



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 24, 2024 – 09:00 AM EDT

PDB ID : 6XU0
Title : Archaeoglobus fulgidus Argonaute protein with DNA oligoduplex 5'-pATCGTGGCCACGAT
Authors : Grazulis, S.; Zaremba, M.
Deposited on : 2020-01-17
Resolution : 1.90 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 2.37.1
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.37.1

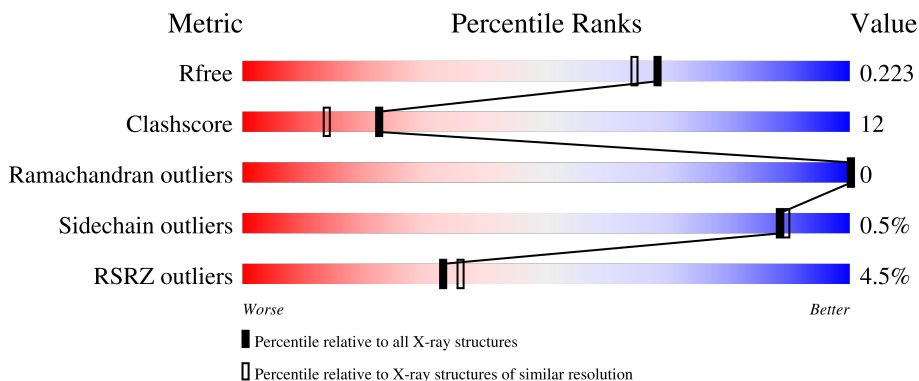
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (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 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	441	
1	B	441	
2	P	14	
2	Q	14	
2	R	14	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	S	14	

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
6	GOL	A	505	-	-	X	-

2 Entry composition i

There are 9 unique types of molecules in this entry. The entry contains 7662 atoms, of which 6 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 Piwi protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	402	3304	2136	547	612	9	0	6	0
1	B	401	3307	2134	6	547	610	10	0	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP A0A101DYI0
A	-12	GLY	-	expression tag	UNP A0A101DYI0
A	-11	SER	-	expression tag	UNP A0A101DYI0
A	-10	SER	-	expression tag	UNP A0A101DYI0
A	-9	HIS	-	expression tag	UNP A0A101DYI0
A	-8	HIS	-	expression tag	UNP A0A101DYI0
A	-7	HIS	-	expression tag	UNP A0A101DYI0
A	-6	HIS	-	expression tag	UNP A0A101DYI0
A	-5	HIS	-	expression tag	UNP A0A101DYI0
A	-4	HIS	-	expression tag	UNP A0A101DYI0
A	-3	SER	-	expression tag	UNP A0A101DYI0
A	-2	GLN	-	expression tag	UNP A0A101DYI0
A	-1	ASP	-	expression tag	UNP A0A101DYI0
A	0	PRO	-	expression tag	UNP A0A101DYI0
B	-13	MET	-	initiating methionine	UNP A0A101DYI0
B	-12	GLY	-	expression tag	UNP A0A101DYI0
B	-11	SER	-	expression tag	UNP A0A101DYI0
B	-10	SER	-	expression tag	UNP A0A101DYI0
B	-9	HIS	-	expression tag	UNP A0A101DYI0
B	-8	HIS	-	expression tag	UNP A0A101DYI0
B	-7	HIS	-	expression tag	UNP A0A101DYI0
B	-6	HIS	-	expression tag	UNP A0A101DYI0
B	-5	HIS	-	expression tag	UNP A0A101DYI0
B	-4	HIS	-	expression tag	UNP A0A101DYI0
B	-3	SER	-	expression tag	UNP A0A101DYI0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	GLN	-	expression tag	UNP A0A101DYI0
B	-1	ASP	-	expression tag	UNP A0A101DYI0
B	0	PRO	-	expression tag	UNP A0A101DYI0

