Full wwPDB X-ray Structure Validation Report ⓘ

Sep 10, 2023 – 04:30 AM EDT

PDB ID	:	4ILG
Title	:	Crystal structure of Aar2p in complex with the Prp8p RNaseH and Jab1/MPN domains
Authors	:	Weber, G.; Heroven, A.C.; Santos, K.F.; Wahl, M.C.
Deposited on	:	2012-12-31
Resolution	:	2.10 Å(reported)

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	:	2.35.1
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
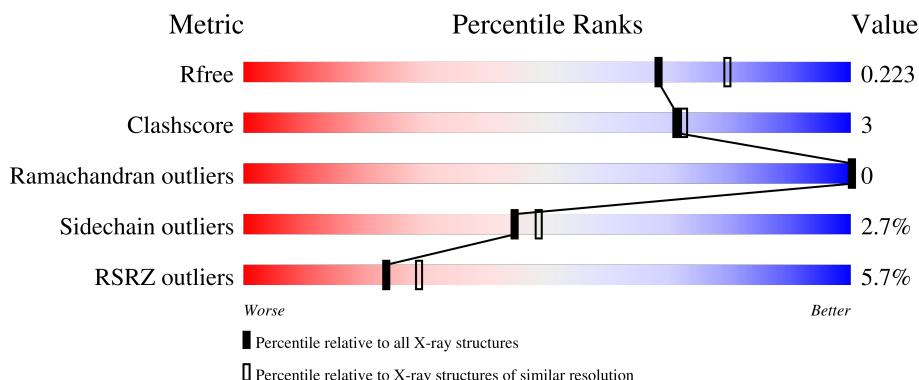
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION
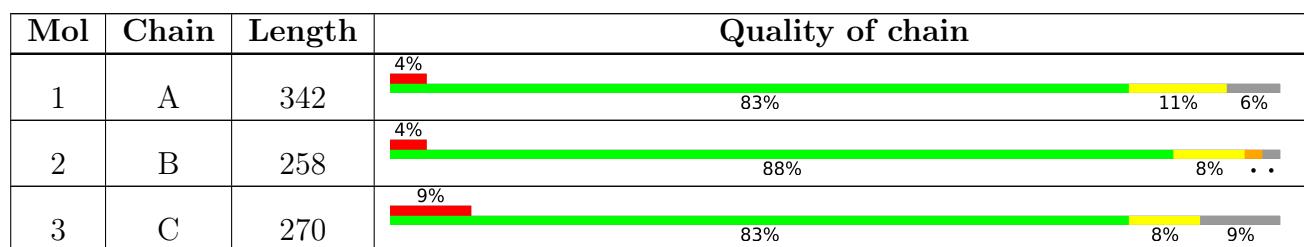
The reported resolution of this entry is 2.10 Å.

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	Percentile Ranks	Value
Rfree		0.223
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		2.7%
RSRZ outliers		5.7%

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}	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	342	4% — 83% — 11% — 6%
2	B	258	4% — 88% — 8% — · ·
3	C	270	9% — 83% — 8% — 9%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 7524 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 A1 cistron-splicing factor AAR2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	321	Total 2710	C 1731	N 444	O 514	S 21	0	7	0

There are 15 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	153	SER	LEU	engineered mutation	UNP P32357
A	154	SER	LYS	engineered mutation	UNP P32357
A	?	-	LEU	deletion	UNP P32357
A	?	-	GLN	deletion	UNP P32357
A	?	-	LYS	deletion	UNP P32357
A	?	-	ALA	deletion	UNP P32357
A	?	-	GLY	deletion	UNP P32357
A	?	-	LYS	deletion	UNP P32357
A	?	-	MET	deletion	UNP P32357
A	?	-	GLU	deletion	UNP P32357
A	?	-	ALA	deletion	UNP P32357
A	?	-	LYS	deletion	UNP P32357
A	?	-	ASN	deletion	UNP P32357
A	?	-	GLU	deletion	UNP P32357
A	?	-	ASP	deletion	UNP P32357

