



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 20, 2023 – 09:47 PM JST

PDB ID : 7DLV  
Title : shrimp dUTPase in complex with StI  
Authors : Ma, Q.; Wang, F.  
Deposited on : 2020-11-30  
Resolution : 2.52 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

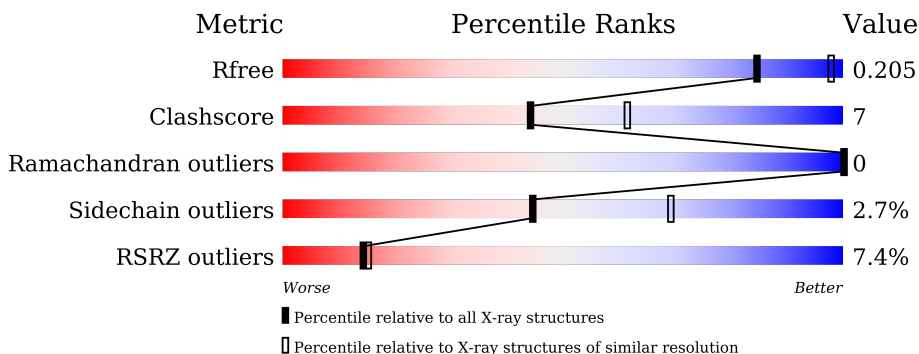
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

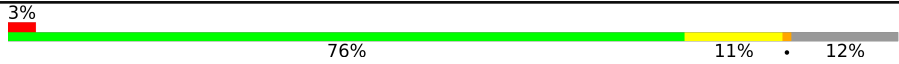


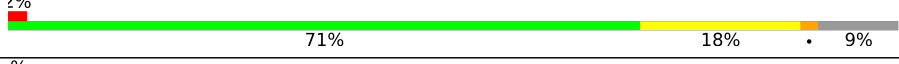
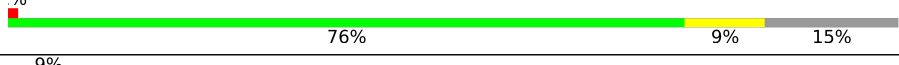

The reported resolution of this entry is 2.52 Å.

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	5743 (2.54-2.50)
Clashscore	141614	6463 (2.54-2.50)
Ramachandran outliers	138981	6335 (2.54-2.50)
Sidechain outliers	138945	6337 (2.54-2.50)
RSRZ outliers	127900	5630 (2.54-2.50)

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 of 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	149	 3% 76% 11% • 12%
1	B	149	 5% 77% 14% • 8%
1	C	149	 3% 74% 17% 9%
1	G	149	 2% 71% 18% • 9%
1	H	149	 % 76% 9% 15%
1	I	149	 9% 69% 21% • 9%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	D	157	<p>4% 78% 13% 8%</p>
2	E	157	<p>18% 70% 18% 10%</p>
2	F	157	<p>4% 76% 15% 9%</p>
2	J	157	<p>4% 82% 11% 7%</p>
2	K	157	<p>4% 78% 13% 8%</p>
2	L	157	<p>22% 71% 18% 9%</p>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 13439 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 shrimp dUTPase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	131	1029	657	178	190	4	0	0	0
1	B	137	1065	679	184	198	4	0	0	0
1	C	136	1061	677	183	197	4	0	0	0
1	G	136	1061	677	183	197	4	0	0	0
1	H	127	993	637	171	181	4	0	0	0
1	I	136	1061	677	183	197	4	0	0	0

- Molecule 2 is a protein called Orf20.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	144	1191	766	191	232	2	0	0	0
2	E	141	1169	751	188	228	2	0	0	0
2	F	143	1183	760	190	231	2	0	0	0
2	J	146	1205	774	193	236	2	0	0	0
2	K	144	1191	766	191	232	2	0	0	0
2	L	143	1183	760	190	231	2	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

