



# wwPDB X-ray Structure Validation Summary Report

Oct 26, 2024 – 01:06 PM EDT

PDB ID : 6APJ  
Title : Crystal Structure of human ST6GALNAC2  
Authors : Forouhar, F.; Moremen, K.W.; Northeast Structural Genomics Consortium (NESG); Tong, L.  
Deposited on : 2017-08-17  
Resolution : 3.10 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](mailto: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  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](#) ) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 2022.3.0, CSD as543be (2022)  
Xtriage (Phenix) : 1.20.1  
EDS : 3.0  
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)  
CCP4 : 9.0.003 (Gargrove)  
Density-Fitness : 1.0.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.39

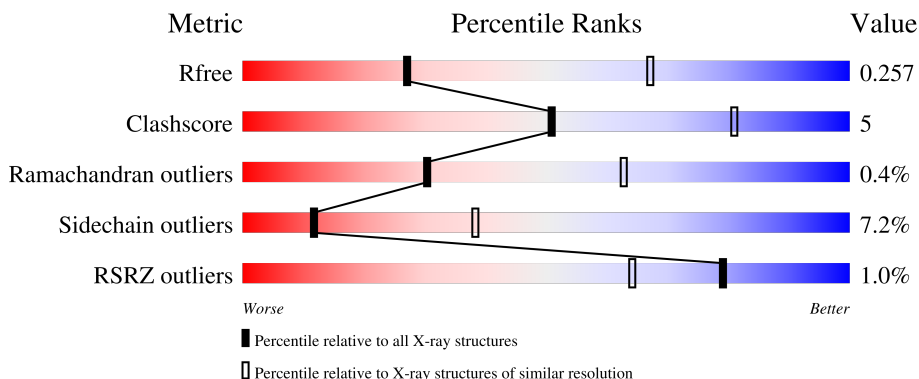
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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))
$R_{free}$	164625	1351 (3.10-3.10)
Clashscore	180529	1454 (3.10-3.10)
Ramachandran outliers	177936	1391 (3.10-3.10)
Sidechain outliers	177891	1391 (3.10-3.10)
RSRZ outliers	164620	1351 (3.10-3.10)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	374	
1	B	374	
1	C	374	
1	D	374	
1	E	374	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	374	 <p>A horizontal bar chart representing the quality of the chain. The bar is divided into three segments: a green segment on the left labeled '66%', a yellow segment in the middle labeled '11%', and a grey segment on the right labeled '21%'. A small red square is at the beginning of the bar, and a small black dot is at the end of the grey segment. A '%' symbol is positioned above the start of the bar.</p>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 13939 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.

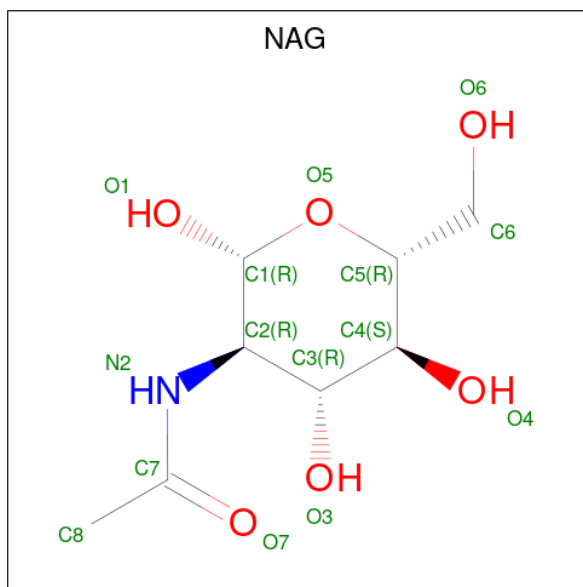
- Molecule 1 is a protein called Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	291	Total 2325	C 1507	N 404	O 406	S 4	Se 4	0	0	0
1	B	285	Total 2273	C 1472	N 395	O 398	S 4	Se 4	0	0	0
1	C	289	Total 2309	C 1497	N 400	O 404	S 4	Se 4	0	0	0
1	D	288	Total 2297	C 1488	N 399	O 402	S 4	Se 4	0	0	0
1	E	288	Total 2307	C 1497	N 401	O 400	S 5	Se 4	0	0	0
1	F	296	Total 2372	C 1540	N 412	O 411	S 5	Se 4	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	initiating methionine	UNP Q9UJ37
B	1	MSE	-	initiating methionine	UNP Q9UJ37
C	1	MSE	-	initiating methionine	UNP Q9UJ37
D	1	MSE	-	initiating methionine	UNP Q9UJ37
E	1	MSE	-	initiating methionine	UNP Q9UJ37
F	1	MSE	-	initiating methionine	UNP Q9UJ37

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).

