



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2018 – 11:49 am GMT

PDB ID : 4K5N  
Title : Phosphonic Arginine Mimetics as Inhibitors of the M1 Aminopeptidases from Plasmodium falciparum  
Authors : McGowan, S.  
Deposited on : 2013-04-15  
Resolution : 1.91 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](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 : 1.7.3 (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

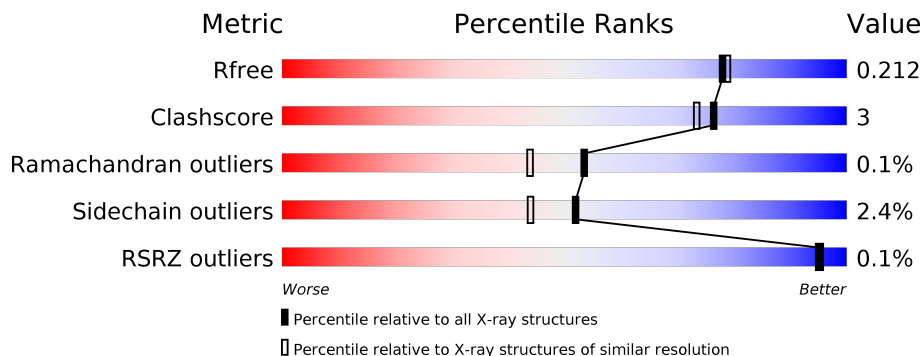
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.91 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	6904 (1.94-1.90)
Clashscore	122126	7577 (1.94-1.90)
Ramachandran outliers	120053	7491 (1.94-1.90)
Sidechain outliers	120020	7491 (1.94-1.90)
RSRZ outliers	108989	6759 (1.94-1.90)

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  $\geq 3$ , 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	895	 89% 9% ..

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 8291 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 M1 family aminopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	889	Total	C	N	O	S	0	8	0
			7237	4662	1173	1373	29			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	213	GLN	ASN	ENGINEERED MUTATION	UNP O96935
A	223	GLN	ASN	ENGINEERED MUTATION	UNP O96935
A	378	PRO	HIS	ENGINEERED MUTATION	UNP O96935
A	501	GLN	ASN	ENGINEERED MUTATION	UNP O96935
A	745	GLN	ASN	ENGINEERED MUTATION	UNP O96935
A	795	GLN	ASN	ENGINEERED MUTATION	UNP O96935
A	1069	GLN	ASN	ENGINEERED MUTATION	UNP O96935
A	1085	HIS	-	EXPRESSION TAG	UNP O96935
A	1086	HIS	-	EXPRESSION TAG	UNP O96935
A	1087	HIS	-	EXPRESSION TAG	UNP O96935
A	1088	HIS	-	EXPRESSION TAG	UNP O96935
A	1089	HIS	-	EXPRESSION TAG	UNP O96935
A	1090	HIS	-	EXPRESSION TAG	UNP O96935

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

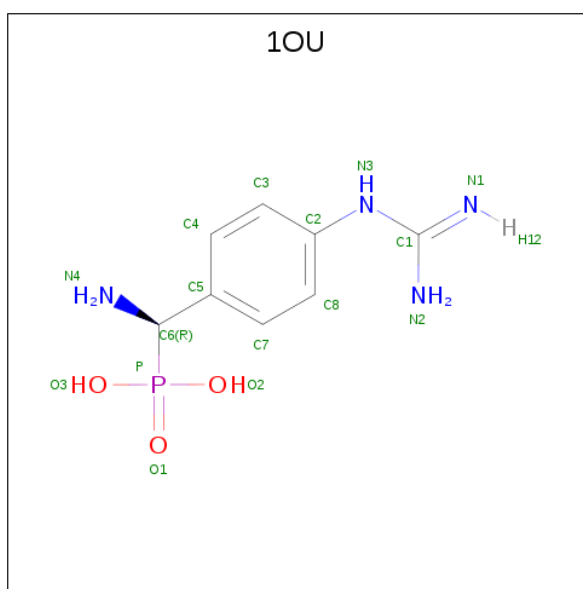
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total	Mg	0	0
			3	3		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

