WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Jun 20, 2016 – 09:23 AM EDT

PDB ID : 4ZH5
Title : Crystal structure of Endoglucanase from Perinereis brevicirris with Cellobiose
Authors : Fewings, R.S.; Swiderska, A.; Sanchez-Weatherby, J.; Sorensen, T.-M.; Schnorr, K.M.; Kneale, G.G.; McGeehan, J.E.
Deposited on : 2015-04-24
Resolution : 1.35 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027790
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027790

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 1.35 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	Percentile Ranks	Value
Rfree		0.149
Clashscore		4
Ramachandran outliers		0.2%
Sidechain outliers		1.2%
RSRZ outliers		0.4%

Worse   *Better*

∎ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
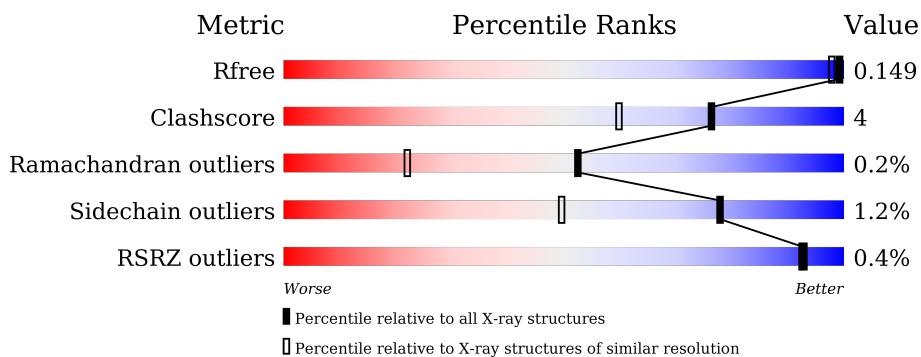
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2199 (1.40-1.32)
Clashscore	102246	2337 (1.40-1.32)
Ramachandran outliers	100387	2280 (1.40-1.32)
Sidechain outliers	100360	2279 (1.40-1.32)
RSRZ outliers	91569	2199 (1.40-1.32)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. 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	426	93% 6%
1	B	426	94% 6%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 7961 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 Endoglucanase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	425	Total 3407	C 2139	N 581	O 672	S 15	0	16	0
1	B	425	Total 3382	C 2124	N 576	O 668	S 14	0	13	0

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total 1	Ca 1	0	0
2	A	1	Total 1	Ca 1	0	0

- Molecule 3 is BETA-D-GLUCOSE (three-letter code: BGC) (formula: $C_6H_{12}O_6$).
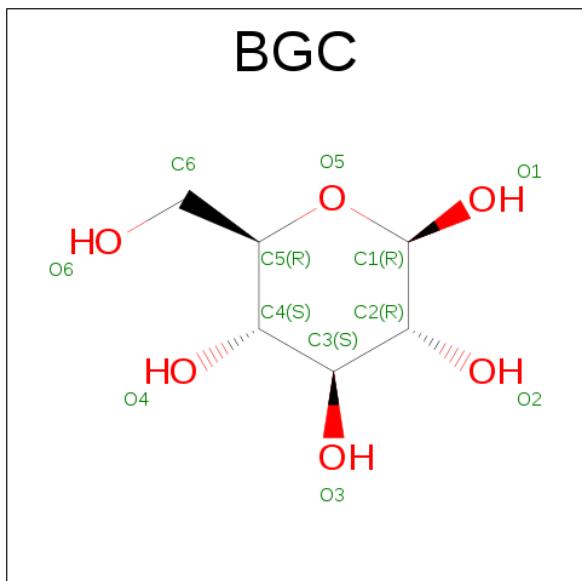
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 12	C 6	O 6	0	0
3	A	1	Total 11	C 6	O 5	0	0
3	B	1	Total 12	C 6	O 6	0	0
3	B	1	Total 11	C 6	O 5	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total 2	Na 2	0	0
4	A	3	Total 3	Na 3	0	0

