wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 04:44 am BST

PDB ID	:	4KTB
Title	:	The crystal structure of posible asymmetric diadenosine tetraphosphate (Ap(4)A) hydrolases from Jonesia denitrificans DSM 20603
Authors	:	Tan, K.; Kim, Y.; Wu, R.; Clancy, S.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2013-05-20
Resolution	:	1.94 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

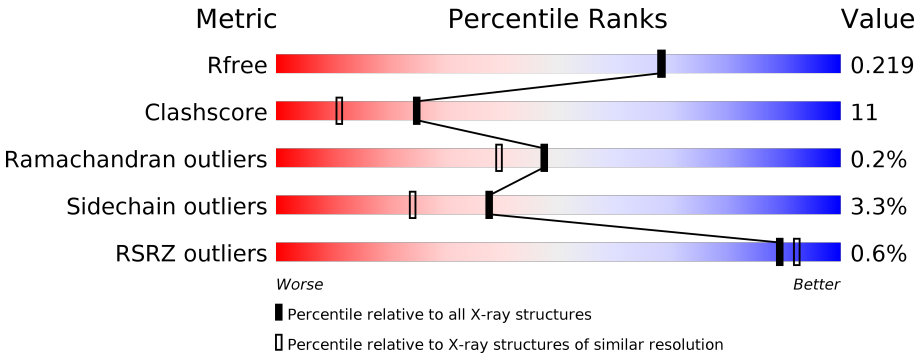
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.94 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	4310 (1.96-1.92)
Clashscore	141614	1023 (1.94-1.94)
Ramachandran outliers	138981	1007 (1.94-1.94)
Sidechain outliers	138945	1007 (1.94-1.94)
RSRZ outliers	127900	4250 (1.96-1.92)

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	194	61% 22% 18%
1	B	194	% 64% 15% • 19%
1	C	194	% 58% 24% • 18%
1	D	194	% 57% 25% • 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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CL	A	204	-	-	X	-
3	CL	A	205	-	-	X	-
3	CL	A	208	-	-	X	-
3	CL	B	203	-	-	X	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5516 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 Putative uncharacterized protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	160	Total	C	N	O	S	Se	0	1	0
			1243	788	221	226	2	6			
1	B	158	Total	C	N	O	S	Se	0	2	0
			1237	784	219	225	2	7			
1	C	160	Total	C	N	O	S	Se	0	2	0
			1256	796	224	227	2	7			
1	D	159	Total	C	N	O	S	Se	0	2	0
			1243	788	221	226	2	6			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	SER	-	EXPRESSION TAG	UNP C7R2I1
A	0	ASN	-	EXPRESSION TAG	UNP C7R2I1
A	1	ALA	-	EXPRESSION TAG	UNP C7R2I1
B	-1	SER	-	EXPRESSION TAG	UNP C7R2I1
B	0	ASN	-	EXPRESSION TAG	UNP C7R2I1
B	1	ALA	-	EXPRESSION TAG	UNP C7R2I1
C	-1	SER	-	EXPRESSION TAG	UNP C7R2I1
C	0	ASN	-	EXPRESSION TAG	UNP C7R2I1
C	1	ALA	-	EXPRESSION TAG	UNP C7R2I1
D	-1	SER	-	EXPRESSION TAG	UNP C7R2I1
D	0	ASN	-	EXPRESSION TAG	UNP C7R2I1
D	1	ALA	-	EXPRESSION TAG	UNP C7R2I1

