wwPDB X-ray Structure Validation Summary Report (i)

Sep 16, 2023 – 04:31 PM EDT

PDB ID	:	4PF1
Title	:	Crystal structure of aminopeptidase from marine sediment archaeon Thaumarchaeota archaeon
Authors	:	Michalska, K.; Chhor, G.; Fayman, K.; Endres, M.; Jedrzejczak, R.; Babnigg, G.; Steen, A.; Lloyd, K.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2014-04-25
Resolution	:	2.10 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 (i) 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 (i)) were used in the production of this report:

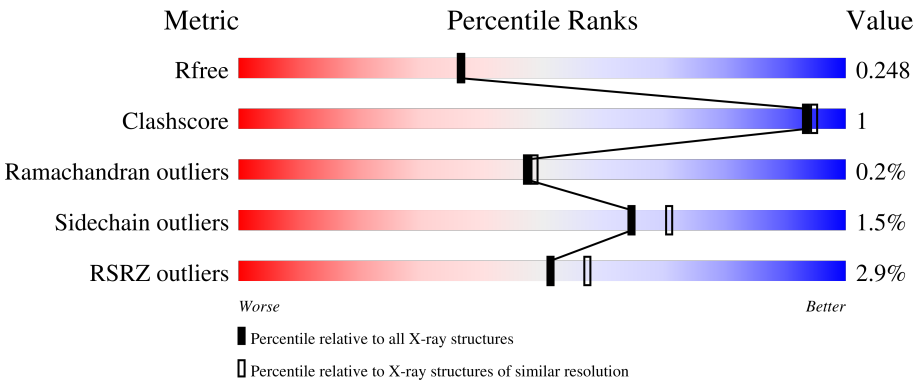
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 <=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	626	3% 92% 6% •
1	B	626	3% 92% 5% •
1	C	626	4% 93% 5% •
1	D	626	3% 92% 5% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 20920 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.

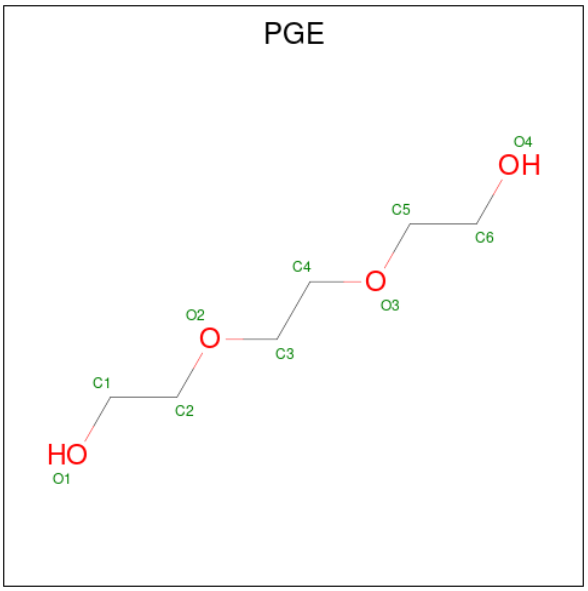
- Molecule 1 is a protein called Peptidase S15/CocE/NonD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	611	4938	3154	847	915	22	0	3	0
1	B	610	4909	3138	838	910	23	0	1	0
1	C	610	4901	3133	837	909	22	0	0	0
1	D	611	4918	3143	842	911	22	0	1	0

There are 12 discrepancies between the modelled and reference sequences:

