



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 01:48 am BST

PDB ID : 4FJC  
Title : Structure of the SAGA Ubp8/Sgf11(1-72, Delta-ZnF)/Sus1/Sgf73 DUB module  
Authors : Samara, N.L.; Ringel, A.E.; Wolberger, C.  
Deposited on : 2012-06-11  
Resolution : 2.83 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

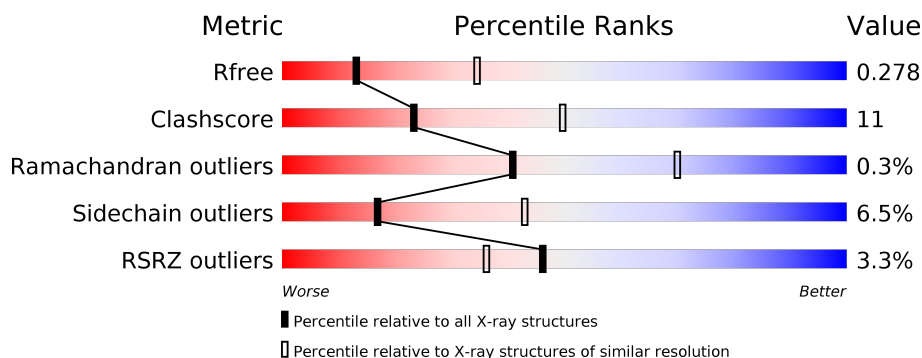
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.83 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	3617 (2.84-2.80)
Clashscore	141614	4060 (2.84-2.80)
Ramachandran outliers	138981	3978 (2.84-2.80)
Sidechain outliers	138945	3980 (2.84-2.80)
RSRZ outliers	127900	3552 (2.84-2.80)

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 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	476	<div> <div>5%</div> <div> <div></div> <div>70%</div> <div>22%</div> <div>6%</div> </div> </div>
1	E	476	<div> <div>%</div> <div> <div></div> <div>67%</div> <div>21%</div> <div>9%</div> </div> </div>
2	B	96	<div> <div></div> <div> <div></div> <div>76%</div> <div>15%</div> <div>5%</div> </div> </div>
2	F	96	<div> <div>5%</div> <div> <div></div> <div>65%</div> <div>26%</div> <div>9%</div> </div> </div>
3	C	99	<div> <div>%</div> <div> <div></div> <div>33%</div> <div>8%</div> <div>57%</div> </div> </div>
3	G	99	<div> <div>3%</div> <div> <div></div> <div>30%</div> <div>10%</div> <div>59%</div> </div> </div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	96	
4	H	96	

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
5	ZN	E	502	-	-	X	-

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 10581 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 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	447	Total	C	N	O	S	0	0	0
			3575	2267	613	659	36			
1	E	434	Total	C	N	O	S	0	0	0
			3483	2214	593	641	35			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	EXPRESSION TAG	UNP P50102
A	-3	ALA	-	EXPRESSION TAG	UNP P50102
A	-2	ALA	-	EXPRESSION TAG	UNP P50102
A	-1	ALA	-	EXPRESSION TAG	UNP P50102
A	0	ALA	-	EXPRESSION TAG	UNP P50102
E	-4	GLY	-	EXPRESSION TAG	UNP P50102
E	-3	ALA	-	EXPRESSION TAG	UNP P50102
E	-2	ALA	-	EXPRESSION TAG	UNP P50102
E	-1	ALA	-	EXPRESSION TAG	UNP P50102
E	0	ALA	-	EXPRESSION TAG	UNP P50102

- Molecule 2 is a protein called Protein SUS1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	92	Total	C	N	O	S	0	0	0
			743	464	122	155	2			
2	F	87	Total	C	N	O	S	0	0	0
			707	444	116	145	2			

- Molecule 3 is a protein called SAGA-associated factor 11.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	43	Total	C	N	O	0	0	0
			342	213	57	72			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	G	41	Total	C	N	O	0	0	0
			324	203	55	66			

