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

wwPDB X-ray Structure Validation Summary Report (i)

Oct 19, 2023 – 09:08 AM EDT

PDB ID	:	2QUM
Title	:	Crystal structure of D-tagatose 3-epimerase from Pseudomonas cichorii with D-tagatose
Authors	:	Yoshida, H.; Yamada, M.; Nishitani, T.; Takada, G.; Izumori, K.; Kamitori, S.
Deposited on	:	2007-08-06
Resolution	:	2.28 Å(reported)

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

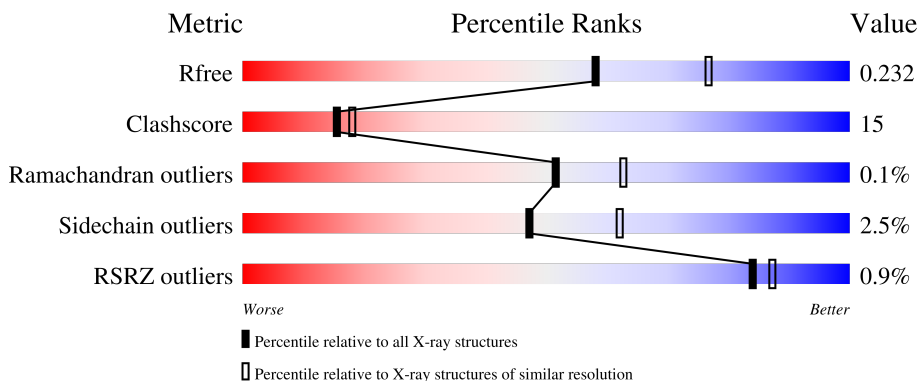|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.28 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		0.232
Clashscore		15
Ramachandran outliers		0.1%
Sidechain outliers		2.5%
RSRZ outliers		0.9%

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$_{free}$	130704	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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	290	.% 72% 27% •
1	B	290	.% 72% 25% •
1	C	290	.% 74% 26% •
1	D	290	2% 71% 28% •

WORLDWIDE
wwPDB
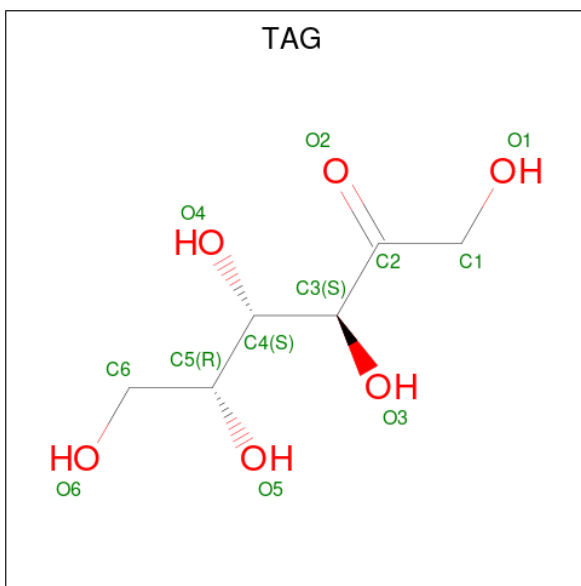PROTEIN DATA BANK

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 9966 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 D-tagatose 3-epimerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	290	Total	C	N	O	S	0	0	0
			2288	1451	392	426	19			
1	B	290	Total	C	N	O	S	0	0	0
			2288	1451	392	426	19			
1	C	290	Total	C	N	O	S	0	0	0
			2288	1451	392	426	19			
1	D	290	Total	C	N	O	S	0	0	0
			2288	1451	392	426	19			

- Molecule 2 is D-tagatose (three-letter code: TAG) (formula: $C_6H_{12}O_6$).

