



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 12:37 am BST

PDB ID : 2IXP  
Title : Crystal structure of the Pp2A phosphatase activator Ypa1 PTPA1 in complex with model substrate  
Authors : Leulliot, N.; Vicentini, G.; Jordens, J.; Quevillon-Cheruel, S.; Schiltz, M.; Barford, D.; Van Tilbeurgh, H.; Goris, J.  
Deposited on : 2006-07-09  
Resolution : 2.80 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

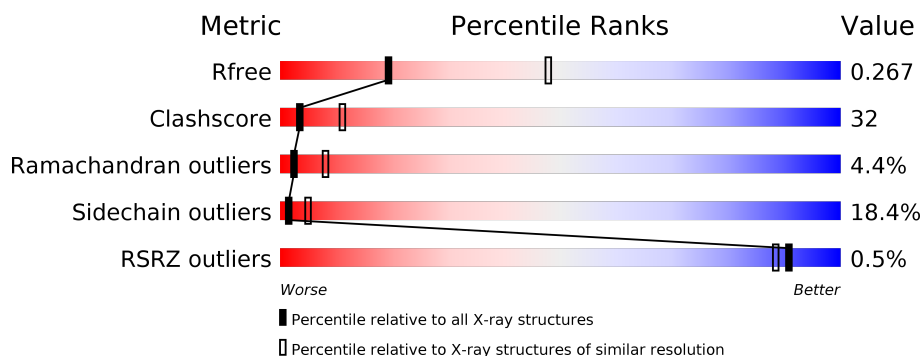
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	323	<div> <div>%</div> <div> <div></div> <div>41%</div> <div>42%</div> <div>12%</div> <div>• •</div> </div> </div>
1	B	323	<div> <div>44%</div> <div>40%</div> <div>12%</div> <div>• •</div> </div>
1	C	323	<div> <div>40%</div> <div>44%</div> <div>12%</div> <div>• •</div> </div>
1	D	323	<div> <div>%</div> <div>42%</div> <div>43%</div> <div>9%</div> <div>• •</div> </div>
2	F	6	<div> <div>100%</div> </div>
2	G	6	<div> <div>50%</div> <div>50%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	H	6	
2	I	6	

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	A	1320	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10394 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 SERINE/THREONINE-PROTEIN PHOSPHATASE 2A ACTIVATOR 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	310	Total	C	N	O	S	0	0	0
			2524	1642	421	447	14			
1	B	316	Total	C	N	O	S	0	0	0
			2576	1675	433	454	14			
1	C	316	Total	C	N	O	S	0	0	0
			2576	1675	433	454	14			
1	D	310	Total	C	N	O	S	0	0	0
			2524	1642	421	447	14			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	318	HIS	-	expression tag	UNP P40454
A	319	HIS	-	expression tag	UNP P40454
A	320	HIS	-	expression tag	UNP P40454
A	321	HIS	-	expression tag	UNP P40454
A	322	HIS	-	expression tag	UNP P40454
A	323	HIS	-	expression tag	UNP P40454
B	318	HIS	-	expression tag	UNP P40454
B	319	HIS	-	expression tag	UNP P40454
B	320	HIS	-	expression tag	UNP P40454
B	321	HIS	-	expression tag	UNP P40454
B	322	HIS	-	expression tag	UNP P40454
B	323	HIS	-	expression tag	UNP P40454
C	318	HIS	-	expression tag	UNP P40454
C	319	HIS	-	expression tag	UNP P40454
C	320	HIS	-	expression tag	UNP P40454
C	321	HIS	-	expression tag	UNP P40454
C	322	HIS	-	expression tag	UNP P40454
C	323	HIS	-	expression tag	UNP P40454
D	318	HIS	-	expression tag	UNP P40454
D	319	HIS	-	expression tag	UNP P40454

