Full wwPDB X-ray Structure Validation Report (i)

May 16, 2020 – 03:11 am BST

PDB ID	:	5ITO
Title	:	Structure of the periplasmic binding protein M117N-NocT from A. tumefaciens in complex with octopine
Authors	:	Vigouroux, A.; Morera, S.
Deposited on	:	2016-03-17
Resolution	:	2.35 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 following versions of software and data (see references (1)) were used in the production of this report:

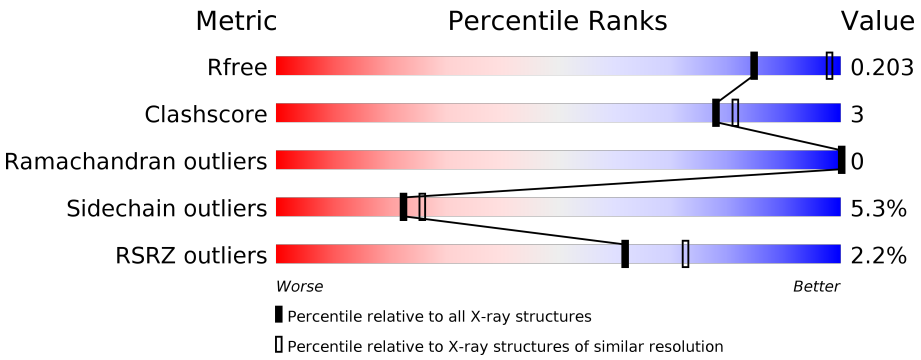
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.35 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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 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	265	% 87% 9% •
1	B	265	3% 85% 11% • •

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	PEG	A	301	-	-	X	X

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 4101 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.

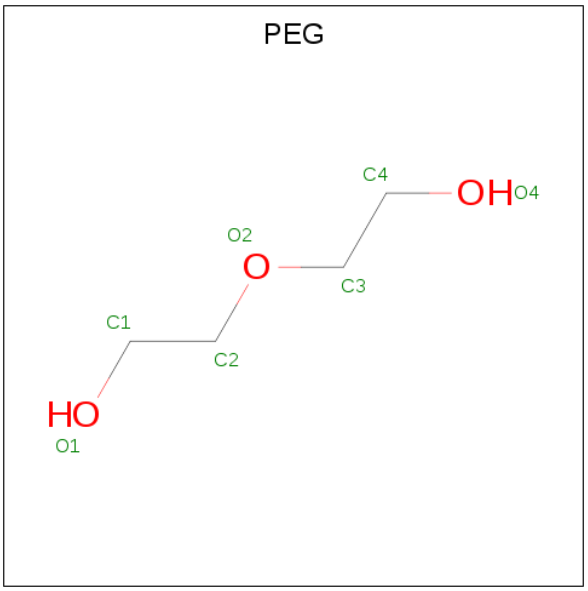
- Molecule 1 is a protein called Nopaline-binding periplasmic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	255	1957	1251	320	375	11	0	0	0
1	B	255	1958	1253	321	373	11	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

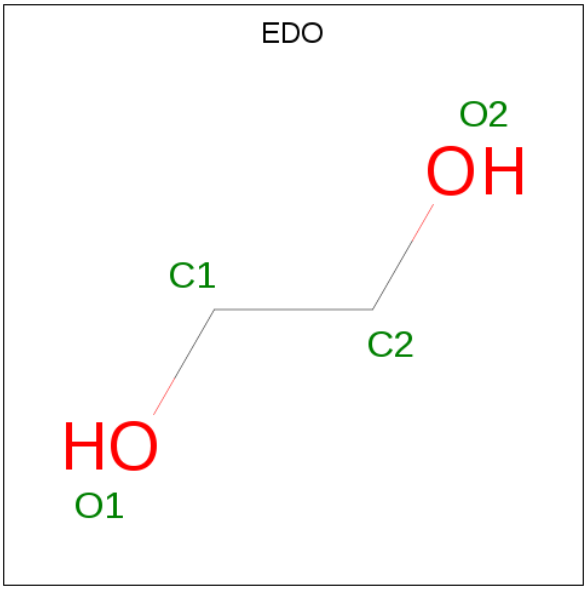
Chain	Residue	Modelled	Actual	Comment	Reference
A	25	MET	-	initiating methionine	UNP P35120
A	117	ASN	MET	conflict	UNP P35120
A	284	HIS	-	expression tag	UNP P35120
A	285	HIS	-	expression tag	UNP P35120
A	286	HIS	-	expression tag	UNP P35120
A	287	HIS	-	expression tag	UNP P35120
A	288	HIS	-	expression tag	UNP P35120
A	289	HIS	-	expression tag	UNP P35120
B	25	MET	-	initiating methionine	UNP P35120
B	117	ASN	MET	conflict	UNP P35120
B	284	HIS	-	expression tag	UNP P35120
B	285	HIS	-	expression tag	UNP P35120
B	286	HIS	-	expression tag	UNP P35120
B	287	HIS	-	expression tag	UNP P35120
B	288	HIS	-	expression tag	UNP P35120
B	289	HIS	-	expression tag	UNP P35120

