W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2017 – 07:59 pm GMT

PDB ID   :   1KJR
Title   :   Crystal Structure of the human galectin-3 CRD in complex with a 3'-derivative of N-Acetyllactosamine
Authors   :   Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J.M.; Nilsson, U.J.
Deposited on   :   2001-12-05
Resolution   :   1.55 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

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 1.55 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	Percentile Ranks	Value
Rfree		0.395
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		2.4%
RSRZ outliers		61.6%

Worse — *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
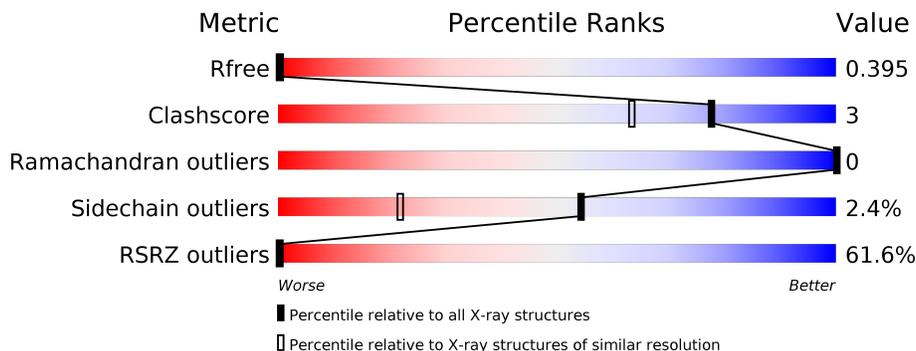
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1088 (1.56-1.56)
Clashscore	112137	1132 (1.56-1.56)
Ramachandran outliers	110173	1110 (1.56-1.56)
Sidechain outliers	110143	1108 (1.56-1.56)
RSRZ outliers	101464	1089 (1.56-1.56)

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. 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	146	58% / 86% / 8% • 5%

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	BEK	A	500	-	X	-	-

WORLDWIDE wwPDB PROTEIN DATA BANK

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 1384 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 Galectin-3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	138	Total 1144	C 728	N 207	O 206	S 3	0	5	0

- Molecule 2 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	2	Total 25	C 14	F 4	N 1	O 6	0	0

- Molecule 3 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).
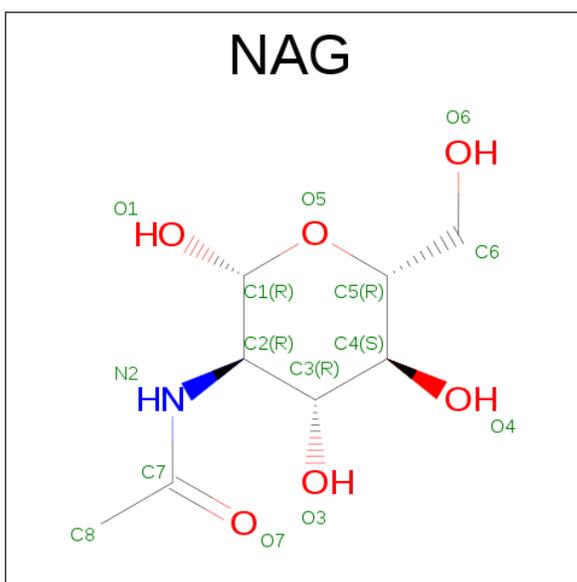
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total 15	C 8	N 1	O 6	0	0

• Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total 1	Cl 1	0	0

• Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	199	Total 199	O 199	0	0

3 Residue-property plots ⓘ

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: Galectin-3

Chain A:
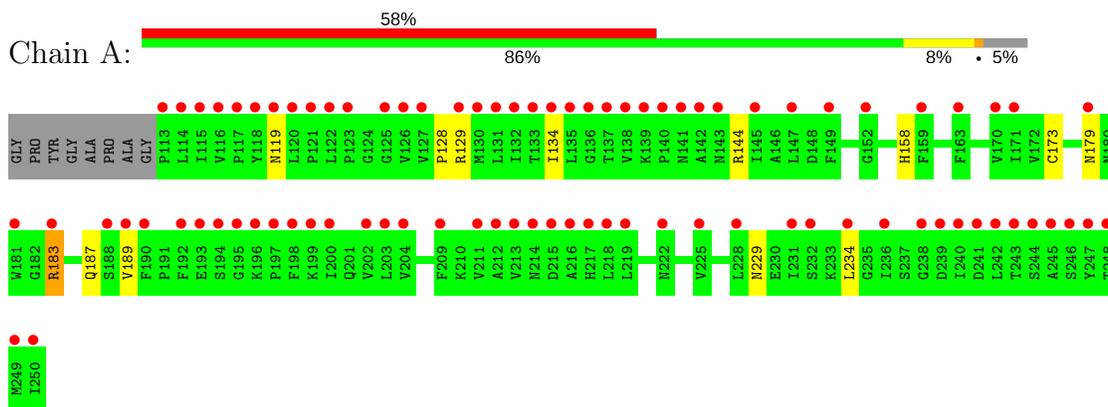
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	37.60Å 58.40Å 64.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.55 29.20 – 1.57	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-1.55) 96.3 (29.20-1.57)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.29 (at 1.57Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.182 , 0.212 0.367 , 0.395	Depositor DCC
R_{free} test set	946 reflections (4.87%)	DCC
Wilson B-factor (Å^2)	18.0	Xtriage
Anisotropy	0.254	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 79.8	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	1384	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.59% 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: BEK, GAL, NAG, 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	0.89	0/1170	0.99	1/1586 (0.1%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	183	ARG	NE-CZ-NH1	9.41	125.00	120.30

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	1144	0	1148	8	0
2	A	25	0	12	1	0
3	A	15	0	14	0	0
4	A	1	0	0	0	0
5	A	199	0	0	2	0
All	All	1384	0	1174	8	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 (8) 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:119:ASN:HB3	5:A:700:HOH:O	1.70	0.91
1:A:134[A]:ILE:HD11	1:A:234:LEU:HD21	1.81	0.63
1:A:187:GLN:HE21	1:A:189:VAL:H	1.52	0.57
1:A:187:GLN:NE2	1:A:189:VAL:H	2.10	0.48
1:A:119:ASN:ND2	5:A:786:HOH:O	2.26	0.45
1:A:144:ARG:HG3	2:A:500:BEK:H83	1.99	0.45
1:A:158:HIS:O	1:A:173:CYS:HA	2.19	0.42
1:A:128:PRO:O	1:A:129:ARG:HB2	2.21	0.41

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	141/146 (97%)	139 (99%)	2 (1%)	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	129/127 (102%)	125 (97%)	4 (3%)	45	13

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	179[A]	ASN
1	A	179[B]	ASN
1	A	183	ARG
1	A	229	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	187	GLN
1	A	229	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

2 carbohydrates 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	$\#|Z| > 2$
2	BEK	A	500	2	15,15,15	7.97	14 (93%)	20,22,22	2.98	5 (25%)
2	GAL	A	501	3,2	10,10,12	0.58	0	10,13,17	0.62	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
2	BEK	A	500	2	-	0/6/6/6	0/1/1/1
2	GAL	A	501	3,2	-	0/2/15/22	0/1/1/1

All (14) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	500	BEK	F4-C4	-14.87	1.10	1.35
2	A	500	BEK	F6-C6	-14.39	1.11	1.35
2	A	500	BEK	O5-C5	-11.93	1.16	1.38
2	A	500	BEK	F3-C3	-9.81	1.19	1.35
2	A	500	BEK	F7-C7	-9.52	1.19	1.35
2	A	500	BEK	C2-C1	-5.56	1.40	1.51
2	A	500	BEK	C1-N	2.19	1.37	1.33
2	A	500	BEK	C4-C3	2.68	1.42	1.37
2	A	500	BEK	C2-C3	3.59	1.45	1.38
2	A	500	BEK	C6-C7	3.90	1.44	1.37
2	A	500	BEK	C5-C4	4.54	1.47	1.39
2	A	500	BEK	C5-C6	5.07	1.49	1.39
2	A	500	BEK	O-C1	5.59	1.35	1.24
2	A	500	BEK	C2-C7	6.79	1.50	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	500	BEK	O-C1-N	-5.90	114.19	122.58
2	A	500	BEK	C2-C7-C6	-2.29	118.93	122.11
2	A	500	BEK	C7-C2-C3	3.08	120.79	116.22
2	A	500	BEK	C8-O5-C5	5.54	129.99	114.81
2	A	500	BEK	O-C1-C2	8.95	131.38	119.72

