wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 04:41 PM GMT

PDB ID	:	4FNU
Title	:	Crystal structure of GH36 alpha-galactosidase AgaA A355E D478A from Geobacillus stearothermophilus in complex with stachyose
Authors	:	Merceron, R.; Foucault, M.; Haser, R.; Mattes, R.; Watzlawick, H.; Gouet, P.
Deposited on	:	2012-06-20
Resolution	:	3.60 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

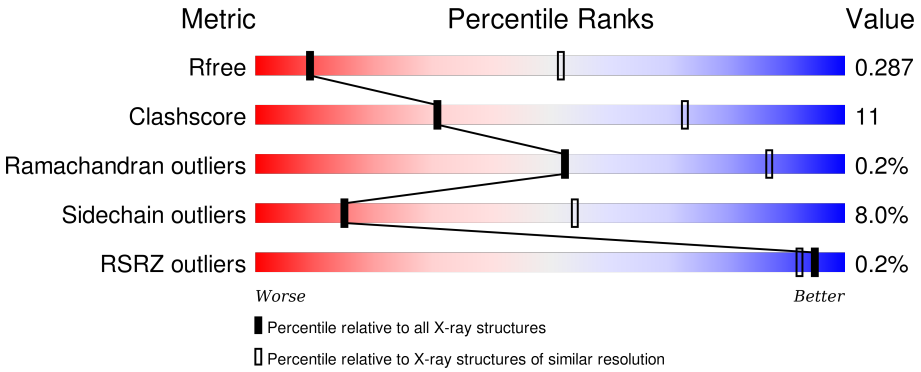
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.60 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}	91344	1408 (3.80-3.40)
Clashscore	102246	1010 (3.74-3.46)
Ramachandran outliers	100387	1007 (3.76-3.44)
Sidechain outliers	100360	1007 (3.76-3.44)
RSRZ outliers	91569	1003 (3.78-3.42)

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	729	66% 28%
1	B	729	69% 26%
1	C	729	% 69% 27%
1	D	729	71% 26%

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	FRU	C	804	-	-	-	X
2	FRU	D	804	-	-	-	X

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 23332 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 Alpha-galactosidase AgaA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			
1	B	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			
1	C	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			
1	D	718	Total	C	N	O	S	0	0	0
			5788	3675	1020	1070	23			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	355	GLU	ALA	ENGINEERED MUTATION	UNP Q9ALJ4
A	478	ALA	ASP	ENGINEERED MUTATION	UNP Q9ALJ4
B	355	GLU	ALA	ENGINEERED MUTATION	UNP Q9ALJ4
B	478	ALA	ASP	ENGINEERED MUTATION	UNP Q9ALJ4
C	355	GLU	ALA	ENGINEERED MUTATION	UNP Q9ALJ4
C	478	ALA	ASP	ENGINEERED MUTATION	UNP Q9ALJ4
D	355	GLU	ALA	ENGINEERED MUTATION	UNP Q9ALJ4
D	478	ALA	ASP	ENGINEERED MUTATION	UNP Q9ALJ4

- Molecule 2 is a polymer of unknown type called SUGAR (4-MER).

