



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 25, 2020 – 08:08 am BST

PDB ID : 6IW6  
Title : Crystal structure of the Lin28-interacting module of human TUT4  
Authors : Yamashita, S.; Tomita, K.  
Deposited on : 2018-12-04  
Resolution : 2.40 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 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

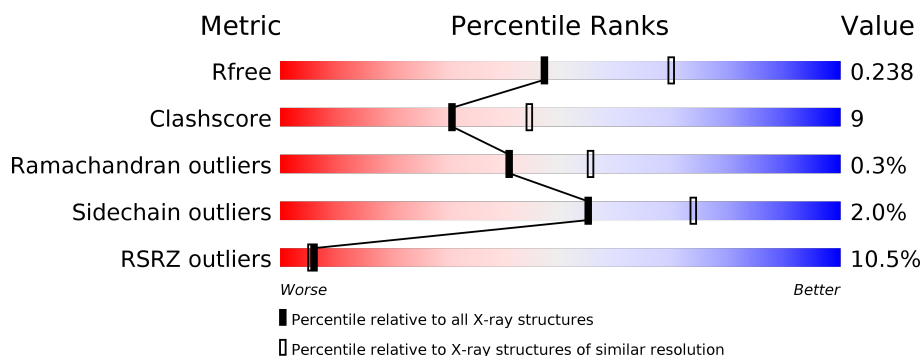
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	448	<div> <div>8%</div> <div>75%</div> <div>21%</div> <div>..</div> </div>
1	B	448	<div> <div>13%</div> <div>78%</div> <div>18%</div> <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	1005	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7285 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 Terminal uridylyltransferase 4, Terminal uridylyltransferase 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	438	Total	C	N	O	S	0	0	0
			3567	2286	620	641	20			
1	B	438	Total	C	N	O	S	0	0	0
			3567	2286	620	641	20			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	252	MET	-	expression tag	UNP Q5TAX3
A	724	LEU	-	expression tag	UNP Q5TAX3
A	725	GLU	-	expression tag	UNP Q5TAX3
A	726	HIS	-	expression tag	UNP Q5TAX3
A	727	HIS	-	expression tag	UNP Q5TAX3
A	728	HIS	-	expression tag	UNP Q5TAX3
A	729	HIS	-	expression tag	UNP Q5TAX3
A	730	HIS	-	expression tag	UNP Q5TAX3
A	731	HIS	-	expression tag	UNP Q5TAX3
B	252	MET	-	expression tag	UNP Q5TAX3
B	724	LEU	-	expression tag	UNP Q5TAX3
B	725	GLU	-	expression tag	UNP Q5TAX3
B	726	HIS	-	expression tag	UNP Q5TAX3
B	727	HIS	-	expression tag	UNP Q5TAX3
B	728	HIS	-	expression tag	UNP Q5TAX3
B	729	HIS	-	expression tag	UNP Q5TAX3
B	730	HIS	-	expression tag	UNP Q5TAX3
B	731	HIS	-	expression tag	UNP Q5TAX3

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

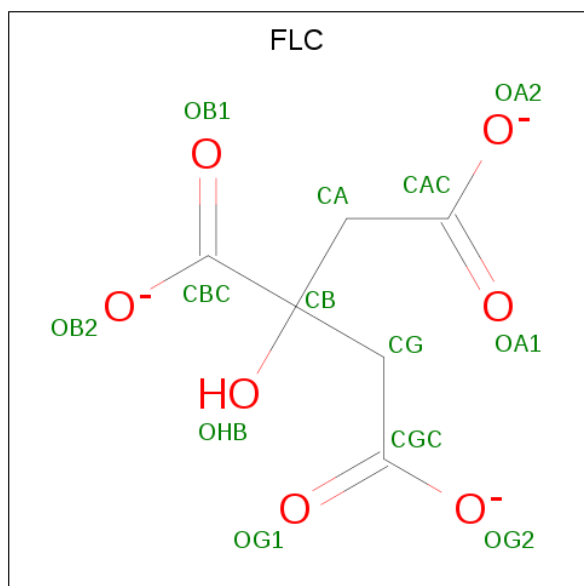
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Zn	0	0
			2	2		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	2	Total	Zn	0	0
			2	2		