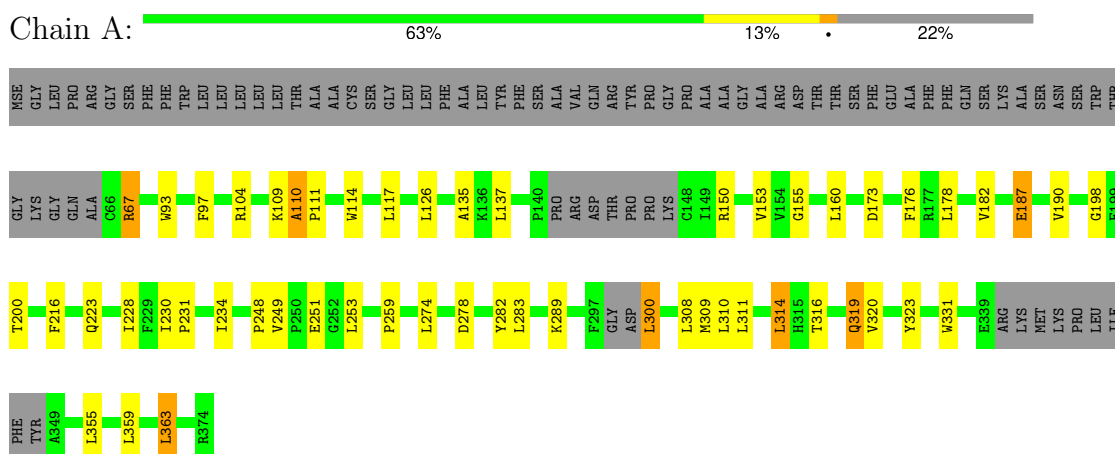


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	Total 14	8	1	5	0	0
2	A	1	Total 14	8	1	5	0	0
2	E	1	Total 14	8	1	5	0	0
2	F	1	Total 14	8	1	5	0	0

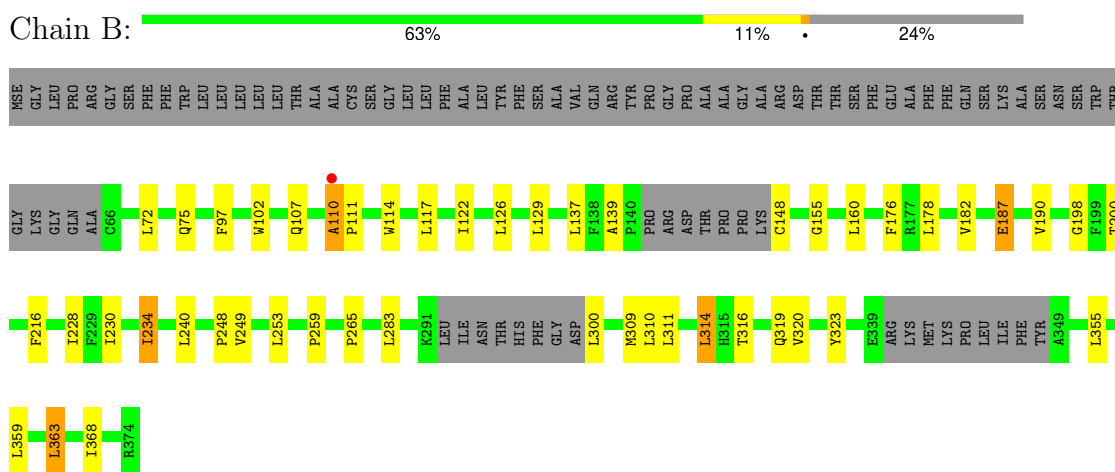
### 3 Residue-property plots [i](#)

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.