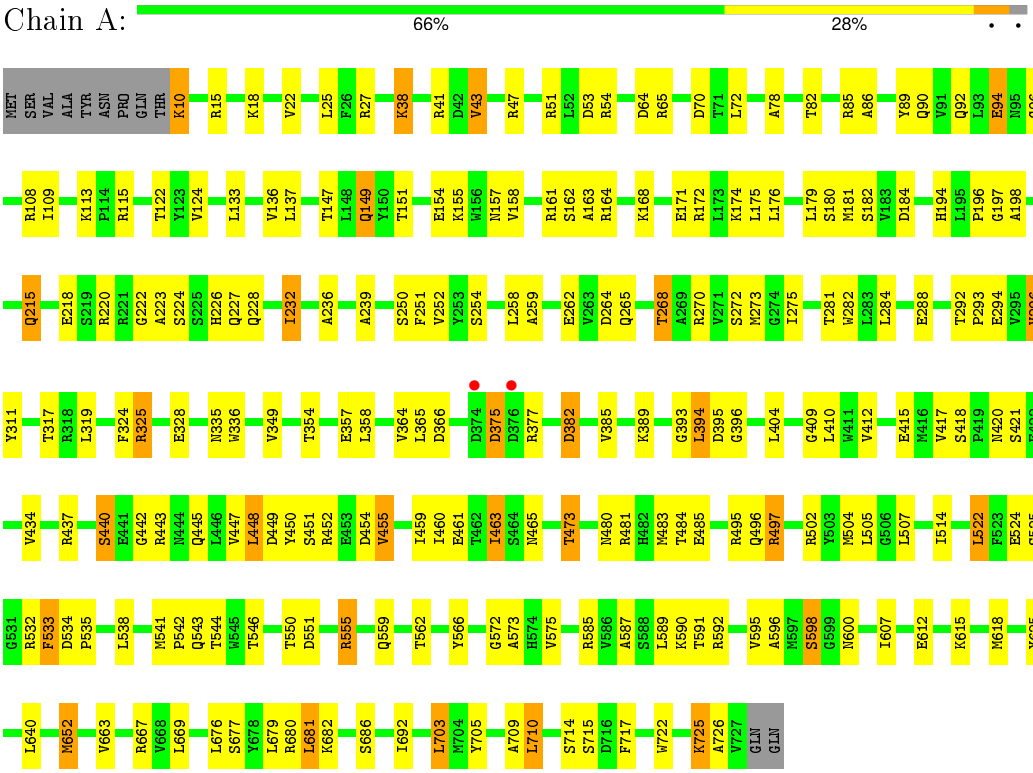
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	4	Total	C	O	0	0
			45	24	21		
2	B	4	Total	C	O	0	0
			45	24	21		
2	C	4	Total	C	O	0	0
			45	24	21		
2	D	4	Total	C	O	0	0
			45	24	21		

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.

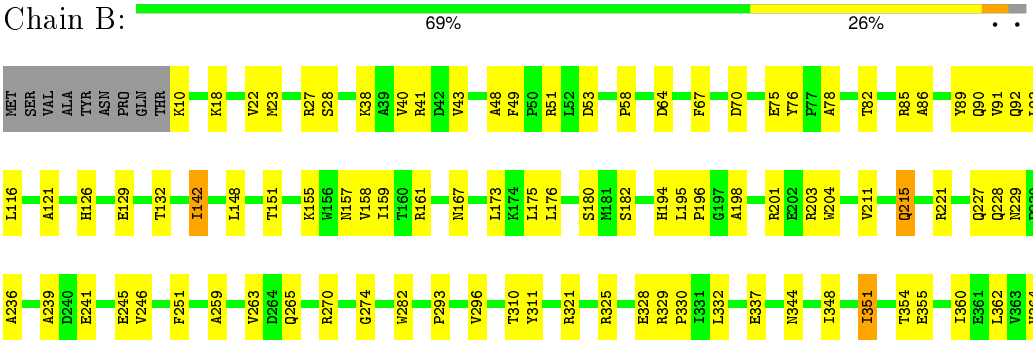
- Molecule 1: Alpha-galactosidase AgaA

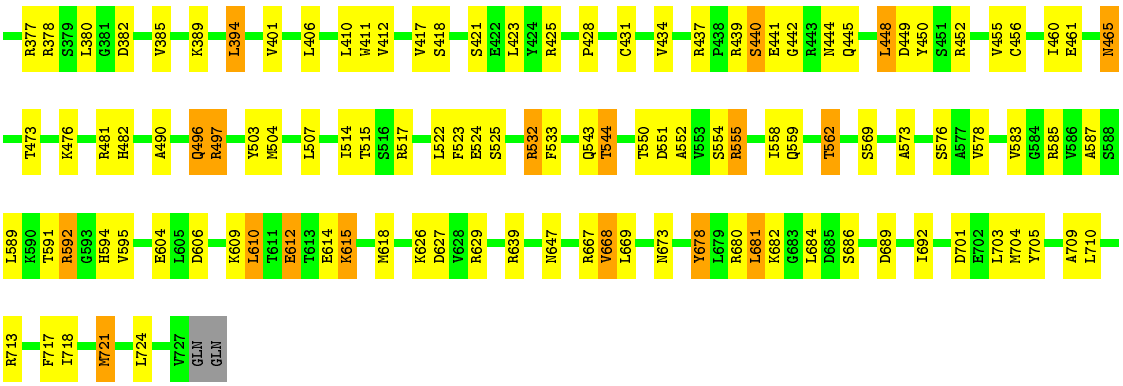
Chain A:



