



Full wwPDB X-ray Structure Validation Report ⓘ

Jul 24, 2018 – 10:35 AM EDT

PDB ID : 5Y1T
Title : Crystal structure of Plasmodium falciparum aminopeptidase N in complex with (S)-2-(3-(2,3-dimethylbenzyl)ureido)-N-hydroxy-4-methylpentanamide
Authors : Marapaka, A.K.; Zhang, Y.; Addlagatta, A.
Deposited on : 2017-07-21
Resolution : 2.14 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 : rb-20031172
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) : rb-20031172

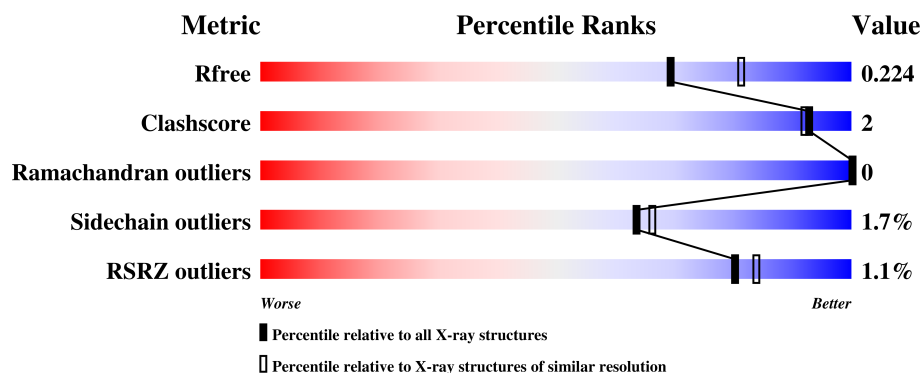
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.14 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	2128 (2.16-2.12)
Clashscore	122126	2253 (2.16-2.12)
Ramachandran outliers	120053	2223 (2.16-2.12)
Sidechain outliers	120020	2222 (2.16-2.12)
RSRZ outliers	108989	2086 (2.16-2.12)

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	914	<div> <div></div> <div>91%</div> <div>5% •</div> </div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7737 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	890	Total	C	N	O	S	0	4	0
			7326	4704	1193	1402	27			

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	172	MET	-	expression tag	UNP O96935
A	173	GLY	-	expression tag	UNP O96935
A	174	SER	-	expression tag	UNP O96935
A	175	SER	-	expression tag	UNP O96935
A	176	HIS	-	expression tag	UNP O96935
A	177	HIS	-	expression tag	UNP O96935
A	178	HIS	-	expression tag	UNP O96935
A	179	HIS	-	expression tag	UNP O96935
A	180	HIS	-	expression tag	UNP O96935
A	181	HIS	-	expression tag	UNP O96935
A	182	SER	-	expression tag	UNP O96935
A	183	SER	-	expression tag	UNP O96935
A	184	GLY	-	expression tag	UNP O96935
A	185	LEU	-	expression tag	UNP O96935
A	186	VAL	-	expression tag	UNP O96935
A	187	PRO	-	expression tag	UNP O96935
A	188	ARG	-	expression tag	UNP O96935
A	189	GLY	-	expression tag	UNP O96935
A	190	SER	-	expression tag	UNP O96935
A	191	HIS	-	expression tag	UNP O96935
A	192	MET	-	expression tag	UNP O96935
A	193	ALA	-	expression tag	UNP O96935
A	194	SER	-	expression tag	UNP O96935

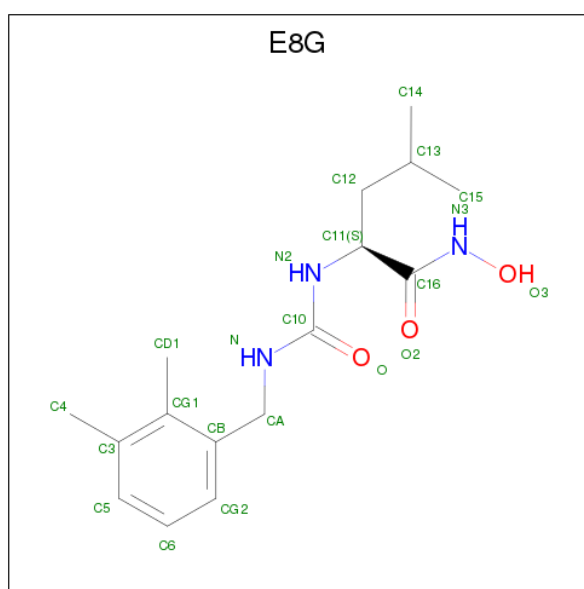
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

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 (2S)-2-[(2,3-dimethylphenyl)methylcarbamoylamino]-4-methyl-N-oxidanyl-pentanamide (three-letter code: E8G) (formula: $C_{16}H_{25}N_3O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			22	16	3	3		

