



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 03:03 PM GMT

PDB ID : 4BE1
Title : PFV intasome with inhibitor XZ-116
Authors : Hare, S.; Cherepanov, P.
Deposited on : 2012-10-08
Resolution : 2.71 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
A user guide is available at
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

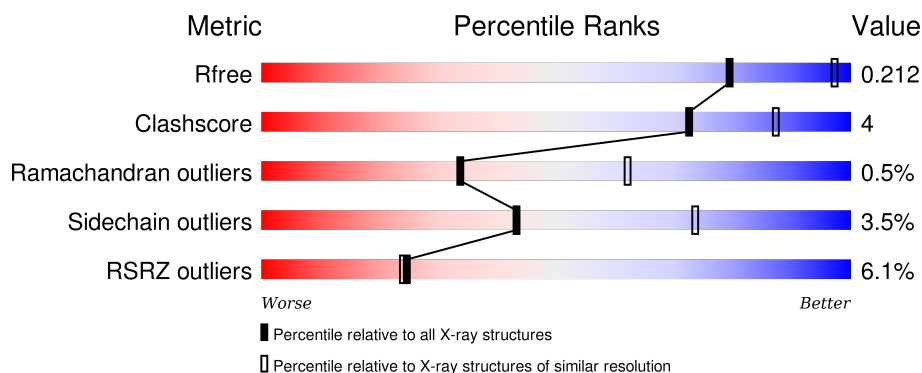
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.71 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}	91344	2439 (2.74-2.70)
Clashscore	102246	2771 (2.74-2.70)
Ramachandran outliers	100387	2726 (2.74-2.70)
Sidechain outliers	100360	2727 (2.74-2.70)
RSRZ outliers	91569	2443 (2.74-2.70)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	395	<div> <div>4%</div> <div>77%</div> <div>15%</div> <div>7%</div> </div>
1	B	395	<div> <div>5%</div> <div>42%</div> <div>5%</div> <div>53%</div> </div>
2	C	19	<div> <div>5%</div> <div>68%</div> <div>21%</div> <div>11%</div> </div>
3	D	17	<div> <div>76%</div> <div>18%</div> <div>6%</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
7	GOL	A	1380	-	-	-	X
7	GOL	B	1301	-	-	-	X

2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 5369 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 INTEGRASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	368	Total	C	N	O	S	0	0	0
			2920	1872	512	532	4			
1	B	184	Total	C	N	O	S	0	0	0
			1417	919	230	267	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP P14350
A	-1	PRO	-	EXPRESSION TAG	UNP P14350
A	0	GLY	-	EXPRESSION TAG	UNP P14350
A	217	SER	GLY	VARIANT	UNP P14350
A	218	GLY	SER	VARIANT	UNP P14350
B	-2	GLY	-	EXPRESSION TAG	UNP P14350
B	-1	PRO	-	EXPRESSION TAG	UNP P14350
B	0	GLY	-	EXPRESSION TAG	UNP P14350
B	217	SER	GLY	VARIANT	UNP P14350
B	218	GLY	SER	VARIANT	UNP P14350

- Molecule 2 is a DNA chain called 19 NUCLEOTIDE PREPROCESSED PFV DONOR DNA (NON-TRANSFERRED STRAND).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	19	Total	C	N	O	P	0	0	0
			387	187	68	114	18			

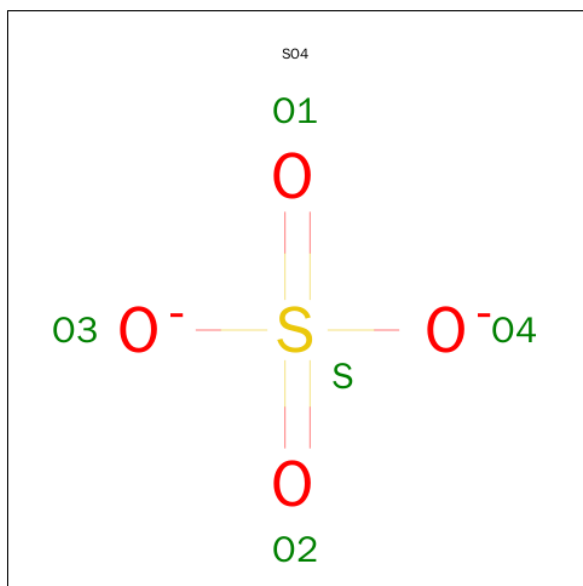
- Molecule 3 is a DNA chain called 17 NUCLEOTIDE PREPROCESSED PFV DONOR DNA (TRANSFERRED STRAND).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	17	Total	C	N	O	P	0	0	0
			345	166	65	98	16			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

