Full wwPDB X-ray Structure Validation Report ⓘ

Jan 24, 2023 – 01:16 PM EST

PDB ID	:	3LJY
Title	:	Crystal structure of putative adhesin (YP_001304413.1) from Parabacteroides distasonis ATCC 8503 at 2.41 A resolution
Authors	:	Joint Center for Structural Genomics (JCSG)
Deposited on	:	2010-01-26
Resolution	:	2.41 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 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:

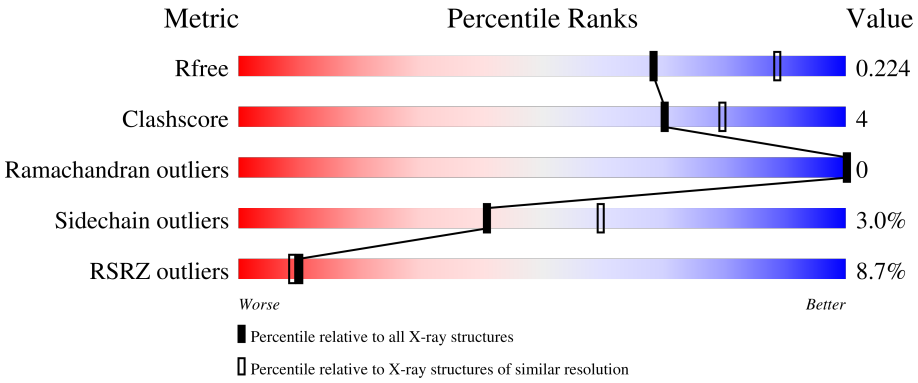
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.41 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	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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	243	4% 86% 13%
1	B	243	17% 87% 11%
1	C	243	4% 86% 12%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 5606 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.

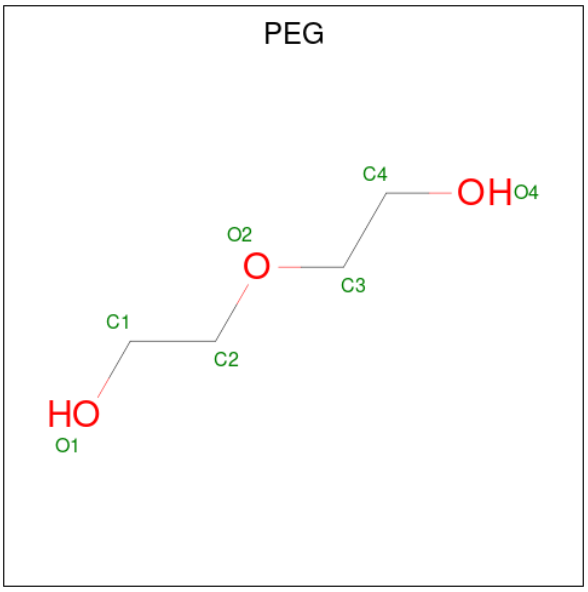
- Molecule 1 is a protein called putative adhesin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	239	Total	C	N	O	S	Se	0	6	0
			1822	1140	310	368	2	2			
1	B	238	Total	C	N	O	S	Se	0	3	0
			1771	1108	302	357	2	2			
1	C	239	Total	C	N	O	S	Se	0	6	0
			1835	1150	317	364	2	2			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	GLY	-	expression tag	UNP A6LGH7
B	20	GLY	-	expression tag	UNP A6LGH7
C	20	GLY	-	expression tag	UNP A6LGH7

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).