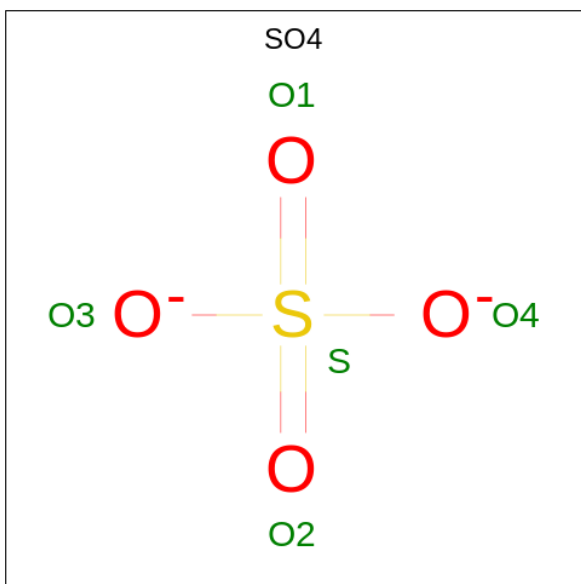
Chain	Residue	Modelled	Actual	Comment	Reference
D	-1	GLY	-	expression tag	UNP Q9F0J8

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	ALA	-	expression tag	UNP Q9F0J8
E	-1	GLY	-	expression tag	UNP Q9F0J8
E	0	ALA	-	expression tag	UNP Q9F0J8
F	-1	GLY	-	expression tag	UNP Q9F0J8
F	0	ALA	-	expression tag	UNP Q9F0J8
J	-1	GLY	-	expression tag	UNP Q9F0J8
J	0	ALA	-	expression tag	UNP Q9F0J8
K	-1	GLY	-	expression tag	UNP Q9F0J8
K	0	ALA	-	expression tag	UNP Q9F0J8
L	-1	GLY	-	expression tag	UNP Q9F0J8
L	0	ALA	-	expression tag	UNP Q9F0J8

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



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

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	H	1	Total O S 5 4 1	0	0
3	J	1	Total O S 5 4 1	0	0
3	K	1	Total O S 5 4 1	0	0


- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Ca 1 1	0	0
4	H	1	Total Ca 1 1	0	0

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


These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: shrimp dUTPase

Chain A: 




- Molecule 1: shrimp dUTPase

Chain B: 



ASN  
GLN

- Molecule 1: shrimp dUTPase

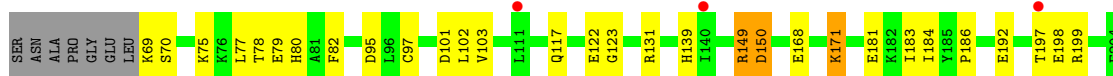
Chain C: 



E201  
F204  
GLY  
THR  
GLY  
ASN  
GLN

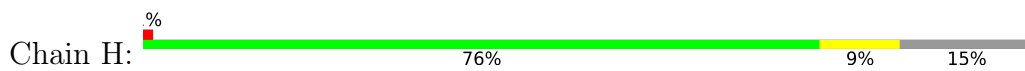
- Molecule 1: shrimp dUTPase

Chain G: 