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total 1	Cl 1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	593	Total 593	O 593	0	0
6	B	525	Total 525	O 525	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Endoglucanase
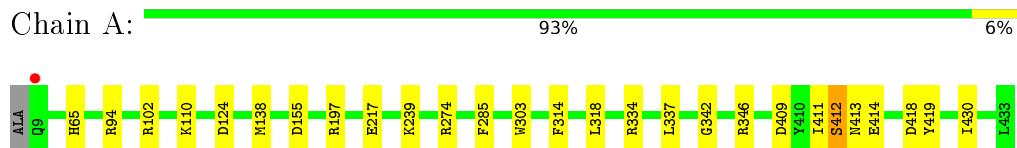
- Molecule 1: Endoglucanase
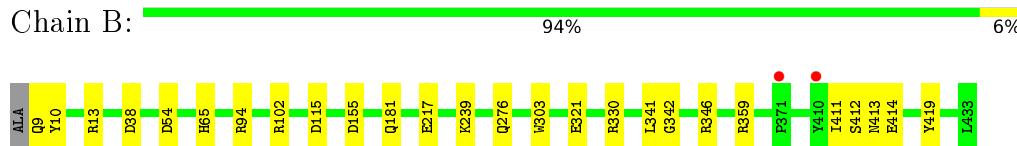
4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	76.15Å 108.17Å 112.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	16.98 – 1.35 16.98 – 1.35	Depositor EDS
% Data completeness (in resolution range)	97.7 (16.98-1.35) 97.8 (16.98-1.35)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.94 (at 1.35Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.115 , 0.148 0.115 , 0.149	Depositor DCC
R_{free} test set	9867 reflections (5.22%)	DCC
Wilson B-factor (Å^2)	9.3	Xtriage
Anisotropy	0.529	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 48.5	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.017 for -h,l,k	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	7961	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.98% 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: NA, CA, BGC, CL

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	1/3520 (0.0%)	0.90	14/4800 (0.3%)
1	B	0.72	2/3498 (0.1%)	0.91	13/4771 (0.3%)
All	All	0.74	3/7018 (0.0%)	0.91	27/9571 (0.3%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	414	GLU	CD-OE2	5.86	1.32	1.25
1	B	217	GLU	CD-OE1	5.73	1.31	1.25
1	B	414	GLU	CD-OE2	5.47	1.31	1.25

All (27) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	102	ARG	NE-CZ-NH2	-10.46	115.07	120.30
1	A	102	ARG	NE-CZ-NH2	-10.35	115.12	120.30
1	B	102	ARG	NE-CZ-NH1	8.81	124.71	120.30
1	A	102	ARG	NE-CZ-NH1	8.26	124.43	120.30
1	A	94	ARG	NE-CZ-NH2	-7.92	116.34	120.30
1	A	274	ARG	NE-CZ-NH2	-7.38	116.61	120.30
1	B	94	ARG	NE-CZ-NH1	6.27	123.44	120.30
1	A	419	TYR	CB-CG-CD1	6.22	124.73	121.00
1	B	94	ARG	NE-CZ-NH2	-6.21	117.19	120.30
1	B	115	ASP	CB-CG-OD1	6.02	123.72	118.30
1	A	274	ARG	NE-CZ-NH1	5.95	123.28	120.30
1	B	54	ASP	CB-CG-OD1	5.91	123.61	118.30
1	A	285	PHE	CB-CG-CD1	5.77	124.84	120.80
1	A	197	ARG	NE-CZ-NH2	-5.68	117.46	120.30
1	B	359	ARG	NE-CZ-NH2	5.68	123.14	120.30
1	B	419	TYR	CB-CG-CD1	5.57	124.34	121.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	334	ARG	NE-CZ-NH1	5.49	123.05	120.30
1	B	419	TYR	CB-CG-CD2	-5.43	117.74	121.00
1	B	341	LEU	CB-CG-CD1	5.36	120.11	111.00
1	A	124	ASP	CB-CG-OD1	5.20	122.98	118.30
1	B	10	TYR	CB-CG-CD2	5.16	124.10	121.00
1	B	13	ARG	NE-CZ-NH2	-5.12	117.74	120.30
1	A	94	ARG	NE-CZ-NH1	5.10	122.85	120.30
1	B	38	ASP	CB-CG-OD1	5.08	122.87	118.30
1	A	138	MET	CG-SD-CE	5.07	108.31	100.20
1	A	197	ARG	NE-CZ-NH1	5.07	122.83	120.30
1	A	414	GLU	N-CA-CB	5.03	119.65	110.60