- Molecule 3 is CITRATE ANION (three-letter code: FLC) (formula:  $C_6H_5O_7$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			13	6	7		
3	B	1	Total	C	O	0	0
			13	6	7		

- 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 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	A	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0

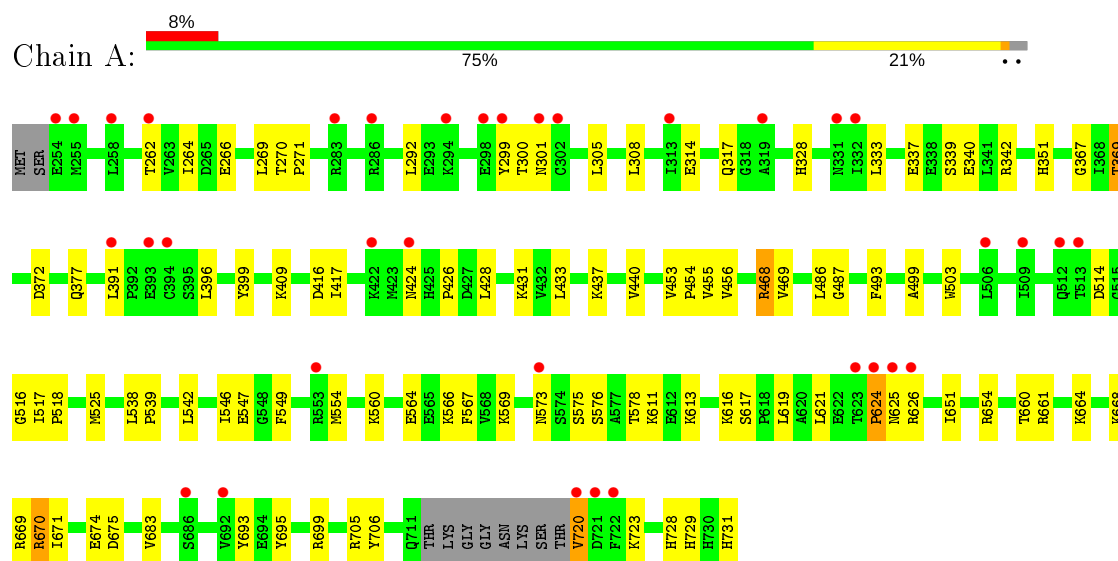
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	33	Total O 33 33	0	0
5	B	56	Total O 56 56	0	0

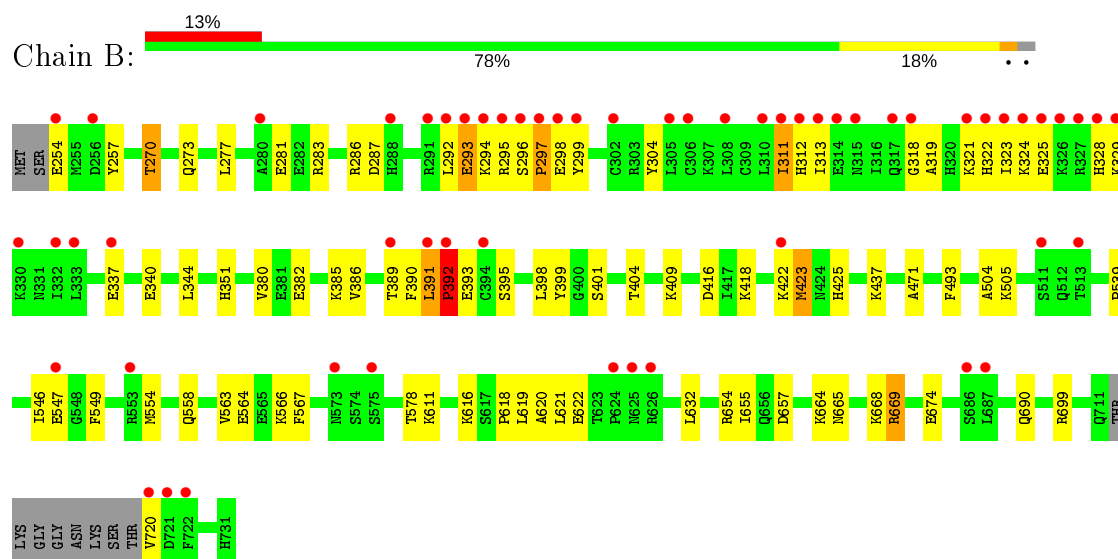
### 3 Residue-property plots [i](#)

