wwPDB X-ray Structure Validation Summary Report (i)

Aug 2, 2023 – 11:20 PM EDT

PDB ID	:	1HSJ
Title	:	SARR MBP FUSION STRUCTURE
Authors	:	Zhang, G.
Deposited on	:	2000-12-26
Resolution	:	2.30 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:

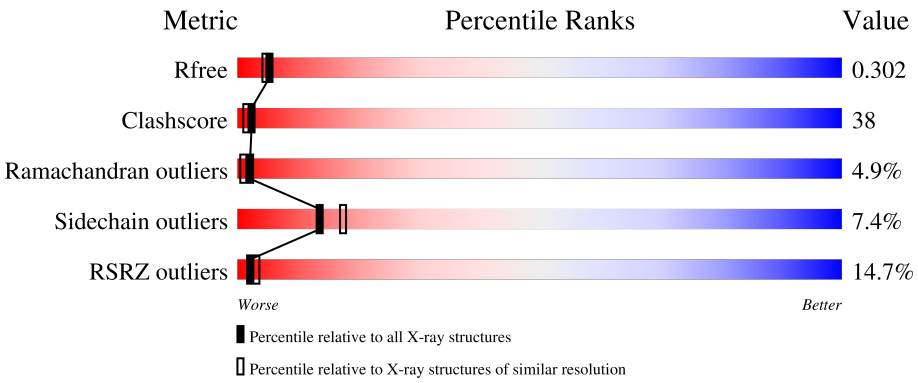
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.34
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.34

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 2.30 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	487	16%; 46%; 46%; 7%
1	B	487	13%; 47%; 45%; 7%

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	GLC	A	671	-	-	X	-
2	GLC	A	672	-	-	X	-
2	GLC	B	674	-	-	X	-

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7730 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.

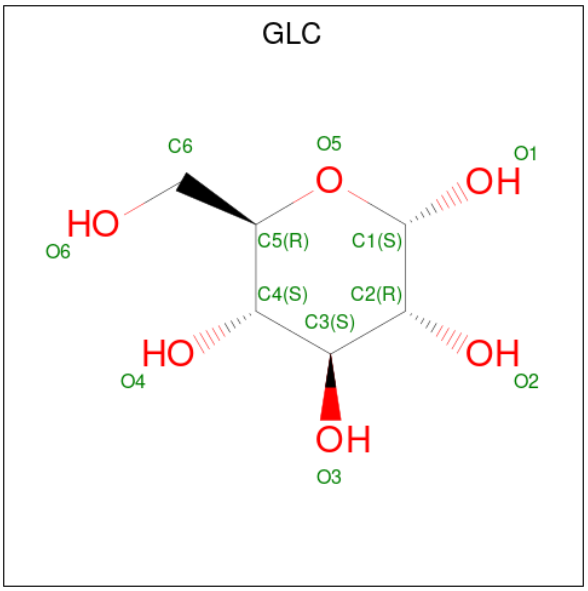
- Molecule 1 is a protein called FUSION PROTEIN CONSISTING OF STAPHYLOCOCCUS ACCESSORY REGULATOR PROTEIN R AND MALTOSE BINDING PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	487	3841	2471	628	734	8	0	0	0
1	B	487	3841	2471	628	734	8	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	362	ALA	-	linker	UNP P0AEY0
A	363	ALA	-	linker	UNP P0AEY0
A	364	ALA	-	linker	UNP P0AEY0
A	365	GLN	-	linker	UNP P0AEY0
A	366	THR	-	linker	UNP P0AEY0
A	367	ASN	-	linker	UNP P0AEY0
A	368	ALA	-	linker	UNP P0AEY0
A	369	ALA	-	linker	UNP P0AEY0
A	370	ALA	-	linker	UNP P0AEY0
A	371	GLU	-	linker	UNP P0AEY0
A	372	PHE	-	linker	UNP P0AEY0
B	362	ALA	-	linker	UNP P0AEY0
B	363	ALA	-	linker	UNP P0AEY0
B	364	ALA	-	linker	UNP P0AEY0
B	365	GLN	-	linker	UNP P0AEY0
B	366	THR	-	linker	UNP P0AEY0
B	367	ASN	-	linker	UNP P0AEY0
B	368	ALA	-	linker	UNP P0AEY0
B	369	ALA	-	linker	UNP P0AEY0
B	370	ALA	-	linker	UNP P0AEY0
B	371	GLU	-	linker	UNP P0AEY0
B	372	PHE	-	linker	UNP P0AEY0

- Molecule 2 is alpha-D-glucopyranose (three-letter code: GLC) (formula: $C_6H_{12}O_6$).



