wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 15, 2023 – 12:39 AM EDT

PDB ID	:	1S5E
Title	:	Cholera holotoxin, Crystal form 1
Authors	:	O'Neal, C.J.; Amaya, E.I.; Jobling, M.G.; Holmes, R.K.; Hol, W.G.
Deposited on	:	2004-01-20
Resolution	:	1.90 Å(reported)

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

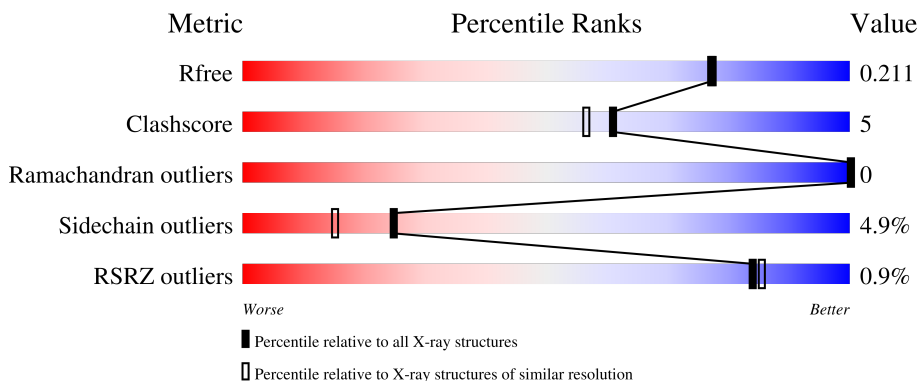|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 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 1.90 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		0.211
Clashscore		5
Ramachandran outliers		0
Sidechain outliers		4.9%
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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	240	4% / 84% / 12% · ·
1	B	240	% / 82% / 10% · 5%
2	D	103	93% / 6% ·
2	E	103	84% / 15% ·
2	F	103	93% / · · ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	103	91% 6%
2	H	103	% 84% 12%
2	J	103	% 89% 9%
2	K	103	% 79% 19%
2	L	103	91% 6%
2	M	103	92% 6%
2	N	103	91% 6%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 12438 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 Cholera enterotoxin, A chain precursor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	232	Total	C	N	O	S	0	0	0
			1813	1138	325	345	5			
1	B	227	Total	C	N	O	S	0	0	0
			1786	1119	318	344	5			

- Molecule 2 is a protein called cholera toxin B protein (CTB).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	103	Total	C	N	O	S	0	0	0
			801	503	138	155	5			
2	E	103	Total	C	N	O	S	0	0	0
			810	508	141	156	5			
2	F	103	Total	C	N	O	S	0	0	0
			813	510	142	156	5			
2	G	102	Total	C	N	O	S	0	0	0
			803	506	139	153	5			
2	H	103	Total	C	N	O	S	0	0	0
			809	508	140	156	5			
2	J	103	Total	C	N	O	S	0	0	0
			811	510	140	156	5			
2	K	103	Total	C	N	O	S	0	0	0
			808	507	140	156	5			
2	L	103	Total	C	N	O	S	0	0	0
			801	502	138	156	5			
2	M	103	Total	C	N	O	S	0	0	0
			805	505	140	155	5			
2	N	103	Total	C	N	O	S	0	0	0
			804	504	139	156	5			

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Na 1	0	0
3	B	1	Total 1	Na 1	0	0

- Molecule 4 is beta-D-galactopyranose (three-letter code: GAL) (formula: $C_6H_{12}O_6$).
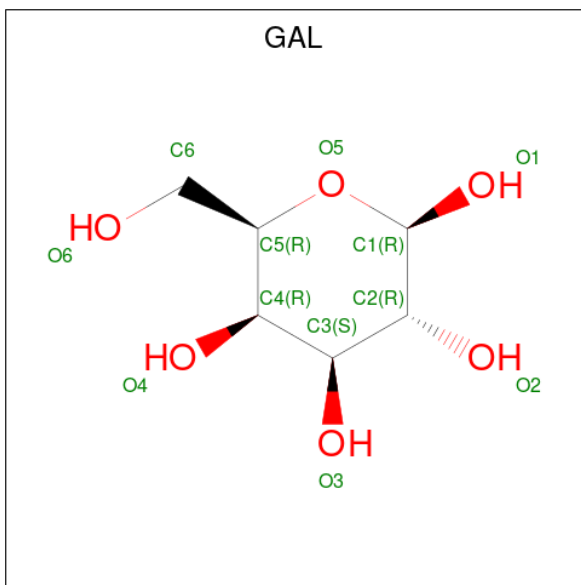
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	D	1	Total 12	C 6	O 6	0	0
4	J	1	Total 12	C 6	O 6	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	103	Total 103	O 103	0	0
5	D	57	Total 57	O 57	0	0
5	E	62	Total 62	O 62	0	0
5	F	52	Total 52	O 52	0	0
5	G	52	Total 52	O 52	0	0
5	H	51	Total 51	O 51	0	0

Continued on next page...