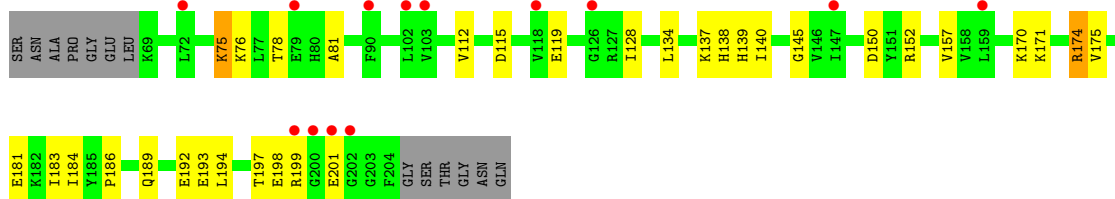


GLY  
SER  
THR  
GLY  
GLY  
ASN  
GLN

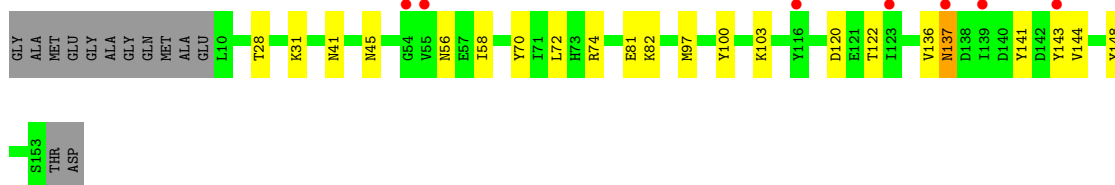
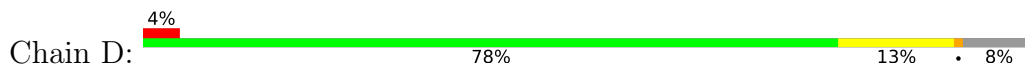
- Molecule 1: shrimp dUTPase



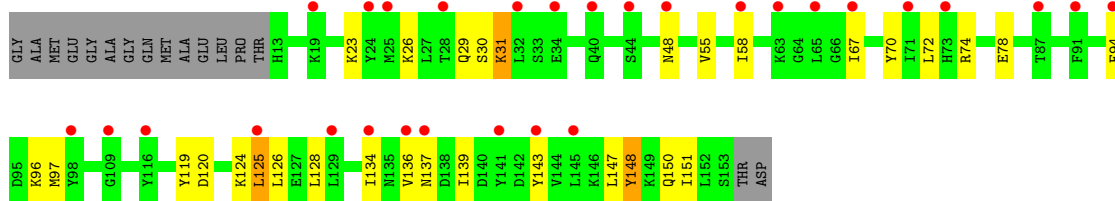
• Molecule 1: shrimp dUTPase



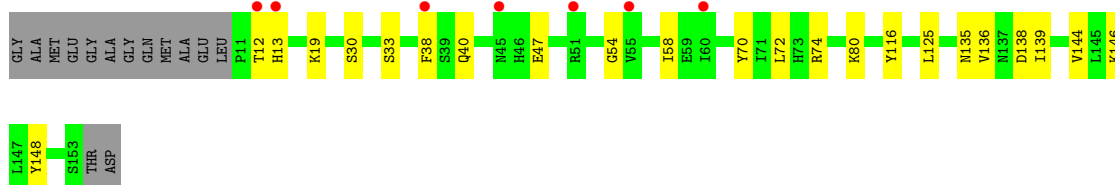
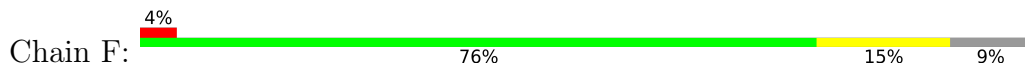
• Molecule 2: Orf20



• Molecule 2: Orf20

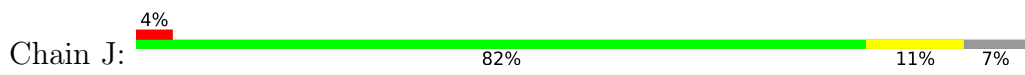


• Molecule 2: Orf20

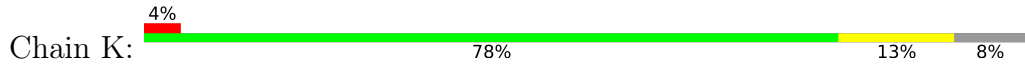


• Molecule 2: Orf20

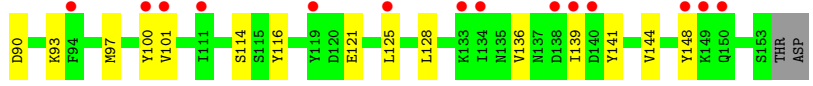
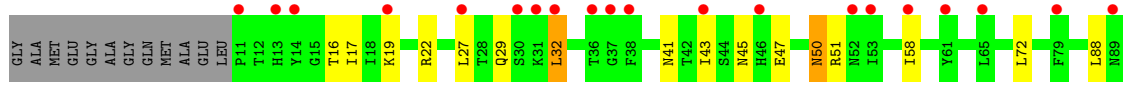