- Molecule 2 is a protein called Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	252	Total 2129	C 1364	N 359	O 396	S 10	0	14	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1833	GLY	-	expression tag	UNP P33334

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1834	ALA	-	expression tag	UNP P33334
B	1835	MET	-	expression tag	UNP P33334
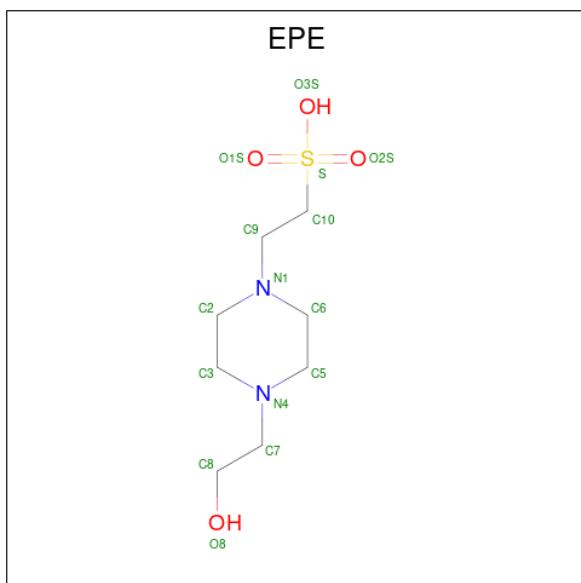
- Molecule 3 is a protein called Pre-mRNA-splicing factor 8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	245	Total 1981	C 1279	N 320	O 376	S 6	0	4	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	2144	GLY	-	expression tag	UNP P33334
C	2145	ALA	-	expression tag	UNP P33334
C	2146	MET	-	expression tag	UNP P33334

- Molecule 4 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).

EPE

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	C	1	Total 15	C 8	N 2	O 4	S 1	0	0
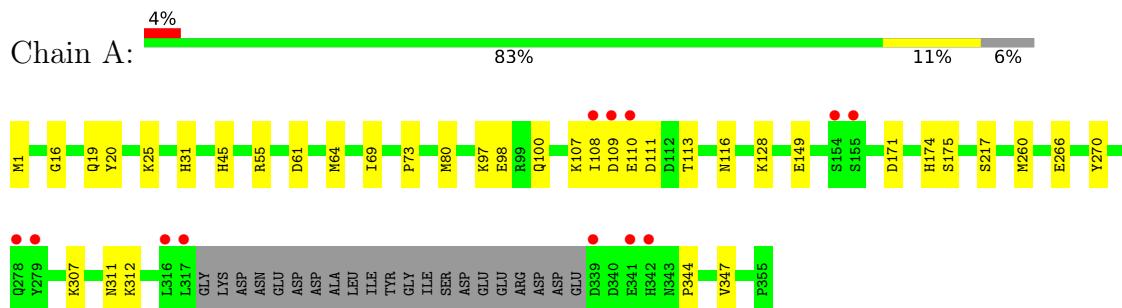
- Molecule 5 is water.

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	245	Total O 245 245	0	0
5	B	241	Total O 241 241	0	0
5	C	203	Total O 203 203	0	0
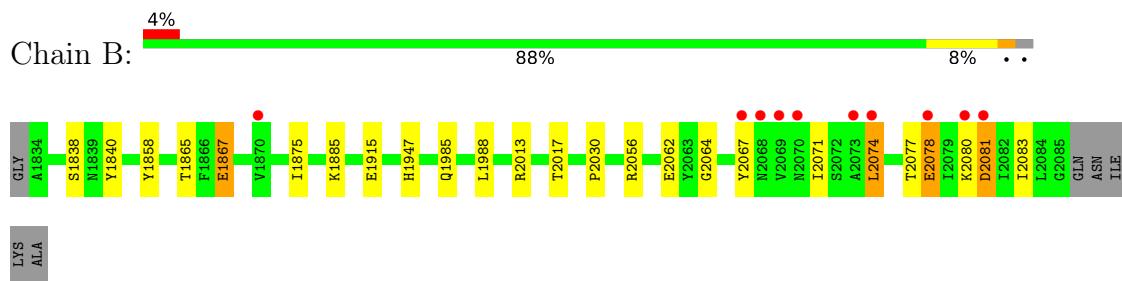
3 Residue-property plots ⓘ