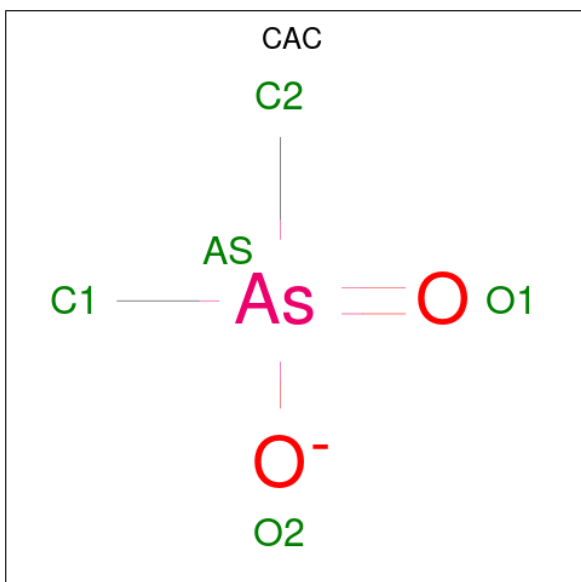
- Molecule 2 is a DNA chain called 5'-D(*AP*TP*CP*GP*TP*GP*GP*CP*CP*AP*CP*GP*AP*TP)-3'.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
2	R	7	Total	C	N	O	P	0	0	0
			147	69	27	44	7			
2	S	6	Total	C	N	O	P	0	0	0
			122	58	23	35	6			
2	P	7	Total	C	N	O	P	0	0	0
			147	69	27	44	7			
2	Q	7	Total	C	N	O	P	0	0	0
			141	67	26	41	7			

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

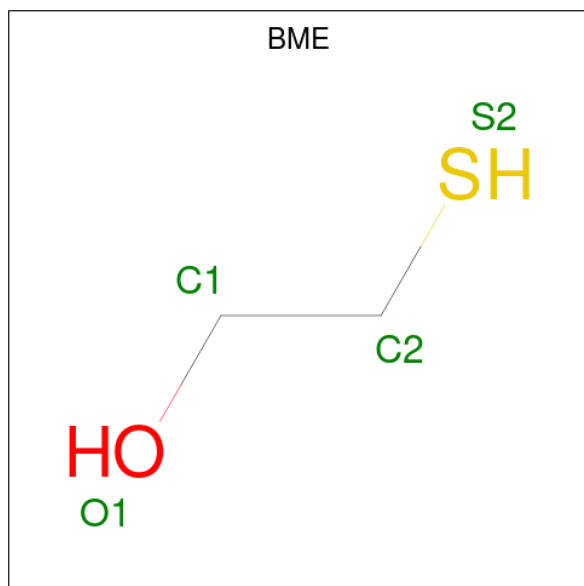
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Mg	0	0
			1	1		
3	B	1	Total	Mg	0	0
			1	1		

- Molecule 4 is CACODYLATE ION (three-letter code: CAC) (formula: C₂H₆AsO₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	As	C	O		
4	A	1	5	1	2	2	0	0
4	A	1	5	1	2	2	0	0
4	B	1	5	1	2	2	0	0

- Molecule 5 is BETA-MERCAPTOETHANOL (three-letter code: BME) (formula: C₂H₆OS).



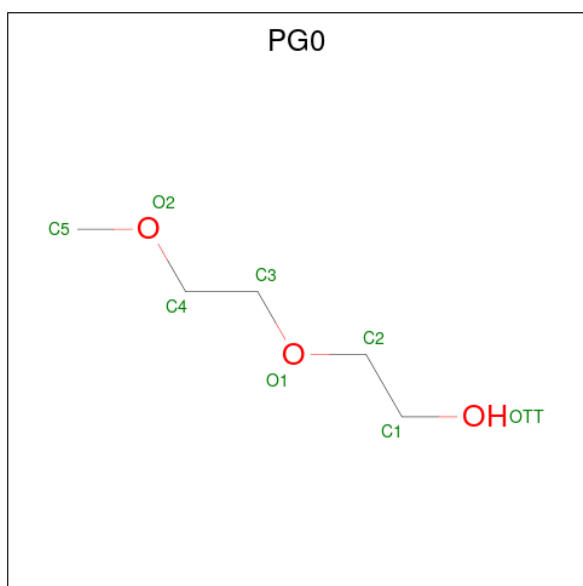
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	S		
5	A	1	4	2	1	1	0	0
5	B	1	4	2	1	1	0	0

