wwPDB X-ray Structure Validation Summary Report (i)

Sep 10, 2023 – 10:20 PM EDT

PDB ID	:	4JTH
Title	:	Crystal structure of F114R/R117Q mutant of 3-deoxy-D-manno-octulosonate 8-phosphate synthase (KDO8PS) from Neisseria meningitidis
Authors	:	Allison, T.M.; Cochrane, F.C.; Jameson, G.B.; Parker, E.J.
Deposited on	:	2013-03-23
Resolution	:	2.00 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

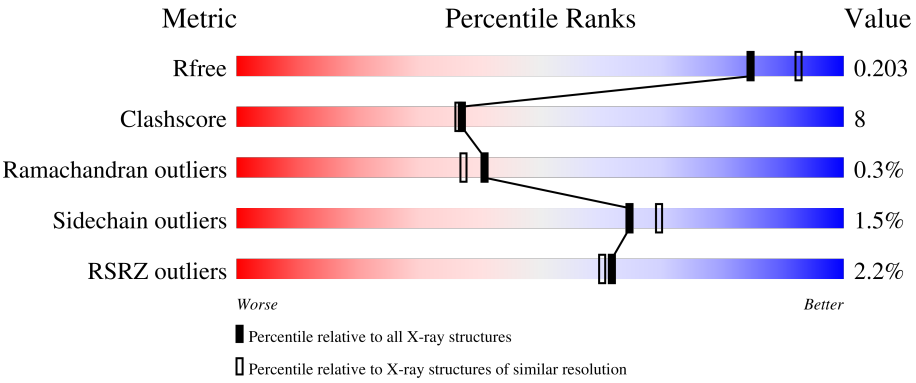
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 <=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	280	4% 83% 6% 11%
1	B	280	% 75% 13% • 10%
1	C	280	% 81% 11% • 8%
1	D	280	2% 80% 10% • 9%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 8774 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 2-dehydro-3-deoxyphosphooctonate aldolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	250	Total	C	N	O	S	0	5	0
			1953	1254	326	362	11			
1	B	251	Total	C	N	O	S	0	10	0
			1994	1280	336	366	12			
1	C	259	Total	C	N	O	S	0	6	0
			2020	1296	341	372	11			
1	D	256	Total	C	N	O	S	0	6	0
			2000	1281	341	367	11			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	114	ARG	PHE	engineered mutation	UNP Q9JZ55
A	117	GLN	ARG	engineered mutation	UNP Q9JZ55
B	114	ARG	PHE	engineered mutation	UNP Q9JZ55
B	117	GLN	ARG	engineered mutation	UNP Q9JZ55
C	114	ARG	PHE	engineered mutation	UNP Q9JZ55
C	117	GLN	ARG	engineered mutation	UNP Q9JZ55
D	114	ARG	PHE	engineered mutation	UNP Q9JZ55
D	117	GLN	ARG	engineered mutation	UNP Q9JZ55

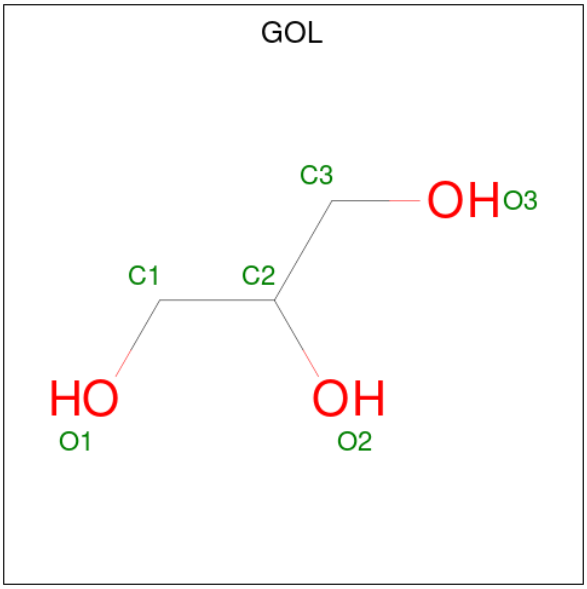
- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Cl	0	0
			1	1		
2	C	2	Total	Cl	0	0
			2	2		
2	D	2	Total	Cl	0	0
			2	2		

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total Na 1 1	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	D	1	Total C O 6 3 3	0	0

- Molecule 5 is water.

