Full wwPDB X-ray Structure Validation Report (i)

Jun 18, 2024 – 01:07 AM EDT

PDB ID	:	2YNK
Title	:	Wzi, an Outer Membrane Protein Involved in Group 1 Capsule Assembly in Escherichia coli, is a Carbohydrate Binding Beta-Barrel
Authors	:	Bushell, S.R.; Mainprize, I.L.; Wear, M.A.; Lou, H.; Kong, L.; Davis, B.; Bayley, H.; Whitfield, C.; Naismith, J.H.
Deposited on	:	2012-10-16
Resolution	:	2.64 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 (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:

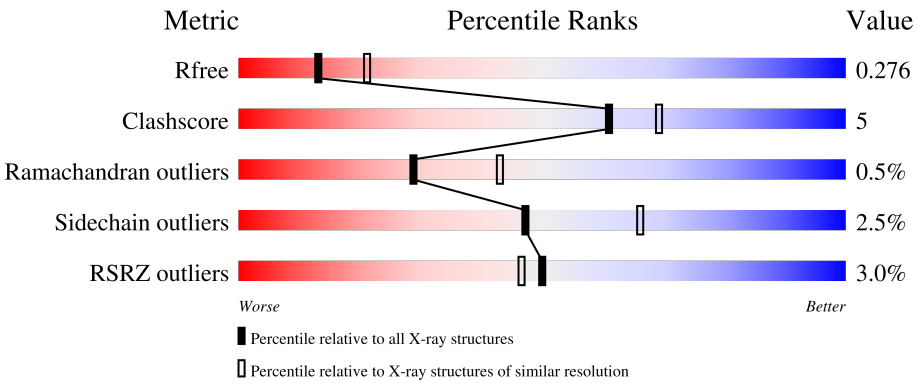
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.64 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	1426 (2.66-2.62)
Clashscore	141614	1472 (2.66-2.62)
Ramachandran outliers	138981	1446 (2.66-2.62)
Sidechain outliers	138945	1446 (2.66-2.62)
RSRZ outliers	127900	1408 (2.66-2.62)

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	456	3% 83% 13% •

2 Entry composition (i)

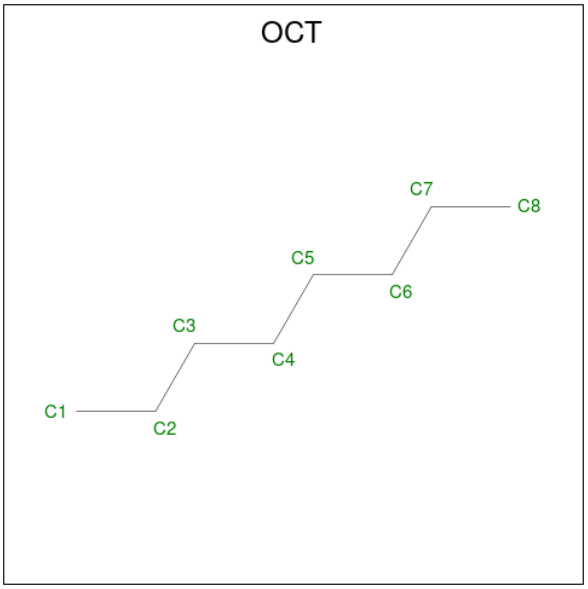
There are 5 unique types of molecules in this entry. The entry contains 3652 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 WZI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	441	3476	2202	610	655	9	0	0	0

- Molecule 2 is N-OCTANE (three-letter code: OCT) (formula: C_8H_{18}).