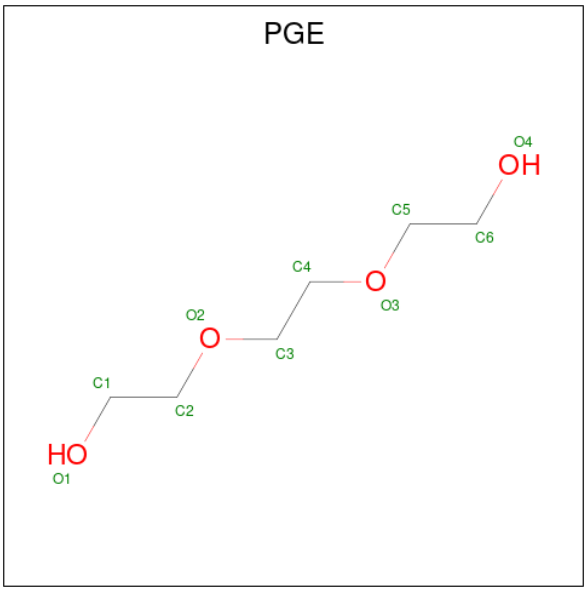


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total C O 7 4 3	0	0
2	B	1	Total C O 7 4 3	0	0
2	C	1	Total C O 7 4 3	0	0
2	C	1	Total C O 7 4 3	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total Cl 1 1	0	0

- Molecule 4 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	1	Total C O 10 6 4	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	59	Total O 59 59	0	0

Continued on next page...

Continued from previous page...

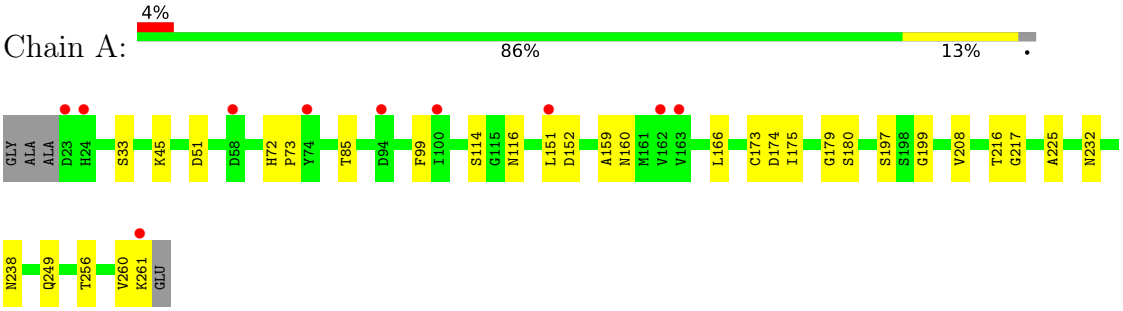
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	13	Total O 13 13	0	0
5	C	67	Total O 67 67	0	0

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.

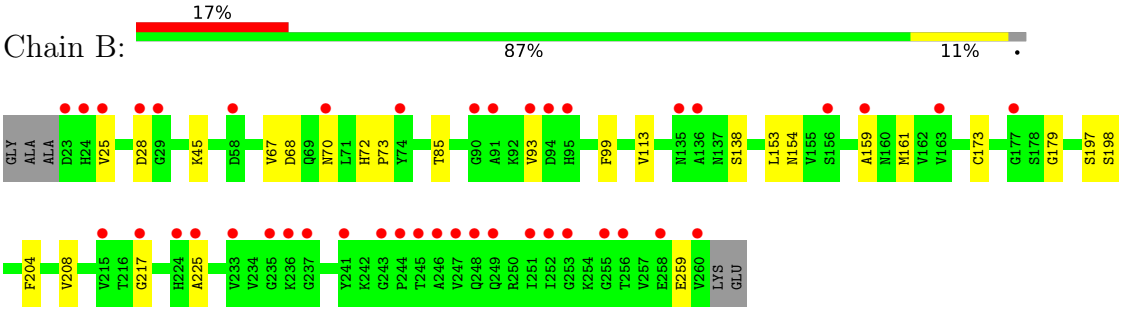
- Molecule 1: putative adhesin

Chain A:



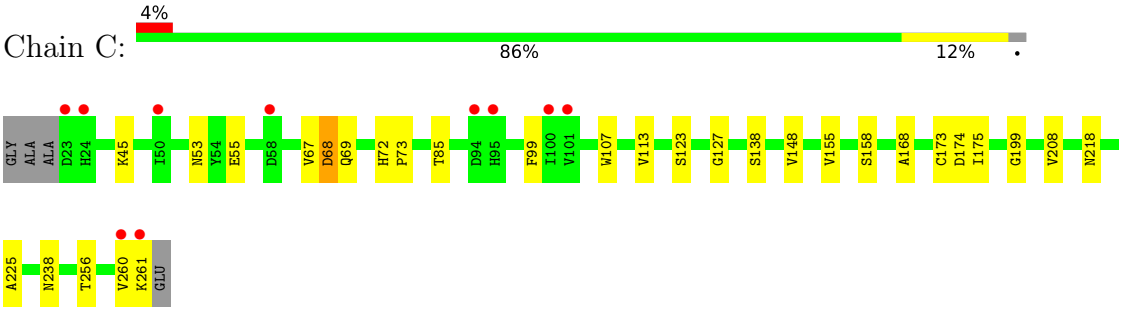
- Molecule 1: putative adhesin

Chain B:



- Molecule 1: putative adhesin

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	65.95Å 55.33Å 120.61Å 90.00° 105.40° 90.00°	Depositor
Resolution (Å)	29.76 – 2.41 29.77 – 2.40	Depositor EDS
% Data completeness (in resolution range)	97.5 (29.76-2.41) 97.6 (29.77-2.40)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	1.88 (at 2.42Å)	Xtriage
Refinement program	REFMAC 5.2.0019, PHENIX	Depositor
R, R_{free}	0.189 , 0.225 0.191 , 0.224	Depositor DCC
R_{free} test set	1628 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	46.5	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 51.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.014 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5606	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.95% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CL, PGE, PEG

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	0.75	1/1843 (0.1%)	0.79	2/2493 (0.1%)
1	B	0.64	5/1792 (0.3%)	0.65	0/2430
1	C	0.73	1/1858 (0.1%)	0.77	0/2513
All	All	0.71	7/5493 (0.1%)	0.74	2/7436 (0.0%)

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	173	CYS	CB-SG	-8.26	1.68	1.82
1	B	70	ASN	CG-ND2	7.13	1.50	1.32
1	B	259	GLU	CD-OE1	6.67	1.32	1.25
1	A	173	CYS	CB-SG	-6.57	1.71	1.82
1	B	259	GLU	CD-OE2	6.40	1.32	1.25
1	B	173	CYS	CB-SG	-5.65	1.72	1.81
1	B	28	ASP	CB-CG	5.57	1.63	1.51

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	51	ASP	CB-CG-OD1	5.28	123.05	118.30
1	A	152	ASP	CB-CG-OD1	5.24	123.02	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	1822	0	1796	11	0
1	B	1771	0	1721	12	0
1	C	1835	0	1828	16	0
2	B	14	0	20	0	0
2	C	14	0	20	0	0
3	C	1	0	0	0	0
4	C	10	0	14	3	0
5	A	59	0	0	0	0
5	B	13	0	0	0	0
5	C	67	0	0	0	0
All	All	5606	0	5399	39	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 4.

