



Full wwPDB X-ray Structure Validation Report ⓘ

Sep 17, 2023 – 03:25 PM EDT

PDB ID : 4YIS
Title : Crystal Structure of LAGLIDADG Meganuclease I-CpaMI Bound to Uncleaved DNA
Authors : Hallinan, J.P.; Kaiser, B.K.; Stoddard, B.L.
Deposited on : 2015-03-02
Resolution : 2.89 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 ⓘ 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 : 2.35.1
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

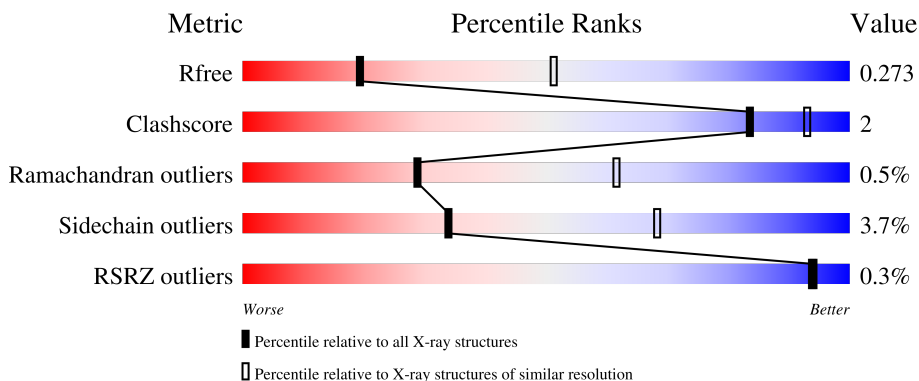
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.89 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	295	
1	B	295	
2	C	28	
2	E	28	
3	D	28	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	28	 82% 18%

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
5	EDO	D	101	-	-	-	X

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 6912 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 Meganuclease I-CpaMI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	295	Total	C	N	O	S	0	0	0
			2334	1512	392	428	2			
1	B	284	Total	C	N	O	S	0	0	0
			2236	1444	375	415	2			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	50	ASN	LEU	conflict	UNP O20960
A	100	GLN	TRP	conflict	UNP O20960
A	114	LYS	LEU	conflict	UNP O20960
A	115	ASN	LEU	conflict	UNP O20960
A	161	ASN	PHE	conflict	UNP O20960
A	264	GLN	LEU	conflict	UNP O20960
A	267	GLN	ILE	conflict	UNP O20960
B	50	ASN	LEU	conflict	UNP O20960
B	100	GLN	TRP	conflict	UNP O20960
B	114	LYS	LEU	conflict	UNP O20960
B	115	ASN	LEU	conflict	UNP O20960
B	161	ASN	PHE	conflict	UNP O20960
B	264	GLN	LEU	conflict	UNP O20960
B	267	GLN	ILE	conflict	UNP O20960

- Molecule 2 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	E	28	Total	C	N	O	P	0	0	0
			559	268	101	163	27			
2	C	28	Total	C	N	O	P	0	0	0
			559	268	101	163	27			

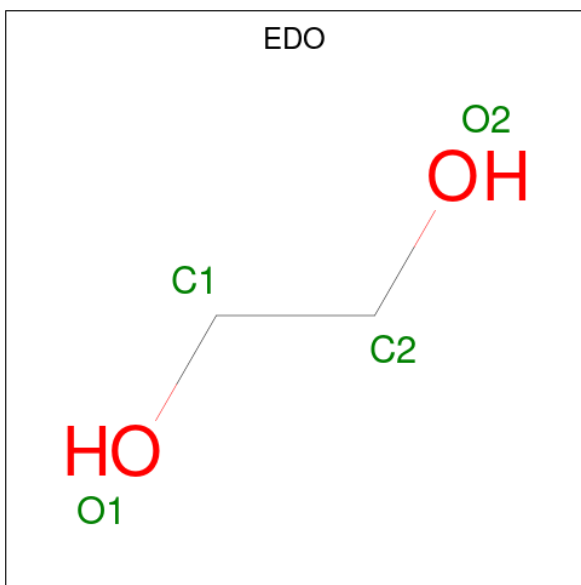
- Molecule 3 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	F	28	Total	C	N	O	P	0	0	0
			581	277	107	170	27			
3	D	28	Total	C	N	O	P	0	0	0
			581	277	107	170	27			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	2	2		
5	B	1	Total	C	O	0	0
			4	2	2		
5	E	1	Total	C	O	0	0
			4	2	2		
5	F	1	Total	C	O	0	0
			4	2	2		
5	D	1	Total	C	O	0	0
			4	2	2		


- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	16	Total O 16 16	0	0
6	B	11	Total O 11 11	0	0
6	E	5	Total O 5 5	0	0
6	F	3	Total O 3 3	0	0
6	C	1	Total O 1 1	0	0
6	D	1	Total O 1 1	0	0

3 Residue-property plots [i](#)


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.