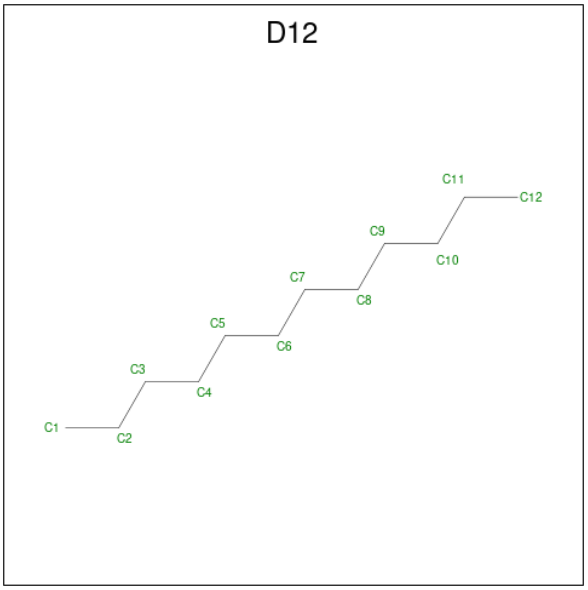
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	C		
2	A	1	8	8	0	0
2	A	1	8	8	0	0
2	A	1	8	8	0	0
2	A	1	8	8	0	0
2	A	1	8	8	0	0

Continued on next page...

Continued from previous page...

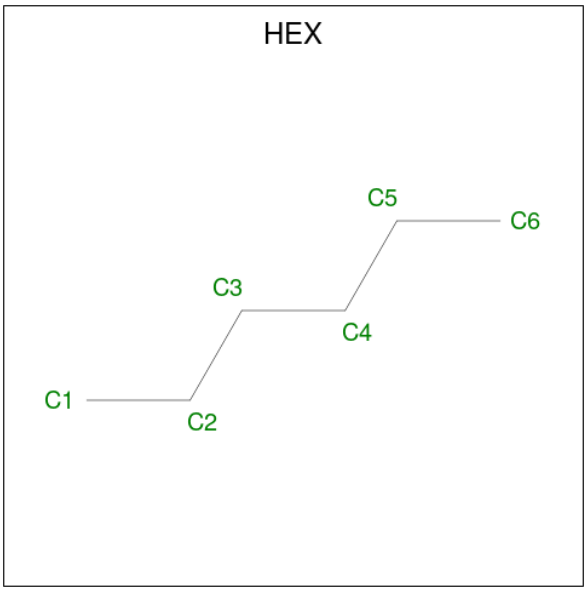
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	C		
2	A	1	8	8	0	0

- Molecule 3 is DODECANE (three-letter code: D12) (formula: $C_{12}H_{26}$).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	C		
3	A	1	12	12	0	0
			Total	C		
3	A	1	12	12	0	0
			Total	C		
3	A	1	12	12	0	0
			Total	C		
3	A	1	12	12	0	0
			Total	C		
3	A	1	12	12	0	0
			Total	C		
3	A	1	12	12	0	0
			Total	C		
3	A	1	12	12	0	0

- Molecule 4 is HEXANE (three-letter code: HEX) (formula: C_6H_{14}).



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

- Molecule 5 is water.

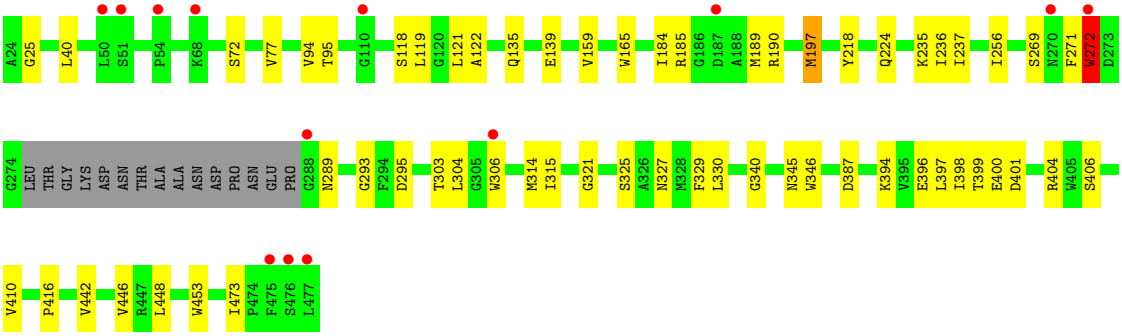
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	14	Total O 14 14	0	0