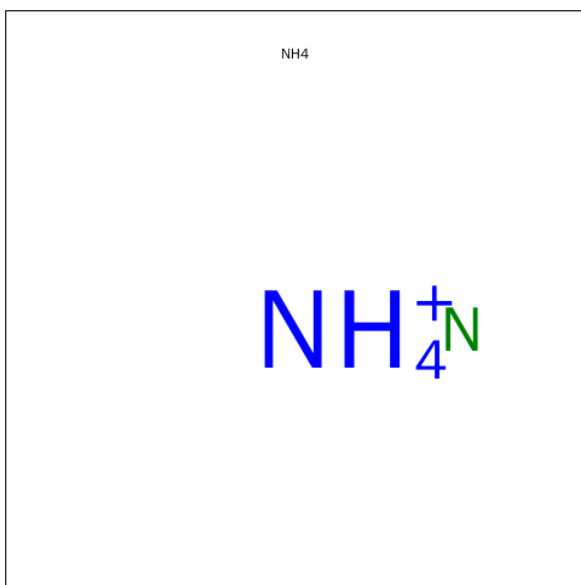
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Zn	0	0
			1	1		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



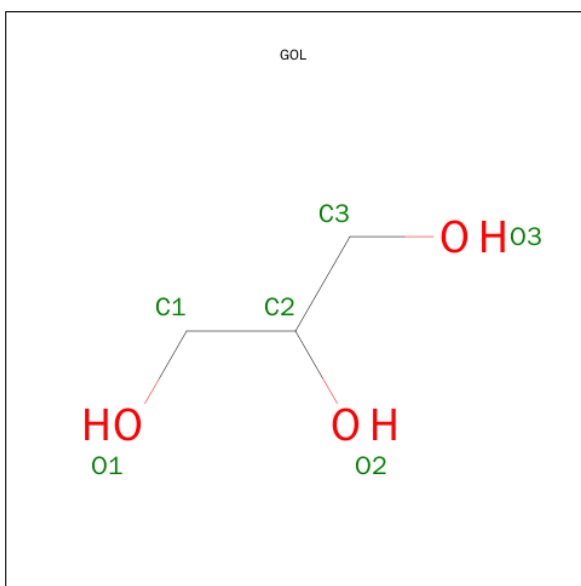
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		
5	A	1	Total	O	S	0	0
			5	4	1		
5	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 6 is AMMONIUM ION (three-letter code: NH4) (formula: H₄N).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total N 1 1	0	0

- Molecule 7 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

Continued on next page...

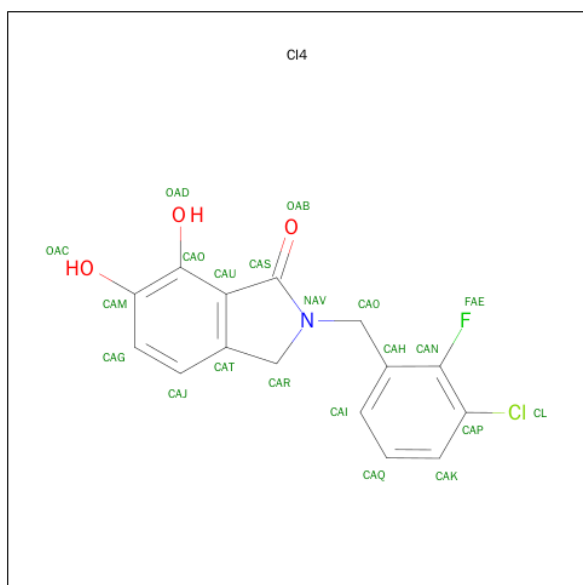
Continued from previous page...