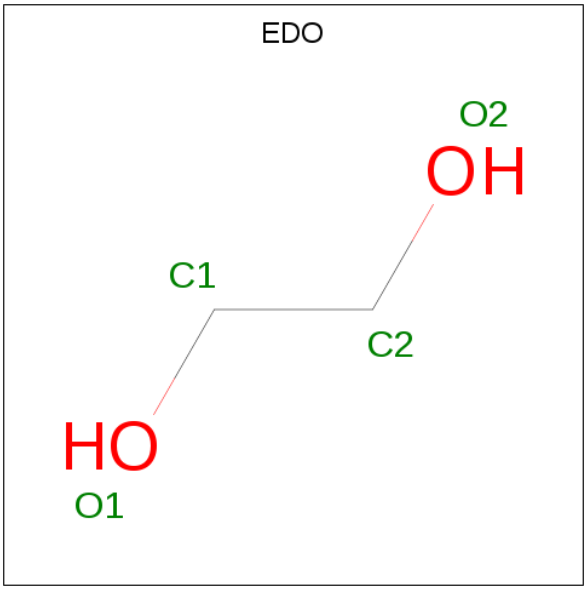
- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Na	0	0
			1	1		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	5	Total Cl 5 5	0	0
3	A	10	Total Cl 10 10	0	0
3	D	5	Total Cl 5 5	0	0
3	C	2	Total Cl 2 2	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

- Molecule 5 is water.

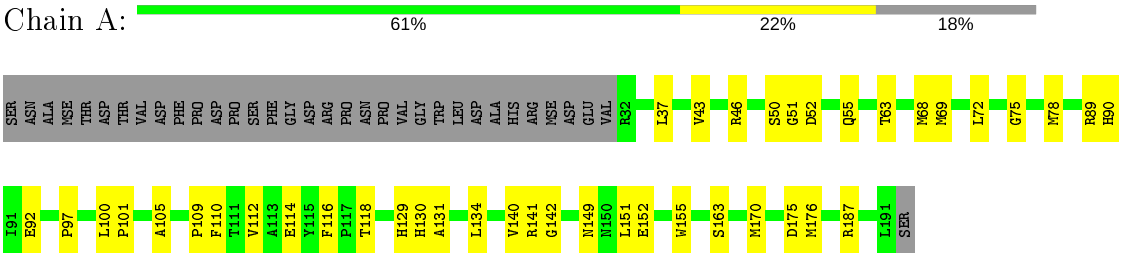
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	139	Total O 139 139	0	0
5	B	140	Total O 140 140	0	0
5	C	111	Total O 111 111	0	0
5	D	120	Total O 120 120	0	0

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.

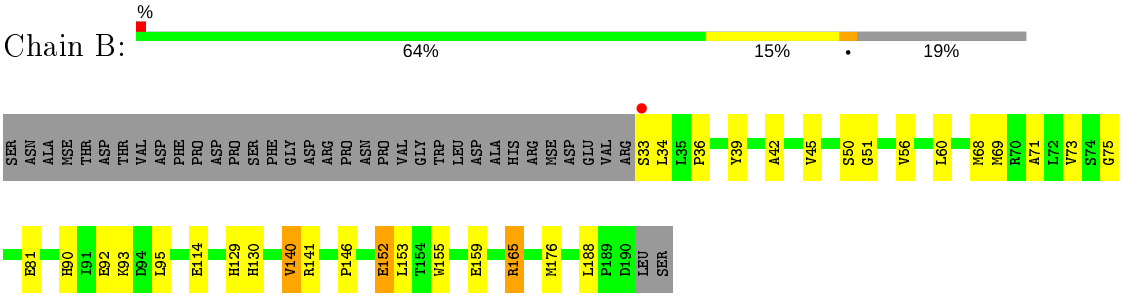
- Molecule 1: Putative uncharacterized protein

Chain A:



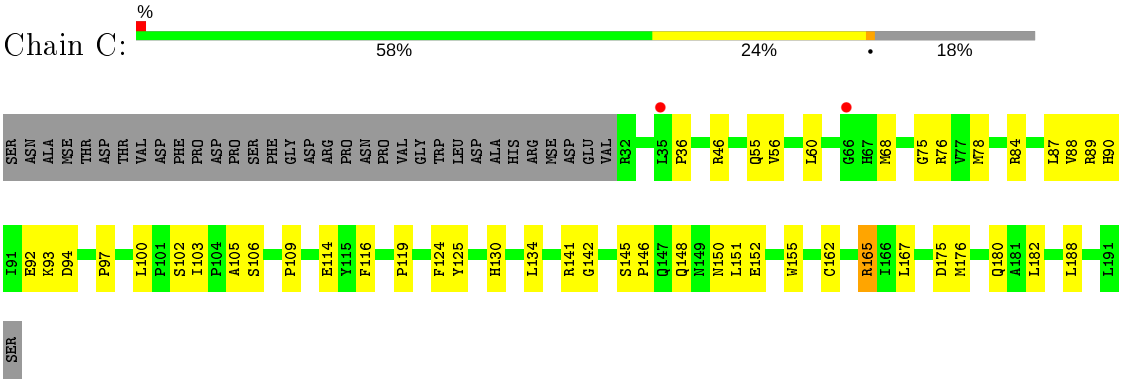
- Molecule 1: Putative uncharacterized protein