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: WZI

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	136.66Å 152.08Å 95.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	59.37 – 2.64 38.02 – 2.64	Depositor EDS
% Data completeness (in resolution range)	99.2 (59.37-2.64) 99.2 (38.02-2.64)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.83 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.245 , 0.279 0.248 , 0.276	Depositor DCC
R_{free} test set	1482 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	59.1	Xtriage
Anisotropy	0.484	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 47.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	3652	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.67% 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: OCT, D12, HEX

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	0.62	3/3569 (0.1%)	0.63	0/4836

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	272	TRP	CD2-CE2	5.14	1.47	1.41
1	A	453	TRP	CD2-CE2	5.08	1.47	1.41
1	A	306	TRP	CD2-CE2	5.06	1.47	1.41

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	3476	0	3270	33	0
2	A	48	0	108	0	0
3	A	84	0	182	2	0
4	A	30	0	70	0	0
5	A	14	0	0	1	0
All	All	3652	0	3630	33	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 (33) 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:303:THR:HG22	1:A:304:LEU:HD23	1.73	0.71
1:A:314:MSE:HE3	1:A:329:PHE:CZ	2.32	0.64
1:A:25:GLY:HA2	1:A:165:TRP:CZ2	2.35	0.61
1:A:399:THR:HG22	1:A:401:ASP:H	1.66	0.60
1:A:314:MSE:HE2	1:A:327:ASN:ND2	2.17	0.59
1:A:446:VAL:HG13	1:A:473:ILE:HG13	1.86	0.58
1:A:314:MSE:HE2	1:A:327:ASN:CG	2.25	0.57
1:A:40:LEU:HD11	1:A:77:VAL:HG12	1.87	0.56
1:A:289:ASN:ND2	5:A:2010:HOH:O	2.22	0.56
1:A:122:ALA:HB2	1:A:135:GLN:OE1	2.06	0.55
1:A:237:ILE:HD13	1:A:271:PHE:CE1	2.43	0.54
1:A:94:VAL:HG22	1:A:121:LEU:HD22	1.91	0.53
1:A:397:LEU:HD22	3:A:1005:D12:H52	1.92	0.52
1:A:159:VAL:HG13	3:A:1003:D12:H92	1.90	0.52
1:A:95:THR:O	1:A:119:LEU:HD12	2.11	0.50
1:A:237:ILE:HD13	1:A:271:PHE:HE1	1.75	0.49
1:A:293:GLY:HA3	1:A:315:ILE:HD13	1.94	0.49
1:A:236:ILE:CG2	1:A:256:ILE:HD12	2.42	0.49
1:A:387:ASP:HB2	1:A:416:PRO:HD2	1.95	0.48
1:A:197:MSE:HE2	1:A:218:TYR:OH	2.14	0.47
1:A:165:TRP:O	1:A:197:MSE:HA	2.15	0.46
1:A:398:ILE:HG23	1:A:404:ARG:HG2	1.97	0.46
1:A:314:MSE:HE3	1:A:329:PHE:CE2	2.50	0.45
1:A:345:ASN:HB2	1:A:396:GLU:HG3	1.99	0.45
1:A:139:GLU:OE1	1:A:190:ARG:NH2	2.47	0.43
1:A:224:GLN:OE1	1:A:235:LYS:NZ	2.52	0.43
1:A:399:THR:HG22	1:A:400:GLU:N	2.33	0.43
1:A:330:LEU:C	1:A:330:LEU:HD23	2.40	0.42
1:A:346:TRP:HA	1:A:394:LYS:O	2.20	0.41
1:A:269:SER:O	1:A:272:TRP:CD1	2.74	0.41
1:A:94:VAL:HG22	1:A:121:LEU:CD2	2.50	0.41
1:A:184:ILE:HG23	1:A:185:ARG:N	2.36	0.40
1:A:442:VAL:HG21	1:A:448:LEU:HD22	2.04	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	437/456 (96%)	411 (94%)	24 (6%)	2 (0%)	29 43