- Molecule 1: Alpha-galactosidase AgaA

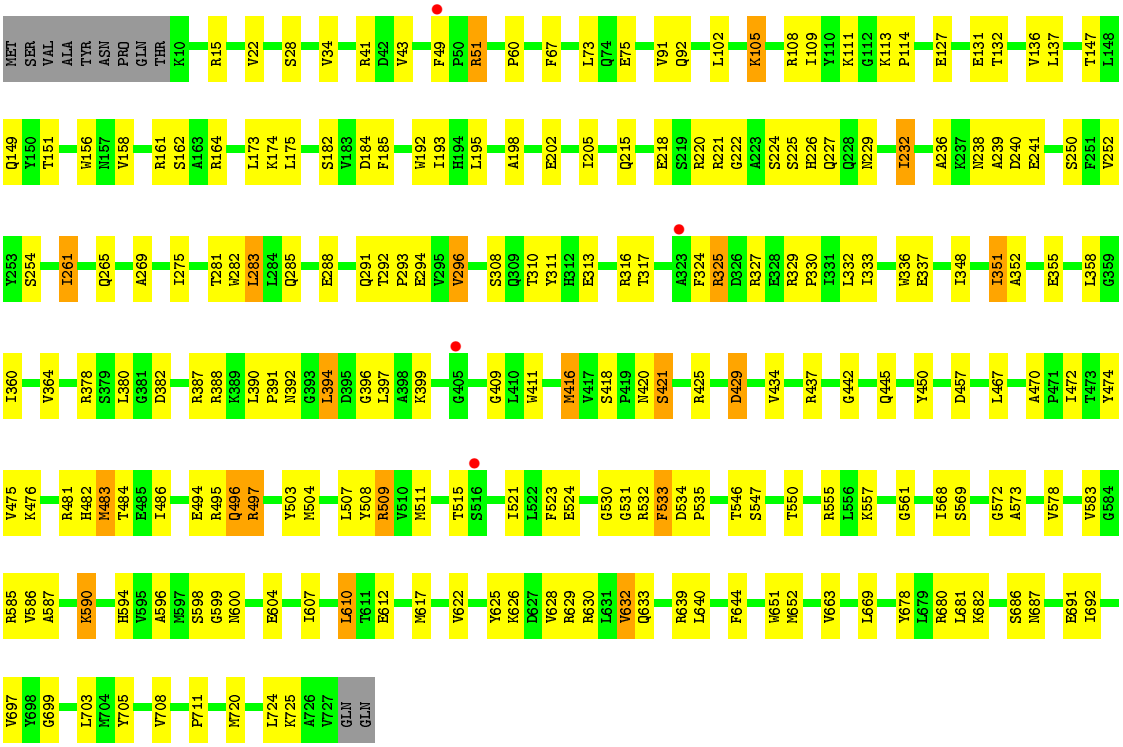
Chain B:





