wwPDB X-ray Structure Validation Summary Report (i)

Feb 6, 2015 – 12:02 PM GMT

PDB ID	:	2QK9
Title	:	Human RNase H catalytic domain mutant D210N in complex with 18-mer RNA/DNA hybrid
Authors	:	Nowotny, M.; Gaidamakov, S.A.; Ghirlando, R.; Cerritelli, S.M.; Crouch, R.J.; Yang, W.
Deposited on	:	2007-07-10
Resolution	:	2.55 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at http://wwpdb.org/ValidationPDFNotes.html

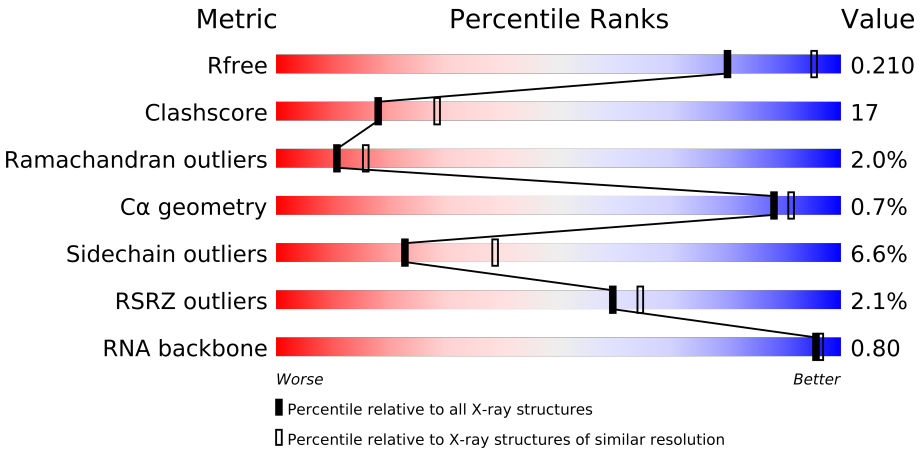
The following versions of software and data (see references) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.17 November 2013
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk24548
Percentile statistics	:	23426
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk24548

1 Overall quality at a glance

The reported resolution of this entry is 2.55 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}	77520	3957 (2.58-2.50)
Clashscore	88313	4689 (2.58-2.50)
Ramachandran outliers	86584	4597 (2.58-2.50)
Cα geometry	86677	4600 (2.58-2.50)
Sidechain outliers	86556	4599 (2.58-2.50)
RSRZ outliers	77580	3958 (2.58-2.50)
RNA backbone	2044	1045 (3.08-2.00)

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 on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	B	18	
2	C	18	
3	A	154	

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

Mol	Type	Chain	Res	Chirality	Geometry	Electron density
5	NA	A	9001	-	-	X
6	FLC	A	1001	-	-	X
7	16D	C	1004	-	-	X
8	GOL	A	1002	-	-	X
8	GOL	A	1003	-	-	X

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 2112 atoms, of which 0 are hydrogen and 0 are deuterium.

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 RNA chain called 5'-R(*AP*GP*UP*GP*CP*GP*AP*CP*AP*CP*CP*UP*GP*AP*UP*UP*CP*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	B	18	377	170	66	124	17	0	0	0

- Molecule 2 is a DNA chain called 5'-D(*GP*GP*AP*AP*TP*CP*AP*GP*GP*TP*GP*TP*CP*GP*CP*AP*CP*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	18	369	176	70	106	17	0	0	0

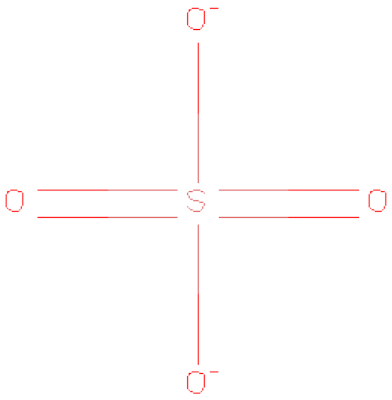
- Molecule 3 is a protein called Ribonuclease H1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	153	1188	739	224	218	7	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	133	GLY	-	EXPRESSION TAG	UNP O60930
A	134	SER	-	EXPRESSION TAG	UNP O60930
A	135	HIS	-	EXPRESSION TAG	UNP O60930
A	210	ASN	ASP	ENGINEERED	UNP O60930

