



Full wwPDB X-ray Structure Validation Report ⓘ

Apr 25, 2022 – 01:03 am BST

PDB ID : 7OAH
Title : conserved hypothetical protein residues 311-335 from Candidatus Magnetomorphum sp. HK-1 fused to GCN4 adaptors, mutant beta2/A
Authors : Adlakha, J.; Albrecht, R.; Hartmann, M.D.
Deposited on : 2021-04-19
Resolution : 1.69 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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.28
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.28

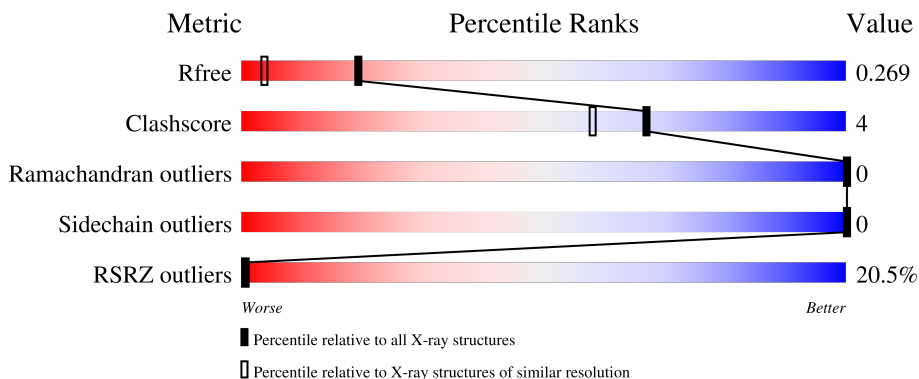
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.69 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	86	
1	B	86	
1	C	86	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4483 atoms, of which 2243 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 General control transcription factor GCN4, conserved hypothetical protein residues 311-335 from Candidatus Magnetomorum sp. HK-1 fused to GCN4 adaptors, mutant beta2/A, General control transcription factor GCN4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
1	A	81	Total	C	H	N	O	S	10	4	0
			1430	440	741	115	131	3			
1	B	81	Total	C	H	N	O	S	10	4	0
			1429	439	741	116	130	3			
1	C	82	Total	C	H	N	O	S	10	2	0
			1427	437	743	116	128	3			

There are 69 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	278	GLY	-	expression tag	UNP P03069
A	279	GLY	-	expression tag	UNP P03069
A	280	GLY	-	expression tag	UNP P03069
A	281	SER	-	expression tag	UNP P03069
A	282	GLY	-	expression tag	UNP P03069
A	286	ILE	LEU	engineered mutation	UNP P03069
A	288	MET	ASP	engineered mutation	UNP P03069
A	290	ILE	VAL	engineered mutation	UNP P03069
A	293	ILE	LEU	engineered mutation	UNP P03069
A	297	ILE	ASN	engineered mutation	UNP P03069
A	300	ILE	LEU	engineered mutation	UNP P03069
A	304	ILE	VAL	engineered mutation	UNP P03069
A	307	ILE	LEU	engineered mutation	UNP P03069
A	321	ALA	TYR	engineered mutation	UNP A0A0N0D484
A	327	ALA	TYR	engineered mutation	UNP A0A0N0D484
A	339	ILE	LEU	engineered mutation	UNP P03069
A	341	TRP	ASP	engineered mutation	UNP P03069
A	343	ILE	VAL	engineered mutation	UNP P03069
A	346	ILE	LEU	engineered mutation	UNP P03069
A	350	ILE	ASN	engineered mutation	UNP P03069
A	353	ILE	LEU	engineered mutation	UNP P03069

Continued on next page...

Continued from previous page...