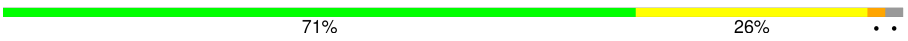
• Molecule 1: Alpha-galactosidase AgaA

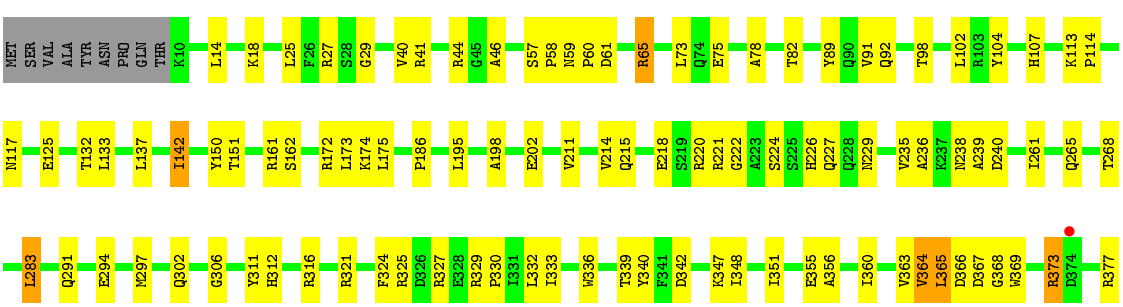
Chain C:

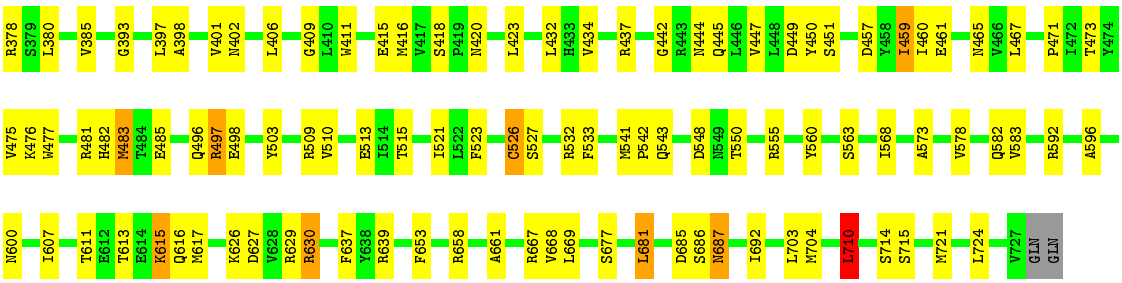


• Molecule 1: Alpha-galactosidase AgaA

Chain D:






R378 S379 L380 V385 G393 L397 A398 V401 N402 L406 G409 L410 W411 E415 N416 V417 S418 P419 N420 L423 L432 H433 V434 R437 G442 R443 N444 Q445 L446 V447 L448 D449 Y450 S451 D457 Y458 T459 I460 E461 N465 V466 L467 P471 T472 T473 Y474
V475 K476 W477 R481 H482 M483 T484 E485 Q496 R497 E498 Y503 R509 V510 E513 I514 T515 I521 L522 F523 C526 S527 R532 F533 M541 P542 Q543 D548 N549 T550 R555 Y560 S563 I568 A573 V578 Q582 V583 R592 A596
N600 I607 T611 E612 T613 E614 K615 Q616 M617 K626 D627 V628 R629 R630 F637 Y638 R639 F653 R658 A661 R667 V668 L669 S677 L681 D685 S686 N687 I692 L703 M704 L710 S714 S715 M721 L724 V727 GLN GLN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	154.07Å 154.07Å 238.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.92 – 3.60 19.92 – 3.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.92-3.60) 99.7 (19.92-3.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$< I/\sigma(I) >$ [1]	3.07 (at 3.62Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.238 , 0.282 0.250 , 0.287	Depositor DCC
R_{free} test set	1906 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	93.0	Xtriage
Anisotropy	0.223	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 19.4	EDS
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 38121 reflections	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	23332	wwPDB-VP
Average B, all atoms (Å^2)	101.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.35% 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: GLA, GLC, FRU

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	0.22	0/5935	0.42	0/8046
1	B	0.22	0/5935	0.40	0/8046
1	C	0.22	0/5935	0.41	0/8046
1	D	0.22	0/5935	0.42	1/8046 (0.0%)
All	All	0.22	0/23740	0.41	1/32184 (0.0%)

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	D	710	LEU	CA-CB-CG	5.87	128.81	115.30

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	5788	0	5621	147	0
1	B	5788	0	5621	130	0
1	C	5788	0	5621	139	0
1	D	5788	0	5621	124	0
2	A	45	0	39	4	0
2	B	45	0	39	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	45	0	39	2	0
2	D	45	0	39	0	0
All	All	23332	0	22640	501	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 11.

