	0.20	5.62	39,53,68,68	0
3	SO4	B	1401	5/5	0.98	0.23	4.68	47,58,72,77	0
5	IOD	A	1404	1/1	1.00	0.21	3.88	58,58,58,58	1
5	IOD	C	1404	1/1	0.99	0.18	2.76	59,59,59,59	1
3	SO4	A	1401	5/5	0.99	0.17	1.56	44,60,65,83	0
3	SO4	A	1402	5/5	0.99	0.21	0.77	44,45,55,65	0
4	ZN	C	1402	1/1	0.95	0.17	0.37	56,56,56,56	0
5	IOD	C	1403	1/1	0.98	0.11	-0.97	68,68,68,68	1
5	IOD	B	1402	1/1	1.00	0.11	-3.96	45,45,45,45	1

## 6.5 Other polymers [i](#)

There are no such residues in this entry.