



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 03:13 pm GMT

PDB ID : 4BPT  
Title : Structural and thermodynamic insight into phenylalanine hydroxylase from the human pathogen *Legionella pneumophila*  
Authors : Leiros, H.-K.S.; Flydal, M.I.; Martinez, A.  
Deposited on : 2013-05-28  
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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

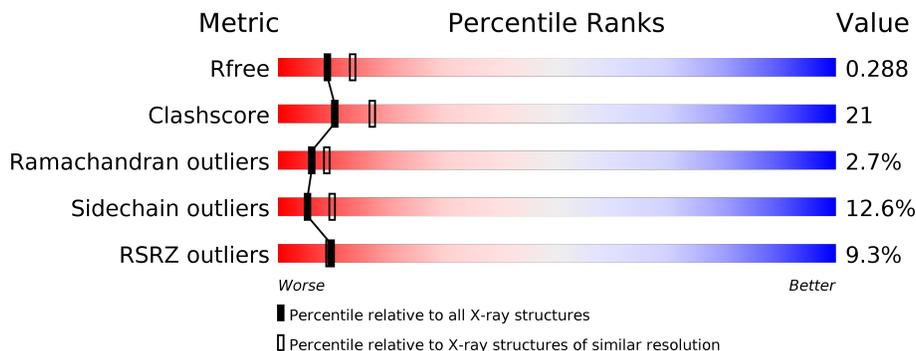
# 1 Overall quality at a glance

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))
$R_{free}$	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	272	 57% 26% 8% 7%
1	B	272	 56% 26% 8% 7%
1	C	272	 17% 45% 29% 8% 1% 15%
1	D	272	 15% 50% 26% 8% 1% 13%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	PEG	A	1260	-	-	X	-
2	PEG	A	1261	-	-	X	-
2	PEG	A	1262	-	-	X	-
2	PEG	B	1260	-	-	X	-

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8155 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.

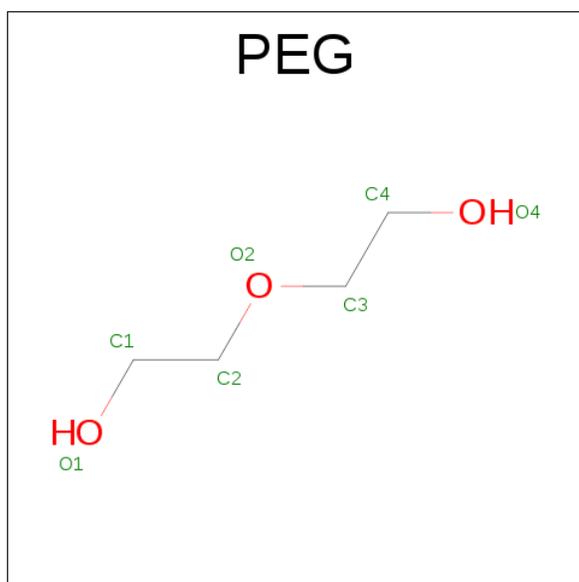
- Molecule 1 is a protein called PHENYLALANINE-4-HYDROXYLASE (PAH) (PHE-4-MO NOOXYGENASE).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	252	Total 2062	C 1345	N 331	O 377	S 9	5	1	0
1	B	252	Total 2053	C 1340	N 330	O 374	S 9	10	0	0
1	C	232	Total 1900	C 1246	N 302	O 343	S 9	193	0	0
1	D	237	Total 1943	C 1272	N 311	O 351	S 9	135	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	60	THR	ILE	CONFLICT	UNP I7HM43
B	60	THR	ILE	CONFLICT	UNP I7HM43
C	60	THR	ILE	CONFLICT	UNP I7HM43
D	60	THR	ILE	CONFLICT	UNP I7HM43