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	3407	0	3116	37	0
1	B	3382	0	3092	30	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	23	0	21	0	0
3	B	23	0	21	0	0
4	A	3	0	0	0	0
4	B	2	0	0	0	0
5	B	1	0	0	0	0
6	A	593	0	0	10	0
6	B	525	0	0	11	0
All	All	7961	0	6250	50	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 (50) 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:337[B]:LEU:CD1	1:A:430[B]:ILE:HD12	1.30	1.60
1:A:337[B]:LEU:HD12	1:A:430[B]:ILE:CD1	1.09	1.54
1:A:337[B]:LEU:CD1	1:A:430[B]:ILE:CD1	1.80	1.53
1:A:303[A]:TRP:CZ2	1:B:303[A]:TRP:CZ2	2.10	1.37
1:A:303[A]:TRP:HZ2	1:B:303[A]:TRP:CE2	1.48	1.28
1:A:303[A]:TRP:CZ2	1:B:303[A]:TRP:CE2	2.22	1.27
1:A:303[A]:TRP:CE2	1:B:303[A]:TRP:HZ2	1.55	1.23
1:A:303[A]:TRP:CE2	1:B:303[A]:TRP:CZ2	2.27	1.19
1:A:303[A]:TRP:CZ2	1:B:303[A]:TRP:NE1	2.24	1.05
1:A:337[B]:LEU:HD13	1:A:430[B]:ILE:HD12	1.06	1.05
1:A:303[A]:TRP:NE1	1:B:303[A]:TRP:CZ2	2.29	1.00
1:A:337[B]:LEU:CD1	1:A:430[B]:ILE:HD11	1.66	0.96
1:B:321[A]:GLU:OE2	1:B:330:ARG:NH2	2.07	0.86
1:A:412[A]:SER:OG	6:A:601:HOH:O	1.98	0.81
1:B:9:GLN:N	6:B:602:HOH:O	2.19	0.76
1:A:411:ILE:O	6:A:603:HOH:O	2.04	0.75
1:A:217[A]:GLU:OE1	6:A:602:HOH:O	2.03	0.75
1:A:303[A]:TRP:CE2	1:B:303[A]:TRP:CE2	2.66	0.75
1:A:303[A]:TRP:CE2	1:B:303[A]:TRP:NE1	2.58	0.72
1:A:337[B]:LEU:HD12	1:A:430[B]:ILE:HD11	0.69	0.68
1:A:303[B]:TRP:CD1	6:A:603:HOH:O	2.45	0.68
1:B:303[B]:TRP:CD1	6:B:604:HOH:O	2.47	0.68
1:A:303[A]:TRP:NE1	1:B:303[A]:TRP:CE2	2.61	0.68
1:A:303[A]:TRP:NE1	1:B:303[A]:TRP:HZ2	1.77	0.67
1:A:303[A]:TRP:CZ3	1:A:411:ILE:HG12	2.29	0.67
1:A:303[A]:TRP:HZ2	1:B:303[A]:TRP:CZ2	1.74	0.67
1:A:303[A]:TRP:HZ2	1:B:303[A]:TRP:NE1	1.75	0.65
1:B:412[A]:SER:OG	6:B:601:HOH:O	2.09	0.62
1:B:239:LYS:HE2	6:B:631:HOH:O	2.01	0.60
1:A:303[A]:TRP:CZ2	1:B:303[A]:TRP:HZ2	1.78	0.57
1:A:217[A]:GLU:OE2	6:A:604:HOH:O	2.18	0.55
1:A:409:ASP:CG	1:A:412[A]:SER:HB2	2.29	0.53
1:B:303[A]:TRP:CZ3	1:B:411:ILE:HG12	2.44	0.53
1:B:303[B]:TRP:HD1	6:B:604:HOH:O	1.90	0.52
1:B:181:GLN:HG3	6:B:942:HOH:O	2.10	0.51
1:A:303[A]:TRP:NE1	1:B:303[A]:TRP:NE1	2.61	0.49
1:A:337[B]:LEU:HD13	1:A:430[B]:ILE:CD1	1.91	0.48
1:A:217[B]:GLU:OE1	6:A:605:HOH:O	2.20	0.47
1:A:239:LYS:HE2	6:A:626:HOH:O	2.14	0.47
1:A:110[A]:LYS:NZ	6:A:607:HOH:O	2.38	0.47
1:B:321[B]:GLU:HG2	6:B:835:HOH:O	2.16	0.46
1:A:65:HIS:HE1	6:A:868:HOH:O	1.99	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:65:HIS:HD2	6:B:874:HOH:O	2.00	0.45
1:B:181:GLN:CG	6:B:942:HOH:O	2.64	0.45
1:A:303[B]:TRP:HD1	6:A:603:HOH:O	1.92	0.45
1:B:65:HIS:HE1	6:B:849:HOH:O	2.00	0.44
1:B:276[B]:GLN:NE2	6:B:616:HOH:O	2.50	0.44
1:A:411:ILE:CD1	1:B:411:ILE:HD11	2.48	0.42
1:A:314:PHE:CZ	1:A:318:LEU:HD11	2.54	0.41
1:A:411:ILE:HD11	1:B:411:ILE:CD1	2.50	0.41