- Molecule 5 is [(R)-amino(4-carbamimidamidophenyl)methyl]phosphonic acid (three-letter code: 1OU) (formula:  $C_8H_{13}N_4O_3P$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	0	0
			16	8	4	3	1		

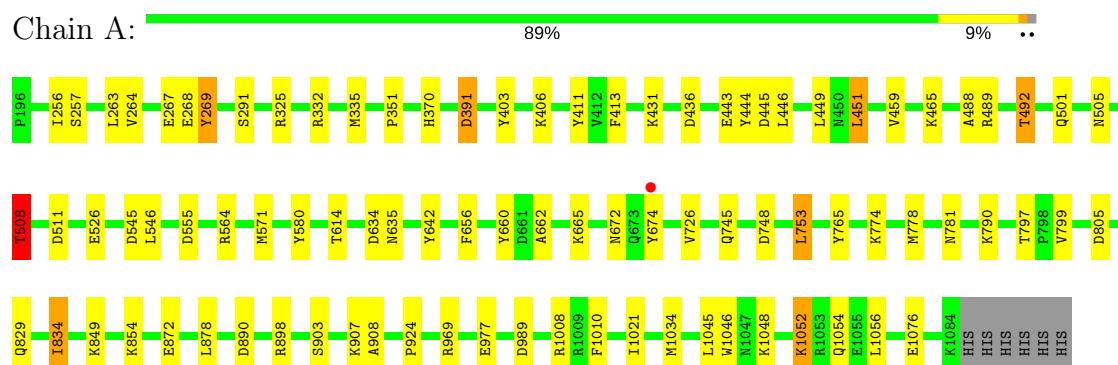
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1022	Total	O	0	0
			1022	1022		

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

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: M1 family aminopeptidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.76Å 109.45Å 118.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	80.45 – 1.91 80.45 – 1.91	Depositor EDS
% Data completeness (in resolution range)	99.8 (80.45-1.91) 99.8 (80.45-1.91)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.06 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.161 , 0.210 0.163 , 0.212	Depositor DCC
$R_{free}$ test set	3936 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.4	Xtriage
Anisotropy	0.344	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 47.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	8291	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% 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: ZN, GOL, MG, 1OU

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	1.17	8/7424 (0.1%)	0.95	16/10053 (0.2%)

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	765	TYR	CD1-CE1	6.39	1.49	1.39
1	A	799	VAL	CB-CG1	6.23	1.66	1.52
1	A	580	TYR	CD2-CE2	5.70	1.47	1.39
1	A	977	GLU	CG-CD	5.48	1.60	1.51
1	A	526	GLU	CG-CD	5.34	1.59	1.51
1	A	642	TYR	CD2-CE2	5.25	1.47	1.39
1	A	660	TYR	CD1-CE1	5.18	1.47	1.39
1	A	849	LYS	CE-NZ	5.18	1.62	1.49

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	325	ARG	NE-CZ-NH2	-8.81	115.89	120.30
1	A	451	LEU	CB-CG-CD1	8.31	125.13	111.00
1	A	748	ASP	CB-CG-OD1	8.26	125.74	118.30
1	A	1008	ARG	NE-CZ-NH2	-8.03	116.28	120.30
1	A	508	THR	N-CA-CB	-6.74	97.49	110.30
1	A	898	ARG	NE-CZ-NH1	6.00	123.30	120.30
1	A	564	ARG	NE-CZ-NH1	5.92	123.26	120.30
1	A	391	ASP	CB-CG-OD2	-5.88	113.01	118.30
1	A	431	LYS	CD-CE-NZ	-5.81	98.35	111.70
1	A	748	ASP	CB-CG-OD2	-5.71	113.16	118.30
1	A	391	ASP	CB-CG-OD1	5.49	123.24	118.30
1	A	898	ARG	NE-CZ-NH2	-5.27	117.66	120.30
1	A	753	LEU	CA-CB-CG	-5.24	103.25	115.30