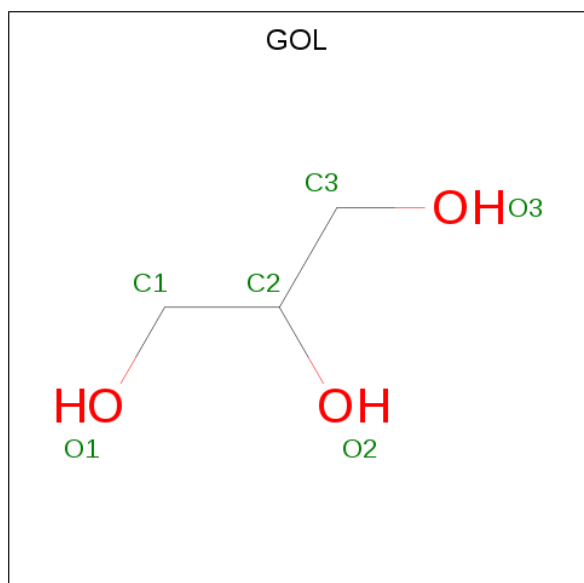
- Molecule 4 is a protein called SAGA-associated factor 73.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	90	Total	C	N	O	S	0	0	0
			717	452	120	141	4			
4	H	77	Total	C	N	O	S	0	0	0
			624	399	104	117	4			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	1	Total	Zn	0	0
			1	1		
5	A	6	Total	Zn	0	0
			6	6		
5	D	1	Total	Zn	0	0
			1	1		
5	E	6	Total	Zn	0	0
			6	6		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

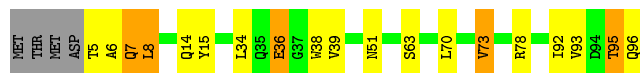
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	16	Total	O	0	0
			16	16		
7	C	1	Total	O	0	0
			1	1		
7	D	5	Total	O	0	0
			5	5		
7	E	20	Total	O	0	0
			20	20		
7	F	2	Total	O	0	0
			2	2		
7	H	2	Total	O	0	0
			2	2		



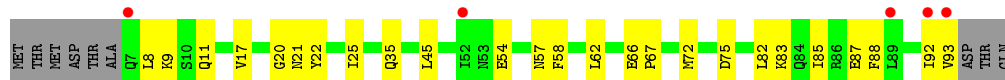
- Molecule 2: Protein SUS1

Chain B: 




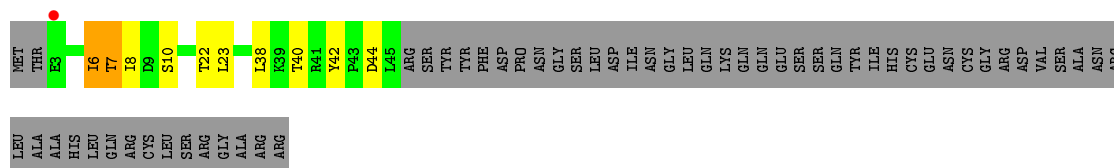
- Molecule 2: Protein SUS1

Chain F: 



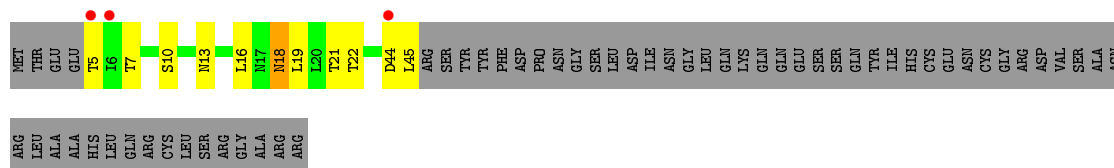
- Molecule 3: SAGA-associated factor 11

Chain C: 