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	439/426 (103%)	429 (98%)	9 (2%)	1 (0%)	52	22
1	B	436/426 (102%)	426 (98%)	9 (2%)	1 (0%)	52	22
All	All	875/852 (103%)	855 (98%)	18 (2%)	2 (0%)	52	22

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	342	GLY
1	B	342	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	347/331 (105%)	341 (98%)	6 (2%)	68	33
1	B	344/331 (104%)	341 (99%)	3 (1%)	84	61
All	All	691/662 (104%)	682 (99%)	9 (1%)	78	45

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

Mol	Chain	Res	Type
1	A	155	ASP
1	A	346	ARG
1	A	412[A]	SER
1	A	412[B]	SER
1	A	413	ASN
1	A	418	ASP
1	B	155	ASP
1	B	346	ARG
1	B	413	ASN

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

Mol	Chain	Res	Type
1	A	9	GLN
1	A	65	HIS
1	A	116	ASN
1	B	65	HIS
1	B	116	ASN
1	B	180	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 12 ligands modelled in this entry, 8 are monoatomic - leaving 4 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	#$|Z| > 2$
3	BGC	A	502	3	12,12,12	0.72	0	17,17,17	1.41	2 (11%)
3	BGC	A	503	3	11,11,12	0.87	0	15,15,17	1.47	2 (13%)
3	BGC	B	501	3	12,12,12	0.75	0	17,17,17	0.91	0
3	BGC	B	502	3	11,11,12	0.38	0	15,15,17	1.34	2 (13%)

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	BGC	A	502	3	-	0/2/22/22	0/1/1/1
3	BGC	A	503	3	-	0/2/19/22	0/1/1/1
3	BGC	B	501	3	-	0/2/22/22	0/1/1/1
3	BGC	B	502	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	503	BGC	O3-C3-C2	-4.15	102.40	110.01
3	A	502	BGC	C4-C3-C2	-4.01	103.41	110.79
3	B	502	BGC	O3-C3-C2	-3.37	103.83	110.01
3	A	502	BGC	O5-C1-C2	-2.09	106.34	110.00
3	B	502	BGC	O4-C4-C3	2.03	114.94	110.36
3	A	503	BGC	O3-C3-C4	2.47	115.94	110.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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(Å²)	Q<0.9
1	A	425/426 (99%)	-0.50	1 (0%)	95	94	5, 9, 17, 57	0
1	B	425/426 (99%)	-0.34	2 (0%)	91	92	6, 13, 25, 34	0
All	All	850/852 (99%)	-0.42	3 (0%)	93	92	5, 11, 23, 57	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	410	TYR	4.2
1	B	371	PRO	2.7
1	A	9	GLN	2.5

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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.

Full wwPDB X-ray Structure Validation Report

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å²)	Q<0.9
3	BGC	B	501	12/12	0.95	0.09	0.77	11,15,20,22	12
3	BGC	A	503	11/12	0.92	0.10	0.65	14,17,23,23	11
3	BGC	B	502	11/12	0.96	0.07	0.45	11,13,16,18	11
3	BGC	A	502	12/12	0.94	0.09	0.43	15,17,22,24	12
2	CA	B	503	1/1	1.00	0.04	-0.92	6,6,6,6	0
2	CA	A	501	1/1	1.00	0.04	-2.46	7,7,7,7	0
4	NA	A	506	1/1	1.00	0.03	-2.68	16,16,16,16	0
5	CL	B	504	1/1	1.00	0.03	-4.67	16,16,16,16	0
4	NA	A	505	1/1	0.99	0.22	-	29,29,29,29	0
4	NA	B	506	1/1	0.92	0.09	-	41,41,41,41	0
4	NA	A	504	1/1	0.99	0.13	-	34,34,34,34	0
4	NA	B	505	1/1	0.98	0.14	-	33,33,33,33	0

6.5 Other polymers (i)

There are no such residues in this entry.