



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 19, 2023 – 10:55 AM EDT

PDB ID : 2RGK
Title : Functional annotation of Escherichia coli yihS-encoded protein
Authors : Itoh, T.; Mikami, B.; Hashimoto, W.; Murata, K.
Deposited on : 2007-10-03
Resolution : 2.50 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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

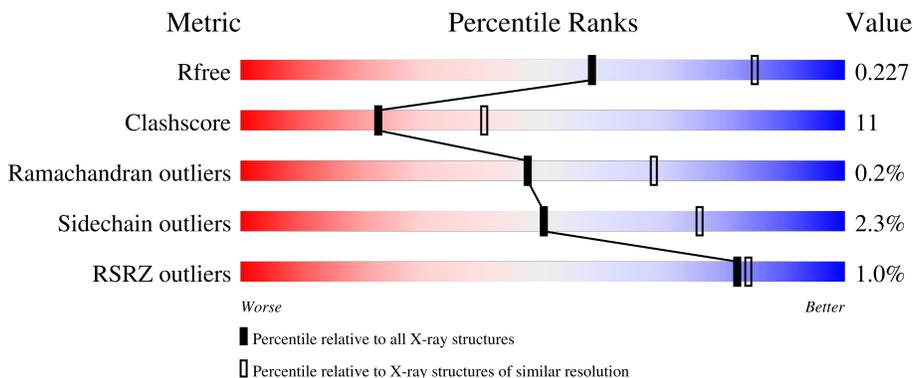
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 $\leq 5\%$. 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	421	 76% 20% ..
1	B	421	 79% 18% ..
1	C	421	 75% 20% ..
1	D	421	 75% 21% ..
1	E	421	 76% 20% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	421	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into segments: a small red segment at the start, followed by a large green segment labeled '77%', then a yellow segment labeled '19%', and finally a small grey segment at the end. Above the bar is a '%' symbol and below it are two dots '••'.</p>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 21377 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 Uncharacterized sugar isomerase yihS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	409	3317	2128	575	598	16	0	0	0
1	B	409	3317	2128	575	598	16	0	0	0
1	C	409	3317	2128	575	598	16	0	0	0
1	D	409	3317	2128	575	598	16	0	0	0
1	E	409	3317	2128	575	598	16	0	0	0
1	F	409	3317	2128	575	598	16	0	0	0

There are 48 discrepancies between the modelled and reference sequences:

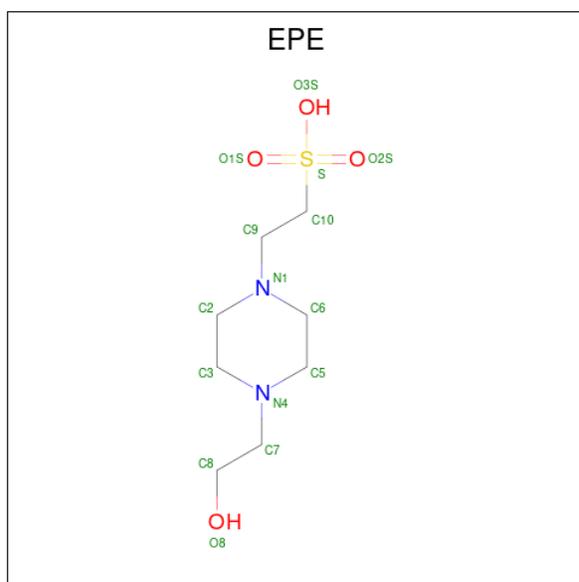
Chain	Residue	Modelled	Actual	Comment	Reference
A	414	LEU	-	expression tag	UNP P32140
A	415	GLU	-	expression tag	UNP P32140
A	416	HIS	-	expression tag	UNP P32140
A	417	HIS	-	expression tag	UNP P32140
A	418	HIS	-	expression tag	UNP P32140
A	419	HIS	-	expression tag	UNP P32140
A	420	HIS	-	expression tag	UNP P32140
A	421	HIS	-	expression tag	UNP P32140
B	414	LEU	-	expression tag	UNP P32140
B	415	GLU	-	expression tag	UNP P32140
B	416	HIS	-	expression tag	UNP P32140
B	417	HIS	-	expression tag	UNP P32140
B	418	HIS	-	expression tag	UNP P32140
B	419	HIS	-	expression tag	UNP P32140
B	420	HIS	-	expression tag	UNP P32140
B	421	HIS	-	expression tag	UNP P32140
C	414	LEU	-	expression tag	UNP P32140