• Molecule 2: Orf20



• Molecule 2: Orf20



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.70Å 135.88Å 135.63Å 90.00° 105.79° 90.00°	Depositor
Resolution (Å)	94.13 – 2.52 94.13 – 2.52	Depositor EDS
% Data completeness (in resolution range)	98.8 (94.13-2.52) 97.9 (94.13-2.52)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.10 (at 2.51Å)	Xtrriage
Refinement program	PHENIX 1.13_2998	Depositor
R, $R_{free}$	0.172 , 0.201 0.175 , 0.205	Depositor DCC
$R_{free}$ test set	4411 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	58.7	Xtrriage
Anisotropy	0.398	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 39.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	0.449 for h,-k,-h-l	Xtrriage
Reported twinning fraction	0.480 for -h,-k,h+l	Depositor
Outliers	0 of 86385 reflections	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	13439	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% 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: CA, 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	0.28	0/1049	0.48	0/1414
1	B	0.26	0/1086	0.44	0/1462
1	C	0.27	0/1082	0.47	0/1457
1	G	0.27	0/1082	0.44	0/1457
1	H	0.25	0/1013	0.44	0/1366
1	I	0.28	0/1082	0.47	0/1457
2	D	0.25	0/1216	0.42	0/1638
2	E	0.27	0/1193	0.49	0/1605
2	F	0.25	0/1208	0.41	0/1626
2	J	0.25	0/1230	0.40	0/1657
2	K	0.27	0/1216	0.43	0/1638
2	L	0.29	0/1208	0.47	0/1626
All	All	0.27	0/13665	0.45	0/18403

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	1029	0	1041	13	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	1065	0	1068	16	0
1	C	1061	0	1065	18	0
1	G	1061	0	1065	25	0
1	H	993	0	1009	9	0
1	I	1061	0	1065	30	0
2	D	1191	0	1176	12	0
2	E	1169	0	1151	21	0
2	F	1183	0	1166	14	0
2	J	1205	0	1187	10	0
2	K	1191	0	1176	16	0
2	L	1183	0	1166	22	0
3	A	5	0	0	0	0
3	B	5	0	0	0	0
3	C	5	0	0	1	0
3	F	5	0	0	0	0
3	G	5	0	0	0	0
3	H	10	0	0	0	0
3	J	5	0	0	0	0
3	K	5	0	0	0	0
4	A	1	0	0	0	0
4	H	1	0	0	0	0
All	All	13439	0	13335	176	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 7.

The worst 5 of 176 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:G:192:GLU:C	1:I:75:LYS:HZ1	1.72	0.93
1:G:199:ARG:NH1	2:K:102:ASN:OD1	2.11	0.84
1:G:192:GLU:O	1:I:75:LYS:NZ	2.10	0.83
2:L:121:GLU:O	2:L:125:LEU:HD13	1.79	0.82
1:I:137:LYS:HD3	2:L:114:SER:HA	1.62	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	129/149 (87%)	125 (97%)	4 (3%)	0	100	100
1	B	135/149 (91%)	129 (96%)	6 (4%)	0	100	100
1	C	134/149 (90%)	128 (96%)	6 (4%)	0	100	100
1	G	134/149 (90%)	130 (97%)	4 (3%)	0	100	100
1	H	125/149 (84%)	121 (97%)	4 (3%)	0	100	100
1	I	134/149 (90%)	129 (96%)	5 (4%)	0	100	100
2	D	142/157 (90%)	137 (96%)	5 (4%)	0	100	100
2	E	139/157 (88%)	134 (96%)	5 (4%)	0	100	100
2	F	141/157 (90%)	135 (96%)	6 (4%)	0	100	100
2	J	144/157 (92%)	141 (98%)	3 (2%)	0	100	100
2	K	142/157 (90%)	138 (97%)	4 (3%)	0	100	100
2	L	141/157 (90%)	138 (98%)	3 (2%)	0	100	100
All	All	1640/1836 (89%)	1585 (97%)	55 (3%)	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	110/121 (91%)	107 (97%)	3 (3%)	44	69
1	B	112/121 (93%)	111 (99%)	1 (1%)	78	91