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



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

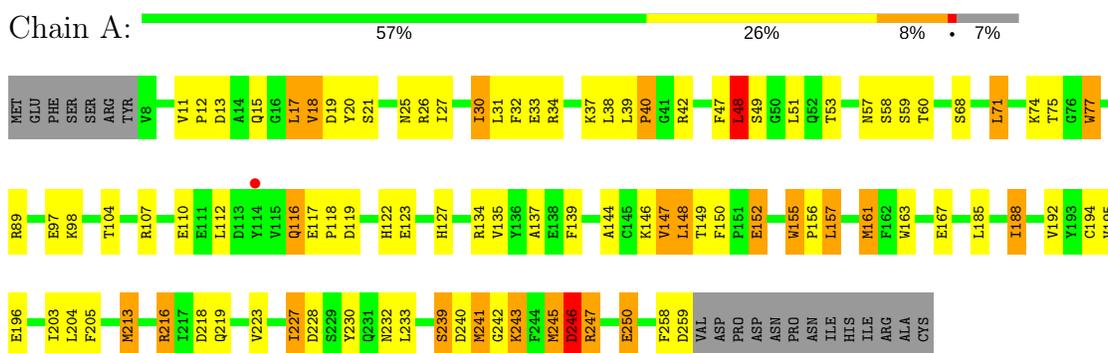
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	56	Total O 56 56	0	0
3	B	51	Total O 51 51	0	0
3	C	27	Total O 27 27	0	0
3	D	21	Total O 21 21	0	0

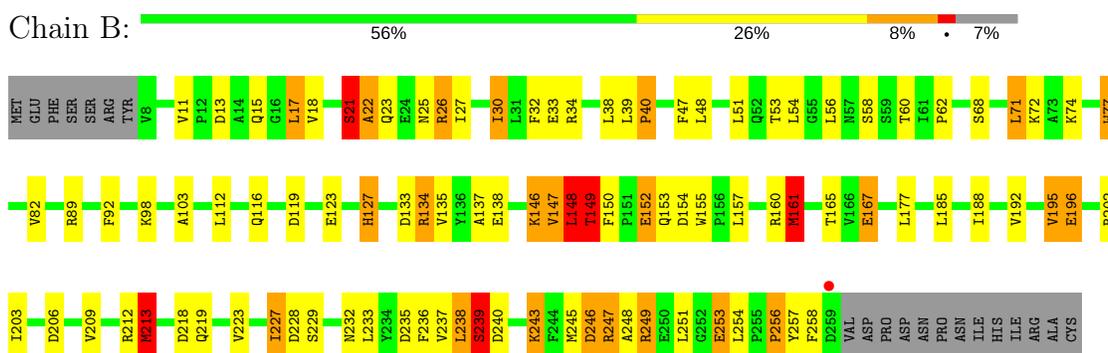
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

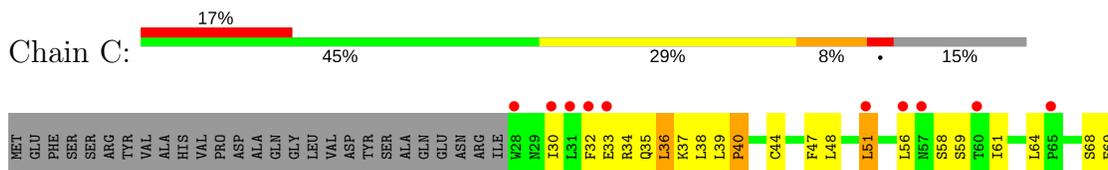
- Molecule 1: PHENYLALANINE-4-HYDROXYLASE (PAH) (PHE-4-MONOOXYGENASE)

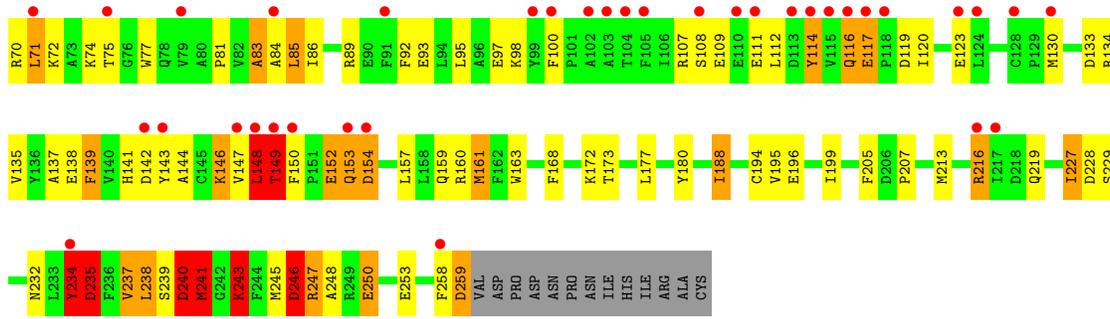


- Molecule 1: PHENYLALANINE-4-HYDROXYLASE (PAH) (PHE-4-MONOOXYGENASE)