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	415	GLU	-	expression tag	UNP P32140
C	416	HIS	-	expression tag	UNP P32140
C	417	HIS	-	expression tag	UNP P32140
C	418	HIS	-	expression tag	UNP P32140
C	419	HIS	-	expression tag	UNP P32140
C	420	HIS	-	expression tag	UNP P32140
C	421	HIS	-	expression tag	UNP P32140
D	414	LEU	-	expression tag	UNP P32140
D	415	GLU	-	expression tag	UNP P32140
D	416	HIS	-	expression tag	UNP P32140
D	417	HIS	-	expression tag	UNP P32140
D	418	HIS	-	expression tag	UNP P32140
D	419	HIS	-	expression tag	UNP P32140
D	420	HIS	-	expression tag	UNP P32140
D	421	HIS	-	expression tag	UNP P32140
E	414	LEU	-	expression tag	UNP P32140
E	415	GLU	-	expression tag	UNP P32140
E	416	HIS	-	expression tag	UNP P32140
E	417	HIS	-	expression tag	UNP P32140
E	418	HIS	-	expression tag	UNP P32140
E	419	HIS	-	expression tag	UNP P32140
E	420	HIS	-	expression tag	UNP P32140
E	421	HIS	-	expression tag	UNP P32140
F	414	LEU	-	expression tag	UNP P32140
F	415	GLU	-	expression tag	UNP P32140
F	416	HIS	-	expression tag	UNP P32140
F	417	HIS	-	expression tag	UNP P32140
F	418	HIS	-	expression tag	UNP P32140
F	419	HIS	-	expression tag	UNP P32140
F	420	HIS	-	expression tag	UNP P32140
F	421	HIS	-	expression tag	UNP P32140

- Molecule 2 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: C₈H₁₈N₂O₄S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	Total 15	8	2	4	1	0	0
2	B	1	Total 15	8	2	4	1	0	0
2	C	1	Total 15	8	2	4	1	0	0
2	D	1	Total 15	8	2	4	1	0	0
2	E	1	Total 15	8	2	4	1	0	0
2	F	1	Total 15	8	2	4	1	0	0

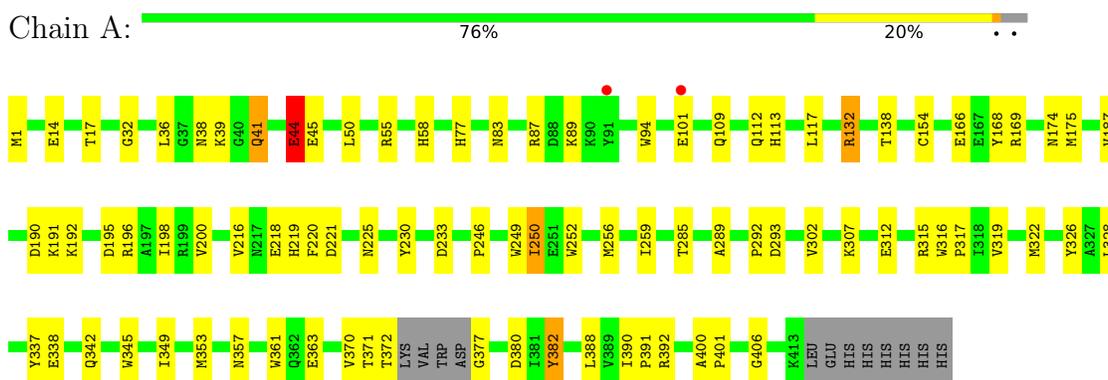
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	247	Total 247	247	0	0
3	B	213	Total 213	213	0	0
3	C	214	Total 214	214	0	0
3	D	244	Total 244	244	0	0
3	E	224	Total 224	224	0	0
3	F	243	Total 243	243	0	0