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



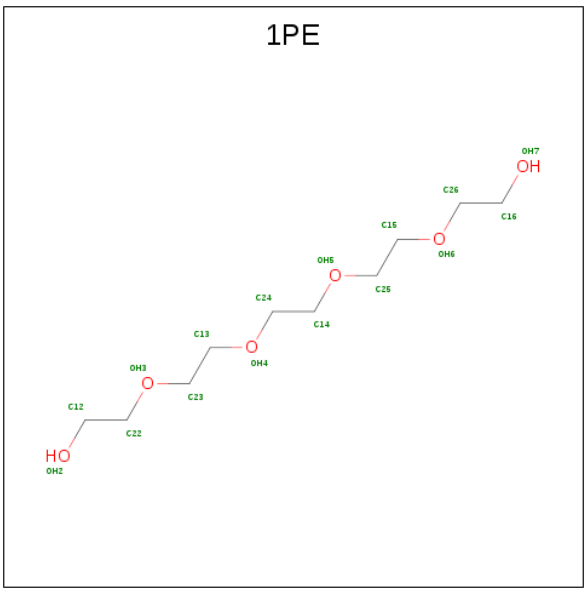
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	7	4	3	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



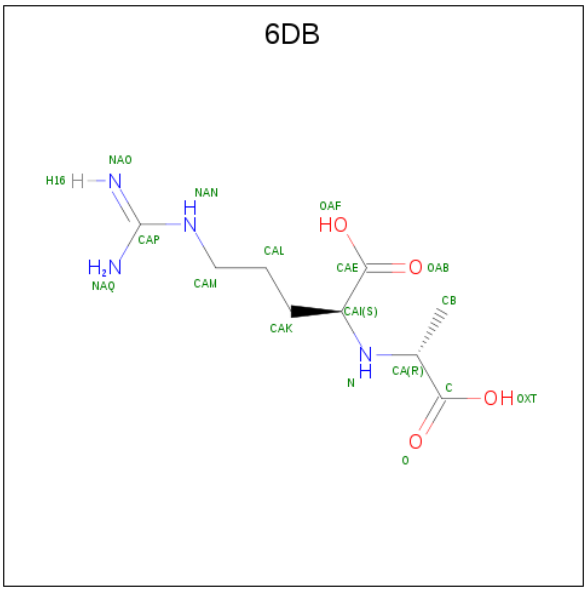
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	4	2	2	0	0
3	B	1	4	2	2	0	0
3	B	1	4	2	2	0	0

- Molecule 4 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula: $C_{10}H_{22}O_6$).