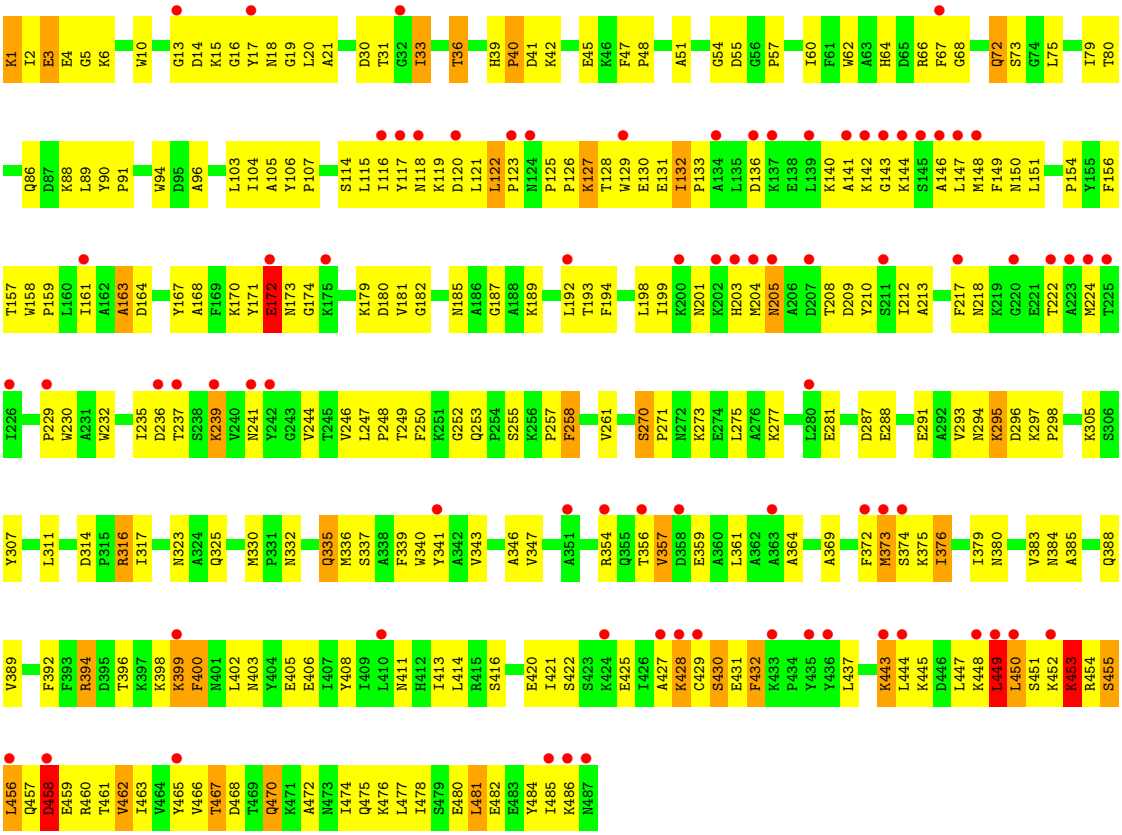
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	12	6	6	0	0
			Total	C	O		
2	A	1	12	6	6	0	0
			Total	C	O		
2	B	1	12	6	6	0	0
			Total	C	O		
2	B	1	12	6	6	0	0

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.

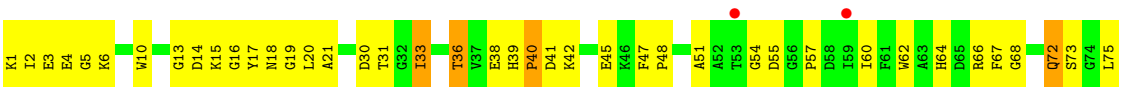
• Molecule 1: FUSION PROTEIN CONSISTING OF STAPHYLOCOCCUS ACCESSORY REGULATOR PROTEIN R AND MALTOSE BINDING PROTEIN