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

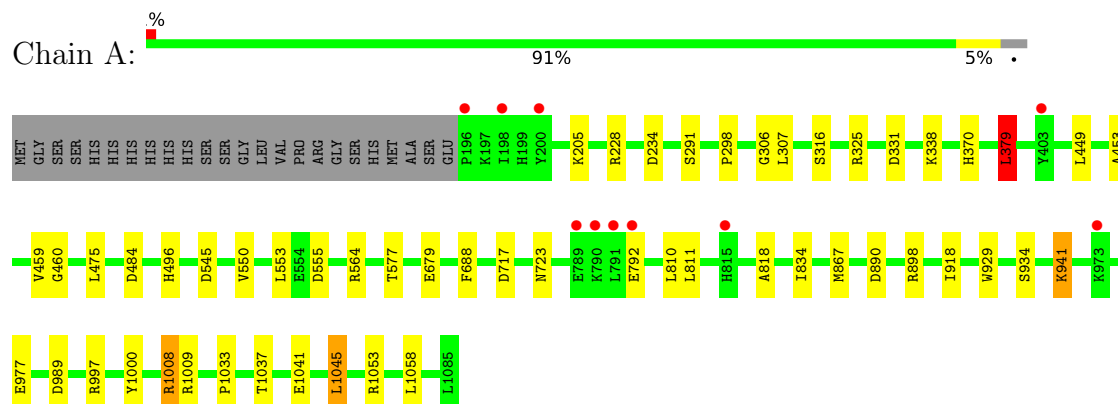
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	361	Total	O	0	0
			361	361		

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, α , β , γ	74.45Å 109.28Å 112.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 2.14 24.97 – 2.14	Depositor EDS
% Data completeness (in resolution range)	98.5 (25.00-2.14) 98.6 (24.97-2.14)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	12.62 (at 2.13Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.164 , 0.223 0.173 , 0.224	Depositor DCC
R_{free} test set	2521 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	25.7	Xtriage
Anisotropy	0.013	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 39.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.009 for -h,l,k	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7737	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.77% 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: GOL, ZN, MG, E8G

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	0.87	2/7501 (0.0%)	0.88	13/10139 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	449	LEU	N-CA	6.08	1.58	1.46
1	A	291	SER	CB-OG	5.42	1.49	1.42

All (13) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	564	ARG	NE-CZ-NH2	-6.79	116.91	120.30
1	A	228	ARG	NE-CZ-NH1	6.65	123.63	120.30
1	A	379	LEU	CA-CB-CG	6.15	129.45	115.30
1	A	325	ARG	NE-CZ-NH2	-5.61	117.50	120.30
1	A	717	ASP	CB-CG-OD1	5.59	123.33	118.30
1	A	1009	ARG	NE-CZ-NH1	5.39	123.00	120.30
1	A	484	ASP	CB-CG-OD1	5.33	123.09	118.30
1	A	234	ASP	CB-CG-OD1	5.30	123.07	118.30
1	A	1008	ARG	NE-CZ-NH1	5.23	122.92	120.30
1	A	898	ARG	NE-CZ-NH2	-5.17	117.72	120.30
1	A	929	TRP	CA-CB-CG	5.10	123.39	113.70
1	A	1053	ARG	NE-CZ-NH1	5.05	122.82	120.30
1	A	331	ASP	CB-CG-OD2	-5.04	113.77	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	7326	0	7256	20	1
2	A	1	0	0	0	0
3	A	3	0	0	0	0
4	A	22	0	0	3	0
5	A	24	0	32	4	0
6	A	361	0	0	10	0
All	All	7737	0	7288	24	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 2.

All (24) 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:679:GLU:OE2	6:A:1202:HOH:O	2.14	0.66
1:A:997:ARG:O	6:A:1203:HOH:O	2.14	0.66
5:A:1109:GOL:H11	6:A:1242:HOH:O	1.96	0.65
1:A:890[A]:ASP:OD1	6:A:1204:HOH:O	2.15	0.64
1:A:811:LEU:HB3	1:A:867:MET:SD	2.45	0.57
5:A:1107:GOL:H2	6:A:1541:HOH:O	2.05	0.56
5:A:1107:GOL:O3	5:A:1107:GOL:O1	2.21	0.56
1:A:496:HIS:CD2	5:A:1107:GOL:O3	2.59	0.56
1:A:834:ILE:HD13	1:A:1045:LEU:HD22	1.89	0.54
1:A:545:ASP:CG	6:A:1201:HOH:O	2.48	0.52
1:A:553:LEU:C	1:A:553:LEU:HD23	2.29	0.52
1:A:550:VAL:HG12	1:A:577:THR:HG21	1.93	0.51
1:A:338:LYS:HG3	1:A:379:LEU:HD23	1.96	0.48
1:A:1000:TYR:N	6:A:1203:HOH:O	2.19	0.48
1:A:997:ARG:C	6:A:1203:HOH:O	2.52	0.47
1:A:459:VAL:HG13	4:A:1105:E8G:C6	2.45	0.46
1:A:453:ALA:HB2	1:A:475:LEU:HD23	1.97	0.45
4:A:1105:E8G:C6	6:A:1493:HOH:O	2.66	0.43
1:A:918:ILE:HD11	1:A:941:LYS:HG2	1.99	0.43
1:A:460:GLY:H	4:A:1105:E8G:C15	2.32	0.43
1:A:306:GLY:O	1:A:316:SER:HA	2.19	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:818:ALA:HB3	6:A:1204:HOH:O	2.19	0.41
1:A:1033:PRO:O	1:A:1037:THR:HG23	2.20	0.41
1:A:298:PRO:HG3	1:A:307:LEU:HD23	2.02	0.40