TAG

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		
2	B	1	Total	C	O	0	0
			12	6	6		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	C	1	Total 12	C 6	O 6	0	0
2	D	1	Total 12	C 6	O 6	0	0

- Molecule 3 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

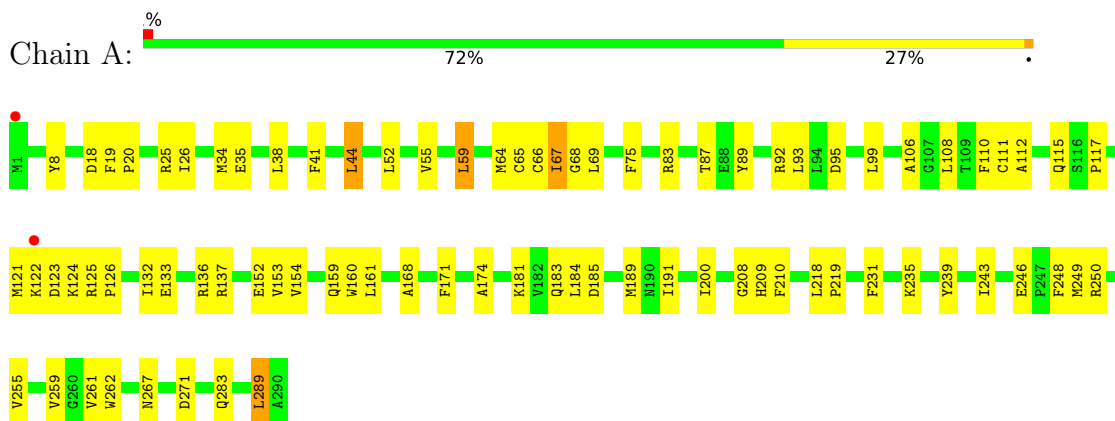| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Mn 1	0	0
3	B	1	Total 1	Mn 1	0	0
3	C	1	Total 1	Mn 1	0	0
3	D	1	Total 1	Mn 1	0	0

- Molecule 4 is water.

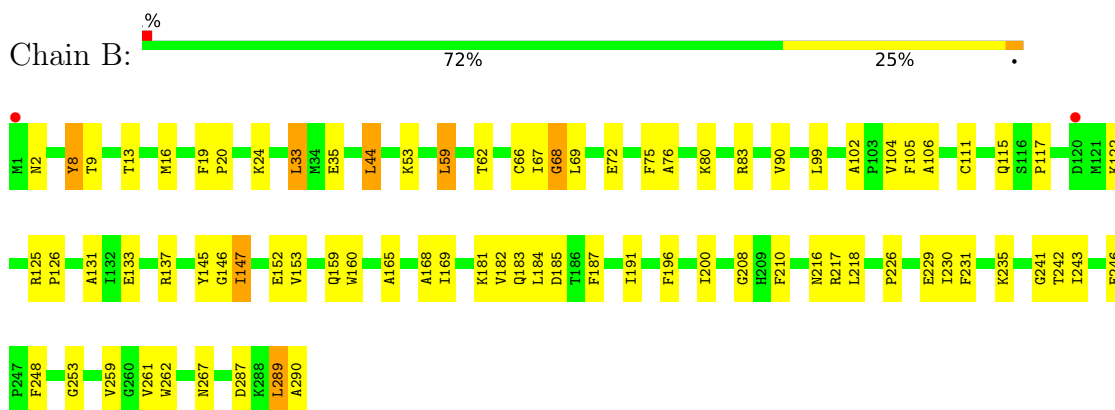| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	162	Total 162	O 162	0	0
4	B	226	Total 226	O 226	0	0
4	C	196	Total 196	O 196	0	0
4	D	178	Total 178	O 178	0	0