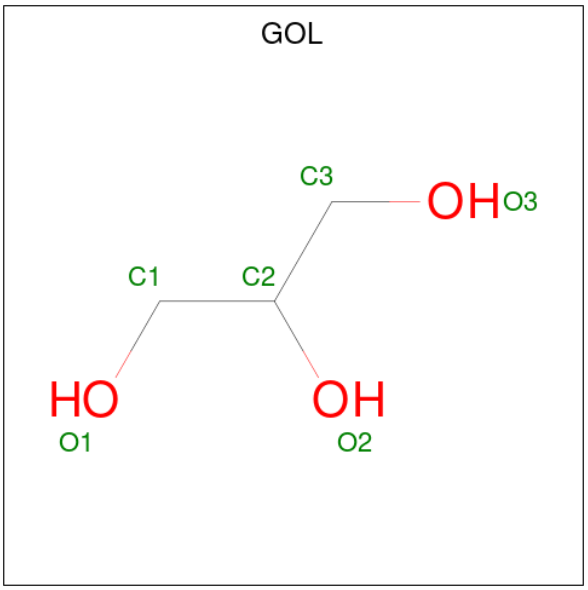
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP M7TVE7
A	-1	ASN	-	expression tag	UNP M7TVE7
A	0	ALA	-	expression tag	UNP M7TVE7
B	-2	SER	-	expression tag	UNP M7TVE7
B	-1	ASN	-	expression tag	UNP M7TVE7
B	0	ALA	-	expression tag	UNP M7TVE7
C	-2	SER	-	expression tag	UNP M7TVE7
C	-1	ASN	-	expression tag	UNP M7TVE7
C	0	ALA	-	expression tag	UNP M7TVE7
D	-2	SER	-	expression tag	UNP M7TVE7
D	-1	ASN	-	expression tag	UNP M7TVE7
D	0	ALA	-	expression tag	UNP M7TVE7

- Molecule 2 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	10	6	4	0	0
			Total	C	O		
2	B	1	10	6	4	0	0
			Total	C	O		
2	D	1	10	6	4	0	0
			Total	C	O		
2	D	1	10	6	4	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 4 is water.

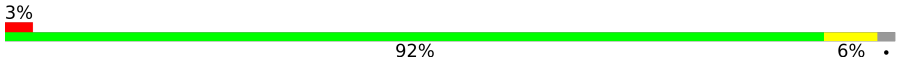
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	300	300	300	0	0
			Total	O		
4	B	287	287	287	0	0
			Total	O		
4	C	247	247	247	0	0
			Total	O		
4	D	290	290	290	0	0

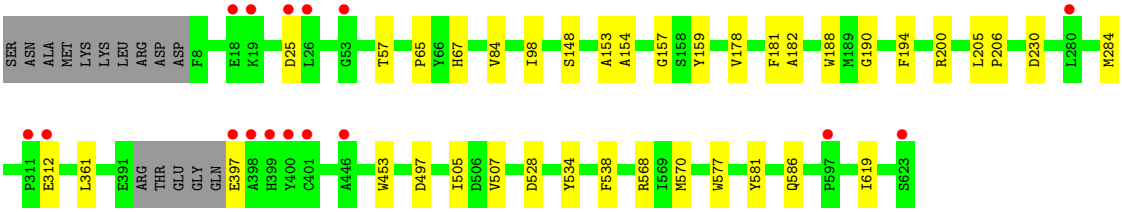
3 Residue-property plots (i)

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: Peptidase S15/CocE/NonD

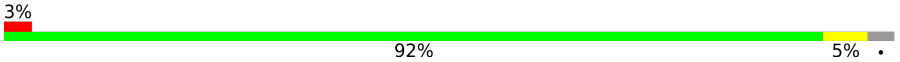
Chain A:

