Full wwPDB X-ray Structure Validation Report ⓘ

Feb 21, 2024 – 08:37 AM EST

PDB ID	:	4O96
Title	:	2.60 Angstrom resolution crystal structure of a protein kinase domain of type III effector NleH2 (ECs1814) from Escherichia coli O157:H7 str. Sakai
Authors	:	Anderson, S.M.; Halavaty, A.S.; Wawrzak, Z.; Kudritska, M.; Skarina, T.; Yim, V.; Savchenko, A.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on	:	2014-01-01
Resolution	:	2.60 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:

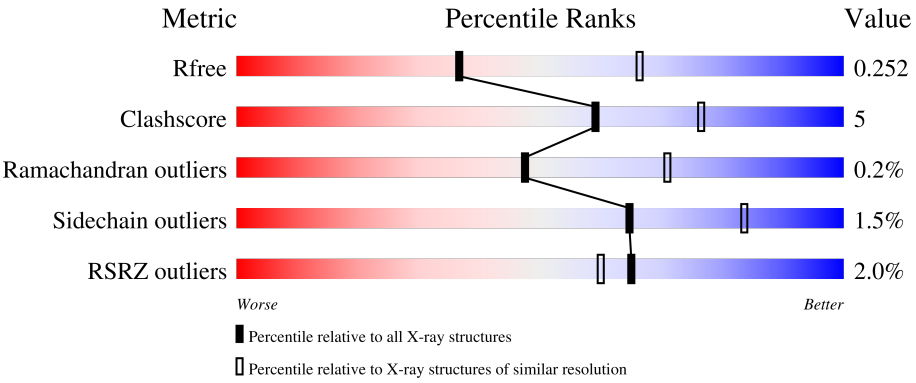
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	166	2% 89% 8%
1	B	166	4% 89% 9%
1	C	166	87% 10%
1	D	166	2% 86% 10%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5400 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.

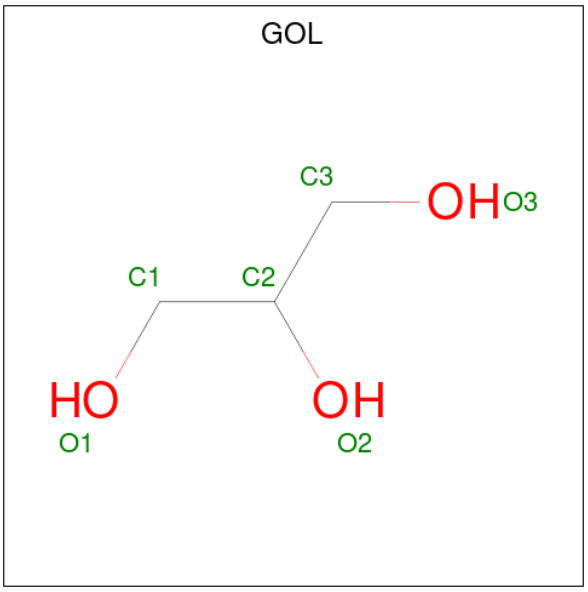
- Molecule 1 is a protein called type III effector protein kinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	163	Total	C	N	O	S	0	0	0
			1284	805	213	259	7			
1	B	163	Total	C	N	O	S	0	2	0
			1304	816	218	263	7			
1	C	163	Total	C	N	O	S	0	2	0
			1306	817	221	261	7			
1	D	163	Total	C	N	O	S	0	2	0
			1305	817	220	261	7			

There are 4 discrepancies between the modelled and reference sequences:

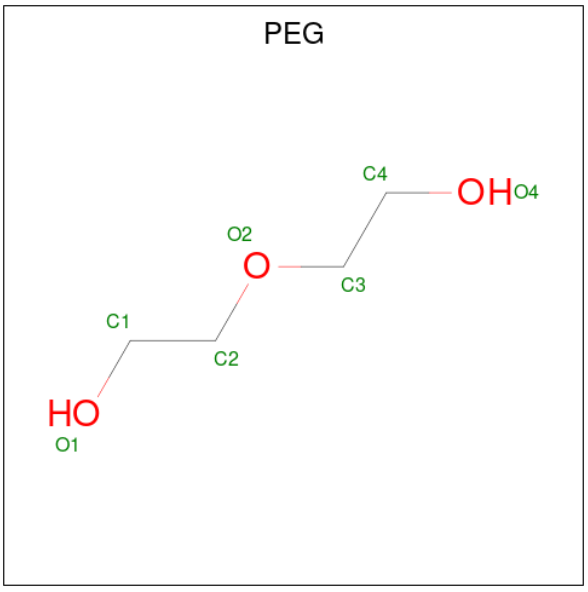
Chain	Residue	Modelled	Actual	Comment	Reference
A	138	GLY	-	expression tag	UNP Q8XAL6
B	138	GLY	-	expression tag	UNP Q8XAL6
C	138	GLY	-	expression tag	UNP Q8XAL6
D	138	GLY	-	expression tag	UNP Q8XAL6