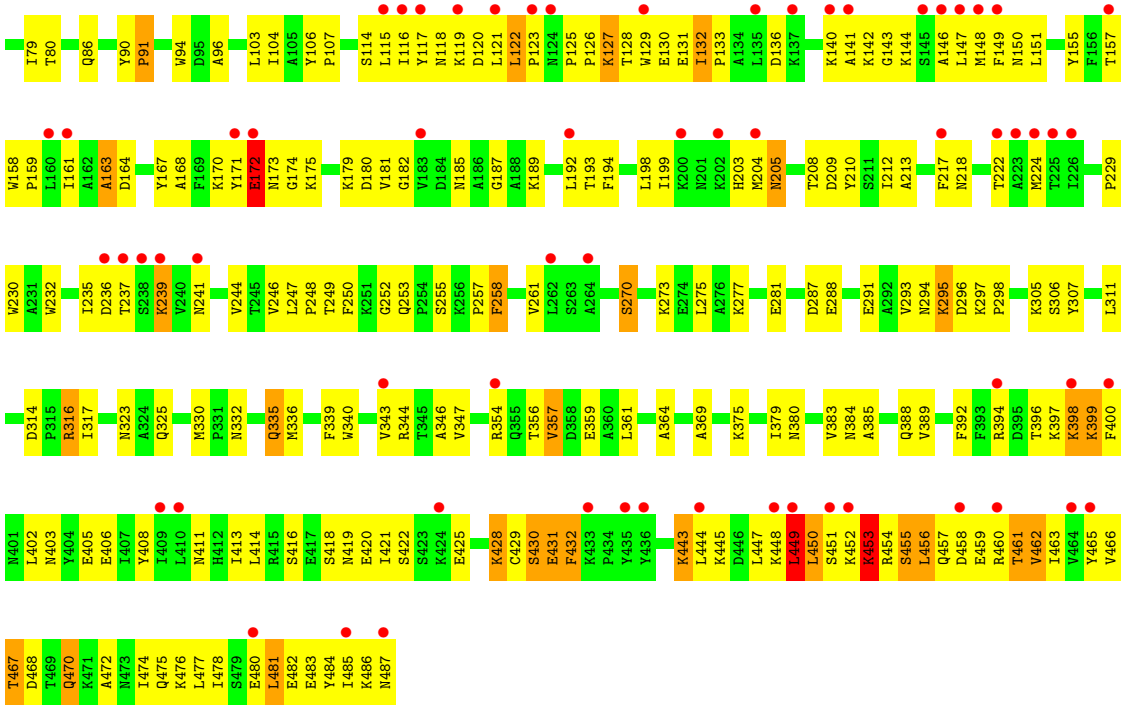
Chain A:



• Molecule 1: FUSION PROTEIN CONSISTING OF STAPHYLOCOCCUS ACCESSORY REGULATOR PROTEIN R AND MALTOSE BINDING PROTEIN

Chain B:




I79 T80 Q86 Y90 P91 W94 D95 A96 L103 I104 A105 Y106 P107 S114 L115 I116 Y117 N118 K119 D120 L121 L122 P123 N124 P125 P126 K127 T128 W129 E130 E131 I132 P133 A134 L135 D136 K137 K140 A141 K142 G143 K144 S145 A146 L147 M148 F149 N150 L151 Y155 F156 T157
W158 P159 L160 I161 A162 A163 D164 Y167 A168 F169 K170 Y171 E172 N173 G174 K175 K179 D180 V181 G182 V183 D184 N185 A186 G187 A188 K189 L192 T193 F194 L198 I199 K200 N201 K202 H203 M204 N205 T208 D209 Y210 S211 I212 A213 F217 N218 T222 A223 M224 T225 I226 P229
W230 A231 W232 I235 D236 T237 S238 K239 V240 N241 V244 T245 V246 L247 P248 T249 F250 K251 G252 Q253 P254 S255 R256 P257 F258 V261 L262 S263 A264 S270 K273 E274 L275 A276 K277 E281 D287 E288 E291 A292 V293 N294 K295 D296 K297 P298 K305 S306 Y307 L311
D314 P315 R316 I317 N323 A324 Q325 M330 P331 N332 Q335 M336 F339 W340 V343 R344 T345 A346 V347 R354 Q355 T356 V357 D358 E359 A360 L361 A364 A369 K375 I379 N380 V383 N384 A385 Q388 V389 F392 F393 R394 D395 T396 K397 K398 K399 F400
N401 L402 N403 Y404 E405 E406 I407 Y408 I409 L410 N411 H412 I413 L414 R415 S416 E417 S418 N419 E420 I421 S422 S423 K424 E425 K428 C429 S430 E431 F432 K433 P434 Y435 Y436 K443 L444 K445 D446 L447 K448 L449 L450 S451 K452 K453 R454 S455 L456 Q457 D458 E459 R460 T461 V462 I463 V464 Y465 V466
T467 D468 T469 Q470 K471 A472 N473 I474 Q475 K476 L477 I478 S479 E480 L481 E482 E483 Y484 I485 K486 N487