All (39) 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:C:55:GLU:HG3	4:C:6:PGE:H52	1.53	0.88
1:C:55:GLU:CG	4:C:6:PGE:H52	2.27	0.64
1:A:160:ASN:HD22	1:A:180:SER:HB3	1.67	0.59
1:A:160:ASN:ND2	1:A:180:SER:HB3	2.19	0.58
1:B:197:SER:OG	1:B:198:SER:N	2.36	0.57
1:A:175:ILE:HG21	1:A:199:GLY:HA3	1.87	0.56
1:C:55:GLU:HG2	1:C:123:SER:HB3	1.86	0.56
1:C:260:VAL:HG12	1:C:261:LYS:HG3	1.90	0.53
1:A:45:LYS:HB2	1:A:85:THR:HG22	1.91	0.53
1:A:260:VAL:O	1:A:261:LYS:HB2	2.10	0.51
1:B:153:LEU:HD22	1:B:161:MSE:SE	2.61	0.51
1:C:155:VAL:HG22	1:C:175:ILE:HD13	1.91	0.51
1:B:113:VAL:HG12	1:B:138:SER:HB3	1.92	0.51
1:C:53:ASN:ND2	4:C:6:PGE:H2	2.26	0.50
1:A:238:ASN:ND2	1:A:256:THR:OG1	2.43	0.49
1:B:67:VAL:HG22	1:B:68:ASP:N	2.28	0.49
1:B:67:VAL:HG22	1:B:68:ASP:H	1.76	0.48
1:B:197:SER:O	1:B:217:GLY:HA3	2.14	0.48
1:B:45:LYS:HB2	1:B:85:THR:HG22	1.96	0.48
1:B:72:HIS:N	1:B:73:PRO:HD2	2.28	0.48
1:B:159:ALA:O	1:B:179:GLY:HA3	2.15	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:113:VAL:HG12	1:C:138[B]:SER:HB2	1.96	0.47
1:B:25:VAL:HG23	1:B:93:VAL:HG22	1.95	0.47
1:C:175:ILE:HG21	1:C:199:GLY:HA3	1.97	0.47
1:C:72:HIS:N	1:C:73:PRO:HD2	2.30	0.46
1:C:238:ASN:ND2	1:C:256:THR:OG1	2.39	0.46
1:A:72:HIS:N	1:A:73:PRO:HD2	2.30	0.46
1:A:151:LEU:HD22	1:A:166:LEU:HD13	1.96	0.46
1:C:208:VAL:O	1:C:225:ALA:HA	2.16	0.46
1:B:208:VAL:O	1:B:225:ALA:HA	2.15	0.45
1:C:107[B]:TRP:CH2	1:C:127:GLY:HA3	2.52	0.45
1:B:25:VAL:CG2	1:B:93:VAL:HG22	2.47	0.45
1:C:45:LYS:HB2	1:C:85:THR:HG22	1.99	0.43
1:C:155:VAL:CG2	1:C:175:ILE:HD13	2.48	0.43
1:A:197:SER:O	1:A:217:GLY:HA3	2.18	0.42
1:A:208:VAL:O	1:A:225:ALA:HA	2.20	0.42
1:A:159:ALA:O	1:A:179:GLY:HA3	2.21	0.41
1:C:148:VAL:O	1:C:168:ALA:HA	2.22	0.40
1:C:67:VAL:HG22	1:C:68:ASP:H	1.86	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	243/243 (100%)	236 (97%)	7 (3%)	0	100	100
1	B	239/243 (98%)	232 (97%)	7 (3%)	0	100	100
1	C	243/243 (100%)	236 (97%)	7 (3%)	0	100	100
All	All	725/729 (100%)	704 (97%)	21 (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	197/199 (99%)	187 (95%)	10 (5%)	24	37
1	B	187/199 (94%)	184 (98%)	3 (2%)	62	78
1	C	200/199 (100%)	194 (97%)	6 (3%)	41	59
All	All	584/597 (98%)	565 (97%)	19 (3%)	41	56

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

Mol	Chain	Res	Type
1	A	33[A]	SER
1	A	33[B]	SER
1	A	99	PHE
1	A	114[A]	SER
1	A	114[B]	SER
1	A	116	ASN
1	A	174	ASP
1	A	216	THR
1	A	232	ASN
1	A	249	GLN
1	B	99	PHE
1	B	154	ASN
1	B	204	PHE
1	C	68	ASP
1	C	69	GLN
1	C	99	PHE
1	C	158	SER
1	C	174	ASP
1	C	218	ASN

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

Mol	Chain	Res	Type
1	A	160	ASN
1	A	238	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	238	ASN
1	C	116	ASN
1	C	160	ASN
1	C	238	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)