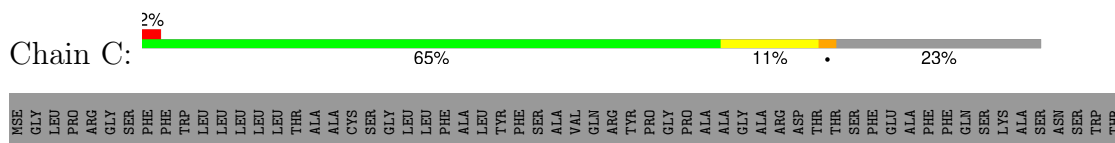
- Molecule 1: Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2

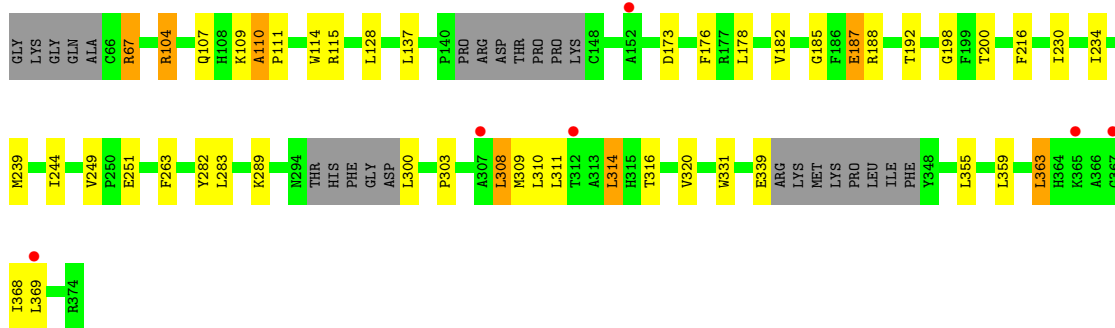


- Molecule 1: Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2

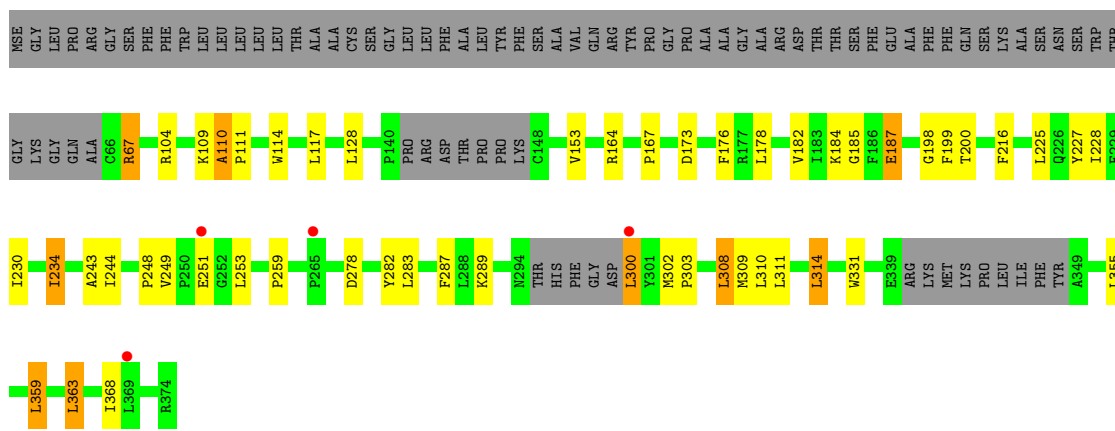


- Molecule 1: Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2

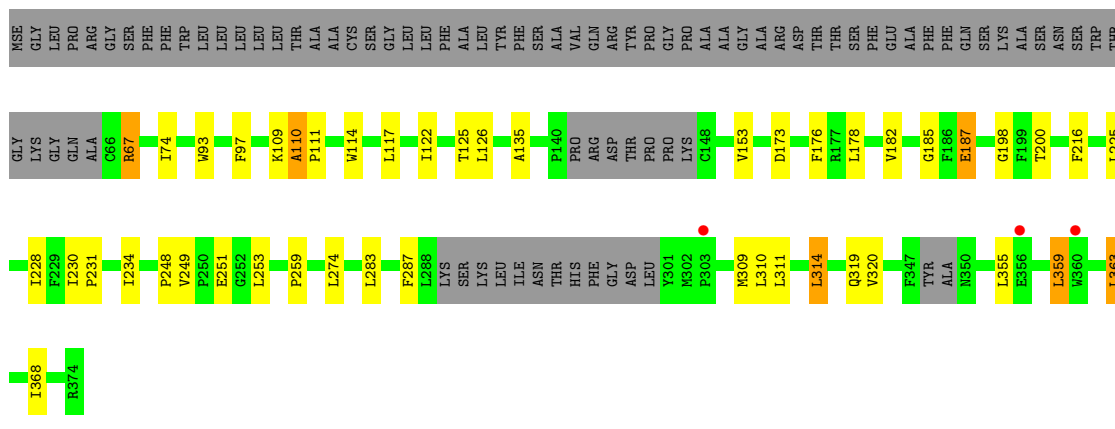




• Molecule 1: Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2

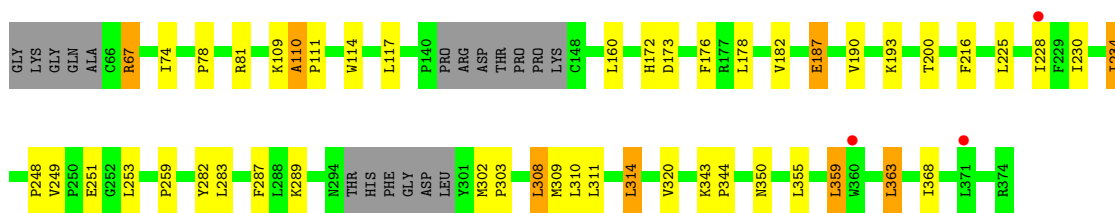


• Molecule 1: Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2



• Molecule 1: Alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 2







## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.35Å 71.80Å 134.25Å 98.78° 101.92° 103.48°	Depositor
Resolution (Å)	44.28 – 3.10 44.28 – 3.10	Depositor EDS
% Data completeness (in resolution range)	93.2 (44.28-3.10) 93.2 (44.28-3.10)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.72 (at 3.12Å)	Xtrriage
Refinement program	PHENIX, REFMAC, CNS	Depositor
R, $R_{free}$	0.201 , 0.252 0.208 , 0.257	Depositor DCC
$R_{free}$ test set	4186 reflections (9.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	91.0	Xtrriage