3 Residue-property plots ⓘ

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.
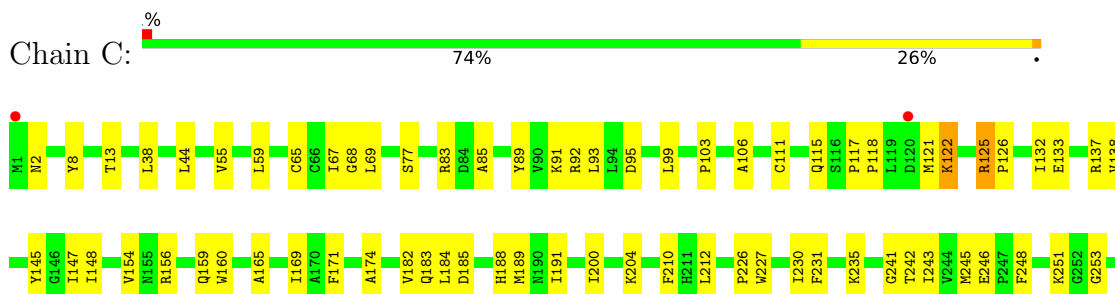
- Molecule 1: D-tagatose 3-epimerase
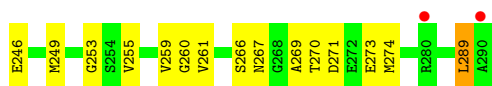

- Molecule 1: D-tagatose 3-epimerase

- Molecule 1: D-tagatose 3-epimerase

V259 G260 V261 W262 R263 N267 R280 L289 A290

• Molecule 1: D-tagatose 3-epimerase

Chain D:

2%

71%   28%

M1 N2 Y8 T9 T13 M16 V17 D18 F19 P20 K24 R25 I36 S37 L38 G39 E40 F41 L44 S45 D46 A47 R50 E51 L52 V55 L59 C66 I67 G68 L69 D74 F75 A76 V90 L99 A106 G107 L108 T109 F110 C111 A112 W113 P114 Q115

S116 P117 P118 M121 R125 P126 A131 K140 Y145 G146 I147 E152 V153 V154 N155 R156 Q159 W160 A165 I169 K181 V182 Q183 L184 D185 H188 M189 N190 I191 R197 L201 A202 C203 G208 N216 R217 L218 P219 F231 K235 I243

E246 M249 G253 S254 V255 V259 G260 V261 S266 N267 C268 A269 T270 D271 E272 E273 M274 R280 L289 A290

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	49.41Å 139.43Å 92.86Å 90.00° 104.04° 90.00°	Depositor
Resolution (Å)	41.31 – 2.28 41.30 – 2.26	Depositor EDS
% Data completeness (in resolution range)	96.7 (41.31-2.28) 95.9 (41.30-2.26)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.12 (at 2.27Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.185 , 0.239 0.177 , 0.232	Depositor DCC
R_{free} test set	5431 reflections (9.71%)	wwPDB-VP
Wilson B-factor (Å^2)	17.1	Xtriage
Anisotropy	0.389	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 41.1	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.053 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9966	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.62% 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 (i)

5.1   Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: TAG, MN

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	0.32	0/2339	0.58	0/3155
1	B	0.34	0/2339	0.59	2/3155 (0.1%)
1	C	0.33	0/2339	0.60	0/3155
1	D	0.32	0/2339	0.59	0/3155
All	All	0.33	0/9356	0.59	2/12620 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	68	GLY	N-CA-C	-5.24	100.00	113.10
1	B	33	LEU	CA-CB-CG	5.09	127.01	115.30

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	2288	0	2245	75	0
1	B	2288	0	2245	65	0
1	C	2288	0	2245	69	0
1	D	2288	0	2245	64	0
2	A	12	0	12	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	12	0	12	2	0
2	C	12	0	12	2	0
2	D	12	0	12	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	162	0	0	5	0
4	B	226	0	0	4	0
4	C	196	0	0	6	0
4	D	178	0	0	3	0
All	All	9966	0	9028	270	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 15.