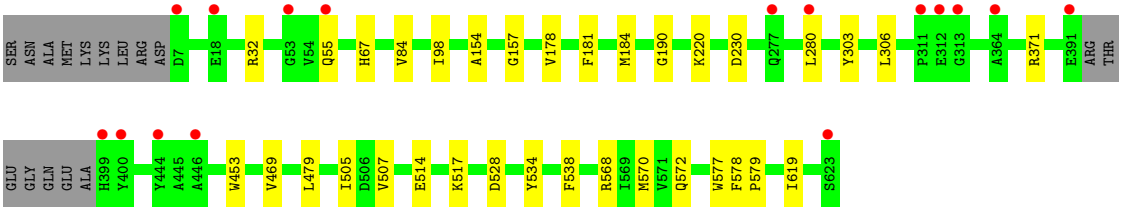




• Molecule 1: Peptidase S15/CocE/NonD

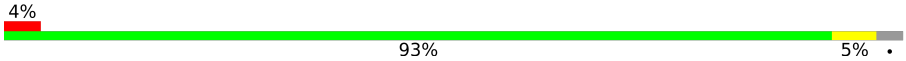
Chain B:

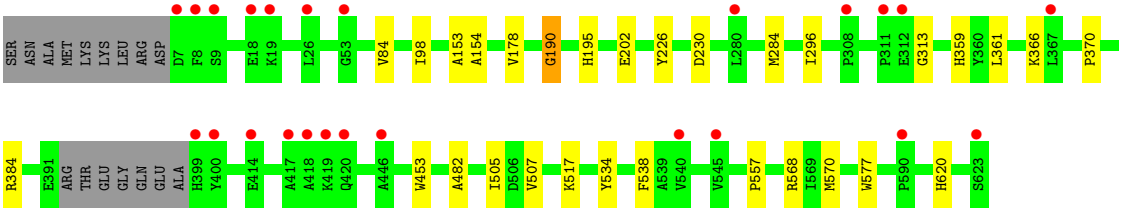




• Molecule 1: Peptidase S15/CocE/NonD

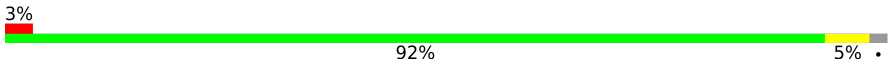
Chain C:

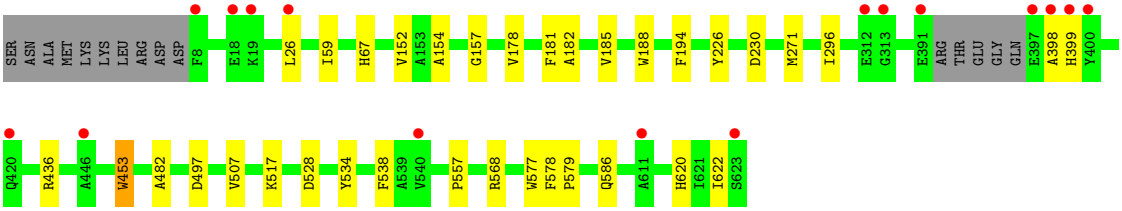




• Molecule 1: Peptidase S15/CocE/NonD

Chain D:




SER
ASN
ALA
MET
LYS
LYS
LEU
ARG
ASP
ASP
F8
E18
K19
L26
I59
H67
V152
A153
A154
G157
V178
F181
A182
V185
W188
F194
Y226
D230
M271
I296
E312
G313
E391
ARG
THR
GLU
GLY
GLN
E397
A398
H399
Y400
Q420
R436
A446
W453
A482
D497
V507
K517
D528
Y534
F538
A539
V540
P557
R568
W577
F578
P579
Q586
A611
H620
I621
I622
S623

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	118.43Å 108.14Å 120.38Å 90.00° 95.08° 90.00°	Depositor
Resolution (Å)	29.69 – 2.10 29.69 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.69-2.10) 99.6 (29.69-2.10)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.10 (at 2.10Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.205 , 0.229 0.222 , 0.248	Depositor DCC
R_{free} test set	2166 reflections (1.23%)	wwPDB-VP
Wilson B-factor (Å^2)	20.5	Xtriage
Anisotropy	0.196	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 49.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.020 for l,-k,h	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	20920	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.66% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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: GOL, PGE