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

- Molecule 8 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	B	1	Total	Mg	0	0
			1	1		
8	A	2	Total	Mg	0	0
			2	2		

- Molecule 9 is 2-(3-CHLORO-2-FLUOROBENZYL)-6,7-DIHYDROXY-2,3-DIHYDRO-1H-ISOINDOL-1-ONE (three-letter code: CI4) (formula: C₁₅H₁₁ClFNO₃).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
9	A	1	Total	C	Cl	F	N	O	0	0
			21	15	1	1	1	3		

- Molecule 10 is water.


Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	163	Total	O	0	0
			163	163		
10	B	37	Total	O	0	0
			37	37		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	C	18	Total	O	0	0
			18	18		
10	D	17	Total	O	0	0
			17	17		

- Molecule 3: 17 NUCLEOTIDE PREPROCESSED PFV DONOR DNA (TRANSFERRED STRAND)

Chain D:  76% 18% 6%



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	161.03Å 161.03Å 123.94Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.92 – 2.71 49.26 – 2.71	Depositor EDS
% Data completeness (in resolution range)	97.9 (50.92-2.71) 97.9 (49.26-2.71)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.07 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.180 , 0.213 0.178 , 0.212	Depositor DCC
R_{free} test set	2215 reflections (5.32%)	DCC
Wilson B-factor (Å ²)	63.1	Xtriage
Anisotropy	0.293	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 45.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 43973 reflections	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5369	wwPDB-VP
Average B, all atoms (Å ²)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.31% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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, MG, ZN, NH4, SO4, CI4

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	0.66	1/2999 (0.0%)	0.76	1/4094 (0.0%)
1	B	0.66	1/1457 (0.1%)	0.72	0/1996
2	C	0.54	0/433	1.12	6/667 (0.9%)
3	D	0.46	0/387	1.13	4/595 (0.7%)
All	All	0.64	2/5276 (0.0%)	0.82	11/7352 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	81	TRP	CD2-CE2	5.21	1.47	1.41
1	B	242	TRP	CD2-CE2	5.12	1.47	1.41

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	1	DT	P-O3'-C3'	9.53	131.14	119.70
2	C	1	DA	P-O3'-C3'	8.45	129.84	119.70
2	C	2	DT	P-O3'-C3'	6.41	127.39	119.70
3	D	6	DA	P-O3'-C3'	6.36	127.33	119.70
3	D	12	DA	P-O3'-C3'	6.25	127.20	119.70

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	2920	0	2945	31	0
1	B	1417	0	1373	8	0
2	C	387	0	218	3	0
3	D	345	0	193	1	0
4	A	1	0	0	0	0
5	A	10	0	0	0	0
5	B	5	0	0	1	0
6	A	1	0	0	0	0
7	A	18	0	24	3	0
7	B	6	0	8	0	0
8	A	2	0	0	0	0
8	B	1	0	0	0	0
9	A	21	0	9	0	0
10	A	163	0	0	2	1
10	B	37	0	0	0	0
10	C	18	0	0	0	0
10	D	17	0	0	0	0
All	All	5369	0	4770	42	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 4.

The worst 5 of 42 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:258:SER:O	1:B:261:LEU:O	2.11	0.68
2:C:1:DA:H2''	2:C:2:DT:H5'	1.77	0.67
1:A:162:SER:HB2	7:A:1378:GOL:H2	1.79	0.64
1:A:69:ARG:HD2	1:A:70:GLU:OE2	2.01	0.61
1:A:51:GLN:HA	1:A:54:ARG:HD2	1.83	0.59

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)
10:A:2012:HOH:O	10:A:2109:HOH:O[8_554]	2.08	0.12

5.3 Torsion angles

5.3.1 Protein backbone

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	366/395 (93%)	345 (94%)	18 (5%)	3 (1%)	24	50
1	B	182/395 (46%)	172 (94%)	10 (6%)	0	100	100
All	All	548/790 (69%)	517 (94%)	28 (5%)	3 (0%)	34	62

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	201	GLU
1	A	14	GLN
1	A	359	GLY

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	327/354 (92%)	315 (96%)	12 (4%)	41	71
1	B	153/354 (43%)	148 (97%)	5 (3%)	45	75
All	All	480/708 (68%)	463 (96%)	17 (4%)	43	73

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	293	SER
1	A	351	THR
1	B	149	MET

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	219	LYS
1	B	164	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	17	GLN