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
2	A	500	BEK	1	0

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 1 is 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
3	NAG	A	502	2	15,15,15	0.74	1 (6%)	21,21,21	1.14	1 (4%)

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	NAG	A	502	2	-	0/6/26/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	502	NAG	C1-C2	2.21	1.55	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	502	NAG	C1-C2-C3	-3.63	105.60	110.54

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

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(Å²)	Q<0.9
1	A	138/146 (94%)	2.59	85 (61%)	0	0	11, 14, 22, 30	0

All (85) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	114	LEU	9.5
1	A	115	ILE	7.3
1	A	134[A]	ILE	5.6
1	A	247	TYR	5.5
1	A	250	ILE	5.3
1	A	138	VAL	5.3
1	A	197	PRO	5.2
1	A	118	TYR	5.2
1	A	244[A]	SER	5.2
1	A	242	LEU	5.2
1	A	113	PRO	5.1
1	A	135	LEU	5.0
1	A	137	THR	5.0
1	A	198	PHE	4.7
1	A	119	ASN	4.6
1	A	123	PRO	4.5
1	A	243	THR	4.3
1	A	136	GLY	4.3
1	A	120	LEU	4.2
1	A	214	ASN	4.2
1	A	132	ILE	4.2
1	A	200	ILE	4.1
1	A	122	LEU	4.0
1	A	234	LEU	4.0
1	A	213	VAL	4.0
1	A	142	ALA	3.9
1	A	215	ASP	3.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	192	PHE	3.8
1	A	240	ILE	3.7
1	A	126	VAL	3.7
1	A	183	ARG	3.6
1	A	140	PRO	3.6
1	A	121	PRO	3.5
1	A	236	ILE	3.5
1	A	218	LEU	3.5
1	A	204	VAL	3.5
1	A	131	LEU	3.4
1	A	139	LYS	3.4
1	A	117	PRO	3.4
1	A	149	PHE	3.4
1	A	216	ALA	3.4
1	A	212	ALA	3.3
1	A	209	PHE	3.3
1	A	199	LYS	3.2
1	A	133	THR	3.2
1	A	116	VAL	3.2
1	A	231	ILE	3.2
1	A	163	PHE	3.1
1	A	170	VAL	3.1
1	A	238	GLY	3.1
1	A	193	GLU	3.1
1	A	147	LEU	3.0
1	A	190	PHE	2.9
1	A	246	SER	2.9
1	A	203	LEU	2.8
1	A	196	LYS	2.8
1	A	241	ASP	2.8
1	A	188	SER	2.8
1	A	171	ILE	2.7
1	A	141	ASN	2.6
1	A	159	PHE	2.6
1	A	232[A]	SER	2.6
1	A	189	VAL	2.5
1	A	202	VAL	2.5
1	A	179[A]	ASN	2.4
1	A	249	MET	2.4
1	A	127	VAL	2.4
1	A	181	TRP	2.4
1	A	211	VAL	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	130	MET	2.4
1	A	225	VAL	2.3
1	A	143	ASN	2.3
1	A	222[A]	ASN	2.3
1	A	248	THR	2.3
1	A	194	SER	2.2
1	A	217	HIS	2.2
1	A	245	ALA	2.2
1	A	219	LEU	2.1
1	A	228	LEU	2.1
1	A	145	ILE	2.1
1	A	129	ARG	2.1
1	A	125	GLY	2.1
1	A	152	GLY	2.1
1	A	239	ASP	2.0
1	A	195	GLY	2.0

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 ⓘ

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, 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	LLDF	B-factors($\mathring{A}^2$)	Q<0.9
2	BEK	A	500	15/15	0.82	0.16	0.30	18,22,29,29	0
2	GAL	A	501	10/12	0.88	0.11	-0.80	13,15,18,20	0

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, 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	LLDF	B-factors($Å^2$)	Q<0.9
3	NAG	A	502	15/15	0.77	0.14	0.05	18,25,31,32	0
4	CL	A	300	1/1	0.84	0.14	-	16,16,16,16	0

6.5 Other polymers ⓘ

There are no such residues in this entry.