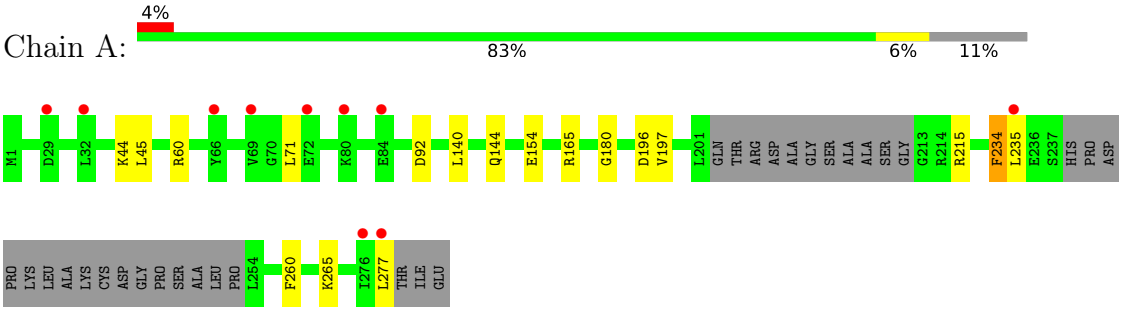
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	162	Total O 162 162	0	0
5	B	224	Total O 224 224	0	0
5	C	215	Total O 215 215	0	0
5	D	194	Total O 194 194	0	0

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.

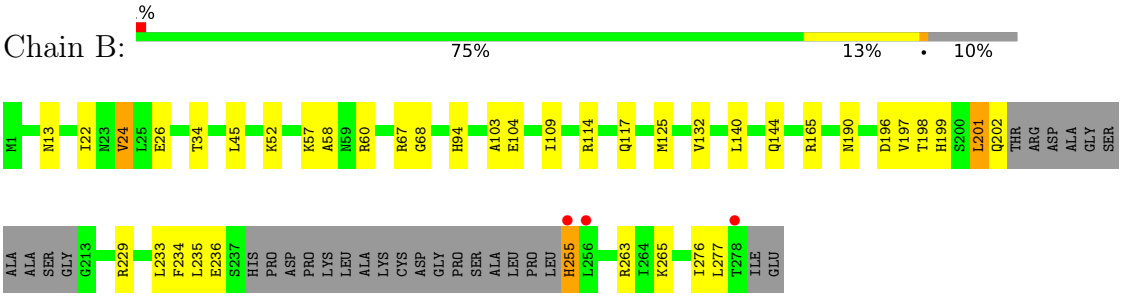
• Molecule 1: 2-dehydro-3-deoxyphosphooctonate aldolase

Chain A:



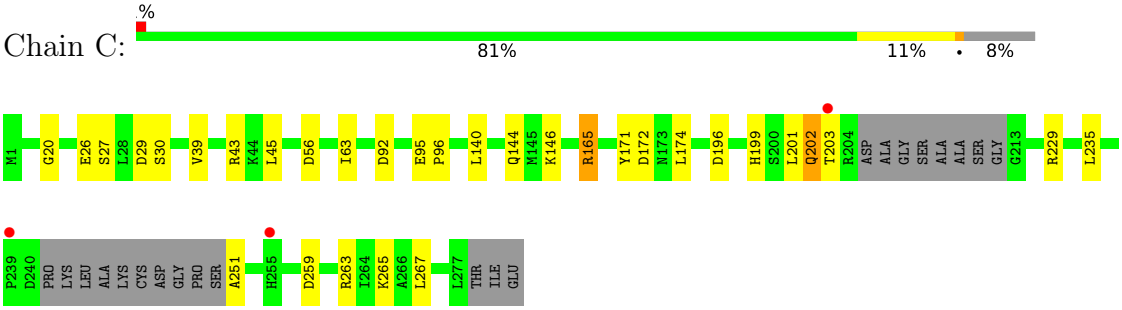
• Molecule 1: 2-dehydro-3-deoxyphosphooctonate aldolase

Chain B:



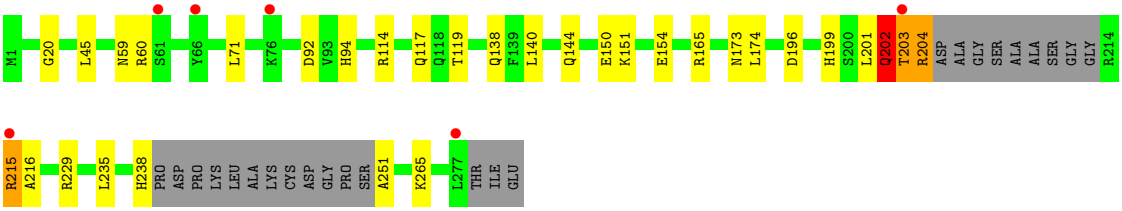
• Molecule 1: 2-dehydro-3-deoxyphosphooctonate aldolase

Chain C:



• Molecule 1: 2-dehydro-3-deoxyphosphooctonate aldolase

Chain D:


M1
G20
L45
N59
R60
S61
Y66
L71
K76
D92
V93
H94
R114
Q117
Q118
T119
Q138
F139
L140
Q144
E150
K151
E154
R165
N173
L174
D196
H199
S200
L201
Q202
T203
R204
ASP
ALA
GLY
SER
ALA
ALA
SER
GLY
GLY
R214