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

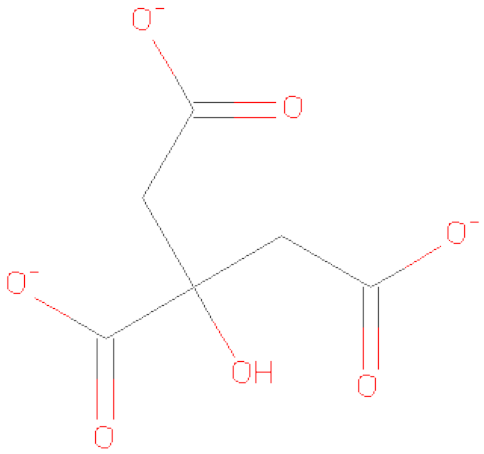


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total O S 5 4 1	0	0
4	B	1	Total O S 5 4 1	0	0
4	C	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0
4	A	1	Total O S 5 4 1	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

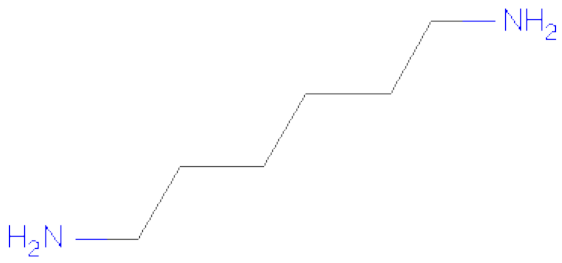
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Na 1 1	0	0

- Molecule 6 is CITRATE ANION (three-letter code: FLC) (formula: $C_6H_5O_7$).



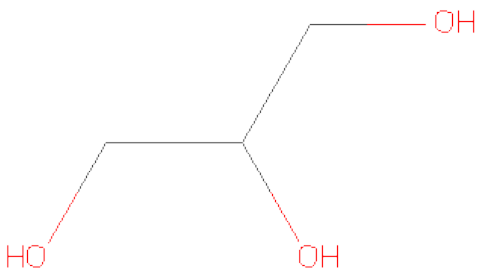
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total C O 13 6 7	0	0

- Molecule 7 is HEXANE-1,6-DIAMINE (three-letter code: 16D) (formula: $C_6H_{16}N_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	1	Total C N 8 6 2	0	0

- Molecule 8 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	A	1	Total C O 6 3 3	0	0
8	A	1	Total C O 6 3 3	0	0

- Molecule 9 is water.

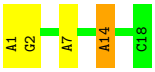
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	B	22	Total O 22 22	0	0
9	C	26	Total O 26 26	0	0
9	A	61	Total O 61 61	0	0

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 errors 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.

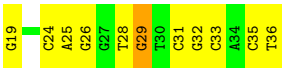
• Molecule 1: 5'-R(*AP*GP*UP*GP*CP*GP*AP*CP*AP*CP*CP*UP*GP*AP*UP*UP*CP*C)-3'

Chain B:



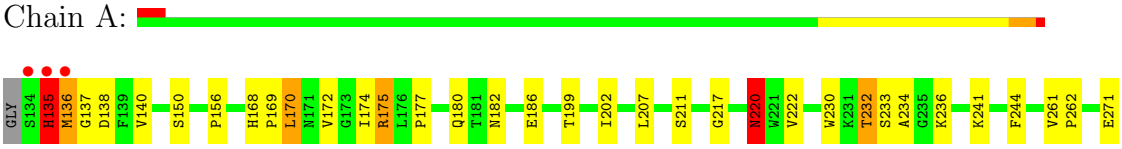
• Molecule 2: 5'-D(*GP*GP*AP*AP*TP*CP*AP*GP*GP*TP*GP*TP*CP*GP*CP*AP*CP*T)-3'