The worst 5 of 501 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:681:LEU:HD11	1:B:703:LEU:HB3	1.61	0.81
1:B:684:LEU:H	1:B:704:MET:HE3	1.45	0.81
1:B:113:LYS:HD2	1:B:151:THR:HG21	1.66	0.78
1:A:113:LYS:HD2	1:A:151:THR:HG21	1.66	0.77
1:C:348:ILE:HD11	1:C:397:LEU:HD11	1.65	0.76

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	716/729 (98%)	683 (95%)	32 (4%)	1 (0%)	56	90
1	B	716/729 (98%)	685 (96%)	30 (4%)	1 (0%)	56	90
1	C	716/729 (98%)	683 (95%)	32 (4%)	1 (0%)	56	90
1	D	716/729 (98%)	682 (95%)	32 (4%)	2 (0%)	46	83
All	All	2864/2916 (98%)	2733 (95%)	126 (4%)	5 (0%)	52	87

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	496	GLN
1	B	496	GLN
1	C	496	GLN
1	D	496	GLN
1	D	526	CYS

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	608/618 (98%)	545 (90%)	63 (10%)	9 42
1	B	608/618 (98%)	555 (91%)	53 (9%)	13 51
1	C	608/618 (98%)	567 (93%)	41 (7%)	20 62
1	D	608/618 (98%)	570 (94%)	38 (6%)	22 64
All	All	2432/2472 (98%)	2237 (92%)	195 (8%)	15 54

5 of 195 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	440	SER
1	B	667	ARG
1	D	532	ARG
1	B	465	ASN
1	B	562	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	291	GLN
1	D	580	ASN
1	C	445	GLN
1	A	291	GLN
1	C	621	GLN

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

16 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	GLA	A	801	2	11,11,12	2.68	2 (18%)	14,15,17	0.95	0
2	GLA	A	802	2	11,11,12	2.62	2 (18%)	14,15,17	1.62	2 (14%)
2	GLC	A	803	2	11,11,12	2.46	2 (18%)	14,15,17	1.38	1 (7%)
2	FRU	A	804	2	11,12,12	1.62	4 (36%)	10,18,18	1.76	2 (20%)
2	GLA	B	801	2	11,11,12	2.74	2 (18%)	14,15,17	1.21	1 (7%)
2	GLA	B	802	2	11,11,12	2.57	3 (27%)	14,15,17	2.40	7 (50%)
2	GLC	B	803	2	11,11,12	2.66	4 (36%)	14,15,17	1.60	3 (21%)
2	FRU	B	804	2	11,12,12	1.62	4 (36%)	10,18,18	2.07	4 (40%)
2	GLA	C	801	2	11,11,12	2.73	2 (18%)	14,15,17	1.31	1 (7%)
2	GLA	C	802	2	11,11,12	2.58	2 (18%)	14,15,17	2.17	5 (35%)
2	GLC	C	803	2	11,11,12	2.39	2 (18%)	14,15,17	2.20	6 (42%)
2	FRU	C	804	2	11,12,12	1.59	2 (18%)	10,18,18	2.00	4 (40%)
2	GLA	D	801	2	11,11,12	2.66	2 (18%)	14,15,17	1.36	1 (7%)
2	GLA	D	802	2	11,11,12	2.47	3 (27%)	14,15,17	2.24	6 (42%)
2	GLC	D	803	2	11,11,12	2.50	2 (18%)	14,15,17	1.72	4 (28%)
2	FRU	D	804	2	11,12,12	1.62	2 (18%)	10,18,18	2.14	4 (40%)