Continued from previous page...

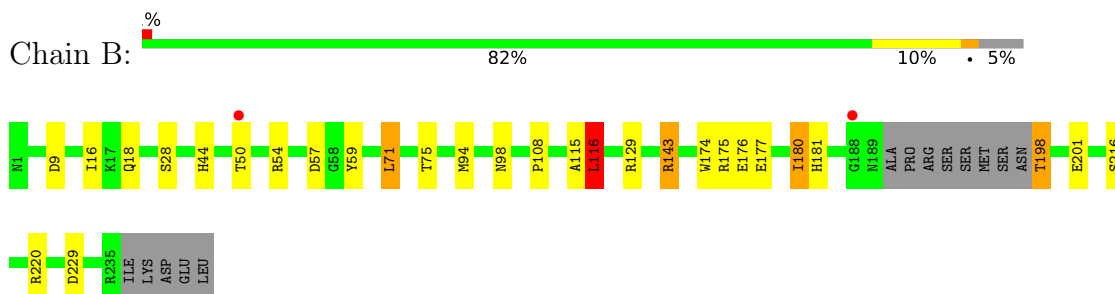| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	106	Total 106	O 106	0	0
5	J	61	Total 61	O 61	0	0
5	K	51	Total 51	O 51	0	0
5	L	51	Total 51	O 51	0	0
5	M	46	Total 46	O 46	0	0
5	N	56	Total 56	O 56	0	0

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: Cholera enterotoxin, A chain precursor
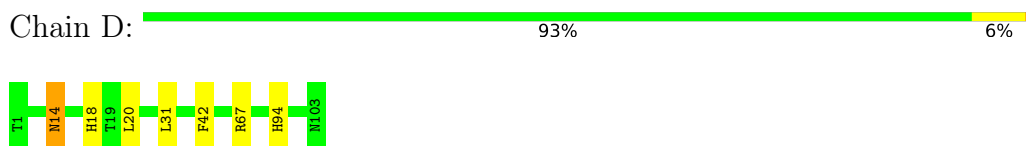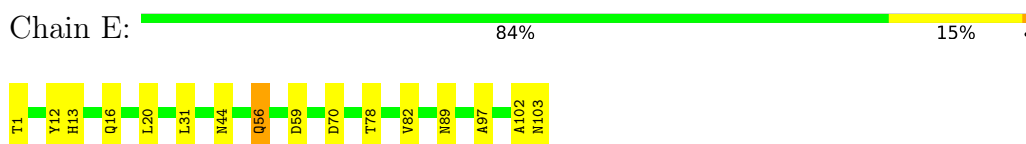
Chain A:

• Molecule 1: Cholera enterotoxin, A chain precursor

Chain B:

• Molecule 2: cholera toxin B protein (CTB)

Chain D:

• Molecule 2: cholera toxin B protein (CTB)

Chain E:

• Molecule 2: cholera toxin B protein (CTB)

Chain F: 93%

T1 C9 T15 H18 T19 L20 L31 A32 H94 N103

• Molecule 2: cholera toxin B protein (CTB)

Chain G: 91%   6%

T1 N4 D7 H18 T19 L20 R35 F48 R67 A97 A102 ASN

• Molecule 2: cholera toxin B protein (CTB)

Chain H: %   84%   12%

T1 C9 Q16 L20 S26 E29 S30 L31 A32 G33 K34 T41 F42 K43 N44 D59 D70 L77 N89 N103

• Molecule 2: cholera toxin B protein (CTB)

Chain J: %   89%   9%

T1 D7 N14 H18 D22 Q56 R67 I74 C86 V87 W88 T92 P93 H94 A98 N103

• Molecule 2: cholera toxin B protein (CTB)

Chain K: %   79%   19%

T1 D7 L8 C9 Y12 Q16 D22 E29 K43 N44 G45 A46 Q56 H57 K62 I65 T78 E79 V82 N89 A97 A102 N103

• Molecule 2: cholera toxin B protein (CTB)