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.
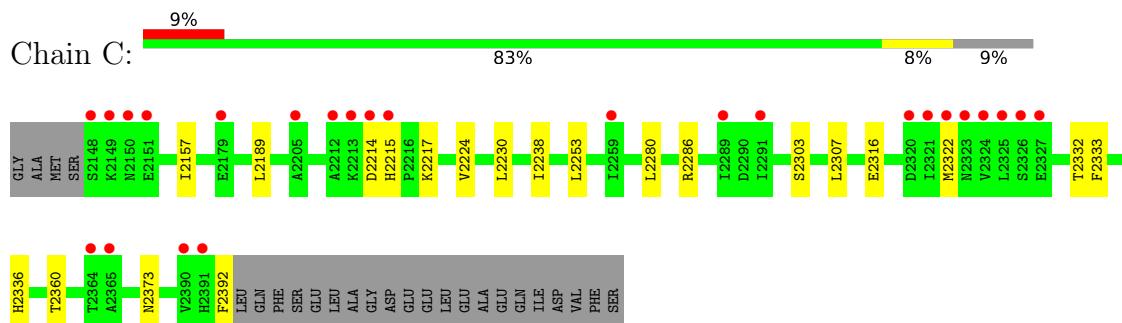
• Molecule 1: A1 cistron-splicing factor AAR2

• Molecule 2: Pre-mRNA-splicing factor 8

• Molecule 3: Pre-mRNA-splicing factor 8

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	84.77Å 63.74Å 110.52Å 90.00° 95.39° 90.00°	Depositor
Resolution (Å)	31.87 – 2.10 31.87 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.7 (31.87-2.10) 99.7 (31.87-2.10)	Depositor EDS
R_{merge}	0.21	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	2.16 (at 2.10Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.180 , 0.228 0.177 , 0.223	Depositor DCC
R_{free} test set	3064 reflections (4.46%)	wwPDB-VP
Wilson B-factor (Å^2)	30.0	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 59.6	EDS
L-test for twinning[2]	$< |L| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7524	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.25% 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: EPE

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	0.49	0/2791	0.57	0/3768
2	B	0.54	0/2186	0.61	0/2961
3	C	0.45	0/2038	0.57	0/2763
All	All	0.49	0/7015	0.58	0/9492

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 ⓘ

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	2710	0	2586	22	1
2	B	2129	0	2173	17	0
3	C	1981	0	1956	10	0
4	C	15	0	17	1	0
5	A	245	0	0	5	1
5	B	241	0	0	3	0
5	C	203	0	0	2	0
All	All	7524	0	6732	47	1

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 3.

