



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 14, 2020 – 06:39 pm BST

PDB ID : 2Y6E  
Title : Structure of the D1D2 domain of USP4, the conserved catalytic domain  
Authors : Luna-Vargas, M.P.A.; Faesen, A.C.; van Dijk, W.J.; Rape, M.; Fish, A.; Sixma, T.K.  
Deposited on : 2011-01-20  
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

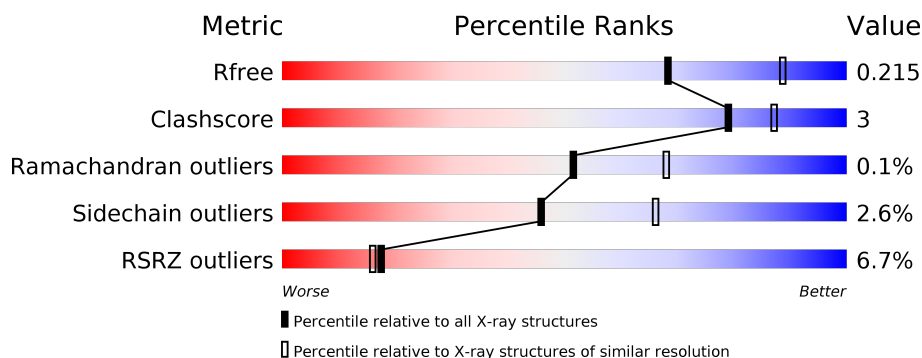
# 1 Overall quality at a glance

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	367	<div> <div>8%</div> <div> <div></div> <div>85%</div> <div>6% • 8%</div> </div> </div>
1	B	367	<div> <div>5%</div> <div> <div></div> <div>86%</div> <div>7% 8%</div> </div> </div>
1	C	367	<div> <div>4%</div> <div> <div></div> <div>85%</div> <div>5% 10%</div> </div> </div>
1	D	367	<div> <div>9%</div> <div> <div></div> <div>80%</div> <div>9% • 11%</div> </div> </div>
1	E	367	<div> <div>5%</div> <div> <div></div> <div>81%</div> <div>6% 13%</div> </div> </div>
1	F	367	<div> <div>5%</div> <div> <div></div> <div>81%</div> <div>6% • 12%</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
1	CME	D	311	-	-	X	-
1	CME	E	311	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 17209 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 CARBOXYL-TERMINAL HYDROLASE 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	337	Total	C	N	O	S	0	1	0
			2734	1750	464	502	18			
1	B	339	Total	C	N	O	S	0	0	0
			2738	1752	465	504	17			
1	C	331	Total	C	N	O	S	0	0	0
			2666	1709	453	487	17			
1	D	328	Total	C	N	O	S	0	0	0
			2646	1697	446	485	18			
1	E	319	Total	C	N	O	S	0	0	0
			2569	1649	432	473	15			
1	F	323	Total	C	N	O	S	0	0	0
			2591	1659	438	479	15			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	294	GLY	-	expression tag	UNP Q13107
A	295	MET	-	expression tag	UNP Q13107
A	491	GLY	-	linker	UNP Q13107
A	492	PRO	-	linker	UNP Q13107
B	294	GLY	-	expression tag	UNP Q13107
B	295	MET	-	expression tag	UNP Q13107
B	491	GLY	-	linker	UNP Q13107
B	492	PRO	-	linker	UNP Q13107
C	294	GLY	-	expression tag	UNP Q13107
C	295	MET	-	expression tag	UNP Q13107
C	491	GLY	-	linker	UNP Q13107
C	492	PRO	-	linker	UNP Q13107
D	294	GLY	-	expression tag	UNP Q13107
D	295	MET	-	expression tag	UNP Q13107
D	491	GLY	-	linker	UNP Q13107
D	492	PRO	-	linker	UNP Q13107
E	294	GLY	-	expression tag	UNP Q13107

*Continued on next page...*

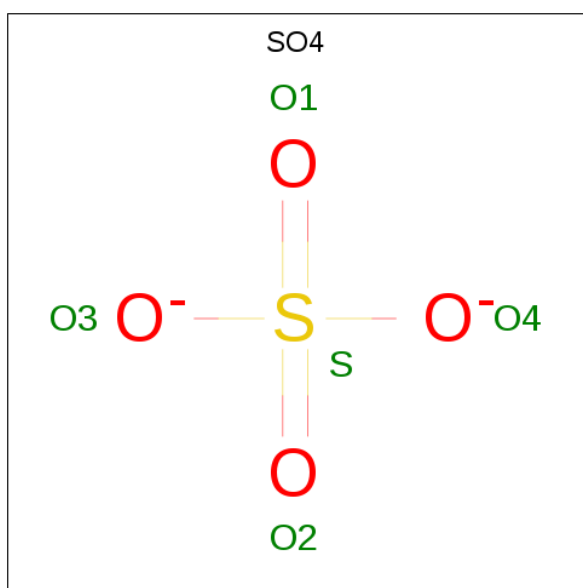
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	295	MET	-	expression tag	UNP Q13107
E	491	GLY	-	linker	UNP Q13107
E	492	PRO	-	linker	UNP Q13107
F	294	GLY	-	expression tag	UNP Q13107
F	295	MET	-	expression tag	UNP Q13107
F	491	GLY	-	linker	UNP Q13107
F	492	PRO	-	linker	UNP Q13107