Chain B:



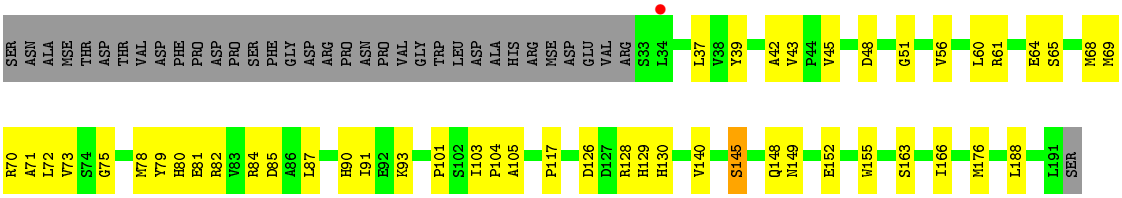
- Molecule 1: Putative uncharacterized protein

Chain C:



- Molecule 1: Putative uncharacterized protein

Chain D:


SER ASN ALA MSE THR ASP THR VAL ASP PHE PRO ASP PRO SER PHE GLY ASP ARG PRO ASN PRO VAL GLY TRP LEU ASP ALA HIS ARG MSE ASP GLU VAL ARG S33 L34 L37 V38 Y39 A42 V43 P44 V45 D48 G51 V56 L60 R61 E64 S65 M68 M69
R70 A71 L72 V73 S74 G75 M78 Y79 H80 E81 R82 V83 R84 D85 A86 L87 H90 I91 E92 K93 P101 S102 I103 P104 A105 P117 D126 D127 R128 H129 H130 V140 S145 Q148 N149 E152 W155 S163 I166 M176 L188 L191 SER

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.49Å 69.81Å 94.31Å 90.00° 90.04° 90.00°	Depositor
Resolution (Å)	47.84 – 1.94 47.84 – 1.94	Depositor EDS
% Data completeness (in resolution range)	90.6 (47.84-1.94) 88.6 (47.84-1.94)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.15 (at 1.94Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, R_{free}	0.173 , 0.224 0.171 , 0.219	Depositor DCC
R_{free} test set	1991 reflections (4.08%)	wwPDB-VP
Wilson B-factor (Å^2)	16.0	Xtriage
Anisotropy	0.089	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 44.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.437 for h,-k,-l	Xtriage
Reported twinning fraction	0.480 for h,-k,-l	Depositor
Outliers	0 of 48781 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5516	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 57.74 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.2663e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: NA, EDO, CL

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	0.38	0/1268	0.60	0/1723
1	B	0.38	0/1262	0.58	0/1713
1	C	0.38	0/1282	0.58	0/1741
1	D	0.37	0/1268	0.59	0/1723
All	All	0.38	0/5080	0.59	0/6900

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	1243	0	1246	28	1
1	B	1237	0	1235	23	0
1	C	1256	0	1259	31	0
1	D	1243	0	1245	34	1
2	A	1	0	0	0	0
3	A	10	0	0	8	0
3	B	5	0	0	3	0
3	C	2	0	0	1	0
3	D	5	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	4	0	6	1	0
5	A	139	0	0	5	0
5	B	140	0	0	5	1
5	C	111	0	0	2	1
5	D	120	0	0	3	0
All	All	5516	0	4991	108	2

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 108 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:D:42:ALA:HB3	1:D:73:VAL:HB	1.70	0.72
1:D:61:ARG:NH1	5:D:417:HOH:O	2.25	0.68
1:B:176[B]:MSE:HE1	1:D:188:LEU:HB3	1.77	0.67
1:A:129:HIS:NE2	5:A:424:HOH:O	2.27	0.66
3:A:203:CL:CL	5:A:434:HOH:O	2.50	0.66