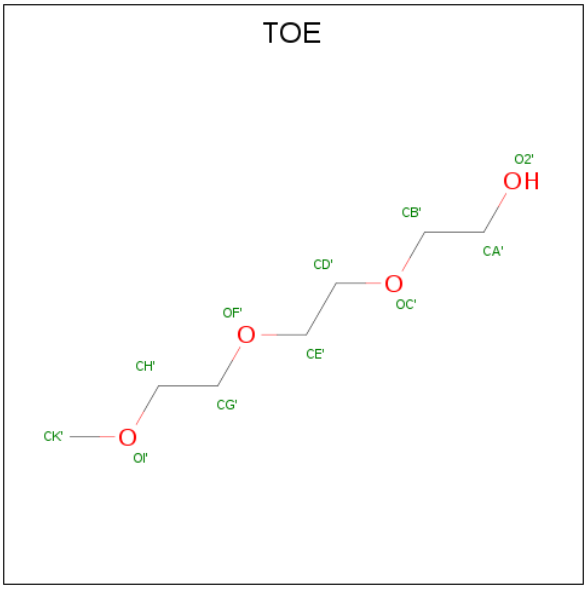
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	A	1	16	10	6	0	0

- Molecule 5 is octopine (three-letter code: 6DB) (formula: $C_9H_{18}N_4O_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	A	1	17	9	4	4	0	0
			Total	C	N	O		
5	B	1	17	9	4	4	0	0

- Molecule 6 is 2-[2-(2-METHOXY-ETHOXY)-ETHOXY]-ETHOXYL (three-letter code: TOE) (formula: $C_7H_{16}O_4$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
6	A	1	11	7	4	0	0
			Total	C	O		
6	A	1	11	7	4	0	0

- Molecule 7 is water.

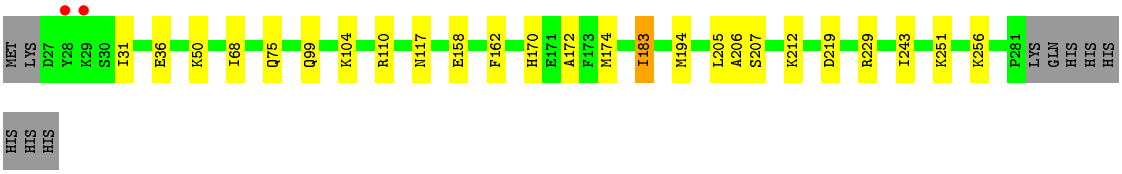
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	70	70	70	0	0
			Total	O		
7	B	25	25	25	0	0

3 Residue-property plots (i)

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.

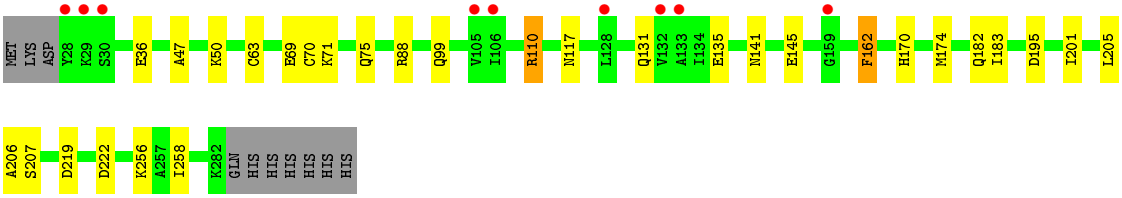
- Molecule 1: Nopaline-binding periplasmic protein

Chain A:



- Molecule 1: Nopaline-binding periplasmic protein

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	113.59Å 113.59Å 37.89Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	25.98 – 2.35 49.19 – 2.35	Depositor EDS
% Data completeness (in resolution range)	100.0 (25.98-2.35) 99.9 (49.19-2.35)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.61 (at 2.34Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.175 , 0.207 0.176 , 0.203	Depositor DCC
R_{free} test set	1142 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	44.6	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 49.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.025 for -h,-k,l 0.024 for h,-h-k,-l 0.015 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4101	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% 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: PEG, 6DB, TOE, EDO, 1PE

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	0.51	0/1995	0.64	0/2695
1	B	0.46	0/1996	0.64	0/2695
All	All	0.48	0/3991	0.64	0/5390

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	1957	0	1967	14	0
1	B	1958	0	1976	10	0
2	A	7	0	10	4	0
3	A	4	0	6	1	0
3	B	8	0	12	1	0
4	A	16	0	22	1	0
5	A	17	0	0	0	0
5	B	17	0	0	0	0
6	A	22	0	32	0	0
7	A	70	0	0	0	0
7	B	25	0	0	0	0
All	All	4101	0	4025	23	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 3.