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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

- Molecule 7 is 2-(2-METHOXYETHOXY)ETHANOL (three-letter code: PG0) (formula: $C_5H_{12}O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	B	1	Total	C	O	0	0
			8	5	3		

- Molecule 8 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	B	1	Total K 1 1	0	0

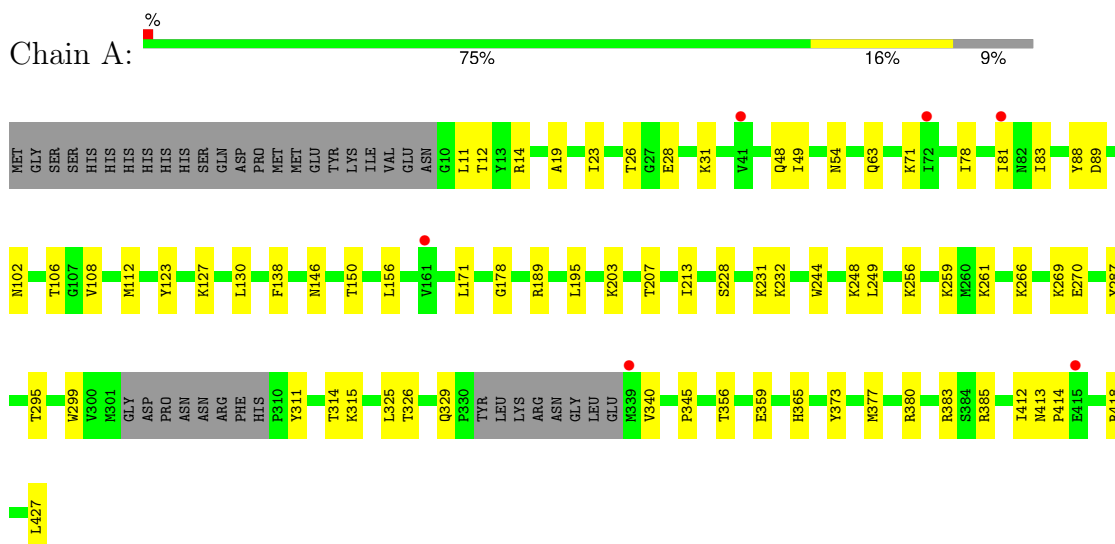
- Molecule 9 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	A	257	Total O 257 257	0	0
9	R	10	Total O 10 10	0	0
9	S	7	Total O 7 7	0	0
9	B	168	Total O 168 168	0	0
9	P	5	Total O 5 5	0	0
9	Q	1	Total O 1 1	0	0