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



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

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



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

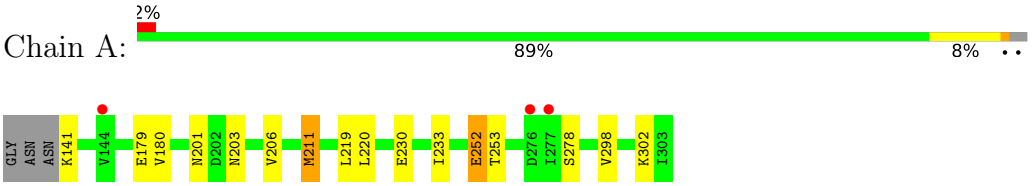
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	40	Total O 40 40	0	0
4	B	64	Total O 64 64	0	0
4	C	37	Total O 37 37	0	0
4	D	29	Total O 29 29	0	1

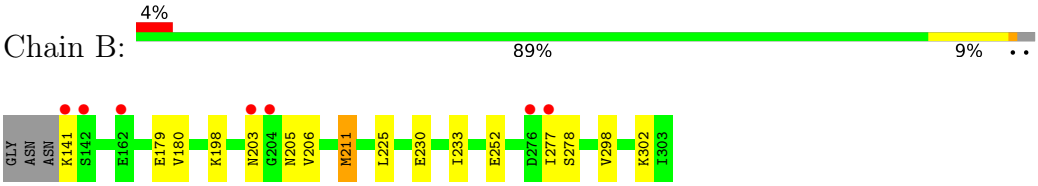
3 Residue-property plots (i)

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.

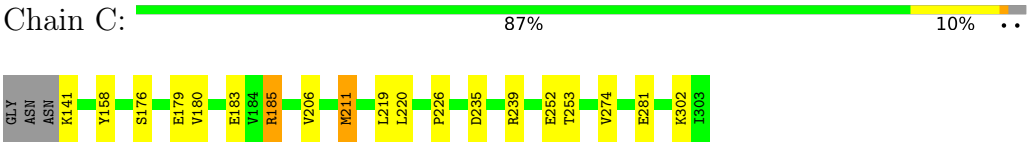
• Molecule 1: type III effector protein kinase



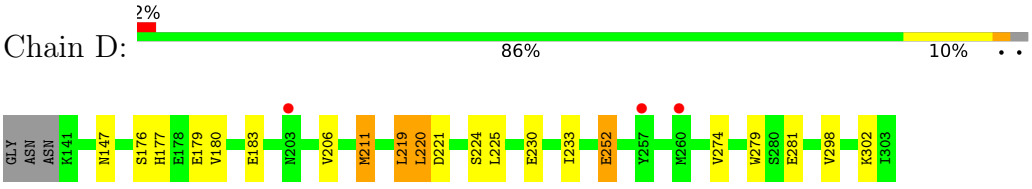
• Molecule 1: type III effector protein kinase



• Molecule 1: type III effector protein kinase



• Molecule 1: type III effector protein kinase



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants a, b, c, α, β, γ	147.25Å 147.25Å 83.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.61 – 2.60 46.57 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.0 (46.61-2.60) 99.1 (46.57-2.60)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.69 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.208 , 0.253 0.209 , 0.252	Depositor DCC
R_{free} test set	1434 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	46.9	Xtriage
Anisotropy	0.427	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 37.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5400	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.15% 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: GOL, 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	0.68	0/1306	0.84	1/1763 (0.1%)
1	B	0.73	0/1326	0.82	0/1789
1	C	0.70	0/1328	0.82	0/1791
1	D	0.63	0/1328	0.83	2/1792 (0.1%)
All	All	0.69	0/5288	0.83	3/7135 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	252	GLU	CB-CA-C	-5.47	99.46	110.40
1	A	252	GLU	CB-CA-C	-5.28	99.85	110.40
1	D	219	LEU	CB-CA-C	5.22	120.11	110.20

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	1284	0	1252	12	0
1	B	1304	0	1269	13	0
1	C	1306	0	1276	16	0
1	D	1305	0	1270	17	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	6	0	8	0	0
2	B	6	0	8	1	0
2	C	6	0	8	0	0
2	D	6	0	8	0	0
3	A	7	0	10	0	0
4	A	40	0	0	1	0
4	B	64	0	0	4	0
4	C	37	0	0	2	0
4	D	29	0	0	5	0
All	All	5400	0	5109	49	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 5.

