



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2018 – 06:39 pm GMT

PDB ID : 5OSF
Title : Crystal structure of Aurora-A kinase in complex with an allosterically binding fragment
Authors : McIntyre, P.J.; Collins, P.M.; von Delft, F.; Bayliss, R.
Deposited on : 2017-08-17
Resolution : 1.89 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at 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 : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

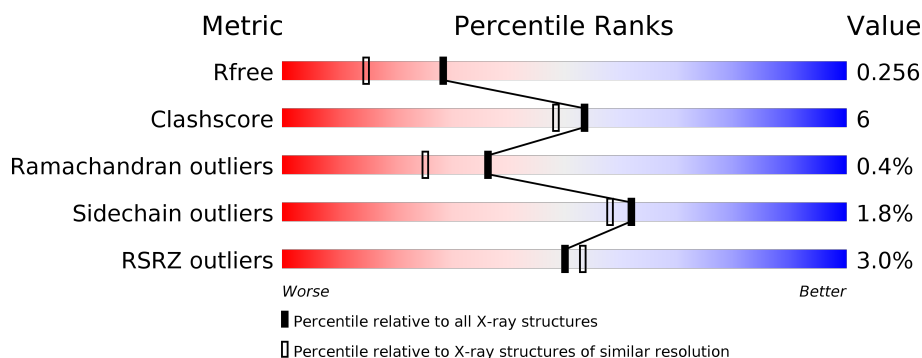
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.89 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}	111664	5502 (1.90-1.90)
Clashscore	122126	6115 (1.90-1.90)
Ramachandran outliers	120053	6048 (1.90-1.90)
Sidechain outliers	120020	6048 (1.90-1.90)
RSRZ outliers	108989	5379 (1.90-1.90)

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	265	<div> <div>3%</div> <div>79%</div> <div>19%</div> <div>.</div> </div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 2272 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 Aurora kinase A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
1	A	265	2141	1374	372	389	1	5	0	0	0

There is a discrepancy between the modelled and reference sequences:

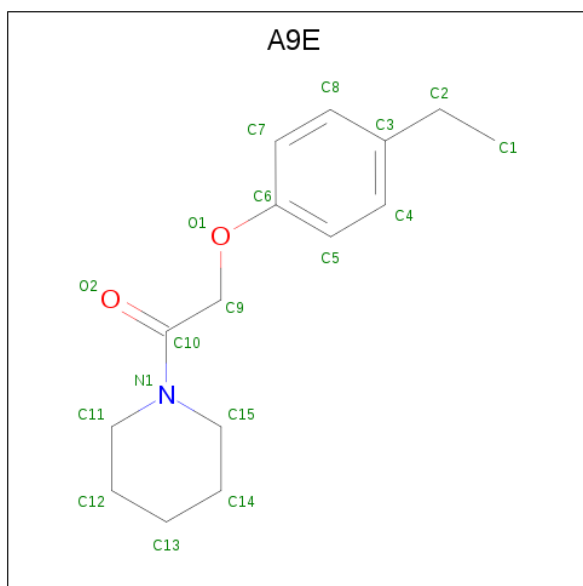
Chain	Residue	Modelled	Actual	Comment	Reference
A	290	ALA	CYS	conflict	UNP O14965

- Molecule 2 is ADENOSINE-5'-DIPHOSPHATE (three-letter code: ADP) (formula: $C_{10}H_{15}N_5O_{10}P_2$).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Mg	0	0
			2	2		

- Molecule 4 is 2-(4-ethylphenoxy)-1-piperidin-1-yl-ethanone (three-letter code: A9E) (formula: C₁₅H₂₁NO₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			18	15	1	2		

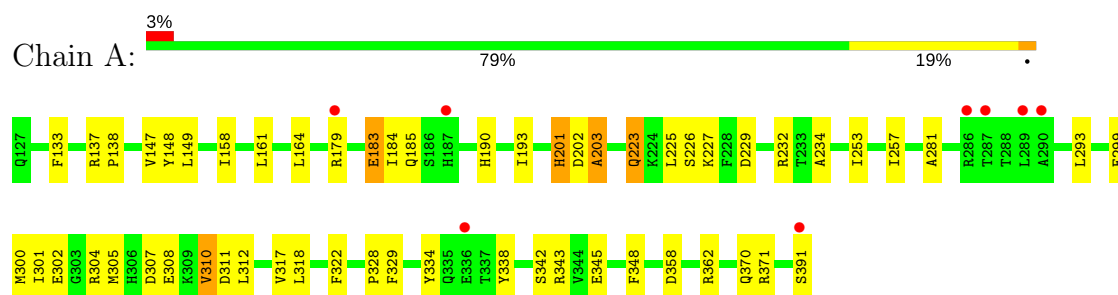
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	84	Total	O	0	0
			84	84		