Anisotropy	0.307	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 70.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.021 for -k,-h,-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	13939	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	100.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.22% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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:  
NAG

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	0.24	0/2388	0.42	0/3232
1	B	0.25	0/2334	0.42	0/3158
1	C	0.24	0/2371	0.42	0/3209
1	D	0.24	0/2358	0.42	0/3191
1	E	0.24	0/2370	0.41	0/3206
1	F	0.24	0/2437	0.42	0/3297
All	All	0.24	0/14258	0.42	0/19293

There are no bond length outliers.

There are no bond angle outliers.

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	2325	0	2281	27	0
1	B	2273	0	2232	23	0
1	C	2309	0	2269	22	0
1	D	2297	0	2260	29	0
1	E	2307	0	2270	23	0
1	F	2372	0	2344	22	0
2	A	28	0	26	0	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	14	0	13	1	0
2	F	14	0	13	1	0
All	All	13939	0	13708	140	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 140 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:182:VAL:HG22	1:A:216:PHE:HB2	1.65	0.78
1:B:182:VAL:HG22	1:B:216:PHE:HB2	1.67	0.75
1:C:182:VAL:HG22	1:C:216:PHE:HB2	1.67	0.74
1:D:182:VAL:HG22	1:D:216:PHE:HB2	1.73	0.70
1:E:228:ILE:HD12	1:E:309:MSE:HE1	1.72	0.70