All (2) 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:D:149:ASN:O	5:C:359:HOH:O[2_646]	2.05	0.15
1:A:149:ASN:OD1	5:B:431:HOH:O[2_657]	2.19	0.01

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	159/194 (82%)	156 (98%)	3 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	158/194 (81%)	154 (98%)	4 (2%)	0	100 100
1	C	160/194 (82%)	158 (99%)	2 (1%)	0	100 100
1	D	159/194 (82%)	154 (97%)	4 (2%)	1 (1%)	25 13
All	All	636/776 (82%)	622 (98%)	13 (2%)	1 (0%)	47 39

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	148	GLN

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	136/159 (86%)	134 (98%)	2 (2%)	65 56
1	B	136/159 (86%)	129 (95%)	7 (5%)	24 9
1	C	138/159 (87%)	134 (97%)	4 (3%)	42 28
1	D	136/159 (86%)	131 (96%)	5 (4%)	34 19
All	All	546/636 (86%)	528 (97%)	18 (3%)	38 24

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	165	ARG
1	C	56	VAL
1	D	65	SER
1	B	140	VAL
1	B	152	GLU

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

Mol	Chain	Res	Type
1	D	129	HIS

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)

Of 24 ligands modelled in this entry, 23 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
4	EDO	A	212	-	3,3,3	0.45	0	2,2,2	0.27	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
4	EDO	A	212	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	212	EDO	1	0

5.7 Other polymers

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

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, 95^{th} 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	155/194 (79%)	-0.44	0 100 100	7, 18, 38, 55	0
1	B	153/194 (78%)	-0.47	1 (0%) 87 91	6, 15, 40, 58	0
1	C	155/194 (79%)	-0.19	2 (1%) 77 81	9, 25, 51, 79	0
1	D	154/194 (79%)	-0.18	1 (0%) 89 92	12, 28, 57, 68	0
All	All	617/776 (79%)	-0.32	4 (0%) 89 92	6, 21, 48, 79	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	34	LEU	3.8
1	C	66	GLY	2.7
1	B	33	SER	2.4
1	C	35	LEU	2.1

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 carbohydrates 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, 95^{th} 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
3	CL	A	208	1/1	0.94	0.12	40,40,40,40	0
4	EDO	A	212	4/4	0.94	0.15	31,31,31,34	0
3	CL	A	204	1/1	0.96	0.16	44,44,44,44	0
3	CL	D	205	1/1	0.96	0.09	29,29,29,29	0
3	CL	D	202	1/1	0.96	0.12	29,29,29,29	0
3	CL	A	210	1/1	0.97	0.06	29,29,29,29	0
3	CL	C	201	1/1	0.98	0.05	23,23,23,23	0
3	CL	D	201	1/1	0.98	0.04	25,25,25,25	0
2	NA	A	201	1/1	0.98	0.06	20,20,20,20	0
3	CL	A	207	1/1	0.98	0.05	27,27,27,27	0
3	CL	A	205	1/1	0.99	0.06	20,20,20,20	0
3	CL	D	203	1/1	0.99	0.09	24,24,24,24	0
3	CL	B	204	1/1	0.99	0.13	30,30,30,30	0
3	CL	A	202	1/1	0.99	0.06	25,25,25,25	0
3	CL	C	202	1/1	0.99	0.04	27,27,27,27	0
3	CL	B	202	1/1	0.99	0.05	23,23,23,23	0
3	CL	A	209	1/1	0.99	0.09	22,22,22,22	0
3	CL	B	201	1/1	0.99	0.05	19,19,19,19	0
3	CL	A	211	1/1	0.99	0.04	30,30,30,30	0
3	CL	D	204	1/1	0.99	0.07	25,25,25,25	0
3	CL	A	206	1/1	0.99	0.05	28,28,28,28	0
3	CL	A	203	1/1	0.99	0.07	25,25,25,25	0
3	CL	B	203	1/1	0.99	0.05	18,18,18,18	0
3	CL	B	205	1/1	1.00	0.06	29,29,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.