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	320	HIS	-	expression tag	UNP P40454
D	321	HIS	-	expression tag	UNP P40454
D	322	HIS	-	expression tag	UNP P40454
D	323	HIS	-	expression tag	UNP P40454

- Molecule 2 is a protein called SIN-ALA-ALA-PRO-LYS-NIT.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	F	6	Total	C	N	O	0	0	0
			43	27	7	9			
2	G	6	Total	C	N	O	0	0	0
			43	27	7	9			
2	H	6	Total	C	N	O	0	0	0
			43	27	7	9			
2	I	6	Total	C	N	O	0	0	0
			43	27	7	9			

- 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	0	1
			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		

*Continued on next page...*

*Continued from previous page...*

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

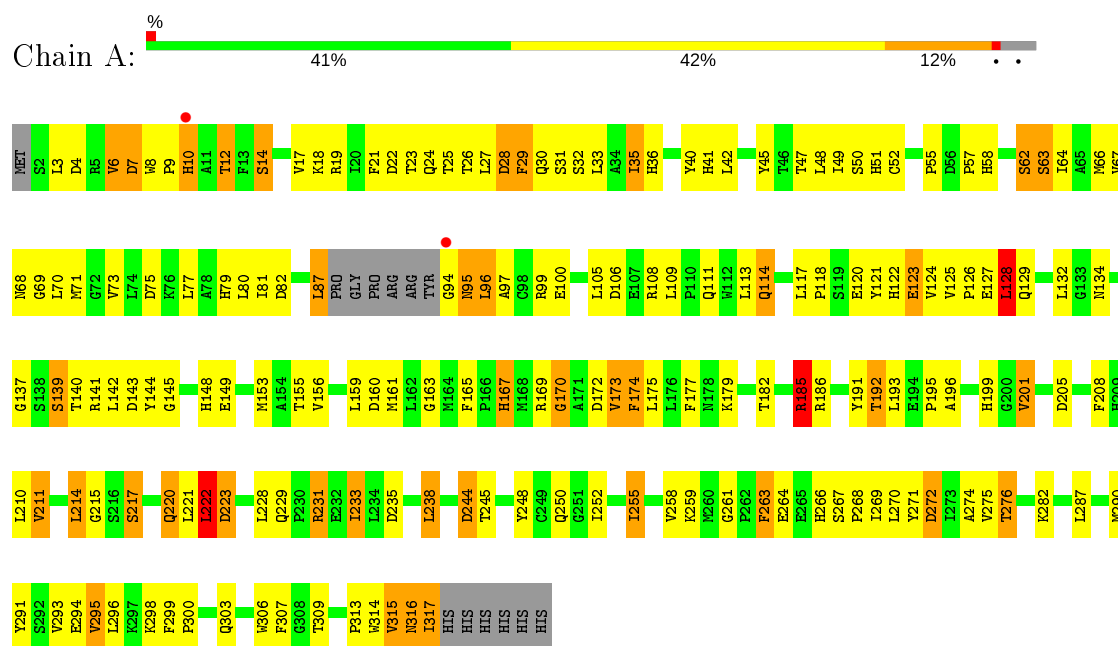
- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Cl	0	0
			1	1		
4	D	1	Total	Cl	0	0
			1	1		

### 3 Residue-property plots

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.