All (23) 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:104:LYS:HA	2:A:301:PEG:H32	1.49	0.92
1:A:212:LYS:HD2	3:A:302:EDO:H22	1.54	0.88
1:A:162:PHE:HB3	1:A:183:ILE:HD12	1.74	0.69
1:A:174:MET:HE2	1:A:205:LEU:HD22	1.75	0.68
1:A:172:ALA:HB2	4:A:303:1PE:H141	1.76	0.68
1:A:104:LYS:HA	2:A:301:PEG:C3	2.27	0.64
1:B:162:PHE:HB3	1:B:183:ILE:HG22	1.83	0.61
1:B:174:MET:HE2	1:B:205:LEU:HD22	1.87	0.56
1:A:31:ILE:HG12	1:A:68:ILE:HD12	1.87	0.56
1:A:104:LYS:HG2	2:A:301:PEG:H32	1.89	0.54
1:A:194:MET:HG3	1:B:47:ALA:O	2.10	0.52
1:B:63:CYS:SG	1:B:70:CYS:HB3	2.51	0.50
1:A:117:ASN:HA	1:A:207:SER:HA	1.95	0.49
1:A:170:HIS:CE1	1:A:206:ALA:HA	2.48	0.49
1:B:117:ASN:HA	1:B:207:SER:HA	1.94	0.49
1:B:222:ASP:HB3	3:B:302:EDO:H11	1.94	0.48
1:A:104:LYS:CA	2:A:301:PEG:H32	2.33	0.47
1:B:170:HIS:CE1	1:B:206:ALA:HA	2.49	0.47
1:B:110:ARG:HD3	1:B:258:ILE:CG2	2.45	0.47
1:A:36:GLU:HA	1:A:75:GLN:O	2.19	0.42
1:B:36:GLU:HA	1:B:75:GLN:O	2.20	0.41
1:A:243:ILE:HD13	1:A:251:LYS:HA	2.02	0.41
1:B:195:ASP:HB3	1:B:201:ILE:HG12	2.03	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	253/265 (96%)	248 (98%)	5 (2%)	0	100	100
1	B	253/265 (96%)	248 (98%)	5 (2%)	0	100	100
All	All	506/530 (96%)	496 (98%)	10 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	209/219 (95%)	201 (96%)	8 (4%)	33	41
1	B	209/219 (95%)	195 (93%)	14 (7%)	16	17
All	All	418/438 (95%)	396 (95%)	22 (5%)	22	26

All (22) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	50	LYS
1	A	99	GLN
1	A	110	ARG
1	A	158	GLU
1	A	183	ILE
1	A	219	ASP
1	A	229	ARG
1	A	256	LYS
1	B	50	LYS
1	B	69	GLU
1	B	71	LYS
1	B	88	ARG
1	B	99	GLN
1	B	110	ARG
1	B	131	GLN
1	B	135	GLU
1	B	141	ASN
1	B	145	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	162	PHE
1	B	182	GLN
1	B	219	ASP
1	B	256	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	273	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

9 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	EDO	B	301	-	3,3,3	0.66	0	2,2,2	0.22	0
5	6DB	A	304	-	10,16,16	1.72	1 (10%)	10,20,20	1.25	2 (20%)
6	TOE	A	305	-	10,10,10	0.63	0	9,9,9	1.58	2 (22%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	1PE	A	303	-	15,15,15	0.70	0	14,14,14	0.42	0
2	PEG	A	301	-	6,6,6	0.41	0	5,5,5	0.37	0
3	EDO	A	302	-	3,3,3	0.49	0	2,2,2	0.26	0
3	EDO	B	302	-	3,3,3	0.66	0	2,2,2	0.15	0
6	TOE	A	306	-	10,10,10	0.54	0	9,9,9	1.77	3 (33%)
5	6DB	B	303	-	10,16,16	1.79	1 (10%)	10,20,20	0.92	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	EDO	B	301	-	-	0/1/1/1	-
5	6DB	A	304	-	-	1/11/19/19	-
6	TOE	A	305	-	-	5/8/8/8	-
4	1PE	A	303	-	-	10/13/13/13	-
2	PEG	A	301	-	-	0/4/4/4	-
3	EDO	A	302	-	-	0/1/1/1	-
3	EDO	B	302	-	-	0/1/1/1	-
6	TOE	A	306	-	-	4/8/8/8	-
5	6DB	B	303	-	-	1/11/19/19	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	303	6DB	CAP-NAN	5.27	1.43	1.33
5	A	304	6DB	CAP-NAN	5.04	1.43	1.33