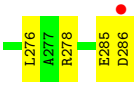
Chain C:



• Molecule 3: Ribonuclease H1

Chain A:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α, β, γ	158.58Å 158.58Å 142.06Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 2.55 49.37 – 2.30	Depositor EDS
% Data completeness (in resolution range)	94.0 (30.00-2.55) 82.9 (49.37-2.30)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.36 (at 2.29Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.190 , 0.216 0.185 , 0.210	Depositor DCC
R_{free} test set	2085 reflections (10.93%)	DCC
Wilson B-factor (Å^2)	41.1	Xtriage
Anisotropy	0.537	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 39.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 26830 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2112	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.07% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: 16D, GOL, FLC, SO4, NA

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	B	0.83	0/420	0.88	1/652 (0.2%)
2	C	0.83	0/414	1.03	2/638 (0.3%)
3	A	0.71	0/1213	0.86	2/1636 (0.1%)
All	All	0.76	0/2047	0.90	5/2926 (0.2%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	29	DG	O5'-P-OP1	-8.64	97.92	105.70
2	C	29	DG	C5'-C4'-O4'	-6.63	96.70	109.30
3	A	175	ARG	NE-CZ-NH1	5.36	122.98	120.30
1	B	14	A	C5'-C4'-C3'	-5.34	107.45	116.00
3	A	135	HIS	N-CA-C	5.25	125.19	111.00

There are no chirality outliers.

There are no planarity outliers.

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	377	0	196	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	369	0	204	20	0
3	A	1188	0	1154	34	0
4	A	20	0	0	0	0
4	B	10	0	0	0	0
4	C	5	0	0	0	0
5	A	1	0	0	0	0
6	A	13	0	5	1	0
7	C	8	0	16	5	0
8	A	12	0	16	0	0
9	A	61	0	0	6	0
9	B	22	0	0	1	0
9	C	26	0	0	1	0
All	All	2112	0	1591	61	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 17.

The worst 5 of 61 close contacts within the same asymmetric unit are listed below. Note that the hydrogen positions were calculated by MolProbity and these may be different from any deposited coordinates. The reason for using them is that the clashscore method is a tool to assess close contacts between **heavy atoms**, and specifically not a validation tool for hydrogen positions. Therefore, the algorithm must be used consistently for every structure.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:C:25:DA:H8	2:C:25:DA:H5'	1.26	0.98
3:A:220:ASN:HD22	3:A:220:ASN:C	1.63	0.98
3:A:186:GLU:HG2	3:A:211:SER:HB2	1.50	0.93
2:C:24:DC:C2'	2:C:25:DA:H5"	2.01	0.90
3:A:232:THR:CG2	3:A:234:ALA:H	1.86	0.89

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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
3	A	151/154 (98%)	145 (96%)	3 (2%)	3 (2%)	11 16

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	135	HIS
3	A	136	MET
3	A	220	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
3	A	121/124 (98%)	113 (93%)	8 (7%)	23 39

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	199	THR
3	A	278	ARG
3	A	232	THR
3	A	170	LEU
3	A	220	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
3	A	203	ASN
3	A	220	ASN
3	A	223	GLN

5.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	B	17/18 (94%)	1 (5%)	0

All (1) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	B	14	A

There are no RNA pucker outliers to report.