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	0.43	0/5091	0.47	0/6935
1	B	0.42	0/5062	0.46	0/6897
1	C	0.42	0/5054	0.46	0/6887
1	D	0.42	0/5071	0.46	0/6909
All	All	0.42	0/20278	0.46	0/27628

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	4938	0	4757	15	0
1	B	4909	0	4729	12	0
1	C	4901	0	4721	14	0
1	D	4918	0	4740	15	0
2	A	10	0	14	0	0
2	B	10	0	14	0	0
2	D	20	0	28	0	0
3	A	30	0	40	0	0
3	B	18	0	24	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	18	0	24	0	0
3	D	24	0	32	1	0
4	A	300	0	0	0	0
4	B	287	0	0	0	0
4	C	247	0	0	0	0
4	D	290	0	0	0	0
All	All	20920	0	19123	54	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 1.

The worst 5 of 54 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:B:220:LYS:HD3	3:D:705:GOL:H12	1.85	0.58
1:C:84:VAL:HG11	1:C:98:ILE:HD13	1.90	0.53
1:D:534:TYR:HB3	1:D:538:PHE:HA	1.90	0.53
1:C:154:ALA:HB3	1:C:178:VAL:HG12	1.91	0.53
1:D:497:ASP:O	1:D:586:GLN:HA	2.10	0.52

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	610/626 (97%)	588 (96%)	20 (3%)	2 (0%)	41	41
1	B	607/626 (97%)	583 (96%)	23 (4%)	1 (0%)	47	49
1	C	606/626 (97%)	578 (95%)	26 (4%)	2 (0%)	41	41
1	D	608/626 (97%)	582 (96%)	25 (4%)	1 (0%)	47	49
All	All	2431/2504 (97%)	2331 (96%)	94 (4%)	6 (0%)	47	49

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	190	GLY
1	C	313	GLY
1	A	182	ALA
1	D	182	ALA
1	A	190	GLY

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	526/536 (98%)	517 (98%)	9 (2%)	60	67
1	B	524/536 (98%)	513 (98%)	11 (2%)	53	59
1	C	523/536 (98%)	519 (99%)	4 (1%)	81	86
1	D	524/536 (98%)	517 (99%)	7 (1%)	69	75
All	All	2097/2144 (98%)	2066 (98%)	31 (2%)	65	71

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	479	LEU
1	D	399	HIS
1	B	528	ASP
1	D	528	ASP
1	D	26	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	437	GLN

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)