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 13 ligands modelled in this entry, 1 is modelled with single atom and 4 are monoatomic - leaving 8 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	$\# Z > 2$
5	SO4	A	1376	-	4,4,4	0.46	0	6,6,6	0.22	0
7	GOL	A	1378	-	5,5,5	0.46	0	5,5,5	0.74	0
7	GOL	A	1379	-	5,5,5	0.29	0	5,5,5	0.72	0
7	GOL	A	1380	-	5,5,5	0.34	0	5,5,5	0.38	0
9	CI4	A	1383	8	22,23,23	1.41	2 (9%)	27,34,34	1.82	5 (18%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
5	SO4	A	1384	-	4,4,4	0.61	0	6,6,6	0.38	0
5	SO4	B	1300	-	4,4,4	0.57	0	6,6,6	0.23	0
7	GOL	B	1301	-	5,5,5	0.43	0	5,5,5	0.32	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
5	SO4	A	1376	-	-	0/0/0/0	0/0/0/0
7	GOL	A	1378	-	-	0/4/4/4	0/0/0/0
7	GOL	A	1379	-	-	0/4/4/4	0/0/0/0
7	GOL	A	1380	-	-	0/4/4/4	0/0/0/0
9	CI4	A	1383	8	-	0/4/16/16	0/3/3/3
5	SO4	A	1384	-	-	0/0/0/0	0/0/0/0
5	SO4	B	1300	-	-	0/0/0/0	0/0/0/0
7	GOL	B	1301	-	-	0/4/4/4	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
9	A	1383	CI4	CAU-CAS	-4.78	1.41	1.49
9	A	1383	CI4	CAR-CAT	2.24	1.53	1.50

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	A	1383	CI4	CAR-CAT-CAU	-3.22	107.69	109.68
9	A	1383	CI4	OAB-CAS-CAU	-2.15	125.93	128.96
9	A	1383	CI4	CA0-NAV-CAR	2.61	126.09	122.26
9	A	1383	CI4	CAK-CAP-CAN	3.14	121.35	119.07
9	A	1383	CI4	CAU-CAS-NAV	5.72	111.73	106.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	A	1378	GOL	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	A	1379	GOL	2	0
5	B	1300	SO4	1	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 ⓘ

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, 95th 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 ²)	Q<0.9
1	A	368/395 (93%)	0.03	17 (4%) 36 35	37, 53, 96, 136	0
1	B	184/395 (46%)	0.14	18 (9%) 10 8	46, 62, 143, 175	0
2	C	19/19 (100%)	-0.18	1 (5%) 30 29	44, 57, 88, 123	0
3	D	17/17 (100%)	-0.23	0 100 100	45, 51, 113, 114	0
All	All	588/826 (71%)	0.05	36 (6%) 25 24	37, 56, 114, 175	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	296	GLN	5.1
1	B	295	LEU	4.7
1	A	8	LEU	4.5
1	B	294	LEU	4.4
1	B	292	LEU	4.2

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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(\AA^2)	Q<0.9
7	GOL	B	1301	6/6	0.89	0.34	3.77	57,69,75,79	0
7	GOL	A	1380	6/6	0.91	0.20	2.99	76,89,100,105	0
5	SO4	A	1384	5/5	0.97	0.19	1.89	80,85,87,93	0
4	ZN	A	393	1/1	1.00	0.15	-0.14	51,51,51,51	0
9	Cl4	A	1383	21/21	0.97	0.16	-0.29	38,48,53,55	0
5	SO4	B	1300	5/5	0.95	0.14	-0.74	78,82,92,100	0
8	MG	A	1382	1/1	0.98	0.09	-2.10	53,53,53,53	0
6	NH4	A	1377	1/1	0.86	0.35	-	32,32,32,32	0
7	GOL	A	1378	6/6	0.81	0.27	-	77,89,95,95	0
7	GOL	A	1379	6/6	0.96	0.16	-	69,77,80,89	0
8	MG	B	1302	1/1	0.88	0.14	-	88,88,88,88	0
5	SO4	A	1376	5/5	0.96	0.20	-	90,93,95,98	0
8	MG	A	1381	1/1	0.99	0.11	-	47,47,47,47	0

6.5 Other polymers [i](#)

There are no such residues in this entry.