Of 6 ligands modelled in this entry, 1 is 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	$\#\|Z\| > 2$
2	PEG	B	2	-	6,6,6	0.70	0	5,5,5	0.51	0
2	PEG	C	5	-	6,6,6	0.61	0	5,5,5	0.44	0
4	PGE	C	6	-	9,9,9	0.75	0	8,8,8	1.22	0
2	PEG	C	3	-	6,6,6	0.55	0	5,5,5	0.37	0
2	PEG	B	4	-	6,6,6	0.72	0	5,5,5	0.79	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	PEG	B	2	-	-	2/4/4/4	-
2	PEG	C	5	-	-	3/4/4/4	-
4	PGE	C	6	-	-	4/7/7/7	-
2	PEG	C	3	-	-	2/4/4/4	-
2	PEG	B	4	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (13) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	6	PGE	C6-C5-O3-C4
2	B	2	PEG	O1-C1-C2-O2
4	C	6	PGE	O3-C5-C6-O4
2	C	5	PEG	C4-C3-O2-C2
4	C	6	PGE	C1-C2-O2-C3
2	C	3	PEG	C1-C2-O2-C3
2	B	4	PEG	C1-C2-O2-C3
2	B	4	PEG	C4-C3-O2-C2
2	C	3	PEG	C4-C3-O2-C2
2	C	5	PEG	C1-C2-O2-C3
2	C	5	PEG	O1-C1-C2-O2
2	B	2	PEG	C1-C2-O2-C3
4	C	6	PGE	O2-C3-C4-O3

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	6	PGE	3	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

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, 95^{th} 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	237/243 (97%)	0.37	10 (4%) 36 34	41, 48, 62, 77	0
1	B	236/243 (97%)	0.87	42 (17%) 1 1	41, 48, 61, 76	0
1	C	237/243 (97%)	0.25	10 (4%) 36 34	41, 48, 62, 77	0
All	All	710/729 (97%)	0.50	62 (8%) 10 9	41, 48, 62, 77	0

All (62) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	253	GLY	6.2
1	B	74	TYR	5.6
1	B	252	ILE	5.4
1	B	251	ILE	5.0
1	B	244	PRO	4.7
1	B	245	THR	4.6
1	B	25	VAL	4.4
1	C	261	LYS	4.3
1	A	24	HIS	4.2
1	B	23	ASP	4.2
1	B	246	ALA	4.1
1	B	260	VAL	4.0
1	B	136	ALA	3.9
1	A	74	TYR	3.6
1	C	260	VAL	3.5
1	B	256	THR	3.5
1	B	235	GLY	3.5
1	B	94	ASP	3.3
1	A	23	ASP	3.0
1	B	248	GLN	3.0
1	B	225	ALA	2.9
1	B	224	HIS	2.9
1	B	237	GLY	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	255	GLY	2.9
1	B	156	SER	2.8
1	B	247	VAL	2.8
1	A	261	LYS	2.8
1	B	95	HIS	2.8
1	B	241	TYR	2.7
1	B	91	ALA	2.7
1	B	28	ASP	2.7
1	B	24	HIS	2.6
1	A	162	VAL	2.6
1	B	243	GLY	2.5
1	B	177	GLY	2.5
1	C	23	ASP	2.5
1	B	258	GLU	2.5
1	B	215	VAL	2.4
1	C	95	HIS	2.4
1	C	50	ILE	2.4
1	B	70	ASN	2.4
1	B	159	ALA	2.4
1	A	58	ASP	2.4
1	A	151	LEU	2.4
1	B	90	GLY	2.4
1	B	163	VAL	2.4
1	C	24	HIS	2.4
1	B	58	ASP	2.3
1	B	249	GLN	2.3
1	B	217	GLY	2.3
1	C	58	ASP	2.3
1	A	163	VAL	2.3
1	C	100	ILE	2.2
1	B	233	VAL	2.2
1	B	135	ASN	2.2
1	B	93	VAL	2.2
1	C	101	VAL	2.2
1	A	94	ASP	2.2
1	C	94	ASP	2.2
1	B	29	GLY	2.1
1	B	236	LYS	2.0
1	A	100	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 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, 95^{th} 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	PEG	C	5	7/7	0.79	0.36	73,76,79,80	0
2	PEG	B	4	7/7	0.81	0.18	63,64,65,65	0
2	PEG	B	2	7/7	0.82	0.48	62,65,68,68	0
4	PGE	C	6	10/10	0.89	0.24	41,47,52,55	0
2	PEG	C	3	7/7	0.91	0.20	53,57,62,62	0
3	CL	C	1	1/1	0.97	0.17	37,37,37,37	0

6.5 Other polymers (i)

There are no such residues in this entry.