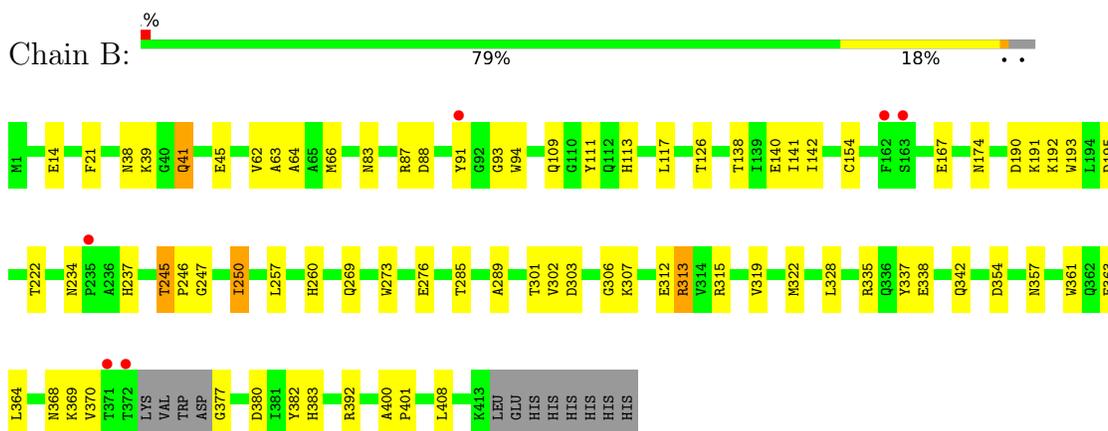
3 Residue-property plots

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.

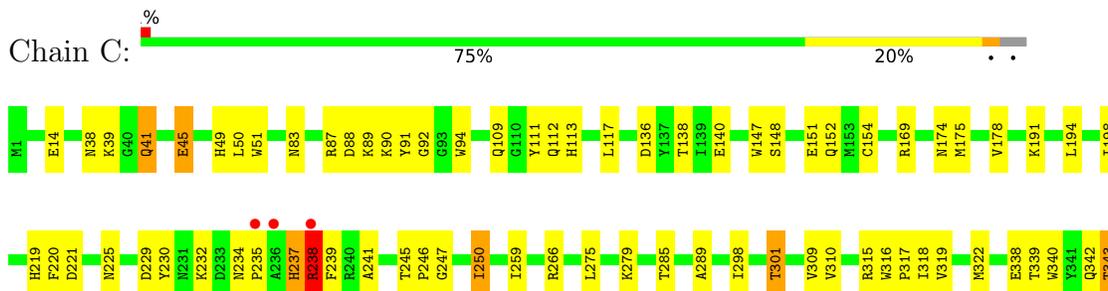
- Molecule 1: Uncharacterized sugar isomerase yihS



- Molecule 1: Uncharacterized sugar isomerase yihS

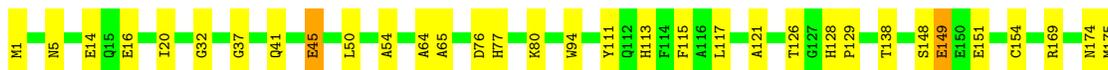
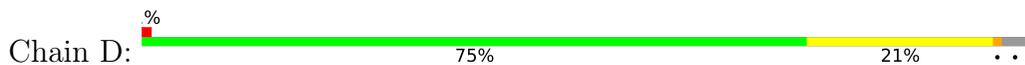