R215
A216
R229
L235
H238
PRO
ASP
PRO
LYS
LEU
ALA
LYS
CYS
ASP
GLY
PRO
SER
A251
K265
L277
THR
ILE
GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	82.15Å 85.99Å 163.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.95 – 2.00 45.95 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.95-2.00) 100.0 (45.95-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	3.81 (at 2.00Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.159 , 0.194 0.167 , 0.203	Depositor DCC
R_{free} test set	3931 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	22.8	Xtriage
Anisotropy	0.410	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 47.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.023 for k,h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8774	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, GOL, 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	0.88	0/1987	0.88	3/2686 (0.1%)
1	B	0.99	1/2038 (0.0%)	0.94	3/2754 (0.1%)
1	C	0.97	2/2066 (0.1%)	0.93	5/2793 (0.2%)
1	D	0.87	0/2049	0.89	4/2769 (0.1%)
All	All	0.93	3/8140 (0.0%)	0.91	15/11002 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	26	GLU	CD-OE2	-5.94	1.19	1.25
1	B	104	GLU	CD-OE2	5.64	1.31	1.25
1	C	171	TYR	CG-CD2	-5.16	1.32	1.39

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	229	ARG	NE-CZ-NH2	-5.98	117.31	120.30
1	A	60	ARG	NE-CZ-NH2	5.84	123.22	120.30
1	D	196	ASP	CB-CG-OD1	5.78	123.50	118.30
1	B	201	LEU	CB-CA-C	-5.72	99.33	110.20
1	C	92	ASP	CB-CG-OD1	5.69	123.42	118.30

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	1953	0	1993	10	0
1	B	1994	0	2048	59	0
1	C	2020	0	2078	22	0
1	D	2000	0	2055	39	0
2	B	1	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	1	0
3	C	1	0	0	0	0
4	D	6	0	8	0	0
5	A	162	0	0	2	0
5	B	224	0	0	4	0
5	C	215	0	0	5	0
5	D	194	0	0	15	0
All	All	8774	0	8182	126	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 8.

The worst 5 of 126 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:259:ASP:HB2	5:C:601:HOH:O	1.45	1.17
1:B:199[B]:HIS:CE1	1:B:236[B]:GLU:HG3	1.91	1.05
1:D:154:GLU:HG2	5:D:585:HOH:O	1.65	0.97
1:C:146:LYS:HD2	5:C:604:HOH:O	1.65	0.97
1:B:109[B]:ILE:HD12	1:B:125[B]:MET:HG3	1.51	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	249/280 (89%)	244 (98%)	5 (2%)	0	100	100
1	B	255/280 (91%)	250 (98%)	5 (2%)	0	100	100
1	C	259/280 (92%)	252 (97%)	6 (2%)	1 (0%)	34	30
1	D	256/280 (91%)	252 (98%)	2 (1%)	2 (1%)	19	13
All	All	1019/1120 (91%)	998 (98%)	18 (2%)	3 (0%)	41	37

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	202	GLN
1	C	229	ARG
1	D	229	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	216/236 (92%)	213 (99%)	3 (1%)	67	72
1	B	223/236 (94%)	219 (98%)	4 (2%)	59	63
1	C	226/236 (96%)	224 (99%)	2 (1%)	78	83
1	D	224/236 (95%)	219 (98%)	5 (2%)	52	55
All	All	889/944 (94%)	875 (98%)	14 (2%)	65	67

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	165	ARG
1	C	202	GLN
1	D	215	ARG
1	D	202	GLN
1	D	203	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	202	GLN
1	D	59	ASN
1	D	217	GLN
1	D	202	GLN
1	B	186	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 7 ligands modelled in this entry, 6 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	GOL	D	303	-	5,5,5	0.21	0	5,5,5	0.67	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	GOL	D	303	-	-	2/4/4/4	-

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
4	D	303	GOL	C1-C2-C3-O3
4	D	303	GOL	O2-C2-C3-O3

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 (i)

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	250/280 (89%)	-0.11	10 (4%)	38	37	13, 27, 57, 80	1 (0%)
1	B	251/280 (89%)	-0.31	3 (1%)	79	78	12, 22, 41, 84	0
1	C	259/280 (92%)	-0.37	3 (1%)	79	78	12, 20, 41, 83	0
1	D	256/280 (91%)	-0.11	6 (2%)	60	59	13, 26, 54, 74	0
All	All	1016/1120 (90%)	-0.23	22 (2%)	62	60	12, 24, 53, 84	1 (0%)

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	255	HIS	6.1
1	A	66	TYR	5.7
1	A	277	LEU	4.9
1	D	203	THR	3.9
1	C	239	PRO	3.7

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 monosaccharides 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
4	GOL	D	303	6/6	0.81	0.29	41,49,51,54	6
3	NA	C	303	1/1	0.95	0.12	28,28,28,28	0
2	CL	D	302	1/1	0.95	0.08	47,47,47,47	0
2	CL	B	301	1/1	0.98	0.09	25,25,25,25	0
2	CL	C	301	1/1	0.99	0.06	24,24,24,24	0
2	CL	C	302	1/1	0.99	0.05	32,32,32,32	0
2	CL	D	301	1/1	0.99	0.06	29,29,29,29	0

6.5 Other polymers (i)

There are no such residues in this entry.