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	451	LEU	CB-CG-CD2	5.16	119.77	111.00
1	A	774	LYS	CD-CE-NZ	-5.13	99.90	111.70
1	A	511	ASP	CB-CG-OD2	5.02	122.82	118.30

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	7237	0	7101	48	0
2	A	1	0	0	0	0
3	A	3	0	0	0	0
4	A	12	0	15	0	0
5	A	16	0	11	0	0
6	A	1022	0	0	20	0
All	All	8291	0	7127	48	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 3.

All (48) 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:834:ILE:HB	6:A:1832:HOH:O	1.36	1.21
1:A:781[A]:ASN:ND2	6:A:2196:HOH:O	1.96	0.99
1:A:890[B]:ASP:OD1	6:A:2190:HOH:O	1.88	0.88
1:A:778[B]:MET:SD	1:A:797:THR:HG21	2.19	0.82
1:A:834:ILE:HG13	1:A:834:ILE:O	1.89	0.72
1:A:969:ARG:HD3	6:A:2058:HOH:O	1.91	0.70
1:A:805:ASP:OD1	6:A:2207:HOH:O	2.11	0.68
1:A:256:ILE:HD11	1:A:269:TYR:CD1	2.28	0.68
1:A:1076:GLU:OE2	6:A:2218:HOH:O	2.13	0.66
1:A:989:ASP:CB	6:A:1494:HOH:O	2.44	0.65
1:A:878:LEU:HD21	1:A:924:PRO:HG2	1.80	0.63

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:672:ASN:ND2	6:A:2192:HOH:O	2.33	0.62
1:A:656:PHE:O	1:A:745:GLN:HG3	2.00	0.61
1:A:263:LEU:HA	1:A:268:GLU:OE1	2.02	0.60
1:A:1034[B]:MET:HE2	6:A:1748:HOH:O	2.02	0.59
1:A:508:THR:HG22	1:A:614:THR:HG22	1.84	0.58
1:A:1034[B]:MET:CE	6:A:1748:HOH:O	2.51	0.58
1:A:726:VAL:HG22	6:A:1725:HOH:O	2.04	0.57
1:A:1076:GLU:HG2	6:A:2193:HOH:O	2.05	0.56
1:A:488:ALA:O	1:A:492:THR:HB	2.08	0.53
1:A:745:GLN:NE2	6:A:1395:HOH:O	2.42	0.52
1:A:508:THR:CG2	1:A:614:THR:HG22	2.40	0.52
1:A:403:TYR:N	1:A:443:GLU:OE2	2.36	0.51
1:A:257:SER:HB2	1:A:291:SER:OG	2.13	0.48
1:A:1046:TRP:O	1:A:1054:GLN:HG2	2.14	0.48
1:A:501:GLN:NE2	1:A:505:ASN:HD22	2.12	0.48
1:A:1052:LYS:HE2	1:A:1056:LEU:HG	1.97	0.47
1:A:834:ILE:HD11	1:A:1048:LYS:HB3	1.97	0.47
1:A:351:PRO:HG2	1:A:391:ASP:HA	1.97	0.46
1:A:444:TYR:CE2	1:A:446:LEU:HB2	2.51	0.46
1:A:674:TYR:CD2	6:A:2069:HOH:O	2.56	0.46
1:A:1010:PHE:CE1	1:A:1021:ILE:HG13	2.51	0.45
1:A:489:ARG:O	1:A:492:THR:HG22	2.17	0.45
1:A:903:SER:O	1:A:907:LYS:HG3	2.17	0.45
1:A:829:GLN:NE2	6:A:1994:HOH:O	2.51	0.44
1:A:790:LYS:HG2	6:A:2033:HOH:O	2.17	0.43
1:A:908:ALA:HA	6:A:2096:HOH:O	2.18	0.43
1:A:445:ASP:N	1:A:445:ASP:OD1	2.49	0.43
1:A:1052:LYS:HA	1:A:1052:LYS:HD2	1.83	0.43
1:A:264:VAL:N	1:A:268:GLU:OE1	2.51	0.43
1:A:465:LYS:HE2	6:A:1581:HOH:O	2.19	0.42
1:A:501:GLN:HA	1:A:501:GLN:HE21	1.84	0.42
1:A:332:ARG:HG3	1:A:335:MET:HG2	2.00	0.42
1:A:781[A]:ASN:CG	6:A:2196:HOH:O	2.46	0.41
1:A:264:VAL:CG2	1:A:267:GLU:HB3	2.51	0.41
1:A:411:TYR:HB3	1:A:413:PHE:CE2	2.56	0.41
1:A:546:LEU:CD1	6:A:2122:HOH:O	2.69	0.40
1:A:662:ALA:O	1:A:665:LYS:HD3	2.21	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	895/895 (100%)	878 (98%)	16 (2%)	1 (0%)	53	44