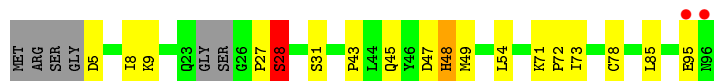
- Molecule 3: SAGA-associated factor 11

Chain G: 



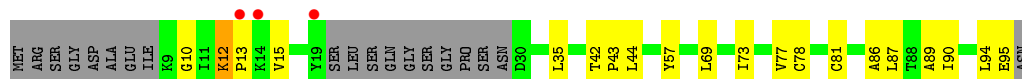
- Molecule 4: SAGA-associated factor 73

Chain D: 



- Molecule 4: SAGA-associated factor 73

Chain H: 





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	76.49 Å 79.20 Å 265.94 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	265.94 – 2.83 44.16 – 2.83	Depositor EDS
% Data completeness (in resolution range)	96.2 (265.94-2.83) 96.3 (44.16-2.83)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 2.81 Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.200 , 0.276 0.202 , 0.278	Depositor DCC
$R_{free}$ test set	1917 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	49.0	Xtriage
Anisotropy	0.387	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 50.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.021 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	10581	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.97% 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: GOL, ZN

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	0.65	0/3652	0.72	1/4918 (0.0%)
1	E	0.63	1/3559 (0.0%)	0.73	0/4793
2	B	0.72	2/749 (0.3%)	0.76	0/1010
2	F	0.50	0/713	0.66	0/960
3	C	0.58	0/343	0.70	0/467
3	G	0.44	0/325	0.67	0/443
4	D	0.69	0/731	0.75	0/986
4	H	0.56	0/637	0.66	0/858
All	All	0.63	3/10709 (0.0%)	0.72	1/14435 (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	1
1	E	0	2
4	D	0	1
All	All	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	36	GLU	CG-CD	5.15	1.59	1.51
1	E	69	TRP	CD2-CE2	5.10	1.47	1.41
2	B	38	TRP	CD2-CE2	5.08	1.47	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	132	ARG	NE-CZ-NH1	5.35	122.97	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	232	GLY	Peptide
4	D	28	SER	Peptide
1	E	227	ASN	Peptide
1	E	366	GLU	Peptide

## 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	3575	0	3491	92	0
1	E	3483	0	3394	102	0
2	B	743	0	749	16	0
2	F	707	0	718	20	0
3	C	342	0	353	9	0
3	G	324	0	341	8	0
4	D	717	0	710	13	0
4	H	624	0	628	21	0
5	A	6	0	0	1	0
5	D	1	0	0	0	0
5	E	6	0	0	2	0
5	H	1	0	0	0	0
6	A	6	0	8	0	0
7	A	16	0	0	1	0
7	C	1	0	0	0	0
7	D	5	0	0	0	0
7	E	20	0	0	1	0
7	F	2	0	0	0	0
7	H	2	0	0	1	0
All	All	10581	0	10392	237	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 11.

The worst 5 of 237 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:246:ILE:HD12	1:A:304:LEU:HD13	1.23	1.16
1:A:156:ILE:HD13	1:A:194:LEU:HD13	1.26	1.14
1:A:246:ILE:CD1	1:A:304:LEU:HD13	1.84	1.06
1:A:292:CYS:HG	5:A:506:ZN:ZN	0.69	0.97
1:E:416:ILE:HD13	1:E:463:LEU:HD23	1.49	0.94

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	437/476 (92%)	407 (93%)	30 (7%)	0	100	100
1	E	422/476 (89%)	391 (93%)	29 (7%)	2 (0%)	29	59
2	B	90/96 (94%)	87 (97%)	3 (3%)	0	100	100
2	F	85/96 (88%)	81 (95%)	3 (4%)	1 (1%)	13	37
3	C	41/99 (41%)	41 (100%)	0	0	100	100
3	G	39/99 (39%)	33 (85%)	6 (15%)	0	100	100
4	D	86/96 (90%)	81 (94%)	5 (6%)	0	100	100
4	H	73/96 (76%)	65 (89%)	7 (10%)	1 (1%)	11	32
All	All	1273/1534 (83%)	1186 (93%)	83 (6%)	4 (0%)	41	70