The worst 5 of 270 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:2:ASN:HB3	1:B:243:ILE:HD13	1.28	1.07
1:A:210:PHE:HB3	1:A:243:ILE:HD13	1.39	1.02
1:C:2:ASN:HB3	1:C:243:ILE:HD13	1.40	0.98
1:A:132:ILE:HD12	1:A:174:ALA:HB3	1.50	0.94
1:A:184:LEU:HD12	1:A:200:ILE:HD13	1.48	0.93

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	
1	A	288/290 (99%)	280 (97%)	8 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	288/290 (99%)	280 (97%)	8 (3%)	0	100	100
1	C	288/290 (99%)	279 (97%)	9 (3%)	0	100	100
1	D	288/290 (99%)	280 (97%)	7 (2%)	1 (0%)	41	49
All	All	1152/1160 (99%)	1119 (97%)	32 (3%)	1 (0%)	51	63

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	74	ASP

5.3.2   Protein sidechains   ⓘ

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	240/240 (100%)	234 (98%)	6 (2%)	47	62
1	B	240/240 (100%)	233 (97%)	7 (3%)	42	56
1	C	240/240 (100%)	234 (98%)	6 (2%)	47	62
1	D	240/240 (100%)	235 (98%)	5 (2%)	53	68
All	All	960/960 (100%)	936 (98%)	24 (2%)	47	62

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	44	LEU
1	C	267	ASN
1	C	125	ARG
1	C	289	LEU
1	B	8	TYR

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	D	2	ASN
1	D	115	GLN
1	D	283	GLN
1	D	267	ASN
1	B	267	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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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	$\#|Z| > 2$
2	TAG	C	291	3	10,11,11	0.54	0	9,14,14	0.61	0
2	TAG	D	291	3	10,11,11	0.69	0	9,14,14	0.56	0
2	TAG	B	291	3	10,11,11	0.62	0	9,14,14	0.57	0
2	TAG	A	291	3	10,11,11	0.62	0	9,14,14	0.50	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	TAG	C	291	3	-	11/16/16/16	-
2	TAG	D	291	3	-	7/16/16/16	-
2	TAG	B	291	3	-	12/16/16/16	-
2	TAG	A	291	3	-	10/16/16/16	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 40 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	291	TAG	C1-C2-C3-C4
2	A	291	TAG	O2-C2-C3-O3
2	A	291	TAG	O2-C2-C3-C4
2	A	291	TAG	O3-C3-C4-O4
2	A	291	TAG	O3-C3-C4-C5

There are no ring outliers.

4 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	291	TAG	2	0
2	D	291	TAG	4	0
2	B	291	TAG	2	0
2	A	291	TAG	3	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, 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	290/290 (100%)	-0.06	2 (0%)	87	90	10, 20, 32, 54	0
1	B	290/290 (100%)	-0.17	2 (0%)	87	90	9, 16, 28, 50	0
1	C	290/290 (100%)	-0.19	2 (0%)	87	90	8, 16, 27, 48	0
1	D	290/290 (100%)	-0.02	5 (1%)	70	75	10, 20, 37, 48	0
All	All	1160/1160 (100%)	-0.11	11 (0%)	84	87	8, 18, 32, 54	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1	MET	4.8
1	C	1	MET	4.2
1	B	120	ASP	3.7
1	C	120	ASP	3.0
1	A	1	MET	2.7

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, 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(Å^2)	Q<0.9
2	TAG	A	291	12/12	0.70	0.38	66,70,71,71	0
2	TAG	B	291	12/12	0.83	0.28	47,51,53,53	0
2	TAG	D	291	12/12	0.83	0.26	40,50,55,58	0
2	TAG	C	291	12/12	0.86	0.27	38,45,50,51	0
3	MN	A	292	1/1	0.99	0.05	26,26,26,26	0
3	MN	C	292	1/1	0.99	0.05	21,21,21,21	0
3	MN	D	292	1/1	0.99	0.08	24,24,24,24	0
3	MN	B	292	1/1	1.00	0.05	23,23,23,23	0

6.5 Other polymers (i)

There are no such residues in this entry.