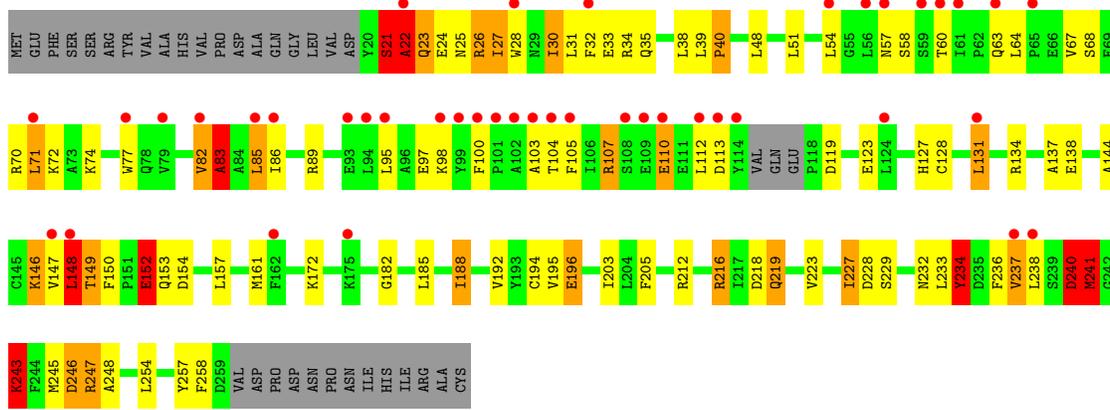


- Molecule 1: PHENYLALANINE-4-HYDROXYLASE (PAH) (PHE-4-MONOOXYGENASE)





• Molecule 1: PHENYLALANINE-4-HYDROXYLASE (PAH) (PHE-4-MONOOXYGENASE )



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.32Å 60.12Å 124.04Å 90.00° 94.07° 90.00°	Depositor
Resolution (Å)	10.00 – 2.50 24.32 – 2.50	Depositor EDS
% Data completeness (in resolution range)	90.4 (10.00-2.50) 90.3 (24.32-2.50)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.35 (at 2.50Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.264 , 0.302 0.252 , 0.288	Depositor DCC
$R_{free}$ test set	1290 reflections (3.18%)	DCC
Wilson B-factor (Å <sup>2</sup> )	38.6	Xtrriage
Anisotropy	0.262	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 31.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	8155	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.43% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: 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	1.44	6/2121 (0.3%)	1.14	11/2885 (0.4%)
1	B	1.70	9/2112 (0.4%)	1.24	16/2873 (0.6%)
1	C	2.82	48/1956 (2.5%)	2.67	78/2659 (2.9%)
1	D	2.11	32/1999 (1.6%)	1.99	44/2714 (1.6%)
All	All	2.07	95/8188 (1.2%)	1.84	149/11131 (1.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	1	4
1	C	1	6
1	D	0	6
All	All	2	17

The worst 5 of 95 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	117	GLU	CG-CD	-48.09	0.79	1.51
1	B	239	SER	CB-OG	-45.23	0.83	1.42
1	A	110	GLU	CB-CG	37.99	2.24	1.52
1	D	63	GLN	CG-CD	37.46	2.37	1.51
1	C	146	LYS	CE-NZ	33.07	2.31	1.49

The worst 5 of 149 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	83	ALA	O-C-N	-38.34	61.35	122.70
1	C	234	TYR	CB-CG-CD2	34.79	141.88	121.00
1	C	117	GLU	CG-CD-OE1	-31.88	54.55	118.30
1	D	234	TYR	CG-CD1-CE1	-29.91	97.37	121.30
1	D	107	ARG	CD-NE-CZ	-29.57	82.20	123.60

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	B	21	SER	CA
1	C	259	ASP	CA

5 of 17 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	147	VAL	Peptide
1	B	147	VAL	Peptide
1	B	148	LEU	Peptide
1	B	21	SER	Peptide
1	B	253	GLU	Sidechain

## 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	2062	0	2010	87	0
1	B	2053	0	2005	89	0
1	C	1900	0	1857	67	0
1	D	1943	0	1900	59	0
2	A	21	0	30	14	0
2	B	21	0	30	8	0
3	A	56	0	0	5	0
3	B	51	0	0	13	0
3	C	27	0	0	3	0
3	D	21	0	0	0	0
All	All	8155	0	7832	316	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 21.