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	459	VAL

### 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	799/827 (97%)	780 (98%)	19 (2%)	52	44

All (19) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	269	TYR
1	A	370	HIS
1	A	406	LYS
1	A	436	ASP
1	A	449	LEU
1	A	451	LEU
1	A	492	THR
1	A	508	THR
1	A	545	ASP
1	A	555	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	571	MET
1	A	634	ASP
1	A	635	ASN
1	A	753	LEU
1	A	834	ILE
1	A	854	LYS
1	A	872	GLU
1	A	1045	LEU
1	A	1052	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	501	GLN
1	A	527	ASN
1	A	829	GLN
1	A	882	ASN
1	A	915	ASN
1	A	1005	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 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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
4	GOL	A	1104	-	5,5,5	0.40	0	5,5,5	1.21	0
4	GOL	A	1105	-	5,5,5	0.97	0	5,5,5	2.00	3 (60%)
5	1OU	A	1106	2	15,16,16	2.35	3 (20%)	19,23,23	1.66	4 (21%)

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
4	GOL	A	1104	-	-	0/4/4/4	0/0/0/0
4	GOL	A	1105	-	-	0/4/4/4	0/0/0/0
5	1OU	A	1106	2	-	0/13/14/14	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1106	1OU	P-C6	-2.46	1.80	1.83
5	A	1106	1OU	P-O2	5.66	1.63	1.54
5	A	1106	1OU	P-O3	5.74	1.64	1.54

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1106	1OU	O2-P-O1	-4.15	102.89	113.43
4	A	1105	GOL	O2-C2-C1	2.13	118.73	109.00
5	A	1106	1OU	C4-C3-C2	2.16	122.85	120.30
5	A	1106	1OU	C4-C5-C6	2.35	123.51	120.75
4	A	1105	GOL	O3-C3-C2	2.36	121.57	110.11
5	A	1106	1OU	O3-P-O2	2.61	114.76	107.60
4	A	1105	GOL	C3-C2-C1	2.87	122.69	111.63

There are no chirality outliers.

There are no torsion outliers.

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 [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	889/895 (99%)	-0.37	1 (0%) 95 95	8, 19, 34, 49	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	674	TYR	2.1

### 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	GOL	A	1105	6/6	0.94	0.14	15,24,26,26	0
3	MG	A	1107	1/1	0.95	0.12	41,41,41,41	0
5	1OU	A	1106	16/16	0.97	0.11	13,21,28,30	0
4	GOL	A	1104	6/6	0.97	0.09	15,16,25,29	0
3	MG	A	1102	1/1	0.98	0.09	31,31,31,31	0
3	MG	A	1103	1/1	0.98	0.12	27,27,27,27	0
2	ZN	A	1101	1/1	1.00	0.06	13,13,13,13	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.