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	64.74Å 70.64Å 75.47Å 65.72° 67.18° 69.58°	Depositor
Resolution (Å)	19.95 – 2.30 19.94 – 2.21	Depositor EDS
% Data completeness (in resolution range)	80.6 (19.95-2.30) 85.1 (19.94-2.21)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	0.03	Depositor
$< I/\sigma(I) >$ [1]	1.26 (at 2.21Å)	Xtriage
Refinement program	CNS 0.9	Depositor
R, R_{free}	0.263 , 0.304 0.266 , 0.302	Depositor DCC
R_{free} test set	2202 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å^2)	49.0	Xtriage
Anisotropy	0.292	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 67.0	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	0.94	EDS
Total number of atoms	7730	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.23% 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: GLC

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	0.46	0/3924	0.66	1/5310 (0.0%)
1	B	0.46	0/3924	0.66	1/5310 (0.0%)
All	All	0.46	0/7848	0.66	2/10620 (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	453	LYS	N-CA-C	-6.02	94.75	111.00
1	A	453	LYS	N-CA-C	-5.95	94.94	111.00

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	3841	0	3851	301	0
1	B	3841	0	3851	303	0
2	A	24	0	24	8	0
2	B	24	0	24	6	0
All	All	7730	0	7750	594	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 38.

The worst 5 of 594 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:1:LYS:HE2	1:B:354:ARG:HB3	1.25	1.16
1:A:451:SER:HB2	1:A:470:GLN:HG2	1.29	1.12
1:B:451:SER:HB2	1:B:470:GLN:HG2	1.28	1.12
1:A:414:LEU:HD12	1:A:478:ILE:HD12	1.33	1.10
1:B:414:LEU:HD12	1:B:478:ILE:HD12	1.28	1.06

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	485/487 (100%)	403 (83%)	57 (12%)	25 (5%)	2 1
1	B	485/487 (100%)	409 (84%)	53 (11%)	23 (5%)	2 1
All	All	970/974 (100%)	812 (84%)	110 (11%)	48 (5%)	2 1

5 of 48 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	168	ALA
1	A	398	LYS
1	A	400	PHE
1	A	462	VAL
1	A	467	THR

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	405/405 (100%)	374 (92%)	31 (8%)	13	16
1	B	405/405 (100%)	376 (93%)	29 (7%)	14	18
All	All	810/810 (100%)	750 (93%)	60 (7%)	13	17

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

Mol	Chain	Res	Type
1	A	458	ASP
1	B	449	LEU
1	B	72	GLN
1	B	443	LYS
1	B	481	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	64	HIS
1	B	218	ASN
1	B	419	ASN
1	B	205	ASN
1	B	323	ASN

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)

4 ligands 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	GLC	A	671	-	12,12,12	0.42	0	17,17,17	0.65	0
2	GLC	B	674	-	12,12,12	0.67	0	17,17,17	0.77	1 (5%)
2	GLC	A	672	-	12,12,12	0.39	0	17,17,17	0.60	0
2	GLC	B	673	-	12,12,12	0.42	0	17,17,17	0.58	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	GLC	A	671	-	-	0/2/22/22	0/1/1/1
2	GLC	B	674	-	-	0/2/22/22	0/1/1/1
2	GLC	A	672	-	-	0/2/22/22	0/1/1/1
2	GLC	B	673	-	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	674	GLC	O4-C4-C3	-2.31	105.00	110.35

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	671	GLC	7	0
2	B	674	GLC	6	0
2	A	672	GLC	6	0
2	B	673	GLC	5	0

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 (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(Å^2)	Q<0.9
1	A	487/487 (100%)	0.88	78 (16%)	1	2	21, 72, 109, 127	0
1	B	487/487 (100%)	0.77	65 (13%)	3	4	27, 69, 108, 127	0
All	All	974/974 (100%)	0.83	143 (14%)	2	3	21, 71, 109, 127	0

The worst 5 of 143 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	147	LEU	9.2
1	A	143	GLY	9.0
1	A	429	CYS	7.9
1	B	224	MET	6.1
1	A	239	LYS	5.7

6.2 Non-standard residues in protein, DNA, RNA chains (i)

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

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	GLC	A	672	12/12	0.88	0.20	71,77,84,84	0
2	GLC	A	671	12/12	0.89	0.17	53,64,76,78	0
2	GLC	B	674	12/12	0.92	0.12	31,50,58,64	0
2	GLC	B	673	12/12	0.94	0.12	30,35,44,44	0

6.5 Other polymers (i)

There are no such residues in this entry.