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 ($\text{RSRZ} > 2$). 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: Aurora kinase A



4 Data and refinement statistics

Property	Value	Source
Space group	P 61 2 2	Depositor
Cell constants a, b, c, α , β , γ	82.20Å 82.20Å 175.21Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.15 – 1.89 45.15 – 1.89	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.15-1.89) 100.0 (45.15-1.89)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.05 (at 1.89Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.202 , 0.254 0.203 , 0.256	Depositor DCC
R_{free} test set	1445 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	47.4	Xtriage
Anisotropy	0.032	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 49.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2272	wwPDB-VP
Average B, all atoms (Å ²)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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, 0.333 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: TPO, A9E, MG, ADP

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	1.50	18/2184 (0.8%)	1.21	14/2960 (0.5%)

All (18) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	223	GLN	CG-CD	8.70	1.71	1.51
1	A	342	SER	CB-OG	-7.14	1.32	1.42
1	A	133	PHE	CE1-CZ	6.81	1.50	1.37
1	A	299	GLU	CD-OE1	6.40	1.32	1.25
1	A	345	GLU	CB-CG	-6.27	1.40	1.52
1	A	310	VAL	CB-CG2	-5.98	1.40	1.52
1	A	348	PHE	CE2-CZ	5.86	1.48	1.37
1	A	234	ALA	CA-CB	5.74	1.64	1.52
1	A	300	MET	CB-CG	-5.69	1.33	1.51
1	A	148	TYR	CE2-CZ	5.68	1.46	1.38
1	A	299	GLU	CG-CD	5.56	1.60	1.51
1	A	183	GLU	CG-CD	5.45	1.60	1.51
1	A	371	ARG	CZ-NH1	5.38	1.40	1.33
1	A	308	GLU	CG-CD	5.28	1.59	1.51
1	A	329	PHE	CE2-CZ	5.16	1.47	1.37
1	A	345	GLU	CD-OE1	5.08	1.31	1.25
1	A	322	PHE	CD1-CE1	5.00	1.49	1.39
1	A	147	VAL	CB-CG2	5.00	1.63	1.52

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	305	MET	CG-SD-CE	15.09	124.34	100.20
1	A	305	MET	CA-CB-CG	-12.73	91.65	113.30
1	A	312	LEU	CB-CG-CD2	7.30	123.41	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	161	LEU	CB-CG-CD2	-6.72	99.57	111.00
1	A	201	HIS	CB-CA-C	6.50	123.40	110.40
1	A	312	LEU	CA-CB-CG	-6.26	100.89	115.30
1	A	229	ASP	CB-CG-OD1	6.25	123.92	118.30
1	A	164	LEU	CB-CG-CD1	6.09	121.36	111.00
1	A	300	MET	CG-SD-CE	-5.95	90.68	100.20
1	A	232	ARG	NE-CZ-NH2	-5.56	117.52	120.30
1	A	149	LEU	CB-CG-CD1	-5.51	101.63	111.00
1	A	312	LEU	CB-CG-CD1	5.33	120.06	111.00
1	A	334	TYR	CB-CG-CD2	-5.33	117.80	121.00
1	A	201	HIS	N-CA-CB	-5.18	101.28	110.60

There are no chirality outliers.

There are no planarity outliers.

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	2141	0	2106	24	0
2	A	27	0	12	1	0
3	A	2	0	0	0	0
4	A	18	0	0	0	0
5	A	84	0	0	4	0
All	All	2272	0	2118	25	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 6.

All (25) 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:370:GLN:NE2	5:A:501:HOH:O	2.07	0.87
1:A:257:ILE:HG22	1:A:318:LEU:HD11	1.63	0.81
1:A:253:ILE:HD11	1:A:281:ALA:HB2	1.63	0.80
1:A:179:ARG:HG3	1:A:179:ARG:HH11	1.49	0.77