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	236	GLN
1	E	230	LEU
2	F	54	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
4	H	87	LEU

### 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	407/429 (95%)	377 (93%)	30 (7%)	13	36
1	E	397/429 (92%)	375 (94%)	22 (6%)	21	50
2	B	87/91 (96%)	80 (92%)	7 (8%)	12	32
2	F	83/91 (91%)	81 (98%)	2 (2%)	49	80
3	C	41/89 (46%)	36 (88%)	5 (12%)	5	14
3	G	39/89 (44%)	35 (90%)	4 (10%)	7	21
4	D	82/86 (95%)	77 (94%)	5 (6%)	18	46
4	H	71/86 (83%)	67 (94%)	4 (6%)	21	49
All	All	1207/1390 (87%)	1128 (94%)	79 (6%)	17	43

5 of 79 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	73	VAL
4	D	31	SER
3	G	44	ASP
2	B	95	THR
3	C	38	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	42	ASN
1	E	110	ASN
2	F	35	GLN
1	E	53	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	70	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 [i](#)

Of 15 ligands modelled in this entry, 14 are monoatomic - leaving 1 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	$\# Z  > 2$
6	GOL	A	507	-	5,5,5	0.51	0	5,5,5	0.64	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
6	GOL	A	507	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	507	GOL	C1-C2-C3-O3
6	A	507	GOL	O2-C2-C3-O3

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 ⓘ

### 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	447/476 (93%)	-0.06	22 (4%) 29 20	22, 41, 97, 119	0
1	E	434/476 (91%)	-0.21	7 (1%) 72 65	27, 44, 82, 110	0
2	B	92/96 (95%)	-0.32	0 100 100	25, 35, 64, 82	0
2	F	87/96 (90%)	0.06	5 (5%) 23 15	36, 60, 98, 114	0
3	C	43/99 (43%)	-0.15	1 (2%) 60 50	26, 30, 70, 103	0
3	G	41/99 (41%)	0.30	3 (7%) 15 8	40, 57, 99, 103	0
4	D	90/96 (93%)	-0.16	2 (2%) 62 52	24, 36, 71, 78	0
4	H	77/96 (80%)	-0.01	3 (3%) 39 29	37, 56, 107, 131	0
All	All	1311/1534 (85%)	-0.11	43 (3%) 46 36	22, 44, 94, 131	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	G	5	THR	8.6
1	A	210	THR	6.3
1	E	211	ASN	4.6
1	A	341	SER	4.3
4	H	19	TYR	4.3

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

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(Å <sup>2</sup> )	Q<0.9
6	GOL	A	507	6/6	0.87	0.24	48,51,56,57	0
5	ZN	A	505	1/1	0.98	0.10	51,51,51,51	0
5	ZN	A	504	1/1	0.98	0.09	54,54,54,54	0
5	ZN	H	101	1/1	0.98	0.05	77,77,77,77	0
5	ZN	A	501	1/1	0.99	0.10	40,40,40,40	0
5	ZN	A	506	1/1	0.99	0.04	87,87,87,87	0
5	ZN	E	504	1/1	0.99	0.09	49,49,49,49	0
5	ZN	E	503	1/1	0.99	0.06	41,41,41,41	0
5	ZN	E	505	1/1	0.99	0.12	40,40,40,40	0
5	ZN	A	503	1/1	0.99	0.11	35,35,35,35	0
5	ZN	E	501	1/1	0.99	0.06	54,54,54,54	0
5	ZN	E	506	1/1	1.00	0.07	37,37,37,37	0
5	ZN	E	502	1/1	1.00	0.09	45,45,45,45	0
5	ZN	D	101	1/1	1.00	0.09	36,36,36,36	0
5	ZN	A	502	1/1	1.00	0.11	28,28,28,28	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.