Chain L: 91%   6%

T1 C9 T15 H18 T19 L20 L31 A32 D59 V82 H94 N103

• Molecule 2: cholera toxin B protein (CTB)

Chain M: 92%   6%

T1 N4 D7 L8 C9 E51 D59 K91 N103

• Molecule 2: cholera toxin B protein (CTB)
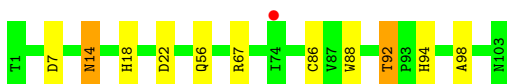
Chain N:

91%   6%   ·

T1 P2 Q3 D7 Q16 L31 N44 H57 I65 L77 N89 N103
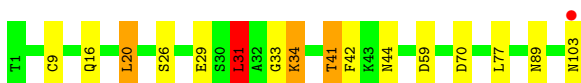
4   Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	59.93Å 108.23Å 122.98Å 90.00° 95.89° 90.00°	Depositor
Resolution (Å)	40.82 – 1.90 49.66 – 1.89	Depositor EDS
% Data completeness (in resolution range)	98.9 (40.82-1.90) 98.8 (49.66-1.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.167 , 0.205 0.176 , 0.211	Depositor DCC
R_{free} test set	6126 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	25.4	Xtriage
Anisotropy	0.475	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 37.3	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	12438	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.39% 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: GAL, 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	A	0.62	0/1865	0.81	5/2530 (0.2%)
1	B	0.60	0/1837	0.82	5/2496 (0.2%)
2	D	0.60	0/815	0.78	0/1104
2	E	0.61	0/824	0.74	1/1114 (0.1%)
2	F	0.65	0/827	0.79	1/1116 (0.1%)
2	G	0.60	0/817	0.76	1/1105 (0.1%)
2	H	0.61	0/823	0.84	4/1113 (0.4%)
2	J	0.62	0/825	0.81	2/1115 (0.2%)
2	K	0.59	0/822	0.75	1/1111 (0.1%)
2	L	0.57	0/815	0.77	2/1104 (0.2%)
2	M	0.57	0/819	0.75	0/1108
2	N	0.59	0/818	0.81	3/1108 (0.3%)
All	All	0.60	0/11907	0.79	25/16124 (0.2%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	M	0	1

There are no bond length outliers.

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	N	31	LEU	CA-CB-CG	6.55	130.37	115.30
2	L	59	ASP	CB-CG-OD2	6.28	123.95	118.30
2	J	7	ASP	CB-CG-OD2	6.03	123.73	118.30
1	A	9	ASP	CB-CG-OD2	5.85	123.57	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	9	ASP	CB-CG-OD2	5.80	123.52	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	M	1	THR	Peptide

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	1813	0	1653	15	0
1	B	1786	0	1634	23	0
2	D	801	0	786	10	0
2	E	810	0	806	10	0
2	F	813	0	813	6	0
2	G	803	0	806	6	0
2	H	809	0	801	8	0
2	J	811	0	807	9	0
2	K	808	0	797	14	0
2	L	801	0	777	6	0
2	M	805	0	793	4	0
2	N	804	0	790	4	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	D	12	0	12	0	0
4	J	12	0	12	2	0
5	A	103	0	0	3	0
5	B	106	0	0	1	0
5	D	57	0	0	0	0
5	E	62	0	0	1	0
5	F	52	0	0	0	0
5	G	52	0	0	2	0
5	H	51	0	0	1	0
5	J	61	0	0	0	0
5	K	51	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	L	51	0	0	0	0
5	M	46	0	0	0	0
5	N	56	0	0	2	0
All	All	12438	0	11287	103	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 5.

The worst 5 of 103 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)
2:K:56:GLN:HE21	2:K:56:GLN:H	1.05	0.97
2:E:56:GLN:HE21	2:E:56:GLN:H	1.07	0.96
2:E:16:GLN:HE21	2:E:89:ASN:HD22	1.14	0.94
2:N:16:GLN:HE21	2:N:89:ASN:HD22	1.15	0.92
2:K:16:GLN:HE21	2:K:89:ASN:HD22	1.18	0.91