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 4,Terminal uridylyltransferase 4



- Molecule 1: Terminal uridylyltransferase 4,Terminal uridylyltransferase 4



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.27Å 127.83Å 168.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.96 – 2.40 49.50 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.8 (19.96-2.40) 99.9 (49.50-2.40)	Depositor EDS
$R_{merge}$	0.19	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.34 (at 2.39Å)	Xtriage
Refinement program	PHENIX 1.12-2829_1309	Depositor
R, $R_{free}$	0.210 , 0.237 0.211 , 0.238	Depositor DCC
$R_{free}$ test set	2401 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.8	Xtriage
Anisotropy	0.284	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 43.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7285	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 42.07 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.1517e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: ZN, FLC, 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	0.39	2/3647 (0.1%)	0.56	1/4927 (0.0%)
1	B	0.42	2/3647 (0.1%)	0.58	2/4927 (0.0%)
All	All	0.40	4/7294 (0.1%)	0.57	3/9854 (0.0%)

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
1	B	0	1
All	All	0	3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	660	THR	C-N	6.67	1.49	1.34
1	B	297	PRO	N-CD	5.34	1.55	1.47
1	A	271	PRO	N-CD	5.29	1.55	1.47
1	B	392	PRO	N-CD	5.13	1.55	1.47

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	391	LEU	C-N-CD	6.07	141.15	128.40
1	A	270	THR	C-N-CD	5.54	140.04	128.40
1	B	423	MET	N-CA-C	5.14	124.89	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	624	PRO	Peptide
1	A	661	ARG	Mainchain
1	B	422	LYS	Peptide

## 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	3567	0	3617	70	1
1	B	3567	0	3617	63	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	13	0	5	2	0
3	B	13	0	5	0	0
4	A	12	0	18	1	1
4	B	20	0	30	5	0
5	A	33	0	0	3	0
5	B	56	0	0	2	0
All	All	7285	0	7292	133	1

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 9.

The worst 5 of 133 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:294:LYS:O	1:B:297:PRO:HG3	1.38	1.22
1:A:455:VAL:HG11	1:A:468:ARG:HB3	1.59	0.84
1:B:294:LYS:O	1:B:297:PRO:CG	2.27	0.77
1:A:668:LYS:O	1:A:669:ARG:HB2	1.86	0.75
1:B:318:GLY:HA2	1:B:321:LYS:HE2	1.71	0.72

All (1) 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:A:670:ARG:NH1	4:A:1004:EDO:O2[6_555]	1.85	0.35

## 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	434/448 (97%)	414 (95%)	20 (5%)	0	100	100
1	B	434/448 (97%)	409 (94%)	22 (5%)	3 (1%)	22	32
All	All	868/896 (97%)	823 (95%)	42 (5%)	3 (0%)	41	55

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	392	PRO
1	B	293	GLU
1	B	669	ARG

### 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	398/406 (98%)	387 (97%)	11 (3%)	43	63
1	B	398/406 (98%)	393 (99%)	5 (1%)	69	84
All	All	796/812 (98%)	780 (98%)	16 (2%)	55	74

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	542	LEU
1	A	575	SER
1	B	311	ILE
1	A	493	PHE
1	B	437	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	312	HIS
1	B	425	HIS