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)
1:A:723:ASN:ND2	1:A:792:GLU:OE1[4_445]	2.16	0.04

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	892/914 (98%)	872 (98%)	20 (2%)	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	826/842 (98%)	812 (98%)	14 (2%)	63	66

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	205	LYS
1	A	370	HIS
1	A	379	LEU
1	A	555	ASP
1	A	688	PHE
1	A	810	LEU
1	A	934	SER
1	A	941	LYS
1	A	977	GLU
1	A	989	ASP
1	A	1008	ARG
1	A	1041	GLU
1	A	1045	LEU
1	A	1058	LEU

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	527	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 9 ligands modelled in this entry, 4 are monoatomic - leaving 5 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	E8G	A	1105	2	22,22,22	1.27	3 (13%)	26,29,29	2.83	14 (53%)
5	GOL	A	1106	-	5,5,5	0.62	0	5,5,5	0.57	0
5	GOL	A	1107	-	5,5,5	0.23	0	5,5,5	0.45	0
5	GOL	A	1108	-	5,5,5	0.40	0	5,5,5	0.59	0
5	GOL	A	1109	-	5,5,5	0.85	0	5,5,5	1.50	1 (20%)

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	E8G	A	1105	2	-	0/19/19/19	0/1/1/1
5	GOL	A	1106	-	-	0/4/4/4	0/0/0/0
5	GOL	A	1107	-	-	0/4/4/4	0/0/0/0
5	GOL	A	1108	-	-	0/4/4/4	0/0/0/0
5	GOL	A	1109	-	-	0/4/4/4	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1105	E8G	C11-N2	-2.76	1.40	1.45
4	A	1105	E8G	CD1-CG1	-2.19	1.47	1.51
4	A	1105	E8G	C10-N	3.06	1.41	1.35

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1105	E8G	CB-CA-N	-6.08	99.79	113.05
4	A	1105	E8G	CG2-CB-CG1	-4.67	116.04	119.47
4	A	1105	E8G	C12-C11-N2	-3.44	102.56	110.57
4	A	1105	E8G	N-C10-N2	-3.11	108.78	115.95
4	A	1105	E8G	CA-CB-CG2	-2.90	114.48	121.00
4	A	1105	E8G	C4-C3-C5	-2.84	114.69	120.30
4	A	1105	E8G	C11-C16-N3	-2.84	110.49	116.76
4	A	1105	E8G	C6-C5-C3	-2.41	117.05	121.11
4	A	1105	E8G	CD1-CG1-CB	-2.08	118.75	120.85
5	A	1109	GOL	C3-C2-C1	2.14	119.86	111.63

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1105	E8G	C4-C3-CG1	2.19	124.72	121.18
4	A	1105	E8G	O2-C16-C11	2.40	125.52	120.44
4	A	1105	E8G	C11-N2-C10	2.48	125.80	120.49
4	A	1105	E8G	O-C10-N	4.67	130.78	122.51
4	A	1105	E8G	CA-N-C10	6.51	132.80	121.52

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1105	E8G	3	0
5	A	1107	GOL	3	0
5	A	1109	GOL	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 [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, 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	890/914 (97%)	-0.18	10 (1%) 80 84	17, 25, 43, 92	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	PRO	4.0
1	A	790	LYS	3.3
1	A	200	TYR	3.1
1	A	791	LEU	2.7
1	A	792	GLU	2.6
1	A	789	GLU	2.4
1	A	403	TYR	2.2
1	A	973	LYS	2.2
1	A	815	HIS	2.1
1	A	198	ILE	2.0

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, 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(\AA^2)	Q<0.9
5	GOL	A	1106	6/6	0.84	0.23	36,42,50,56	0
4	E8G	A	1105	22/22	0.89	0.25	22,51,94,99	0
5	GOL	A	1108	6/6	0.89	0.21	47,50,57,63	0
5	GOL	A	1109	6/6	0.89	0.22	35,48,53,61	0
3	MG	A	1103	1/1	0.91	0.06	43,43,43,43	0
3	MG	A	1102	1/1	0.91	0.06	38,38,38,38	0
5	GOL	A	1107	6/6	0.93	0.28	40,43,54,55	0
3	MG	A	1104	1/1	0.97	0.13	44,44,44,44	0
2	ZN	A	1101	1/1	1.00	0.02	21,21,21,21	0

6.5 Other polymers [i](#)

There are no such residues in this entry.