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	226/240 (94%)	222 (98%)	4 (2%)	0	100	100
1	B	223/240 (93%)	219 (98%)	4 (2%)	0	100	100
2	D	101/103 (98%)	99 (98%)	2 (2%)	0	100	100
2	E	101/103 (98%)	100 (99%)	1 (1%)	0	100	100
2	F	101/103 (98%)	100 (99%)	1 (1%)	0	100	100
2	G	100/103 (97%)	99 (99%)	1 (1%)	0	100	100
2	H	101/103 (98%)	99 (98%)	2 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	J	101/103 (98%)	100 (99%)	1 (1%)	0	100	100
2	K	101/103 (98%)	100 (99%)	1 (1%)	0	100	100
2	L	101/103 (98%)	100 (99%)	1 (1%)	0	100	100
2	M	101/103 (98%)	99 (98%)	2 (2%)	0	100	100
2	N	101/103 (98%)	100 (99%)	1 (1%)	0	100	100
All	All	1458/1510 (97%)	1437 (99%)	21 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	183/204 (90%)	174 (95%)	9 (5%)	25	15
1	B	183/204 (90%)	174 (95%)	9 (5%)	25	15
2	D	85/89 (96%)	84 (99%)	1 (1%)	71	70
2	E	88/89 (99%)	82 (93%)	6 (7%)	16	7
2	F	89/89 (100%)	85 (96%)	4 (4%)	27	18
2	G	87/89 (98%)	85 (98%)	2 (2%)	50	45
2	H	87/89 (98%)	80 (92%)	7 (8%)	12	5
2	J	87/89 (98%)	85 (98%)	2 (2%)	50	45
2	K	87/89 (98%)	80 (92%)	7 (8%)	12	5
2	L	84/89 (94%)	79 (94%)	5 (6%)	19	9
2	M	86/89 (97%)	82 (95%)	4 (5%)	26	16
2	N	86/89 (97%)	82 (95%)	4 (5%)	26	16
All	All	1232/1298 (95%)	1172 (95%)	60 (5%)	25	15

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	77	LEU
2	M	59	ASP
1	B	143	ARG
2	M	9	CYS
2	N	77	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
2	K	16	GLN
2	L	94	HIS
2	K	44	ASN
2	L	18	HIS
2	M	21	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 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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	GAL	D	751	-	12,12,12	0.64	0	17,17,17	0.72	0
4	GAL	J	752	-	12,12,12	0.59	0	17,17,17	2.54	7 (41%)

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	GAL	D	751	-	-	0/2/22/22	0/1/1/1
4	GAL	J	752	-	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	J	752	GAL	C1-O5-C5	-6.06	102.23	113.66
4	J	752	GAL	C1-C2-C3	-4.96	100.02	110.31
4	J	752	GAL	O5-C1-C2	-4.01	103.12	110.28
4	J	752	GAL	O1-C1-C2	2.70	116.63	109.03
4	J	752	GAL	O5-C5-C6	2.38	112.36	106.44

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	J	752	GAL	2	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	232/240 (96%)	-0.09	9 (3%)	39	42	5, 24, 43, 54	0
1	B	227/240 (94%)	0.01	2 (0%)	84	85	6, 11, 21, 31	0
2	D	103/103 (100%)	-0.18	0	100	100	6, 9, 16, 29	1 (0%)
2	E	103/103 (100%)	-0.42	0	100	100	5, 10, 15, 24	0
2	F	103/103 (100%)	-0.39	0	100	100	5, 9, 16, 25	0
2	G	102/103 (99%)	-0.40	0	100	100	6, 11, 16, 22	0
2	H	103/103 (100%)	-0.16	1 (0%)	82	84	15, 19, 26, 38	0
2	J	103/103 (100%)	-0.07	1 (0%)	82	84	5, 8, 15, 27	1 (0%)
2	K	103/103 (100%)	-0.20	1 (0%)	82	84	6, 11, 19, 31	0
2	L	103/103 (100%)	-0.19	0	100	100	7, 12, 18, 27	0
2	M	103/103 (100%)	-0.22	0	100	100	7, 12, 19, 27	0
2	N	103/103 (100%)	-0.32	0	100	100	6, 9, 15, 26	0
All	All	1488/1510 (98%)	-0.19	14 (0%)	84	85	5, 12, 30, 54	2 (0%)

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	34	GLY	5.0
2	H	103	ASN	3.3
1	A	52	PHE	3.1
2	K	103	ASN	2.6
1	A	80	HIS	2.5

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
4	GAL	J	752	12/12	0.74	0.19	40,43,45,45	0
4	GAL	D	751	12/12	0.83	0.20	36,41,44,46	0
3	NA	A	241	1/1	0.99	0.08	23,23,23,23	0
3	NA	B	241	1/1	0.99	0.06	23,23,23,23	0

6.5   Other polymers ⓘ

There are no such residues in this entry.