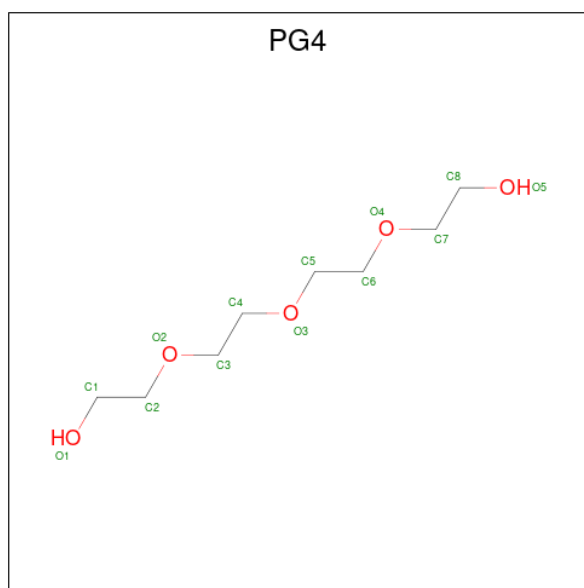
Chain	Residue	Modelled	Actual	Comment	Reference
A	357	ILE	VAL	engineered mutation	UNP P03069
A	360	ILE	LEU	engineered mutation	UNP P03069
B	278	GLY	-	expression tag	UNP P03069
B	279	GLY	-	expression tag	UNP P03069
B	280	GLY	-	expression tag	UNP P03069
B	281	SER	-	expression tag	UNP P03069
B	282	GLY	-	expression tag	UNP P03069
B	286	ILE	LEU	engineered mutation	UNP P03069
B	288	MET	ASP	engineered mutation	UNP P03069
B	290	ILE	VAL	engineered mutation	UNP P03069
B	293	ILE	LEU	engineered mutation	UNP P03069
B	297	ILE	ASN	engineered mutation	UNP P03069
B	300	ILE	LEU	engineered mutation	UNP P03069
B	304	ILE	VAL	engineered mutation	UNP P03069
B	307	ILE	LEU	engineered mutation	UNP P03069
B	321	ALA	TYR	engineered mutation	UNP A0A0N0D484
B	327	ALA	TYR	engineered mutation	UNP A0A0N0D484
B	339	ILE	LEU	engineered mutation	UNP P03069
B	341	TRP	ASP	engineered mutation	UNP P03069
B	343	ILE	VAL	engineered mutation	UNP P03069
B	346	ILE	LEU	engineered mutation	UNP P03069
B	350	ILE	ASN	engineered mutation	UNP P03069
B	353	ILE	LEU	engineered mutation	UNP P03069
B	357	ILE	VAL	engineered mutation	UNP P03069
B	360	ILE	LEU	engineered mutation	UNP P03069
C	278	GLY	-	expression tag	UNP P03069
C	279	GLY	-	expression tag	UNP P03069
C	280	GLY	-	expression tag	UNP P03069
C	281	SER	-	expression tag	UNP P03069
C	282	GLY	-	expression tag	UNP P03069
C	286	ILE	LEU	engineered mutation	UNP P03069
C	288	MET	ASP	engineered mutation	UNP P03069
C	290	ILE	VAL	engineered mutation	UNP P03069
C	293	ILE	LEU	engineered mutation	UNP P03069
C	297	ILE	ASN	engineered mutation	UNP P03069
C	300	ILE	LEU	engineered mutation	UNP P03069
C	304	ILE	VAL	engineered mutation	UNP P03069
C	307	ILE	LEU	engineered mutation	UNP P03069
C	321	ALA	TYR	engineered mutation	UNP A0A0N0D484
C	327	ALA	TYR	engineered mutation	UNP A0A0N0D484
C	339	ILE	LEU	engineered mutation	UNP P03069
C	341	TRP	ASP	engineered mutation	UNP P03069

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	343	ILE	VAL	engineered mutation	UNP P03069
C	346	ILE	LEU	engineered mutation	UNP P03069
C	350	ILE	ASN	engineered mutation	UNP P03069
C	353	ILE	LEU	engineered mutation	UNP P03069
C	357	ILE	VAL	engineered mutation	UNP P03069
C	360	ILE	LEU	engineered mutation	UNP P03069

- Molecule 2 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: C₈H₁₈O₅).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
2	B	1	31	8	18	5	1	0

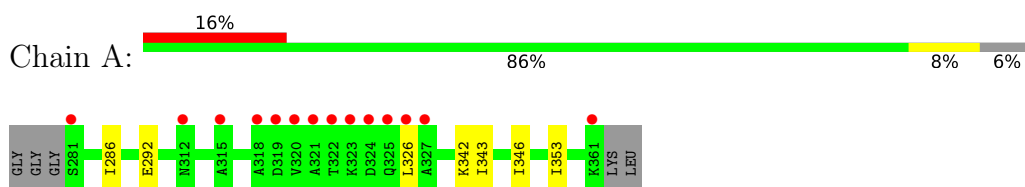
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	56	Total	O	0	0
			56	56		
3	B	54	Total	O	0	0
			54	54		
3	C	56	Total	O	0	0
			56	56		