- Molecule 1: Uncharacterized sugar isomerase yihS

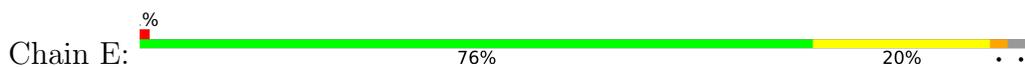




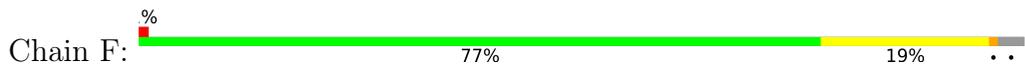
- Molecule 1: Uncharacterized sugar isomerase yihS



- Molecule 1: Uncharacterized sugar isomerase yihS



- Molecule 1: Uncharacterized sugar isomerase yihS



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	245.87Å 142.16Å 94.59Å 90.00° 107.61° 90.00°	Depositor
Resolution (Å)	14.97 – 2.50 48.66 – 2.50	Depositor EDS
% Data completeness (in resolution range)	96.5 (14.97-2.50) 96.6 (48.66-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	12.79 (at 2.51Å)	Xtrriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.180 , 0.236 0.172 , 0.227	Depositor DCC
R_{free} test set	10377 reflections (10.05%)	wwPDB-VP
Wilson B-factor (Å ²)	18.6	Xtrriage
Anisotropy	0.888	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 50.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	21377	wwPDB-VP
Average B, all atoms (Å ²)	20.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.83% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: 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	0.33	0/3421	0.57	0/4657
1	B	0.32	0/3421	0.56	0/4657
1	C	0.33	0/3421	0.57	0/4657
1	D	0.34	0/3421	0.58	0/4657
1	E	0.34	0/3421	0.58	0/4657
1	F	0.34	0/3421	0.57	0/4657
All	All	0.33	0/20526	0.57	0/27942

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 [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	3317	0	3155	78	0
1	B	3317	0	3155	74	0
1	C	3317	0	3155	91	0
1	D	3317	0	3155	71	0
1	E	3317	0	3155	75	0
1	F	3317	0	3155	68	0
2	A	15	0	18	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	15	0	18	1	0
2	C	15	0	18	1	0
2	D	15	0	18	1	0
2	E	15	0	18	3	0
2	F	15	0	18	0	0
3	A	247	0	0	8	0
3	B	213	0	0	7	0
3	C	214	0	0	4	0
3	D	244	0	0	8	0
3	E	224	0	0	6	0
3	F	243	0	0	10	0
All	All	21377	0	19038	447	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 11.

The worst 5 of 447 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:371:THR:HG22	1:A:372:THR:H	1.28	0.99
1:E:379:GLN:H	1:E:379:GLN:NE2	1.62	0.97
1:E:154:CYS:H	1:E:174:ASN:HD21	1.14	0.94
1:B:154:CYS:H	1:B:174:ASN:HD21	1.13	0.93
1:E:379:GLN:H	1:E:379:GLN:HE21	1.17	0.92

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	405/421 (96%)	390 (96%)	14 (4%)	1 (0%)	47 68

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	405/421 (96%)	388 (96%)	17 (4%)	0	100	100
1	C	405/421 (96%)	388 (96%)	13 (3%)	4 (1%)	15	28
1	D	405/421 (96%)	392 (97%)	13 (3%)	0	100	100
1	E	405/421 (96%)	393 (97%)	12 (3%)	0	100	100
1	F	405/421 (96%)	393 (97%)	12 (3%)	0	100	100
All	All	2430/2526 (96%)	2344 (96%)	81 (3%)	5 (0%)	47	68

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	44	GLU
1	C	191	LYS
1	C	235	PRO
1	C	237	HIS
1	C	238	ARG

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	335/347 (96%)	325 (97%)	10 (3%)	41	68
1	B	335/347 (96%)	329 (98%)	6 (2%)	59	81
1	C	335/347 (96%)	328 (98%)	7 (2%)	53	78
1	D	335/347 (96%)	328 (98%)	7 (2%)	53	78
1	E	335/347 (96%)	325 (97%)	10 (3%)	41	68
1	F	335/347 (96%)	328 (98%)	7 (2%)	53	78
All	All	2010/2082 (96%)	1963 (98%)	47 (2%)	50	76