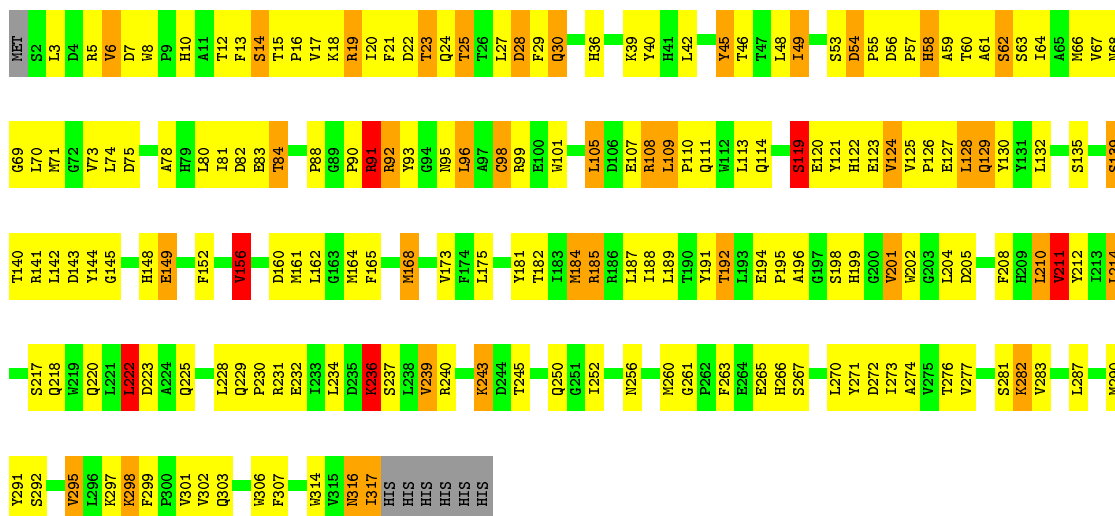
#### • Molecule 1: SERINE/THREONINE-PROTEIN PHOSPHATASE 2A ACTIVATOR 1





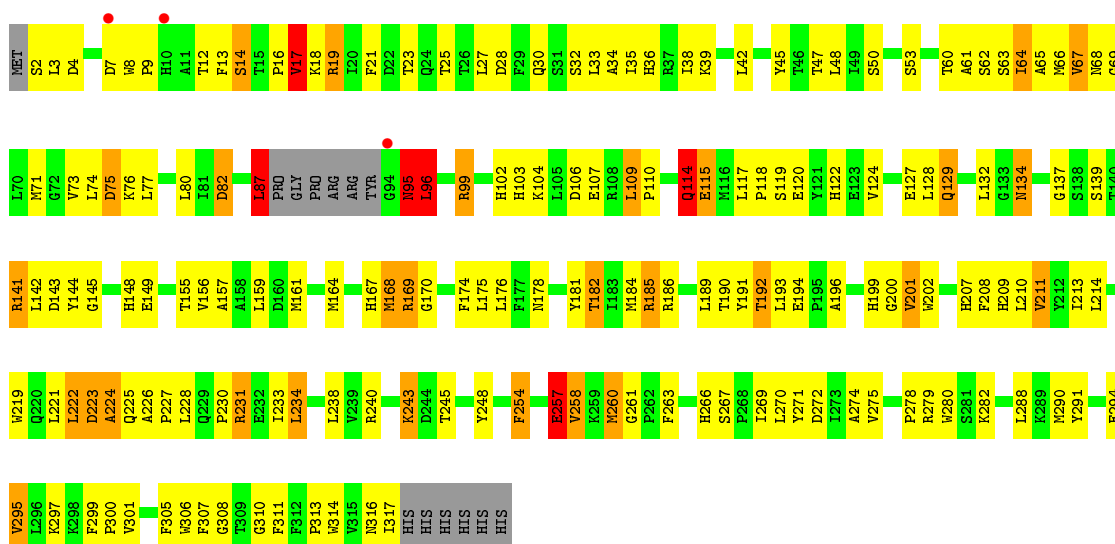
• Molecule 1: SERINE/THREONINE-PROTEIN PHOSPHATASE 2A ACTIVATOR 1

Chain C: 40% 44% 12% . .



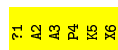
• Molecule 1: SERINE/THREONINE-PROTEIN PHOSPHATASE 2A ACTIVATOR 1

Chain D: 42% 43% 9% . .