All (49) 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:205:ASN:HB2	4:B:533:HOH:O	1.54	1.06
1:C:211:MET:HE1	4:C:505:HOH:O	1.71	0.90
1:D:221:ASP:HB3	4:D:529[B]:HOH:O	1.77	0.84
1:A:211:MET:HE1	4:A:529:HOH:O	1.76	0.83
1:D:211:MET:HE1	4:D:525:HOH:O	1.90	0.72
1:C:226:PRO:HG3	4:C:532:HOH:O	1.97	0.65
1:B:198:LYS:HE3	4:B:529:HOH:O	1.96	0.64
1:A:220:LEU:CD1	1:A:253:THR:HG22	2.30	0.62
1:A:252:GLU:HG2	1:A:298:VAL:HG11	1.84	0.60
1:D:177[A]:HIS:CE1	4:D:523:HOH:O	2.55	0.59
1:A:180:VAL:HG21	1:A:206:VAL:HG11	1.85	0.58
1:D:221:ASP:C	4:D:529[B]:HOH:O	2.41	0.58
1:D:252:GLU:HG2	1:D:298:VAL:HG11	1.86	0.57
1:C:220:LEU:CD1	1:C:253:THR:HG22	2.34	0.57
1:D:274:VAL:HG21	1:D:281:GLU:OE2	2.06	0.55
1:C:274:VAL:HG21	1:C:281:GLU:OE2	2.06	0.55
1:B:252:GLU:HG2	1:B:298:VAL:HG11	1.88	0.55
1:A:220:LEU:HD11	1:A:253:THR:HG22	1.89	0.53
1:D:180:VAL:HG21	1:D:206:VAL:HG11	1.91	0.52
1:C:235:ASP:O	1:C:239[B]:ARG:CD	2.60	0.50
1:C:180:VAL:HG21	1:C:206:VAL:HG11	1.93	0.50
1:C:176:SER:N	1:D:176:SER:HB3	2.28	0.49
1:C:158:TYR:OH	1:D:279:TRP:CZ3	2.63	0.49
1:B:180:VAL:HG21	1:B:206:VAL:HG11	1.95	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:176:SER:HB3	1:D:176:SER:N	2.29	0.48
1:A:179:GLU:OE1	1:B:179:GLU:OE1	2.32	0.47
1:B:203:ASN:HB3	4:B:533:HOH:O	2.15	0.46
1:C:220:LEU:HD11	1:C:253:THR:HG22	1.98	0.45
1:B:203:ASN:CB	4:B:533:HOH:O	2.64	0.45
1:B:211:MET:HE2	2:B:401:GOL:O3	2.17	0.45
1:A:203:ASN:OD1	1:C:185[B]:ARG:HD3	2.18	0.44
1:D:180:VAL:HG11	4:D:526:HOH:O	2.19	0.43
1:A:201:ASN:HD22	1:B:277:ILE:HG22	1.83	0.42
1:A:141:LYS:HE3	1:A:141:LYS:HA	2.02	0.42
1:D:219:LEU:O	1:D:302:LYS:NZ	2.52	0.42
1:A:219:LEU:O	1:A:302:LYS:NZ	2.53	0.42
1:D:179:GLU:OE1	1:D:183:GLU:OE1	2.38	0.42
1:C:179:GLU:OE1	1:C:183:GLU:OE1	2.38	0.41
1:A:201:ASN:ND2	1:B:277:ILE:CG2	2.83	0.41
1:B:230:GLU:O	1:B:233:ILE:HG22	2.21	0.41
1:D:230:GLU:O	1:D:233:ILE:HG22	2.21	0.41
1:B:141:LYS:HE3	1:B:141:LYS:HA	2.03	0.41
1:C:141:LYS:HE3	1:C:141:LYS:HA	2.03	0.41
1:A:230:GLU:O	1:A:233:ILE:HG22	2.21	0.40
1:D:225:LEU:HD12	1:D:302:LYS:HB2	2.03	0.40
1:C:158:TYR:OH	1:D:279:TRP:HZ3	2.03	0.40
1:C:252:GLU:OE2	1:D:220:LEU:HD11	2.22	0.40
1:B:225:LEU:HD12	1:B:302:LYS:HB2	2.04	0.40
1:C:219:LEU:O	1:C:302:LYS:NZ	2.55	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	161/166 (97%)	158 (98%)	2 (1%)	1 (1%)	25 47