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	GLA	A	801	2	-	0/2/19/22	0/1/1/1
2	GLA	A	802	2	-	0/2/19/22	0/1/1/1
2	GLC	A	803	2	-	0/2/19/22	0/1/1/1
2	FRU	A	804	2	-	0/5/24/24	0/1/1/1
2	GLA	B	801	2	-	0/2/19/22	0/1/1/1
2	GLA	B	802	2	-	0/2/19/22	0/1/1/1
2	GLC	B	803	2	-	0/2/19/22	0/1/1/1
2	FRU	B	804	2	-	0/5/24/24	0/1/1/1
2	GLA	C	801	2	-	0/2/19/22	0/1/1/1
2	GLA	C	802	2	-	0/2/19/22	0/1/1/1
2	GLC	C	803	2	-	0/2/19/22	0/1/1/1
2	FRU	C	804	2	-	0/5/24/24	0/1/1/1
2	GLA	D	801	2	-	0/2/19/22	0/1/1/1
2	GLA	D	802	2	-	0/2/19/22	0/1/1/1
2	GLC	D	803	2	-	0/2/19/22	0/1/1/1
2	FRU	D	804	2	-	0/5/24/24	0/1/1/1

The worst 5 of 40 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	801	GLA	C2-C3	-7.94	1.41	1.52
2	B	801	GLA	C2-C3	-7.90	1.41	1.52
2	A	801	GLA	C2-C3	-7.81	1.41	1.52
2	A	802	GLA	C2-C3	-7.67	1.42	1.52
2	D	801	GLA	C2-C3	-7.57	1.42	1.52

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	803	GLC	O5-C1-C2	-2.79	106.34	110.86
2	D	804	FRU	C6-C5-C4	-2.75	108.58	115.08
2	D	802	GLA	O3-C3-C4	-2.74	104.17	110.34
2	B	802	GLA	O2-C2-C3	-2.56	104.97	110.12
2	D	803	GLC	O4-C4-C3	-2.54	104.61	110.34

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	801	GLA	2	0
2	A	802	GLA	2	0
2	C	802	GLA	1	0
2	C	804	FRU	2	0

5.6 Ligand geometry (i)

There are no ligands in this entry.

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

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	A	718/729 (98%)	-0.32	2 (0%) 94 90	61, 101, 145, 173	0
1	B	718/729 (98%)	-0.47	0 100 100	56, 90, 127, 144	0
1	C	718/729 (98%)	-0.27	4 (0%) 90 83	71, 113, 139, 163	0
1	D	718/729 (98%)	-0.42	1 (0%) 95 94	62, 93, 126, 151	0
All	All	2872/2916 (98%)	-0.37	7 (0%) 95 92	56, 100, 136, 173	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	516	SER	2.9
1	C	323	ALA	2.7
1	A	376	ASP	2.7
1	A	374	ASP	2.2
1	C	405	GLY	2.2

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, 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	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
2	FRU	D	804	12/12	0.87	0.31	5.85	102,121,135,143	0
2	FRU	C	804	12/12	0.68	0.52	5.04	99,131,142,146	0
2	GLA	C	802	11/12	0.88	0.27	1.84	99,110,115,119	0
2	GLA	B	802	11/12	0.86	0.28	1.58	104,109,124,127	0
2	GLA	B	801	11/12	0.89	0.23	0.70	95,103,109,110	0
2	GLA	C	801	11/12	0.84	0.25	0.52	114,119,123,125	0
2	GLA	D	802	11/12	0.91	0.18	0.46	94,102,110,110	0
2	GLA	A	801	11/12	0.86	0.26	0.42	105,123,130,131	0
2	GLA	A	802	11/12	0.80	0.29	0.27	106,116,119,121	0
2	GLA	D	801	11/12	0.93	0.19	-0.01	87,98,102,103	0
2	FRU	B	804	12/12	0.81	0.39	-	104,131,137,140	0
2	GLC	B	803	11/12	0.86	0.38	-	122,129,136,138	0
2	FRU	A	804	12/12	0.69	0.51	-	117,138,149,157	0
2	GLC	A	803	11/12	0.80	0.47	-	125,141,146,151	0
2	GLC	C	803	11/12	0.72	0.63	-	124,139,148,150	0
2	GLC	D	803	11/12	0.73	0.33	-	109,129,137,138	0

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.