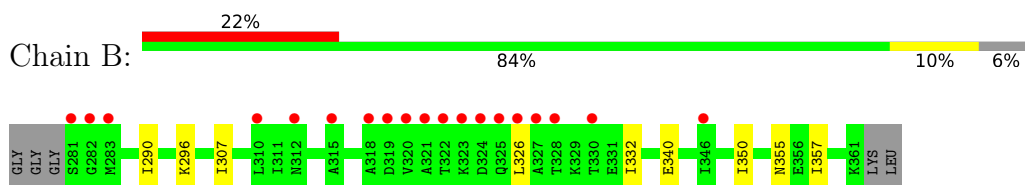
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

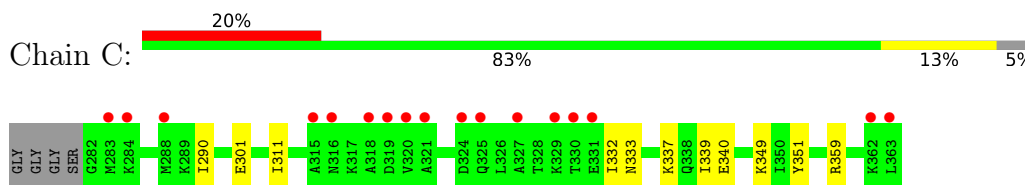
- Molecule 1: General control transcription factor GCN4, conserved hypothetical protein residues 311-335 from *Candidatus Magnetomorum* sp. HK-1 fused to GCN4 adaptors, mutant beta2/A, General control transcription factor GCN4



- Molecule 1: General control transcription factor GCN4, conserved hypothetical protein residues 311-335 from *Candidatus Magnetomorum* sp. HK-1 fused to GCN4 adaptors, mutant beta2/A, General control transcription factor GCN4



- Molecule 1: General control transcription factor GCN4, conserved hypothetical protein residues 311-335 from *Candidatus Magnetomorum* sp. HK-1 fused to GCN4 adaptors, mutant beta2/A, General control transcription factor GCN4



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	26.13Å 38.77Å 132.28Å 90.00° 94.85° 90.00°	Depositor
Resolution (Å)	37.19 – 1.69 37.19 – 1.69	Depositor EDS
% Data completeness (in resolution range)	97.9 (37.19-1.69) 97.9 (37.19-1.69)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.22 (at 1.70Å)	Xtrriage
Refinement program	REFMAC 5.8.0049 2013/06/30	Depositor
R, R_{free}	0.220 , 0.268 0.230 , 0.269	Depositor DCC
R_{free} test set	1456 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	26.2	Xtrriage
Anisotropy	0.102	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.009 for h,-k,-h-l	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	4483	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.41% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: PG4

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	0.38	0/703	0.50	0/937
1	B	0.36	0/702	0.50	0/936
1	C	0.37	0/695	0.55	0/924
All	All	0.37	0/2100	0.52	0/2797

There are no bond length outliers.

There are no bond angle outliers.

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	689	741	740	10	1
1	B	688	741	740	10	3
1	C	684	743	742	9	4
2	B	13	18	18	0	0
3	A	56	0	0	3	0
3	B	54	0	0	0	0
3	C	56	0	0	1	0
All	All	2240	2243	2240	17	4

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.

All (17) 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:292[B]:GLU:HG2	3:A:402:HOH:O	1.77	0.84
1:A:292[B]:GLU:OE2	3:A:402:HOH:O	2.00	0.79
1:A:343:ILE:HD11	1:C:339:ILE:HG23	1.78	0.64
1:B:290[B]:ILE:HD11	1:C:290:ILE:HD11	1.83	0.60
1:B:296:LYS:CE	1:C:301:GLU:OE2	2.53	0.56
1:B:296:LYS:HE2	1:C:301:GLU:OE2	2.07	0.55
1:A:292[B]:GLU:CG	3:A:402:HOH:O	2.44	0.54
1:A:343:ILE:CD1	1:C:339:ILE:HG23	2.42	0.49
1:A:326:LEU:HD22	1:B:326:LEU:HB3	1.95	0.48
1:B:307:ILE:HG23	1:C:311:ILE:HD11	1.97	0.47
1:A:346:ILE:HG23	1:B:350:ILE:HD11	1.97	0.47
1:A:353:ILE:HG23	1:B:357:ILE:HD11	1.97	0.47
1:B:332:ILE:HD11	1:C:332:ILE:HD13	1.96	0.47
1:A:286:ILE:HG23	1:B:290[A]:ILE:HD11	1.99	0.44
1:C:349:LYS:NZ	3:C:404:HOH:O	2.53	0.42
1:A:353:ILE:HG23	1:B:357:ILE:CD1	2.50	0.41
1:C:333:ASN:O	1:C:337:LYS:HD2	2.21	0.41