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	163/166 (98%)	159 (98%)	4 (2%)	0	100 100
1	C	163/166 (98%)	162 (99%)	1 (1%)	0	100 100
1	D	163/166 (98%)	162 (99%)	1 (1%)	0	100 100
All	All	650/664 (98%)	641 (99%)	8 (1%)	1 (0%)	47 71

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	278	SER

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	146/148 (99%)	145 (99%)	1 (1%)	84 94
1	B	148/148 (100%)	146 (99%)	2 (1%)	67 85
1	C	148/148 (100%)	145 (98%)	3 (2%)	55 78
1	D	148/148 (100%)	144 (97%)	4 (3%)	44 71
All	All	590/592 (100%)	580 (98%)	10 (2%)	65 81

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	211	MET
1	B	211	MET
1	B	278	SER
1	C	185[A]	ARG
1	C	185[B]	ARG
1	C	211	MET
1	D	147	ASN
1	D	211	MET
1	D	220	LEU
1	D	224	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

5 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	GOL	C	401	-	5,5,5	0.50	0	5,5,5	0.24	0
2	GOL	B	401	-	5,5,5	0.78	0	5,5,5	1.54	1 (20%)
2	GOL	D	401	-	5,5,5	0.62	0	5,5,5	0.56	0
3	PEG	A	402	-	6,6,6	0.36	0	5,5,5	0.59	0
2	GOL	A	401	-	5,5,5	0.69	0	5,5,5	0.43	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	GOL	C	401	-	-	2/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	B	401	-	-	2/4/4/4	-
2	GOL	D	401	-	-	2/4/4/4	-
3	PEG	A	402	-	-	1/4/4/4	-
2	GOL	A	401	-	-	4/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	401	GOL	O2-C2-C3	-2.20	99.44	109.12

There are no chirality outliers.

All (11) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	401	GOL	O1-C1-C2-C3
2	D	401	GOL	O1-C1-C2-C3
2	A	401	GOL	O1-C1-C2-C3
2	C	401	GOL	O1-C1-C2-C3
2	B	401	GOL	O1-C1-C2-O2
2	C	401	GOL	O1-C1-C2-O2
2	A	401	GOL	O1-C1-C2-O2
2	D	401	GOL	O1-C1-C2-O2
2	A	401	GOL	O2-C2-C3-O3
2	A	401	GOL	C1-C2-C3-O3
3	A	402	PEG	C4-C3-O2-C2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	401	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

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	163/166 (98%)	-0.24	3 (1%) 68 64	37, 53, 74, 108	0
1	B	163/166 (98%)	-0.17	7 (4%) 35 28	29, 45, 92, 146	0
1	C	163/166 (98%)	-0.13	0 100 100	29, 54, 101, 115	0
1	D	163/166 (98%)	0.21	3 (1%) 68 64	38, 71, 115, 139	0
All	All	652/664 (98%)	-0.08	13 (1%) 65 60	29, 55, 103, 146	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	141	LYS	5.1
1	D	260	MET	4.2
1	B	276	ASP	4.2
1	D	257	TYR	4.0
1	B	277	ILE	4.0
1	A	276	ASP	3.9
1	D	203	ASN	3.6
1	B	142	SER	3.2
1	B	203	ASN	3.2
1	B	204	GLY	2.8
1	A	144	VAL	2.6
1	B	162	GLU	2.3
1	A	277	ILE	2.3

6.2 Non-standard residues in protein, DNA, RNA chains

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	GOL	A	401	6/6	0.83	0.26	53,63,67,68	0
2	GOL	C	401	6/6	0.86	0.21	56,61,73,73	0
3	PEG	A	402	7/7	0.87	0.23	67,70,77,79	0
2	GOL	D	401	6/6	0.88	0.20	56,66,67,68	0
2	GOL	B	401	6/6	0.91	0.20	39,43,48,51	0

6.5 Other polymers (i)

There are no such residues in this entry.