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	D	1	Total Zn 1 1	0	0
2	E	1	Total Zn 1 1	0	0
2	B	1	Total Zn 1 1	0	0
2	C	1	Total Zn 1 1	0	0
2	A	1	Total Zn 1 1	0	0
2	F	1	Total Zn 1 1	0	0

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

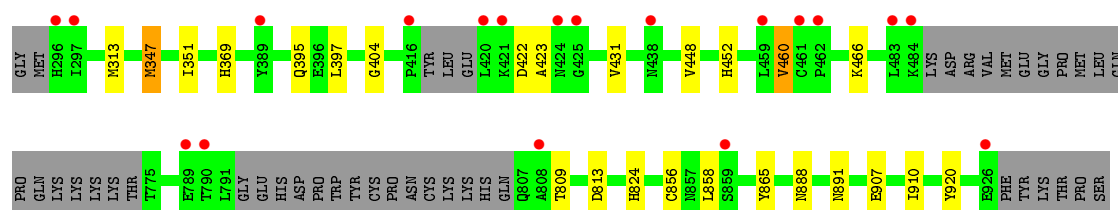


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	C	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	250	Total	O	0	0
			250	250		
4	B	232	Total	O	0	0
			232	232		
4	C	204	Total	O	0	0
			204	204		
4	D	214	Total	O	0	0
			214	214		
4	E	166	Total	O	0	0
			166	166		
4	F	183	Total	O	0	0
			183	183		







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	110.50Å 151.03Å 178.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.59 – 2.40 44.59 – 2.40	Depositor EDS
% Data completeness (in resolution range)	(Not available) (44.59-2.40) 94.8 (44.59-2.40)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.69 (at 2.39Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, $R_{free}$	0.178 , 0.210 0.183 , 0.215	Depositor DCC
$R_{free}$ test set	5583 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.2	Xtriage
Anisotropy	0.108	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 64.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	17209	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.23% 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, CME, 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	0.50	0/2791	0.64	0/3780
1	B	0.50	0/2798	0.63	0/3793
1	C	0.48	0/2723	0.62	0/3691
1	D	0.50	0/2700	0.65	0/3656
1	E	0.47	0/2620	0.63	0/3547
1	F	0.49	0/2641	0.64	0/3577
All	All	0.49	0/16273	0.63	0/22044

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 [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	2734	0	2668	13	0
1	B	2738	0	2646	15	0
1	C	2666	0	2588	12	0
1	D	2646	0	2565	21	0
1	E	2569	0	2501	17	0
1	F	2591	0	2525	13	0
2	A	1	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
2	F	1	0	0	0	0
3	A	5	0	0	0	0
3	C	5	0	0	0	0
4	A	250	0	0	0	0
4	B	232	0	0	1	0
4	C	204	0	0	1	0
4	D	214	0	0	1	0
4	E	166	0	0	1	0
4	F	183	0	0	0	0
All	All	17209	0	15493	87	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 3.

The worst 5 of 87 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:D:311:CME:SD	1:D:311:CME:SG	2.32	1.28
1:B:311:CME:SG	1:B:311:CME:SD	2.42	1.16
1:C:311:CME:HZ3	1:C:392:GLN:HB2	1.30	1.11
1:C:309:ASN:OD1	1:C:311:CME:HE2	1.75	0.85
1:C:311:CME:CZ	1:C:392:GLN:HB2	2.09	0.82

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	331/367 (90%)	324 (98%)	7 (2%)	0	100	100
1	B	332/367 (90%)	322 (97%)	10 (3%)	0	100	100
1	C	322/367 (88%)	316 (98%)	6 (2%)	0	100	100
1	D	315/367 (86%)	307 (98%)	7 (2%)	1 (0%)	41	55
1	E	310/367 (84%)	303 (98%)	6 (2%)	1 (0%)	41	55
1	F	314/367 (86%)	309 (98%)	5 (2%)	0	100	100
All	All	1924/2202 (87%)	1881 (98%)	41 (2%)	2 (0%)	51	68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	925	ASP
1	D	826	LYS

### 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	294/321 (92%)	284 (97%)	10 (3%)	37	56
1	B	293/321 (91%)	287 (98%)	6 (2%)	55	74
1	C	285/321 (89%)	278 (98%)	7 (2%)	47	67
1	D	283/321 (88%)	272 (96%)	11 (4%)	32	50
1	E	274/321 (85%)	270 (98%)	4 (2%)	65	80
1	F	277/321 (86%)	271 (98%)	6 (2%)	52	71
All	All	1706/1926 (89%)	1662 (97%)	44 (3%)	46	66