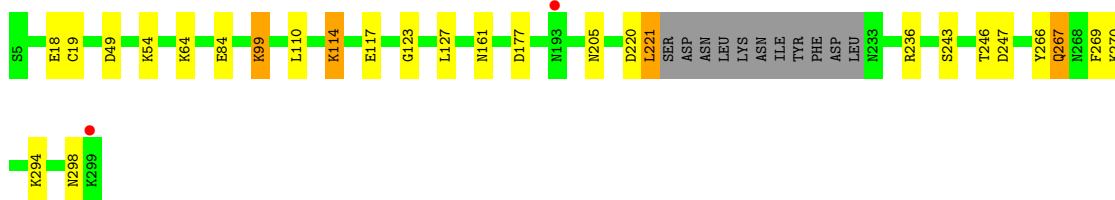
- Molecule 1: Meganuclease I-CpaMI

Chain A:  91% 8%



- Molecule 1: Meganuclease I-CpaMI

Chain B:  87% 8%



- Molecule 2: DNA (28-MER)

Chain E:  75% 25%




- Molecule 2: DNA (28-MER)

Chain C:  71% 29%




- Molecule 3: DNA (28-MER)

Chain F:  82% 18%



- Molecule 3: DNA (28-MER)

Chain D:  82% 18%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	75.87Å 91.80Å 139.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.89 39.27 – 2.78	Depositor EDS
% Data completeness (in resolution range)	94.3 (50.00-2.89) 93.4 (39.27-2.78)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.78 (at 2.77Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.217 , 0.274 0.220 , 0.273	Depositor DCC
R_{free} test set	1197 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	58.1	Xtrriage
Anisotropy	0.590	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 20.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	6912	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.07% 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: CA, EDO

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	0.53	0/2387	0.65	0/3227
1	B	0.51	0/2286	0.65	2/3092 (0.1%)
2	C	0.48	0/625	0.85	0/959
2	E	0.46	0/625	0.86	0/959
3	D	0.51	1/652 (0.2%)	0.95	4/1008 (0.4%)
3	F	0.45	0/652	0.89	1/1008 (0.1%)
All	All	0.50	1/7227 (0.0%)	0.75	7/10253 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	16	DT	P-OP1	7.10	1.61	1.49

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	177	ASP	CB-CG-OD2	-9.07	110.14	118.30
3	D	16	DT	O5'-P-OP1	-7.87	98.62	105.70
1	B	177	ASP	CB-CG-OD1	6.96	124.56	118.30
3	F	18	DG	C1'-O4'-C4'	-5.85	104.25	110.10
3	D	18	DG	C1'-O4'-C4'	-5.52	104.58	110.10
3	D	4	DA	C1'-O4'-C4'	-5.29	104.81	110.10
3	D	7	DG	C1'-O4'-C4'	-5.28	104.83	110.10

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	2334	0	2314	11	1
1	B	2236	0	2192	12	1
2	C	559	0	315	5	0
2	E	559	0	315	4	0
3	D	581	0	319	1	0
3	F	581	0	319	3	0
4	A	2	0	0	0	0
4	B	3	0	0	0	0
5	A	4	0	6	0	0
5	B	4	0	6	0	0
5	D	4	0	6	0	0
5	E	4	0	6	0	0
5	F	4	0	6	0	0
6	A	16	0	0	0	0
6	B	11	0	0	1	0
6	C	1	0	0	0	0
6	D	1	0	0	0	0
6	E	5	0	0	0	0
6	F	3	0	0	1	0
All	All	6912	0	5804	31	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 2.

All (31) 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:B:84:GLU:OE1	1:B:114:LYS:NZ	2.05	0.89
1:B:110:LEU:O	1:B:114:LYS:HG3	1.77	0.84
1:A:110:LEU:O	1:A:114:LYS:HG3	1.91	0.70
3:F:16:DT:OP1	6:F:201:HOH:O	2.10	0.68
1:B:19:CYS:O	1:B:99:LYS:HE2	1.97	0.64
1:B:117:GLU:O	1:B:123:GLY:HA3	1.98	0.64
1:A:117:GLU:O	1:A:123:GLY:HA3	1.97	0.63
2:E:12:DA:H4'	2:E:13:DT:OP1	2.02	0.59