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	283/374 (76%)	264 (93%)	18 (6%)	1 (0%)	30	63
1	B	277/374 (74%)	255 (92%)	20 (7%)	2 (1%)	19	51
1	C	281/374 (75%)	261 (93%)	19 (7%)	1 (0%)	30	63
1	D	280/374 (75%)	258 (92%)	21 (8%)	1 (0%)	30	63
1	E	280/374 (75%)	265 (95%)	14 (5%)	1 (0%)	30	63
1	F	290/374 (78%)	266 (92%)	23 (8%)	1 (0%)	37	68
All	All	1691/2244 (75%)	1569 (93%)	115 (7%)	7 (0%)	30	63

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	110	ALA
1	B	110	ALA
1	C	110	ALA
1	D	110	ALA
1	E	110	ALA

### 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	248/308 (80%)	230 (93%)	18 (7%)	11	37
1	B	242/308 (79%)	225 (93%)	17 (7%)	12	39
1	C	246/308 (80%)	227 (92%)	19 (8%)	10	35
1	D	245/308 (80%)	226 (92%)	19 (8%)	10	35
1	E	246/308 (80%)	231 (94%)	15 (6%)	15	43
1	F	253/308 (82%)	234 (92%)	19 (8%)	11	36
All	All	1480/1848 (80%)	1373 (93%)	107 (7%)	12	38

5 of 107 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	178	LEU
1	D	359	LEU
1	F	289	LYS
1	D	225	LEU
1	D	289	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	319	GLN

### 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 oligosaccharides 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	NAG	A	501	1	14,14,15	0.26	0	17,19,21	0.40	0
2	NAG	F	401	1	14,14,15	0.61	1 (7%)	17,19,21	1.38	2 (11%)
2	NAG	E	401	1	14,14,15	0.67	1 (7%)	17,19,21	1.37	2 (11%)
2	NAG	A	502	1	14,14,15	0.36	0	17,19,21	0.56	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	NAG	A	501	1	-	2/6/23/26	0/1/1/1
2	NAG	F	401	1	-	6/6/23/26	0/1/1/1
2	NAG	E	401	1	-	4/6/23/26	0/1/1/1
2	NAG	A	502	1	-	2/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	401	NAG	C1-C2	2.25	1.55	1.52
2	F	401	NAG	C1-C2	2.01	1.55	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	401	NAG	C2-N2-C7	4.57	129.02	122.90
2	E	401	NAG	C2-N2-C7	4.50	128.93	122.90
2	F	401	NAG	C1-C2-N2	2.30	114.06	110.43
2	E	401	NAG	C1-C2-N2	2.10	113.75	110.43

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	F	401	NAG	O5-C5-C6-O6
2	F	401	NAG	C4-C5-C6-O6
2	E	401	NAG	C8-C7-N2-C2
2	E	401	NAG	O7-C7-N2-C2
2	F	401	NAG	C8-C7-N2-C2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	F	401	NAG	1	0
2	E	401	NAG	1	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	287/374 (76%)	-0.52	0 <b>100</b> <b>100</b>	46, 73, 130, 228	0
1	B	281/374 (75%)	-0.58	1 (0%) <b>89</b> <b>77</b>	45, 74, 123, 229	0
1	C	285/374 (76%)	-0.02	6 (2%) <b>63</b> <b>44</b>	56, 102, 160, 237	0
1	D	284/374 (75%)	-0.11	4 (1%) <b>73</b> <b>56</b>	50, 102, 172, 231	0
1	E	284/374 (75%)	-0.14	3 (1%) <b>77</b> <b>61</b>	59, 111, 174, 268	0
1	F	292/374 (78%)	-0.18	3 (1%) <b>79</b> <b>64</b>	56, 106, 168, 226	0
All	All	1713/2244 (76%)	-0.26	17 (0%) <b>79</b> <b>64</b>	45, 93, 167, 268	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	360	TRP	5.1
1	D	265	PRO	4.0
1	F	228	ILE	3.7
1	F	371	LEU	3.5
1	E	303	PRO	3.1

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
2	NAG	E	401	14/15	0.53	0.12	135,157,168,182	0
2	NAG	F	401	14/15	0.57	0.13	133,171,182,193	0
2	NAG	A	501	14/15	0.74	0.10	123,136,160,171	0
2	NAG	A	502	14/15	0.79	0.09	150,163,173,177	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.