The worst 5 of 316 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:C:44:CYS:SG	1:C:130:MET:HE1	1.57	1.42
1:C:44:CYS:SG	1:C:130:MET:CE	2.13	1.34
1:B:26:ARG:N	3:B:2006:HOH:O	1.71	1.21
1:B:27:ILE:N	3:B:2006:HOH:O	1.73	1.19
1:C:240:ASP:HA	1:C:241:MET:HB2	1.24	1.19

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	251/272 (92%)	227 (90%)	20 (8%)	4 (2%)	11	19
1	B	250/272 (92%)	223 (89%)	21 (8%)	6 (2%)	7	11
1	C	230/272 (85%)	203 (88%)	21 (9%)	6 (3%)	6	9
1	D	233/272 (86%)	206 (88%)	17 (7%)	10 (4%)	3	3
All	All	964/1088 (89%)	859 (89%)	79 (8%)	26 (3%)	6	9

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	241	MET
1	A	247	ARG
1	B	22	ALA
1	B	247	ARG
1	C	84	ALA

### 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	223/241 (92%)	195 (87%)	28 (13%)	5	10
1	B	222/241 (92%)	195 (88%)	27 (12%)	6	10
1	C	206/241 (86%)	178 (86%)	28 (14%)	4	8
1	D	210/241 (87%)	184 (88%)	26 (12%)	5	10
All	All	861/964 (89%)	752 (87%)	109 (13%)	5	9

5 of 109 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	213	MET
1	C	71	LEU
1	D	172	LYS
1	B	227	ILE
1	C	30	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	25	ASN
1	B	35	GLN
1	D	35	GLN
1	A	232	ASN
1	D	25	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](#)

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	PEG	A	1260	-	6,6,6	0.76	0	5,5,5	0.99	0
2	PEG	A	1261	-	6,6,6	0.66	0	5,5,5	0.54	0
2	PEG	A	1262	-	6,6,6	1.44	2 (33%)	5,5,5	0.58	0
2	PEG	B	1260	-	6,6,6	1.02	0	5,5,5	0.70	0
2	PEG	B	1261	-	6,6,6	0.47	0	5,5,5	0.91	0
2	PEG	B	1262	-	6,6,6	0.76	0	5,5,5	1.28	1 (20%)

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	PEG	A	1260	-	-	0/4/4/4	0/0/0/0
2	PEG	A	1261	-	-	0/4/4/4	0/0/0/0
2	PEG	A	1262	-	-	0/4/4/4	0/0/0/0
2	PEG	B	1260	-	-	0/4/4/4	0/0/0/0
2	PEG	B	1261	-	-	0/4/4/4	0/0/0/0
2	PEG	B	1262	-	-	0/4/4/4	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1262	PEG	O2-C3	2.23	1.51	1.42
2	A	1262	PEG	C3-C4	2.61	1.63	1.49

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1262	PEG	C3-O2-C2	2.19	122.79	113.30

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1260	PEG	4	0
2	A	1261	PEG	8	0
2	A	1262	PEG	10	0
2	B	1260	PEG	8	0
2	B	1261	PEG	1	0
2	B	1262	PEG	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	252/272 (92%)	-0.24	1 (0%) 92   92	15, 27, 49, 84	4 (1%)
1	B	252/272 (92%)	-0.22	1 (0%) 92   92	16, 29, 56, 90	5 (1%)
1	C	229/272 (84%)	0.97	46 (20%) 1   1	19, 65, 111, 130	56 (24%)
1	D	235/272 (86%)	0.91	42 (17%) 2   1	21, 64, 93, 115	40 (17%)
All	All	968/1088 (88%)	0.33	90 (9%) 9   9	15, 40, 92, 130	105 (10%)

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	116	GLN	11.4
1	C	99	TYR	9.5
1	C	113	ASP	7.0
1	C	114	TYR	6.7
1	D	114	TYR	6.7

### 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, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q < 0.9
2	PEG	A	1262	7/7	0.89	0.17	1.21	13,23,27,33	0
2	PEG	A	1261	7/7	0.92	0.15	0.47	27,29,34,36	0
2	PEG	B	1262	7/7	0.93	0.13	-0.61	36,39,46,47	0
2	PEG	B	1261	7/7	0.97	0.11	-0.63	24,28,34,36	0
2	PEG	A	1260	7/7	0.96	0.12	-1.16	12,14,17,20	0
2	PEG	B	1260	7/7	0.98	0.09	-1.60	12,12,14,15	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.