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	112/121 (93%)	111 (99%)	1 (1%)	78	91
1	G	112/121 (93%)	106 (95%)	6 (5%)	22	40
1	H	106/121 (88%)	103 (97%)	3 (3%)	43	68
1	I	112/121 (93%)	107 (96%)	5 (4%)	27	48
2	D	133/140 (95%)	130 (98%)	3 (2%)	50	74
2	E	130/140 (93%)	122 (94%)	8 (6%)	18	33
2	F	132/140 (94%)	129 (98%)	3 (2%)	50	74
2	J	134/140 (96%)	133 (99%)	1 (1%)	84	93
2	K	133/140 (95%)	132 (99%)	1 (1%)	81	92
2	L	132/140 (94%)	128 (97%)	4 (3%)	41	66
All	All	1458/1566 (93%)	1419 (97%)	39 (3%)	44	69

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	I	75	LYS
2	L	32	LEU
1	I	139	HIS
1	I	189	GLN
2	L	51	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	D	137	ASN
2	F	40	GLN
2	K	48	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 11 ligands modelled in this entry, 2 are monoatomic - leaving 9 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	301	-	4,4,4	0.14	0	6,6,6	0.07	0
3	SO4	F	201	-	4,4,4	0.15	0	6,6,6	0.05	0
3	SO4	G	301	-	4,4,4	0.14	0	6,6,6	0.05	0
3	SO4	K	201	-	4,4,4	0.14	0	6,6,6	0.06	0
3	SO4	H	302	-	4,4,4	0.14	0	6,6,6	0.06	0
3	SO4	J	201	-	4,4,4	0.16	0	6,6,6	0.07	0
3	SO4	B	301	-	4,4,4	0.14	0	6,6,6	0.05	0
3	SO4	H	301	-	4,4,4	0.14	0	6,6,6	0.06	0
3	SO4	A	301	-	4,4,4	0.14	0	6,6,6	0.06	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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	301	SO4	1	0

## 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	131/149 (87%)	0.64	4 (3%) 49 53	35, 52, 81, 124	0
1	B	137/149 (91%)	0.80	7 (5%) 28 30	43, 59, 84, 127	0
1	C	136/149 (91%)	0.76	5 (3%) 41 45	37, 57, 90, 125	0
1	G	136/149 (91%)	0.68	3 (2%) 62 65	38, 59, 91, 116	0
1	H	127/149 (85%)	0.62	1 (0%) 86 88	35, 50, 71, 92	0
1	I	136/149 (91%)	0.82	13 (9%) 8 8	36, 57, 85, 117	0
2	D	144/157 (91%)	0.75	7 (4%) 29 32	44, 62, 94, 114	0
2	E	141/157 (89%)	1.18	29 (20%) 1 0	53, 86, 115, 150	0
2	F	143/157 (91%)	0.62	7 (4%) 29 32	39, 51, 79, 101	0
2	J	146/157 (92%)	0.60	6 (4%) 37 41	39, 57, 78, 102	0
2	K	144/157 (91%)	0.79	7 (4%) 29 32	40, 60, 98, 117	0
2	L	143/157 (91%)	1.32	34 (23%) 0 0	58, 82, 120, 155	0
All	All	1664/1836 (90%)	0.80	123 (7%) 14 15	35, 60, 102, 155	0

The worst 5 of 123 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	L	30	SER	6.0
2	L	31	LYS	5.1
2	L	133	LYS	4.8
2	L	36	THR	4.6
1	I	199	ARG	4.5

### 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 monosaccharides 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
3	SO4	H	302	5/5	0.89	0.16	93,101,101,103	0
4	CA	A	302	1/1	0.90	0.14	75,75,75,75	0
3	SO4	K	201	5/5	0.95	0.19	65,71,73,78	0
3	SO4	B	301	5/5	0.96	0.12	68,74,77,77	0
3	SO4	J	201	5/5	0.96	0.12	57,60,63,69	0
3	SO4	C	301	5/5	0.96	0.12	78,85,87,93	0
3	SO4	H	301	5/5	0.96	0.18	74,79,82,83	0
3	SO4	F	201	5/5	0.97	0.13	66,68,70,72	0
3	SO4	A	301	5/5	0.97	0.17	50,56,58,59	0
3	SO4	G	301	5/5	0.98	0.11	63,65,69,71	0
4	CA	H	303	1/1	0.99	0.13	67,67,67,67	0

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

There are no such residues in this entry.