All (47) 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)
2:B:2013[B]:ARG:NH1	2:B:2083:ILE:O	2.05	0.88
1:A:108:ILE:HB	1:A:111:ASP:HB2	1.55	0.87
2:B:2078:GLU:HB3	2:B:2081:ASP:HB2	1.59	0.84
1:A:307:LYS:NZ	1:A:311[A]:ASN:OD1	2.28	0.66
1:A:107:LYS:HG3	1:A:109:ASP:H	1.61	0.66
1:A:149:GLU:OE1	5:A:638:HOH:O	2.14	0.65
1:A:97:LYS:O	1:A:100:GLN:HG3	2.00	0.61
2:B:2080:LYS:HD2	2:B:2080:LYS:H	1.67	0.60
2:B:1985[A]:GLN:NE2	5:B:2341:HOH:O	2.20	0.59
1:A:347:VAL:HG21	2:B:1875:ILE:HD11	1.85	0.59
3:C:2189:LEU:HD13	3:C:2224:VAL:HG23	1.87	0.57
2:B:2064:GLY:HA3	2:B:2071:ILE:HD11	1.88	0.56
1:A:128:LYS:NZ	5:A:642:HOH:O	2.35	0.56
3:C:2286:ARG:NH2	5:C:2678:HOH:O	2.38	0.56
1:A:1[A]:MET:HE1	5:A:609:HOH:O	2.04	0.56
1:A:31:HIS:CG	1:A:98:GLU:HG2	2.43	0.53
2:B:1838:SER:C	2:B:1840:TYR:H	2.13	0.51
1:A:31:HIS:CD2	1:A:98:GLU:HG2	2.46	0.51
2:B:2017:THR:HG23	2:B:2062[B]:GLU:HG2	1.94	0.49
4:C:2501:EPE:H82	4:C:2501:EPE:H51	1.55	0.48
1:A:107:LYS:HG3	1:A:109:ASP:N	2.29	0.48
1:A:61:ASP:HB3	1:A:64[A]:MET:SD	2.54	0.48
2:B:2078:GLU:CB	2:B:2081:ASP:HB2	2.39	0.47
1:A:266:GLU:HB2	5:A:450:HOH:O	2.15	0.47
1:A:116:ASN:HB3	1:A:270:TYR:CD2	2.51	0.46
2:B:1865:THR:OG1	2:B:1867:GLU:HG2	2.16	0.45
2:B:1885[B]:LYS:HE3	5:B:2120:HOH:O	2.16	0.45
2:B:2071:ILE:H	2:B:2071:ILE:HD12	1.82	0.45
3:C:2307:LEU:HD12	3:C:2333:PHE:O	2.17	0.45
1:A:344:PRO:HB3	2:B:1858:TYR:CE2	2.53	0.44
2:B:2071:ILE:O	2:B:2074:LEU:HB2	2.17	0.44
3:C:2238[A]:ILE:HG21	3:C:2280:LEU:HD11	2.00	0.44
2:B:1915:GLU:HG2	2:B:1947:HIS:CE1	2.53	0.43
2:B:2056:ARG:HD3	5:B:2303:HOH:O	2.19	0.43
1:A:111:ASP:OD1	1:A:113:THR:OG1	2.16	0.43
3:C:2215:HIS:CE1	3:C:2217[A]:LYS:HB2	2.53	0.43
1:A:69:ILE:HD13	1:A:80:MET:HA	2.01	0.43
3:C:2215:HIS:CE1	3:C:2217[B]:LYS:HB2	2.54	0.42
1:A:171:ASP:HB3	1:A:174[A]:HIS:ND1	2.35	0.42
1:A:108:ILE:O	1:A:110:GLU:N	2.45	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:2214:ASP:OD2	3:C:2215:HIS:N	2.53	0.41
3:C:2336:HIS:HE1	5:C:2692:HOH:O	2.03	0.41
1:A:20:TYR:HB3	5:A:555:HOH:O	2.20	0.41
1:A:217:SER:HB2	1:A:260:MET:HG2	2.03	0.41
2:B:2030:PRO:HG3	3:C:2157:ILE:HD11	2.02	0.41
3:C:2253:LEU:HD23	3:C:2253:LEU:HA	1.87	0.41
1:A:16:GLY:HA3	1:A:45:HIS:CE1	2.57	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:73:PRO:O	5:A:612:HOH:O[2_454]	2.17	0.03

5.3　Torsion angles ⓘ

5.3.1　Protein backbone ⓘ

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	323/342 (94%)	312 (97%)	11 (3%)	0	100	100
2	B	263/258 (102%)	251 (95%)	12 (5%)	0	100	100
3	C	247/270 (92%)	242 (98%)	5 (2%)	0	100	100
All	All	833/870 (96%)	805 (97%)	28 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2　Protein sidechains ⓘ