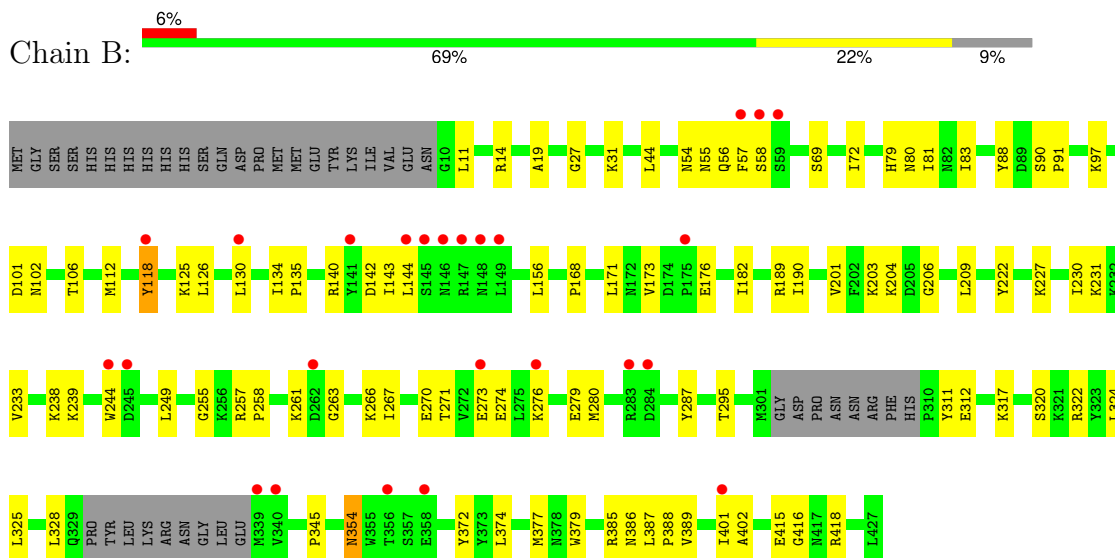
3 Residue-property plots

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: Piwi protein



- Molecule 1: Piwi protein

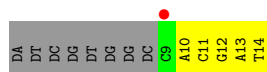


- Molecule 2: 5'-D(*AP*TP*CP*GP*TP*GP*GP*CP*CP*AP*CP*GP*AP*TP)-3'





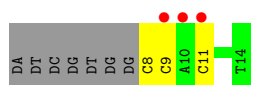
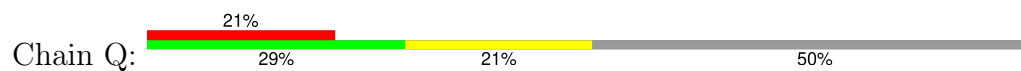
● Molecule 2: 5'-D(*AP*TP*CP*GP*TP*GP*GP*CP*CP*AP*CP*GP*AP*TP)-3'



● Molecule 2: 5'-D(*AP*TP*CP*GP*TP*GP*GP*CP*CP*AP*CP*GP*AP*TP)-3'



● Molecule 2: 5'-D(*AP*TP*CP*GP*TP*GP*GP*CP*CP*AP*CP*GP*AP*TP)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	51.91Å 61.20Å 103.09Å 98.32° 104.96° 100.62°	Depositor
Resolution (Å)	46.44 – 1.90 48.78 – 1.90	Depositor EDS
% Data completeness (in resolution range)	90.5 (46.44-1.90) 91.4 (48.78-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.74 (at 1.90Å)	Xtrriage
Refinement program	PHENIX 1.13_2998	Depositor
R, R_{free}	0.180 , 0.225 0.178 , 0.223	Depositor DCC
R_{free} test set	8488 reflections (9.98%)	wwPDB-VP
Wilson B-factor (Å ²)	35.3	Xtrriage
Anisotropy	0.198	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 54.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	7662	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.77% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: CAC, GOL, K, PG0, MG, BME

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	0.65	0/3385	0.68	0/4593
1	B	0.50	0/3382	0.61	0/4587
2	P	1.06	1/164 (0.6%)	0.91	0/250
2	Q	0.65	0/157	0.90	0/239
2	R	1.40	1/164 (0.6%)	1.03	0/250
2	S	0.80	0/136	0.93	0/207
All	All	0.63	2/7388 (0.0%)	0.68	0/10126

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	R	1	DA	OP3-P	-10.80	1.48	1.61
2	P	1	DA	OP3-P	-9.81	1.49	1.61

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	3304	0	3320	58	0
1	B	3301	6	3307	84	0
2	P	147	0	80	9	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	Q	141	0	79	5	0
2	R	147	0	80	9	0
2	S	122	0	68	9	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	10	0	0	0	0
4	B	5	0	0	1	0
5	A	4	0	6	1	0
5	B	4	0	6	2	0
6	A	6	0	8	4	0
6	B	6	0	8	0	0
7	B	8	0	11	0	0
8	B	1	0	0	0	0
9	A	257	0	0	8	0
9	B	168	0	0	6	0
9	P	5	0	0	1	0
9	Q	1	0	0	0	0
9	R	10	0	0	2	0
9	S	7	0	0	0	0
All	All	7656	6	6973	162	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 12.