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:185:GLN:OE1	5:A:502:HOH:O	2.10	0.69
1:A:179:ARG:NH1	1:A:183:GLU:OE1	2.27	0.66
1:A:358:ASP:O	1:A:362:ARG:HG3	2.03	0.58
1:A:202:ASP:O	1:A:203:ALA:HB3	2.08	0.54
1:A:343:ARG:HD3	5:A:503:HOH:O	2.11	0.49
1:A:370:GLN:CD	5:A:501:HOH:O	2.45	0.49
1:A:179:ARG:NH1	1:A:179:ARG:HG3	2.25	0.48
1:A:301:ILE:HG22	1:A:338:TYR:CE1	2.50	0.47
1:A:317:VAL:HG13	1:A:328:PRO:HD2	1.96	0.47
1:A:225:LEU:O	1:A:226:SER:HB2	2.15	0.46
1:A:301:ILE:HG22	1:A:338:TYR:HE1	1.80	0.45
1:A:310:VAL:HG23	1:A:311:ASP:N	2.33	0.44
1:A:307:ASP:O	1:A:310:VAL:HG22	2.18	0.44
1:A:183:GLU:HG2	1:A:184:ILE:N	2.33	0.44
1:A:302:GLU:OE1	1:A:304:ARG:NH1	2.45	0.43
2:A:401:ADP:H8	2:A:401:ADP:O5'	2.02	0.42
1:A:293:LEU:HD22	1:A:301:ILE:HD13	2.02	0.41
1:A:137:ARG:HA	1:A:138:PRO:HD3	1.92	0.41
1:A:190:HIS:HB3	1:A:193:ILE:HD12	2.03	0.41
1:A:179:ARG:HD2	1:A:179:ARG:O	2.20	0.40
1:A:158:ILE:HD12	1:A:158:ILE:HG23	1.80	0.40

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	262/265 (99%)	251 (96%)	10 (4%)	1 (0%)	36 25

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	203	ALA

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	225/232 (97%)	221 (98%)	4 (2%)	62 58

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	201	HIS
1	A	223	GLN
1	A	227	LYS
1	A	391	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

1 non-standard protein/DNA/RNA residue is 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	TPO	A	288	1	9,10,11	3.19	4 (44%)	11,14,16	1.33	1 (9%)

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	TPO	A	288	1	-	0/8/11/13	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	288	TPO	P-O1P	3.29	1.61	1.50
1	A	288	TPO	P-OG1	4.23	1.67	1.59
1	A	288	TPO	CA-C	5.13	1.57	1.50
1	A	288	TPO	CB-CA	5.25	1.63	1.53

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	288	TPO	O-C-CA	-2.82	118.53	125.09

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 2 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
2	ADP	A	401	3	25,29,29	1.84	6 (24%)	25,45,45	2.46	9 (36%)
4	A9E	A	404	-	19,19,19	0.55	0	24,24,24	0.81	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
2	ADP	A	401	3	-	0/12/32/32	0/3/3/3
4	A9E	A	404	-	-	0/11/19/19	0/2/2/2

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	ADP	C5'-C4'	2.09	1.58	1.51
2	A	401	ADP	PA-O5'	2.10	1.67	1.59
2	A	401	ADP	O2'-C2'	2.39	1.48	1.43
2	A	401	ADP	C5-C4	2.76	1.46	1.40
2	A	401	ADP	C4-N3	3.19	1.40	1.35
2	A	401	ADP	PB-O3A	5.53	1.68	1.60

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	ADP	N3-C2-N1	-8.86	121.28	128.86
2	A	401	ADP	O5'-PA-O1A	-2.82	98.04	109.07
2	A	401	ADP	C4-C5-N7	-2.67	106.83	109.41
2	A	401	ADP	O2'-C2'-C3'	-2.50	103.81	111.83
2	A	401	ADP	O4'-C4'-C3'	-2.01	101.18	105.15
2	A	401	ADP	O3B-PB-O1B	2.52	120.42	110.60
2	A	401	ADP	C2-N1-C6	2.65	123.26	118.75
2	A	401	ADP	C4'-O4'-C1'	2.85	112.80	109.83
2	A	401	ADP	O3B-PB-O2B	3.04	119.60	107.59

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
2	A	401	ADP	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	264/265 (99%)	0.08	8 (3%) 50 53	36, 52, 82, 146	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	290	ALA	6.0
1	A	286	ARG	5.2
1	A	336	GLU	3.8
1	A	391	SER	3.2
1	A	287	THR	3.1
1	A	179	ARG	2.5
1	A	289	LEU	2.4
1	A	187	HIS	2.2

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

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, 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	B-factors(Å ²)	Q<0.9
1	TPO	A	288	11/12	0.82	0.31	80,100,118,118	0

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, 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	B-factors(Å ²)	Q<0.9
3	MG	A	403	1/1	0.62	0.14	56,56,56,56	0
4	A9E	A	404	18/18	0.78	0.30	69,72,76,76	0
2	ADP	A	401	27/27	0.97	0.09	34,35,41,42	0
3	MG	A	402	1/1	0.98	0.04	37,37,37,37	0

6.5 Other polymers [i](#)

There are no such residues in this entry.