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	305/316 (96%)	300 (98%)	5 (2%)	62	69
2	B	242/233 (104%)	235 (97%)	7 (3%)	42	46
3	C	221/237 (93%)	213 (96%)	8 (4%)	35	36
All	All	768/786 (98%)	748 (97%)	20 (3%)	44	50

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	19	GLN
1	A	25	LYS
1	A	55	ARG
1	A	175	SER
1	A	312	LYS
2	B	1867	GLU
2	B	1988	LEU
2	B	2067	TYR
2	B	2074	LEU
2	B	2077	THR
2	B	2078	GLU
2	B	2081	ASP
3	C	2230	LEU
3	C	2303	SER
3	C	2316	GLU
3	C	2322	MET
3	C	2332	THR
3	C	2360	THR
3	C	2373	ASN
3	C	2392	PHE

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 ⓘ

1 ligand is 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	$\#|Z| > 2$
4	EPE	C	2501	-	15,15,15	1.10	1 (6%)	18,20,20	1.81	5 (27%)

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
4	EPE	C	2501	-	-	2/9/19/19	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	2501	EPE	C10-S	3.85	1.83	1.77

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	2501	EPE	C5-N4-C3	4.42	118.77	108.83
4	C	2501	EPE	C7-N4-C5	3.50	120.19	111.23
4	C	2501	EPE	C7-N4-C3	3.01	118.93	111.23

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	2501	EPE	O3S-S-C10	2.60	109.97	105.77
4	C	2501	EPE	O2S-S-C10	2.16	109.51	106.92

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	2501	EPE	C8-C7-N4-C5
4	C	2501	EPE	N4-C7-C8-O8

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	2501	EPE	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, 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	321/342 (93%)	-0.10	12 (3%)	41	48	20, 42, 99, 173	0
2	B	252/258 (97%)	-0.07	10 (3%)	38	44	18, 33, 98, 123	0
3	C	245/270 (90%)	0.17	25 (10%)	6	8	24, 41, 104, 168	0
All	All	818/870 (94%)	-0.01	47 (5%)	23	29	18, 39, 101, 173	0

All (47) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	2325	LEU	7.9
3	C	2321	ILE	7.4
1	A	109	ASP	5.4
1	A	279	TYR	5.2
3	C	2323	ASN	4.7
1	A	278	GLN	4.6
3	C	2324	VAL	4.5
3	C	2214	ASP	4.3
2	B	2073	ALA	4.1
3	C	2326	SER	4.0
3	C	2390	VAL	4.0
3	C	2215	HIS	3.9
1	A	317	LEU	3.7
3	C	2322	MET	3.6
3	C	2391	HIS	3.6
2	B	2081	ASP	3.6
3	C	2327	GLU	3.5
2	B	2067	TYR	3.5
2	B	2068	ASN	3.3
1	A	341	GLU	3.2
2	B	2074	LEU	3.2
1	A	110	GLU	3.0
3	C	2213	LYS	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	108	ILE	2.8
2	B	2078	GLU	2.8
2	B	2070	ASN	2.8
1	A	342	HIS	2.7
1	A	154	SER	2.7
3	C	2151	GLU	2.7
3	C	2150	ASN	2.6
3	C	2149	LYS	2.6
3	C	2365	ALA	2.5
2	B	2069	VAL	2.4
3	C	2212	ALA	2.4
2	B	1870	VAL	2.4
1	A	339	ASP	2.3
2	B	2080	LYS	2.3
1	A	316	LEU	2.3
3	C	2364	THR	2.2
3	C	2179	GLU	2.2
3	C	2205	ALA	2.2
1	A	155	SER	2.2
3	C	2289	ILE	2.1
3	C	2320	ASP	2.1
3	C	2259	ILE	2.1
3	C	2148	SER	2.1
3	C	2291	ILE	2.0

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 ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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
4	EPE	C	2501	15/15	0.96	0.11	43,56,73,80	0

6.5 Other polymers ⓘ

There are no such residues in this entry.