All (4) 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:342:LYS:HZ3	1:C:351:TYR:HH[2_747]	1.15	0.45
1:B:355[B]:ASN:ND2	1:C:340:GLU:OE2[2_747]	2.10	0.10
1:B:355[B]:ASN:HD21	1:C:340:GLU:OE2[2_747]	1.56	0.04
1:B:340:GLU:OE1	1:C:359:ARG:HH11[2_747]	1.60	0.00

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	83/86 (96%)	82 (99%)	1 (1%)	0	100	100
1	B	83/86 (96%)	83 (100%)	0	0	100	100
1	C	82/86 (95%)	82 (100%)	0	0	100	100
All	All	248/258 (96%)	247 (100%)	1 (0%)	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	78/76 (103%)	78 (100%)	0	100	100
1	B	78/76 (103%)	78 (100%)	0	100	100
1	C	77/76 (101%)	77 (100%)	0	100	100
All	All	233/228 (102%)	233 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	312	ASN
1	B	312	ASN
1	B	335	GLN
1	B	352	HIS
1	C	312	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

1 ligand 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
2	PG4	B	401	-	12,12,12	0.36	0	11,11,11	0.42	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	PG4	B	401	-	-	0/10/10/10	-

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.

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

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	81/86 (94%)	1.43	14 (17%) 1 1	16, 28, 53, 84	0
1	B	81/86 (94%)	1.85	19 (23%) 0 0	15, 31, 59, 97	0
1	C	82/86 (95%)	1.49	17 (20%) 1 0	15, 33, 57, 79	0
All	All	244/258 (94%)	1.59	50 (20%) 1 0	15, 31, 59, 97	0

All (50) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	324	ASP	10.7
1	B	323	LYS	10.4
1	B	326	LEU	10.2
1	A	322	THR	9.2
1	C	363	LEU	9.0
1	A	318	ALA	8.1
1	A	324	ASP	7.7
1	B	321	ALA	7.0
1	B	325	GLN	6.7
1	A	323	LYS	6.4
1	B	322	THR	5.5
1	C	324	ASP	5.4
1	C	318	ALA	5.1
1	A	327	ALA	4.8
1	A	281	SER	4.7
1	B	320	VAL	4.7
1	B	327	ALA	4.7
1	C	321	ALA	4.7
1	A	321	ALA	4.7
1	B	318	ALA	4.4
1	B	319	ASP	4.4
1	A	326	LEU	3.8
1	C	362	LYS	3.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	320	VAL	3.6
1	A	319	ASP	3.5
1	C	315	ALA	3.3
1	B	312	ASN	3.3
1	B	328	THR	3.2
1	C	330	THR	3.2
1	B	282	GLY	3.1
1	A	315	ALA	3.1
1	B	281	SER	3.1
1	C	283	MET	3.0
1	B	330	THR	3.0
1	A	320	VAL	2.9
1	C	288	MET	2.8
1	B	315	ALA	2.8
1	C	319	ASP	2.7
1	A	312	ASN	2.7
1	A	325	GLN	2.7
1	B	283	MET	2.4
1	C	329	LYS	2.3
1	B	310	LEU	2.3
1	C	327	ALA	2.3
1	C	284	LYS	2.3
1	C	331	GLU	2.2
1	A	361	LYS	2.1
1	B	346	ILE	2.1
1	C	325	GLN	2.1
1	C	316	ASN	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 monosaccharides 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
2	PG4	B	401	13/13	0.94	0.14	20,24,30,49	1

6.5 Other polymers [i](#)

There are no such residues in this entry.