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	365	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	302	VAL
1	D	382	TYR
1	E	231	ASN
1	E	343	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 74 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	234	ASN
1	F	269	GLN
1	E	269	GLN
1	F	113	HIS
1	B	336	GLN

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](#)

6 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	EPE	B	3002	-	15,15,15	1.05	1 (6%)	18,20,20	1.60	3 (16%)
2	EPE	E	3005	-	15,15,15	0.90	0	18,20,20	1.35	2 (11%)
2	EPE	D	3004	-	15,15,15	1.00	0	18,20,20	1.34	2 (11%)
2	EPE	A	3001	-	15,15,15	1.22	1 (6%)	18,20,20	1.60	3 (16%)
2	EPE	C	3003	-	15,15,15	1.06	1 (6%)	18,20,20	1.38	2 (11%)
2	EPE	F	3006	-	15,15,15	1.08	1 (6%)	18,20,20	5.85	6 (33%)

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	EPE	B	3002	-	-	0/9/19/19	0/1/1/1
2	EPE	E	3005	-	-	0/9/19/19	0/1/1/1
2	EPE	D	3004	-	-	1/9/19/19	0/1/1/1
2	EPE	A	3001	-	-	0/9/19/19	0/1/1/1
2	EPE	C	3003	-	-	1/9/19/19	0/1/1/1
2	EPE	F	3006	-	-	3/9/19/19	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	3001	EPE	C10-S	3.20	1.82	1.77
2	F	3006	EPE	C10-S	2.74	1.81	1.77
2	C	3003	EPE	C10-S	2.73	1.81	1.77
2	B	3002	EPE	C10-S	2.67	1.81	1.77

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	3006	EPE	O3S-S-C10	-13.41	84.08	105.77
2	F	3006	EPE	O3S-S-O1S	-12.24	81.37	111.27
2	F	3006	EPE	O3S-S-O2S	-11.98	82.00	111.27
2	F	3006	EPE	O1S-S-C10	10.32	119.34	106.92
2	A	3001	EPE	O1S-S-C10	4.91	112.83	106.92

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	3006	EPE	C9-C10-S-O1S
2	F	3006	EPE	C9-C10-S-O2S
2	C	3003	EPE	N4-C7-C8-O8
2	F	3006	EPE	N4-C7-C8-O8
2	D	3004	EPE	N4-C7-C8-O8

There are no ring outliers.

5 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	3002	EPE	1	0
2	E	3005	EPE	3	0
2	D	3004	EPE	1	0
2	A	3001	EPE	2	0
2	C	3003	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 [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	409/421 (97%)	-0.31	2 (0%) 91 91	12, 19, 34, 47	0
1	B	409/421 (97%)	-0.11	6 (1%) 73 75	11, 20, 36, 56	0
1	C	409/421 (97%)	-0.25	5 (1%) 79 80	10, 19, 33, 70	0
1	D	409/421 (97%)	-0.28	4 (0%) 82 84	6, 16, 31, 52	0
1	E	409/421 (97%)	-0.20	4 (0%) 82 84	7, 16, 32, 50	0
1	F	409/421 (97%)	-0.32	3 (0%) 87 89	8, 17, 32, 47	0
All	All	2454/2526 (97%)	-0.25	24 (0%) 82 84	6, 18, 33, 70	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	372	THR	4.5
1	C	235	PRO	4.5
1	C	372	THR	4.0
1	C	236	ALA	3.8
1	C	238	ARG	3.4

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(\AA^2)	Q<0.9
2	EPE	B	3002	15/15	0.90	0.35	43,57,65,67	0
2	EPE	A	3001	15/15	0.92	0.23	40,42,49,51	0
2	EPE	F	3006	15/15	0.93	0.27	29,39,48,52	0
2	EPE	D	3004	15/15	0.96	0.24	25,38,44,48	0
2	EPE	C	3003	15/15	0.96	0.23	32,40,48,51	0
2	EPE	E	3005	15/15	0.97	0.27	32,42,44,47	0

6.5 Other polymers [i](#)

There are no such residues in this entry.