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	304	6DB	CB-CA-N	2.32	113.21	108.67
6	A	306	TOE	OC'-CD'-CE'	2.28	120.68	110.39
6	A	306	TOE	OF'-CE'-CD'	2.21	120.35	110.39
6	A	305	TOE	OF'-CE'-CD'	2.21	120.34	110.39
5	A	304	6DB	CAK-CAL-CAM	2.19	118.63	112.05
6	A	306	TOE	CG'-OF'-CE'	2.19	122.78	113.29
6	A	305	TOE	OC'-CD'-CE'	2.08	119.78	110.39

There are no chirality outliers.

All (21) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	304	6DB	CAE-CAI-N-CA
5	B	303	6DB	CAE-CAI-N-CA
4	A	303	1PE	OH4-C13-C23-OH3
4	A	303	1PE	OH6-C15-C25-OH5
6	A	305	TOE	O2'-CA'-CB'-OC'
6	A	305	TOE	OF'-CG'-CH'-OI'
4	A	303	1PE	OH5-C14-C24-OH4
4	A	303	1PE	C23-C13-OH4-C24
6	A	305	TOE	CA'-CB'-OC'-CD'
4	A	303	1PE	C16-C26-OH6-C15
6	A	305	TOE	CH'-CG'-OF'-CE'
4	A	303	1PE	C13-C23-OH3-C22
6	A	306	TOE	O2'-CA'-CB'-OC'
4	A	303	1PE	C14-C24-OH4-C13
6	A	306	TOE	CD'-CE'-OF'-CG'
6	A	305	TOE	CD'-CE'-OF'-CG'
4	A	303	1PE	C12-C22-OH3-C23
6	A	306	TOE	CA'-CB'-OC'-CD'
4	A	303	1PE	C15-C25-OH5-C14
4	A	303	1PE	OH7-C16-C26-OH6
6	A	306	TOE	OF'-CG'-CH'-OI'

There are no ring outliers.

4 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	303	1PE	1	0
2	A	301	PEG	4	0
3	A	302	EDO	1	0
3	B	302	EDO	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	255/265 (96%)	-0.10	2 (0%)	86	91	29, 40, 62, 94	0
1	B	255/265 (96%)	0.27	9 (3%)	44	56	43, 61, 90, 109	0
All	All	510/530 (96%)	0.09	11 (2%)	62	72	29, 51, 84, 109	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	28	TYR	3.7
1	B	132	VAL	2.8
1	B	128	LEU	2.6
1	B	159	GLY	2.4
1	B	106	ILE	2.3
1	B	105	VAL	2.3
1	B	30	SER	2.3
1	A	29	LYS	2.1
1	A	28	TYR	2.1
1	B	29	LYS	2.0
1	B	133	ALA	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 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. 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	PEG	A	301	7/7	0.74	0.47	55,59,62,64	0
3	EDO	B	301	4/4	0.76	0.16	73,73,74,74	0
3	EDO	B	302	4/4	0.81	0.23	53,56,61,64	0
6	TOE	A	306	11/11	0.84	0.19	57,61,62,64	0
4	1PE	A	303	16/16	0.85	0.20	60,66,68,68	0
3	EDO	A	302	4/4	0.91	0.12	58,58,58,59	0
6	TOE	A	305	11/11	0.93	0.18	67,70,76,76	0
5	6DB	B	303	17/17	0.94	0.16	42,48,64,67	0
5	6DB	A	304	17/17	0.97	0.14	30,35,41,42	0

6.5 Other polymers (i)

There are no such residues in this entry.