19 ligands are 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	PGE	B	701	-	9,9,9	0.11	0	8,8,8	0.06	0
3	GOL	B	702	-	5,5,5	0.07	0	5,5,5	0.16	0
2	PGE	D	706	-	9,9,9	0.13	0	8,8,8	0.05	0
3	GOL	A	702	-	5,5,5	0.05	0	5,5,5	0.18	0
3	GOL	A	703	-	5,5,5	0.04	0	5,5,5	0.13	0
3	GOL	A	706	-	5,5,5	0.04	0	5,5,5	0.13	0
3	GOL	C	701	-	5,5,5	0.04	0	5,5,5	0.11	0
3	GOL	D	703	-	5,5,5	0.05	0	5,5,5	0.15	0
3	GOL	B	704	-	5,5,5	0.06	0	5,5,5	0.33	0
3	GOL	A	704	-	5,5,5	0.06	0	5,5,5	0.23	0
3	GOL	A	705	-	5,5,5	0.08	0	5,5,5	0.33	0
3	GOL	C	702	-	5,5,5	0.06	0	5,5,5	0.10	0
3	GOL	D	704	-	5,5,5	0.06	0	5,5,5	0.30	0
3	GOL	D	702	-	5,5,5	0.04	0	5,5,5	0.39	0
2	PGE	D	701	-	9,9,9	0.12	0	8,8,8	0.06	0
3	GOL	D	705	-	5,5,5	0.05	0	5,5,5	0.18	0
2	PGE	A	701	-	9,9,9	0.11	0	8,8,8	0.03	0
3	GOL	C	703	-	5,5,5	0.04	0	5,5,5	0.20	0
3	GOL	B	703	-	5,5,5	0.04	0	5,5,5	0.18	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	PGE	B	701	-	-	5/7/7/7	-
3	GOL	B	702	-	-	0/4/4/4	-
2	PGE	D	706	-	-	3/7/7/7	-
3	GOL	A	702	-	-	0/4/4/4	-
3	GOL	A	703	-	-	1/4/4/4	-
3	GOL	A	706	-	-	0/4/4/4	-
3	GOL	C	701	-	-	0/4/4/4	-
3	GOL	D	703	-	-	0/4/4/4	-
3	GOL	B	704	-	-	0/4/4/4	-
3	GOL	A	704	-	-	2/4/4/4	-
3	GOL	A	705	-	-	2/4/4/4	-
3	GOL	C	702	-	-	0/4/4/4	-
3	GOL	D	704	-	-	0/4/4/4	-
3	GOL	D	702	-	-	2/4/4/4	-
2	PGE	D	701	-	-	3/7/7/7	-
3	GOL	D	705	-	-	0/4/4/4	-
2	PGE	A	701	-	-	5/7/7/7	-
3	GOL	C	703	-	-	2/4/4/4	-
3	GOL	B	703	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 25 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	705	GOL	O1-C1-C2-C3
3	D	702	GOL	O1-C1-C2-C3
2	D	706	PGE	O2-C3-C4-O3
2	A	701	PGE	O2-C3-C4-O3
2	B	701	PGE	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	705	GOL	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	703	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	611/626 (97%)	0.07	16 (2%)	56	61	9, 20, 36, 69	0
1	B	610/626 (97%)	0.12	16 (2%)	56	61	11, 22, 38, 61	0
1	C	610/626 (97%)	0.26	24 (3%)	39	45	13, 25, 44, 69	0
1	D	611/626 (97%)	0.10	16 (2%)	56	61	13, 22, 39, 68	0
All	All	2442/2504 (97%)	0.14	72 (2%)	51	57	9, 22, 40, 69	0

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	398	ALA	5.7
1	C	418	ALA	5.3
1	B	446	ALA	4.8
1	C	311	PRO	4.4
1	D	397	GLU	4.3

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
3	GOL	D	704	6/6	0.78	0.25	31,33,33,34	0
3	GOL	C	703	6/6	0.79	0.19	40,41,42,42	0
2	PGE	A	701	10/10	0.82	0.16	44,46,49,49	0
3	GOL	B	704	6/6	0.84	0.22	29,30,30,31	0
2	PGE	D	706	10/10	0.85	0.20	41,43,46,46	0
3	GOL	A	705	6/6	0.86	0.20	26,28,29,29	0
3	GOL	C	701	6/6	0.86	0.22	43,43,43,44	0
3	GOL	D	705	6/6	0.86	0.28	39,40,41,41	0
3	GOL	C	702	6/6	0.87	0.14	29,32,32,32	0
3	GOL	A	704	6/6	0.87	0.19	31,34,36,38	0
3	GOL	D	703	6/6	0.88	0.19	40,42,43,43	0
3	GOL	B	702	6/6	0.90	0.16	22,28,31,32	0
3	GOL	A	702	6/6	0.91	0.28	28,29,30,31	0
3	GOL	A	703	6/6	0.91	0.18	29,31,32,33	0
2	PGE	D	701	10/10	0.91	0.09	31,33,36,36	0
3	GOL	D	702	6/6	0.92	0.13	26,28,30,30	0
3	GOL	A	706	6/6	0.92	0.26	46,46,47,48	0
2	PGE	B	701	10/10	0.93	0.12	28,29,30,31	0
3	GOL	B	703	6/6	0.93	0.17	31,32,33,34	0

6.5 Other polymers (i)

There are no such residues in this entry.