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:12:DA:H4'	2:C:13:DT:OP1	2.04	0.58
1:A:97:THR:O	1:A:100:GLN:HG2	2.05	0.57
1:A:294:LYS:NZ	2:E:19:DG:OP2	2.37	0.57
2:C:17:DA:H2''	2:C:18:DA:O5'	2.10	0.51
2:E:17:DA:H2''	2:E:18:DA:O5'	2.11	0.50
1:B:246:THR:HG22	1:B:247:ASP:OD1	2.11	0.50
1:A:246:THR:HG22	1:A:247:ASP:OD1	2.12	0.48
1:A:19:CYS:O	1:A:99:LYS:HE2	2.16	0.46
2:E:3:DT:H2'	2:E:4:DA:C8	2.50	0.45
1:B:294:LYS:NZ	2:C:19:DG:OP2	2.49	0.45
1:B:64:LYS:HG2	2:C:7:DC:OP2	2.16	0.45
1:A:189:THR:OG1	3:F:4:DA:N7	2.50	0.44
1:B:266:TYR:O	1:B:269:PHE:HB3	2.18	0.44
2:C:3:DT:H2'	2:C:4:DA:C8	2.53	0.44
1:B:49:ASP:HA	6:B:406:HOH:O	2.18	0.43
1:A:266:TYR:O	1:A:269:PHE:HB3	2.18	0.43
1:A:225:LEU:HD23	1:A:225:LEU:N	2.35	0.42
1:B:236:ARG:NH2	3:D:10:DT:O4	2.53	0.42
1:B:220:ASP:O	1:B:221:LEU:HD23	2.21	0.41
1:A:35:TRP:CD1	1:A:124:LEU:HD21	2.55	0.41
1:A:266:TYR:CE2	1:A:270:LYS:HD2	2.56	0.41
3:F:12:DA:H2''	3:F:13:DA:C8	2.56	0.40
1:B:266:TYR:CE2	1:B:270:LYS:HD2	2.56	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:267:GLN:OE1	1:B:267:GLN:OE1[4_555]	2.15	0.05

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	293/295 (99%)	278 (95%)	13 (4%)	2 (1%)	22	54
1	B	280/295 (95%)	270 (96%)	9 (3%)	1 (0%)	34	66
All	All	573/590 (97%)	548 (96%)	22 (4%)	3 (0%)	29	61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	298	ASN
1	A	224	ASN
1	A	298	ASN

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	252/264 (96%)	244 (97%)	8 (3%)	39	73
1	B	240/264 (91%)	230 (96%)	10 (4%)	30	63
All	All	492/528 (93%)	474 (96%)	18 (4%)	34	68

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	28	ASN
1	A	72	SER
1	A	114	LYS
1	A	189	THR
1	A	221	LEU
1	A	232	LEU
1	A	264	GLN
1	A	267	GLN
1	B	18	GLU
1	B	54	LYS
1	B	99	LYS
1	B	114	LYS
1	B	127	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	161	ASN
1	B	205	ASN
1	B	221	LEU
1	B	243	SER
1	B	267	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	28	ASN
1	A	213	GLN
1	B	57	GLN
1	B	100	GLN
1	B	115	ASN
1	B	161	ASN
1	B	186	ASN
1	B	205	ASN
1	B	213	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](#)

Of 10 ligands modelled in this entry, 5 are monoatomic - leaving 5 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
5	EDO	A	303	-	3,3,3	0.42	0	2,2,2	0.33	0
5	EDO	F	101	-	3,3,3	0.45	0	2,2,2	0.40	0
5	EDO	B	304	-	3,3,3	0.62	0	2,2,2	0.13	0
5	EDO	E	101	-	3,3,3	0.46	0	2,2,2	0.32	0
5	EDO	D	101	-	3,3,3	0.44	0	2,2,2	0.32	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
5	EDO	A	303	-	-	1/1/1/1	-
5	EDO	F	101	-	-	1/1/1/1	-
5	EDO	B	304	-	-	0/1/1/1	-
5	EDO	E	101	-	-	0/1/1/1	-
5	EDO	D	101	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	303	EDO	O1-C1-C2-O2
5	F	101	EDO	O1-C1-C2-O2

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

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	295/295 (100%)	-0.21	0 100 100	25, 38, 58, 95	0
1	B	284/295 (96%)	-0.24	2 (0%) 87 87	28, 42, 54, 88	0
2	C	28/28 (100%)	-0.44	0 100 100	31, 43, 62, 75	0
2	E	28/28 (100%)	-0.42	0 100 100	29, 45, 61, 70	0
3	D	28/28 (100%)	-0.41	0 100 100	37, 53, 65, 94	0
3	F	28/28 (100%)	-0.47	0 100 100	27, 48, 66, 80	0
All	All	691/702 (98%)	-0.26	2 (0%) 94 94	25, 40, 59, 95	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	299	LYS	2.2
1	B	193	ASN	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 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
5	EDO	D	101	4/4	0.69	0.46	95,101,105,105	0
5	EDO	B	304	4/4	0.80	0.42	70,76,78,81	0
4	CA	B	303	1/1	0.80	0.33	77,77,77,77	0
5	EDO	F	101	4/4	0.86	0.41	87,88,88,90	0
5	EDO	E	101	4/4	0.88	0.49	96,99,99,100	0
4	CA	A	301	1/1	0.90	0.24	55,55,55,55	0
5	EDO	A	303	4/4	0.90	0.28	77,80,81,81	0
4	CA	B	301	1/1	0.92	0.21	65,65,65,65	0
4	CA	B	302	1/1	0.95	0.16	75,75,75,75	0
4	CA	A	302	1/1	0.97	0.23	58,58,58,58	0

6.5 Other polymers [i](#)

There are no such residues in this entry.