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	321	GLY
1	A	340	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	357/360 (99%)	348 (98%)	9 (2%)	47 66

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	72	SER
1	A	118	SER
1	A	189	MSE
1	A	197	MSE
1	A	272	TRP
1	A	295	ASP
1	A	325	SER
1	A	406	SER
1	A	410	VAL

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

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)

18 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	$\#\|Z\| > 2$
3	D12	A	1004	-	11,11,11	0.48	0	10,10,10	0.22	0
3	D12	A	1005	-	11,11,11	0.37	0	10,10,10	0.31	0
4	HEX	A	1013	-	5,5,5	0.33	0	4,4,4	0.29	0
4	HEX	A	1018	-	5,5,5	0.31	0	4,4,4	0.37	0
4	HEX	A	1016	-	5,5,5	0.34	0	4,4,4	0.30	0
2	OCT	A	1010	-	7,7,7	0.27	0	6,6,6	0.40	0
2	OCT	A	1014	-	7,7,7	0.33	0	6,6,6	0.34	0
3	D12	A	1002	-	11,11,11	0.38	0	10,10,10	0.35	0
4	HEX	A	1017	-	5,5,5	0.30	0	4,4,4	0.32	0
3	D12	A	1007	-	11,11,11	0.36	0	10,10,10	0.39	0
2	OCT	A	1001	-	7,7,7	0.35	0	6,6,6	0.35	0
2	OCT	A	1009	-	7,7,7	0.30	0	6,6,6	0.41	0
3	D12	A	1006	-	11,11,11	0.36	0	10,10,10	0.36	0
3	D12	A	1008	-	11,11,11	0.32	0	10,10,10	0.40	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	HEX	A	1015	-	5,5,5	0.33	0	4,4,4	0.27	0
2	OCT	A	1012	-	7,7,7	0.35	0	6,6,6	0.31	0
3	D12	A	1003	-	11,11,11	0.38	0	10,10,10	0.28	0
2	OCT	A	1011	-	7,7,7	0.33	0	6,6,6	0.32	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
3	D12	A	1004	-	-	4/9/9/9	-
3	D12	A	1005	-	-	4/9/9/9	-
4	HEX	A	1013	-	-	0/3/3/3	-
4	HEX	A	1018	-	-	0/3/3/3	-
4	HEX	A	1016	-	-	1/3/3/3	-
2	OCT	A	1010	-	-	3/5/5/5	-
2	OCT	A	1014	-	-	3/5/5/5	-
3	D12	A	1002	-	-	4/9/9/9	-
4	HEX	A	1017	-	-	1/3/3/3	-
3	D12	A	1007	-	-	5/9/9/9	-
2	OCT	A	1001	-	-	2/5/5/5	-
2	OCT	A	1009	-	-	0/5/5/5	-
3	D12	A	1006	-	-	4/9/9/9	-
3	D12	A	1008	-	-	2/9/9/9	-
4	HEX	A	1015	-	-	0/3/3/3	-
2	OCT	A	1012	-	-	1/5/5/5	-
3	D12	A	1003	-	-	6/9/9/9	-
2	OCT	A	1011	-	-	0/5/5/5	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (40) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1006	D12	C3-C4-C5-C6
2	A	1001	OCT	C4-C5-C6-C7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	1005	D12	C6-C7-C8-C9
3	A	1004	D12	C6-C7-C8-C9
3	A	1008	D12	C11-C10-C9-C8
3	A	1005	D12	C2-C3-C4-C5
3	A	1006	D12	C5-C6-C7-C8
3	A	1007	D12	C3-C4-C5-C6
2	A	1010	OCT	C3-C4-C5-C6
3	A	1005	D12	C4-C5-C6-C7
3	A	1004	D12	C2-C3-C4-C5
3	A	1003	D12	C6-C7-C8-C9
4	A	1017	HEX	C1-C2-C3-C4
3	A	1003	D12	C4-C5-C6-C7
3	A	1008	D12	C4-C5-C6-C7
3	A	1006	D12	C9-C10-C11-C12
3	A	1003	D12	C5-C6-C7-C8
3	A	1002	D12	C9-C10-C11-C12
2	A	1012	OCT	C1-C2-C3-C4
3	A	1004	D12	C5-C6-C7-C8
2	A	1010	OCT	C2-C3-C4-C5
3	A	1007	D12	C9-C10-C11-C12
2	A	1010	OCT	C5-C6-C7-C8
2	A	1014	OCT	C5-C6-C7-C8
3	A	1005	D12	C5-C6-C7-C8
2	A	1001	OCT	C5-C6-C7-C8
3	A	1006	D12	C6-C7-C8-C9
4	A	1016	HEX	C2-C3-C4-C5
3	A	1007	D12	C5-C6-C7-C8
2	A	1014	OCT	C1-C2-C3-C4
3	A	1003	D12	C9-C10-C11-C12
3	A	1007	D12	C7-C8-C9-C10
3	A	1002	D12	C2-C3-C4-C5
3	A	1002	D12	C11-C10-C9-C8
3	A	1007	D12	C6-C7-C8-C9
3	A	1004	D12	C4-C5-C6-C7
3	A	1003	D12	C7-C8-C9-C10
3	A	1002	D12	C3-C4-C5-C6
3	A	1003	D12	C11-C10-C9-C8
2	A	1014	OCT	C2-C3-C4-C5