• Molecule 2: SIN-ALA-ALA-PRO-LYS-NIT

Chain F: 100%



• Molecule 2: SIN-ALA-ALA-PRO-LYS-NIT



Chain G:  50% 50%



• Molecule 2: SIN-ALA-ALA-PRO-LYS-NIT

Chain H:  50% 33% 17%



• Molecule 2: SIN-ALA-ALA-PRO-LYS-NIT

Chain I:  33% 50% 17%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	86.89Å 86.89Å 410.63Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.12 – 2.80 29.12 – 2.80	Depositor EDS
% Data completeness (in resolution range)	100.0 (29.12-2.80) 99.9 (29.12-2.80)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.86 (at 2.80Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.229 , 0.302 0.235 , 0.267	Depositor DCC
$R_{free}$ test set	2158 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.1	Xtriage
Anisotropy	0.715	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , -24.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.439 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	10394	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	2.0	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 44.70 % 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 1.4694e-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: CL, SIN, NIT, 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	1.14	4/2604 (0.2%)	1.13	10/3543 (0.3%)
1	B	1.11	2/2660 (0.1%)	1.14	6/3621 (0.2%)
1	C	1.12	5/2660 (0.2%)	1.17	8/3621 (0.2%)
1	D	1.10	6/2604 (0.2%)	1.19	16/3543 (0.5%)
2	F	1.20	0/26	1.19	0/34
2	G	2.50	1/26 (3.8%)	1.35	0/34
2	H	2.08	1/26 (3.8%)	1.25	0/34
2	I	2.79	2/26 (7.7%)	1.80	1/34 (2.9%)
All	All	1.13	21/10632 (0.2%)	1.16	41/14464 (0.3%)

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	B	0	1
1	D	0	3
All	All	0	4

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	I	5	LYS	C-O	11.71	1.45	1.23
2	G	5	LYS	C-O	11.16	1.44	1.23
1	C	194	GLU	CG-CD	9.12	1.65	1.51
2	H	5	LYS	C-O	9.04	1.40	1.23
1	C	98	CYS	CB-SG	-8.97	1.67	1.82

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	19	ARG	NE-CZ-NH2	-9.24	115.68	120.30
1	D	185	ARG	NE-CZ-NH1	9.06	124.83	120.30
1	C	236	LYS	CD-CE-NZ	8.64	131.58	111.70
1	B	222	LEU	CA-CB-CG	8.28	134.35	115.30
1	D	87	LEU	CA-CB-CG	8.09	133.90	115.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	138	SER	Peptide
1	D	17	VAL	Peptide
1	D	257	GLU	Peptide
1	D	261	GLY	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	2524	0	2466	159	0
1	B	2576	0	2519	179	0
1	C	2576	0	2519	173	0
1	D	2524	0	2466	151	0
2	F	43	0	39	17	0
2	G	43	0	37	3	0
2	H	43	0	38	5	0
2	I	43	0	38	4	0
3	A	10	0	0	4	0
3	B	5	0	0	1	0
3	D	5	0	0	1	0
4	A	1	0	0	0	0
4	D	1	0	0	0	0
All	All	10394	0	10122	659	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 32.

The worst 5 of 659 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:185:ARG:CB	1:A:185:ARG:CG	1.74	1.65
1:A:233:ILE:CG1	1:A:233:ILE:CD1	1.77	1.61
2:F:5:LYS:C	2:F:6:NIT:HN12	1.18	1.42
1:B:20:ILE:HA	1:B:25:THR:CG2	1.61	1.29
2:F:5:LYS:C	2:F:6:NIT:N1	1.88	1.27

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	306/323 (95%)	257 (84%)	30 (10%)	19 (6%)	1	4
1	B	314/323 (97%)	257 (82%)	41 (13%)	16 (5%)	2	6
1	C	314/323 (97%)	259 (82%)	45 (14%)	10 (3%)	4	13
1	D	306/323 (95%)	252 (82%)	44 (14%)	10 (3%)	4	13
2	F	2/6 (33%)	2 (100%)	0	0	100	100
2	G	2/6 (33%)	2 (100%)	0	0	100	100
2	H	2/6 (33%)	2 (100%)	0	0	100	100
2	I	2/6 (33%)	2 (100%)	0	0	100	100
All	All	1248/1316 (95%)	1033 (83%)	160 (13%)	55 (4%)	2	8