The worst 5 of 162 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)
2:R:4:DG:H2'	2:R:5:DT:H5'	1.58	0.85
1:A:231:LYS:HD3	6:A:505:GOL:H31	1.59	0.84
1:A:11[A]:LEU:HD23	1:A:12:THR:N	1.95	0.82
1:A:11[A]:LEU:HD22	1:A:23:ILE:HG21	1.65	0.79
1:B:266:LYS:O	1:B:270:GLU:HG3	1.84	0.78

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	402/441 (91%)	390 (97%)	12 (3%)	0	100	100
1	B	401/441 (91%)	388 (97%)	13 (3%)	0	100	100
All	All	803/882 (91%)	778 (97%)	25 (3%)	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	376/407 (92%)	374 (100%)	2 (0%)	88	89
1	B	373/407 (92%)	371 (100%)	2 (0%)	88	89
All	All	749/814 (92%)	745 (100%)	4 (0%)	88	89

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

Mol	Chain	Res	Type
1	A	89	ASP
1	A	413	ASN
1	B	118	TYR
1	B	354	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	63	GLN
1	A	146	ASN
1	B	291	HIS

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](#)

Of 11 ligands modelled in this entry, 3 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	$\# Z > 2$
6	GOL	B	506	-	5,5,5	1.42	1 (20%)	5,5,5	0.88	0
5	BME	A	504	-	3,3,3	0.52	0	2,2,2	0.47	0
4	CAC	A	503	-	2,4,4	0.78	0	2,6,6	0.38	0
4	CAC	A	502	-	2,4,4	0.87	0	2,6,6	0.17	0
6	GOL	A	505	-	5,5,5	0.26	0	5,5,5	0.31	0
5	BME	B	504	-	3,3,3	0.38	0	2,2,2	0.35	0
7	PG0	B	501	-	7,7,7	0.96	0	6,6,6	0.53	0
4	CAC	B	503	-	2,4,4	0.85	0	2,6,6	0.25	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
6	GOL	B	506	-	-	0/4/4/4	-
5	BME	A	504	-	-	0/1/1/1	-
6	GOL	A	505	-	-	4/4/4/4	-
5	BME	B	504	-	-	0/1/1/1	-
7	PG0	B	501	-	-	4/5/5/5	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	B	506	GOL	C1-C2	2.12	1.59	1.51

There are no bond angle outliers.

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	505	GOL	C1-C2-C3-O3
7	B	501	PG0	O1-C3-C4-O2
6	A	505	GOL	O2-C2-C3-O3
7	B	501	PG0	OTT-C1-C2-O1
6	A	505	GOL	O1-C1-C2-C3

There are no ring outliers.

4 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	504	BME	1	0
6	A	505	GOL	4	0
5	B	504	BME	2	0
4	B	503	CAC	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues

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	402/441 (91%)	0.03	6 (1%) 73 76	22, 35, 59, 80	0
1	B	401/441 (90%)	0.28	25 (6%) 20 23	29, 44, 74, 110	0
2	P	7/14 (50%)	0.41	1 (14%) 2 2	46, 63, 122, 139	0
2	Q	7/14 (50%)	1.90	3 (42%) 0 0	60, 99, 141, 145	0
2	R	7/14 (50%)	0.24	1 (14%) 2 2	32, 48, 104, 118	0
2	S	6/14 (42%)	0.31	1 (16%) 1 1	40, 66, 105, 109	0
All	All	830/938 (88%)	0.17	37 (4%) 33 36	22, 40, 74, 145	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	57	PHE	7.2
1	B	146	ASN	4.9
1	B	283	ARG	4.9
1	B	273	GLU	4.2
1	B	145	SER	4.0

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
7	PG0	B	501	8/8	0.78	0.23	63,76,91,95	0
4	CAC	A	502	5/5	0.81	0.44	110,111,157,159	0
4	CAC	A	503	5/5	0.84	0.27	70,72,99,160	0
6	GOL	A	505	6/6	0.88	0.28	40,45,59,70	0
6	GOL	B	506	6/6	0.90	0.14	52,59,74,74	0
5	BME	B	504	4/4	0.94	0.14	48,66,69,74	0
5	BME	A	504	4/4	0.95	0.12	34,40,48,57	0
4	CAC	B	503	5/5	0.96	0.12	58,62,91,138	0
3	MG	B	502	1/1	0.96	0.05	39,39,39,39	0
8	K	B	505	1/1	0.96	0.37	83,83,83,83	0
3	MG	A	501	1/1	0.99	0.06	27,27,27,27	0

6.5 Other polymers [i](#)

There are no such residues in this entry.