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	805	HIS
1	D	358	LEU
1	F	422	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	806	GLN
1	C	858	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	805	HIS
1	D	391	GLN
1	F	391	GLN
1	C	296	HIS
1	F	296	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

6 non-standard protein/DNA/RNA residues 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	$\# Z  > 2$
1	CME	E	311	1	8,9,10	1.21	1 (12%)	5,9,11	2.20	1 (20%)
1	CME	C	311	1	8,9,10	1.24	1 (12%)	5,9,11	1.88	2 (40%)
1	CME	A	311	1	8,9,10	1.25	1 (12%)	5,9,11	2.56	2 (40%)
1	CME	B	311	1	8,9,10	1.90	1 (12%)	5,9,11	3.94	3 (60%)
1	CME	D	311	1	8,9,10	1.50	1 (12%)	5,9,11	2.22	2 (40%)
1	CME	F	311	1	8,9,10	0.88	0	5,9,11	1.67	1 (20%)

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
1	CME	E	311	1	-	3/5/8/10	-
1	CME	C	311	1	-	4/5/8/10	-
1	CME	A	311	1	-	2/5/8/10	-
1	CME	B	311	1	-	3/5/8/10	-
1	CME	D	311	1	-	3/5/8/10	-
1	CME	F	311	1	-	4/5/8/10	-

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	311	CME	SD-SG	5.11	2.42	2.03
1	D	311	CME	SD-SG	3.79	2.32	2.03
1	A	311	CME	SD-SG	3.10	2.27	2.03
1	C	311	CME	SD-SG	2.98	2.26	2.03
1	E	311	CME	SD-SG	2.89	2.25	2.03

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	311	CME	CE-SD-SG	6.95	135.46	103.45
1	A	311	CME	CE-SD-SG	4.88	125.91	103.45
1	B	311	CME	CB-SG-SD	4.81	116.30	103.82
1	E	311	CME	CB-SG-SD	4.64	115.85	103.82
1	D	311	CME	CB-SG-SD	3.93	114.00	103.82

There are no chirality outliers.

5 of 19 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	E	311	CME	CA-CB-SG-SD
1	A	311	CME	CE-SD-SG-CB
1	B	311	CME	CZ-CE-SD-SG
1	D	311	CME	CA-CB-SG-SD
1	B	311	CME	CE-SD-SG-CB

There are no ring outliers.

5 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	E	311	CME	6	0
1	C	311	CME	5	0
1	A	311	CME	1	0
1	B	311	CME	2	0
1	D	311	CME	7	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 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	C	1925	-	4,4,4	0.37	0	6,6,6	0.21	0
3	SO4	A	1926	-	4,4,4	0.41	0	6,6,6	0.22	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 ⓘ

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	336/367 (91%)	-0.08	28 (8%) 11 10	23, 39, 116, 140	0
1	B	338/367 (92%)	0.01	17 (5%) 28 27	23, 40, 103, 125	0
1	C	330/367 (89%)	-0.21	16 (4%) 30 29	24, 45, 103, 152	0
1	D	327/367 (89%)	0.09	34 (10%) 6 6	26, 45, 119, 265	0
1	E	318/367 (86%)	-0.16	19 (5%) 21 20	29, 46, 96, 139	0
1	F	322/367 (87%)	-0.22	19 (5%) 22 21	26, 49, 94, 145	0
All	All	1971/2202 (89%)	-0.09	133 (6%) 17 16	23, 45, 103, 265	0

The worst 5 of 133 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	417	TYR	11.3
1	C	798	TYR	10.4
1	A	422	ASP	8.7
1	A	800	PRO	8.1
1	D	800	PRO	8.0

### 6.2 Non-standard residues in protein, DNA, RNA chains [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
1	CME	C	311	10/11	0.88	0.18	38,41,58,59	0
1	CME	D	311	10/11	0.88	0.19	31,37,53,53	0
1	CME	B	311	10/11	0.90	0.21	23,28,44,44	0
1	CME	A	311	10/11	0.90	0.18	27,32,45,47	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
1	CME	F	311	10/11	0.90	0.17	35,43,66,68	0
1	CME	E	311	10/11	0.92	0.18	32,37,56,58	0

### 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	SO4	C	1925	5/5	0.79	0.23	118,121,122,125	0
2	ZN	F	1000	1/1	0.86	0.05	99,99,99,99	0
2	ZN	D	1000	1/1	0.87	0.24	162,162,162,162	0
2	ZN	C	1000	1/1	0.89	0.04	103,103,103,103	0
2	ZN	A	1000	1/1	0.90	0.15	159,159,159,159	0
3	SO4	A	1926	5/5	0.91	0.18	100,104,105,105	0
2	ZN	E	1000	1/1	0.92	0.08	123,123,123,123	0
2	ZN	B	1000	1/1	0.99	0.04	88,88,88,88	0

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

There are no such residues in this entry.