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1005	D12	1	0
3	A	1003	D12	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	432/456 (94%)	0.06	13 (3%) 50 46	40, 55, 78, 107	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	476	SER	3.9
1	A	270	ASN	3.4
1	A	477	LEU	3.0
1	A	54	PRO	2.5
1	A	272	TRP	2.5
1	A	288	GLY	2.4
1	A	68	LYS	2.4
1	A	50	LEU	2.3
1	A	110	GLY	2.2
1	A	187	ASP	2.1
1	A	51	SER	2.1
1	A	306	TRP	2.1
1	A	475	PHE	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	OCT	A	1011	8/8	0.69	0.23	71,75,76,76	0
4	HEX	A	1018	6/6	0.71	0.30	67,70,71,71	0
3	D12	A	1005	12/12	0.77	0.20	61,71,73,75	0
4	HEX	A	1013	6/6	0.77	0.24	64,67,69,70	0
2	OCT	A	1014	8/8	0.77	0.16	75,82,86,86	0
2	OCT	A	1012	8/8	0.81	0.23	71,76,79,81	0
4	HEX	A	1017	6/6	0.83	0.25	63,69,73,73	0
3	D12	A	1002	12/12	0.84	0.22	68,71,74,74	0
3	D12	A	1006	12/12	0.85	0.27	53,61,72,72	0
2	OCT	A	1010	8/8	0.85	0.29	69,74,75,76	0
3	D12	A	1004	12/12	0.86	0.25	45,60,62,63	0
3	D12	A	1003	12/12	0.86	0.22	63,64,73,74	0
3	D12	A	1008	12/12	0.88	0.30	69,73,78,80	0
2	OCT	A	1009	8/8	0.89	0.22	66,75,76,76	0
3	D12	A	1007	12/12	0.91	0.19	45,49,52,53	0
4	HEX	A	1015	6/6	0.91	0.19	62,62,64,64	0
2	OCT	A	1001	8/8	0.92	0.20	54,57,60,60	0
4	HEX	A	1016	6/6	0.92	0.17	52,56,58,58	0

6.5 Other polymers (i)

There are no such residues in this entry.