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 12 ligands modelled in this entry, 1 is monoatomic - leaving 11 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	$\#\|Z\| > 2$
6	FLC	A	1001	-	3,12,12	0.64	0	3,17,17	1.12	0
8	GOL	A	1002	-	5,5,5	0.40	0	5,5,5	0.37	0
8	GOL	A	1003	-	5,5,5	0.49	0	5,5,5	0.37	0
4	SO4	A	1005	-	4,4,4	0.42	0	6,6,6	0.32	0
4	SO4	A	1006	-	4,4,4	0.44	0	6,6,6	0.33	0
4	SO4	A	1010	-	4,4,4	0.39	0	6,6,6	0.19	0
4	SO4	A	1011	-	4,4,4	0.36	0	6,6,6	0.22	0
4	SO4	B	1007	-	4,4,4	0.28	0	6,6,6	0.16	0
4	SO4	B	1008	-	4,4,4	0.31	0	6,6,6	0.11	0
7	16D	C	1004	-	7,7,7	0.63	0	6,6,6	0.64	0
4	SO4	C	1009	-	4,4,4	0.37	0	6,6,6	0.21	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
6	FLC	A	1001	-	-	0/6/16/16	0/0/0/0
8	GOL	A	1002	-	-	0/4/4/4	0/0/0/0
8	GOL	A	1003	-	-	0/4/4/4	0/0/0/0
4	SO4	A	1005	-	-	0/0/0/0	0/0/0/0
4	SO4	A	1006	-	-	0/0/0/0	0/0/0/0
4	SO4	A	1010	-	-	0/0/0/0	0/0/0/0
4	SO4	A	1011	-	-	0/0/0/0	0/0/0/0
4	SO4	B	1007	-	-	0/0/0/0	0/0/0/0
4	SO4	B	1008	-	-	0/0/0/0	0/0/0/0
7	16D	C	1004	-	-	0/5/5/5	0/0/0/0
4	SO4	C	1009	-	-	0/0/0/0	0/0/0/0

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.

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.

5.9 Trace-atom model geometry

5.9.1 Abnormal distances between consecutive trace-atoms (i)

The following table provides a summary of the distances observed between consecutive Cα atoms (in protein chains) or P atoms (in RNA or DNA chains).

Mol	Chain	Analysed	Outliers
1	B	16/18 (88.9%)	0/16 (0.0%)
2	C	16/18 (88.9%)	0/16 (0.0%)

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Outliers
3	A	152/154 (98.7%)	0/152 (0.0%)
All	All	184/190 (96.8%)	0/184 (0.0%)

5.9.2 Cα torsion geometry

The following table provides a summary of the Cα pseudo-geometry for proteins. The "Percentiles" column shows the percent Cα pseudo-geometry outliers as a percentile score with respect to all X-ray entries, followed by that with respect to X-ray 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
3	A	150/154 (97%)	111 (74%)	38 (25%)	1 (1%)	88 98

All (1) Cα pseudo-geometry outliers are listed below:

Mol	Chain	Res	Type
3	A	220	ASN

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	B	18/18 (100%)	-0.96	0 100 100	27, 39, 53, 55	0
2	C	18/18 (100%)	-0.59	0 100 100	33, 39, 59, 64	0
3	A	153/154 (99%)	-0.32	4 (2%) 52 57	25, 39, 62, 99	5 (3%)
All	All	189/190 (99%)	-0.41	4 (2%) 59 64	25, 39, 59, 99	5 (2%)

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	134	SER	4.4
3	A	136	MET	3.9
3	A	286	ASP	3.2
3	A	135	HIS	2.8

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. 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^{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	RSR	LLDF	B-factors(Å^2)	Q<0.9
7	16D	C	1004	8/8	0.49	30.23	67,72,73,73	0
6	FLC	A	1001	13/13	0.43	15.38	72,75,76,78	13
5	NA	A	9001	1/1	0.16	7.69	50,50,50,50	0
8	GOL	A	1002	6/6	0.28	7.57	72,74,78,81	0
8	GOL	A	1003	6/6	0.22	3.92	70,72,74,74	0
4	SO4	A	1005	5/5	0.10	-0.50	81,82,83,83	0
4	SO4	A	1011	5/5	0.08	-	73,74,76,76	0
4	SO4	C	1009	5/5	0.24	-	121,122,122,122	0
4	SO4	B	1008	5/5	0.34	-	140,140,140,141	0
4	SO4	B	1007	5/5	0.26	-	134,134,134,134	0
4	SO4	A	1010	5/5	0.22	-	126,127,127,128	0
4	SO4	A	1006	5/5	0.14	-	70,70,71,72	0

6.5 Other polymers (i)

There are no such residues in this entry.