### 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 [i](#)

Of 14 ligands modelled in this entry, 4 are monoatomic - leaving 10 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	$\# Z  > 2$
4	EDO	B	1008	-	3,3,3	0.47	0	2,2,2	0.30	0
4	EDO	B	1007	-	3,3,3	0.45	0	2,2,2	0.51	0
4	EDO	B	1006	-	3,3,3	0.47	0	2,2,2	0.40	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	EDO	B	1004	-	3,3,3	0.38	0	2,2,2	0.59	0
3	FLC	B	1003	-	3,12,12	1.30	0	3,17,17	1.52	1 (33%)
4	EDO	A	1004	-	3,3,3	0.54	0	2,2,2	0.16	0
4	EDO	A	1005	-	3,3,3	0.45	0	2,2,2	0.38	0
4	EDO	B	1005	-	3,3,3	0.47	0	2,2,2	0.30	0
4	EDO	A	1006	-	3,3,3	0.52	0	2,2,2	0.23	0
3	FLC	A	1003	-	3,12,12	1.01	0	3,17,17	1.47	1 (33%)

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
4	EDO	B	1008	-	-	0/1/1/1	-
4	EDO	B	1007	-	-	0/1/1/1	-
4	EDO	B	1006	-	-	1/1/1/1	-
4	EDO	B	1004	-	-	1/1/1/1	-
3	FLC	B	1003	-	-	1/6/16/16	-
4	EDO	A	1004	-	-	0/1/1/1	-
4	EDO	A	1005	-	-	0/1/1/1	-
4	EDO	B	1005	-	-	0/1/1/1	-
4	EDO	A	1006	-	-	0/1/1/1	-
3	FLC	A	1003	-	-	3/6/16/16	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	1003	FLC	CB-CA-CAC	-2.50	110.98	114.98
3	A	1003	FLC	CB-CA-CAC	-2.31	111.29	114.98

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1003	FLC	CA-CB-CG-CGC
3	A	1003	FLC	CBC-CB-CG-CGC
3	A	1003	FLC	OHB-CB-CG-CGC
4	B	1004	EDO	O1-C1-C2-O2
3	B	1003	FLC	CAC-CA-CB-CG

There are no ring outliers.

5 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	1008	EDO	2	0
4	B	1006	EDO	1	0
4	B	1004	EDO	2	0
4	A	1004	EDO	1	1
3	A	1003	FLC	2	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	438/448 (97%)	0.58	35 (7%) 12 11	34, 66, 112, 171	0
1	B	438/448 (97%)	0.81	57 (13%) 3 3	33, 64, 128, 227	0
All	All	876/896 (97%)	0.69	92 (10%) 6 5	33, 65, 123, 227	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	295	ARG	12.3
1	B	302	CYS	9.1
1	B	721	ASP	7.5
1	B	297	PRO	7.3
1	B	296	SER	6.9

### 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
4	EDO	A	1006	4/4	0.73	0.20	71,72,75,75	0
2	ZN	B	1001	1/1	0.79	0.05	123,123,123,123	0
3	FLC	A	1003	13/13	0.79	0.24	75,82,90,91	0
4	EDO	A	1005	4/4	0.79	0.60	80,81,81,81	0
4	EDO	B	1008	4/4	0.80	0.26	69,72,74,77	0
4	EDO	B	1005	4/4	0.82	0.25	87,87,90,91	0
4	EDO	A	1004	4/4	0.85	0.34	54,58,58,59	0
4	EDO	B	1007	4/4	0.86	0.20	72,72,73,74	0
3	FLC	B	1003	13/13	0.87	0.24	72,78,94,95	0
4	EDO	B	1006	4/4	0.91	0.23	68,70,73,76	0
4	EDO	B	1004	4/4	0.95	0.27	52,55,61,61	0
2	ZN	A	1001	1/1	0.98	0.05	66,66,66,66	0
2	ZN	B	1002	1/1	1.00	0.19	41,41,41,41	0
2	ZN	A	1002	1/1	1.00	0.18	39,39,39,39	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.