5 of 55 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	7	ASP
1	A	96	LEU
1	A	175	LEU
1	A	223	ASP
1	A	303	GLN

### 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	277/289 (96%)	227 (82%)	50 (18%)	1	5
1	B	282/289 (98%)	231 (82%)	51 (18%)	1	5
1	C	282/289 (98%)	224 (79%)	58 (21%)	1	3
1	D	277/289 (96%)	230 (83%)	47 (17%)	2	6
2	F	2/2 (100%)	2 (100%)	0	100	100
2	G	2/2 (100%)	2 (100%)	0	100	100
2	H	2/2 (100%)	1 (50%)	1 (50%)	0	0
2	I	2/2 (100%)	2 (100%)	0	100	100
All	All	1126/1164 (97%)	919 (82%)	207 (18%)	1	5

5 of 207 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	276	THR
1	C	58	HIS
1	D	201	VAL
1	B	285	LYS
1	C	14	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 41 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	316	ASN
1	C	129	GLN
1	D	225	GLN
1	C	58	HIS
1	C	68	ASN

### 5.3.3 RNA ⓘ

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 6 ligands modelled in this entry, 2 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	1318[A]	-	4,4,4	0.33	0	6,6,6	0.69	0
3	SO4	D	1319[B]	-	4,4,4	0.27	0	6,6,6	0.61	0
3	SO4	A	1320	-	4,4,4	0.26	0	6,6,6	0.34	0
3	SO4	B	1319	-	4,4,4	0.19	0	6,6,6	0.49	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.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1318[A]	SO4	1	0
3	D	1319[B]	SO4	1	0
3	A	1320	SO4	3	0
3	B	1319	SO4	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	H	1
2	I	1
2	F	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	5:LYS	C	6:NIT	N1	1.88
1	H	1:SIN	C4	2:ALA	N	1.20
1	I	1:SIN	C4	2:ALA	N	1.19



## 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	310/323 (95%)	-0.13	2 (0%) 89 86	2, 2, 4, 11	0
1	B	316/323 (97%)	-0.16	1 (0%) 94 93	2, 2, 3, 9	0
1	C	316/323 (97%)	-0.17	0 100 100	2, 2, 4, 9	0
1	D	310/323 (95%)	-0.11	3 (0%) 82 77	2, 2, 3, 11	0
2	F	4/6 (66%)	-0.54	0 100 100	2, 2, 2, 2	0
2	G	4/6 (66%)	-0.23	0 100 100	2, 2, 2, 2	0
2	H	4/6 (66%)	-0.28	0 100 100	2, 2, 2, 2	0
2	I	4/6 (66%)	-0.38	0 100 100	2, 2, 3, 6	0
All	All	1268/1316 (96%)	-0.15	6 (0%) 91 88	2, 2, 4, 11	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	10	HIS	3.6
1	A	94	GLY	3.3
1	A	10	HIS	2.9
1	D	94	GLY	2.8
1	B	10	HIS	2.3

### 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(Å <sup>2</sup> )	Q<0.9
4	CL	A	1321	1/1	0.94	0.07	13,13,13,13	0
3	SO4	A	1318[A]	5/5	0.95	0.24	2,2,2,2	5
3	SO4	D	1319[B]	5/5	0.96	0.21	2,2,2,2	5
3	SO4	A	1320	5/5	0.96	0.15	53,54,54,54	0
4	CL	D	1320	1/1	0.97	0.07	7,7,7,7	0
3	SO4	B	1319	5/5	0.98	0.19	2,2,2,2	5

## 6.5 Other polymers [i](#)

There are no such residues in this entry.