



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 15, 2018 – 03:52 PM EST

PDB ID : 4EN3  
Title : Crystal structure of a human Valpha24(-) NKT TCR in complex with CD1d/  
alpha-galactosylceramide  
Authors : Lopez-Sagaseta, J.; Adams, E.J.  
Deposited on : 2012-04-12  
Resolution : 2.57 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030736  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : rb-20030736

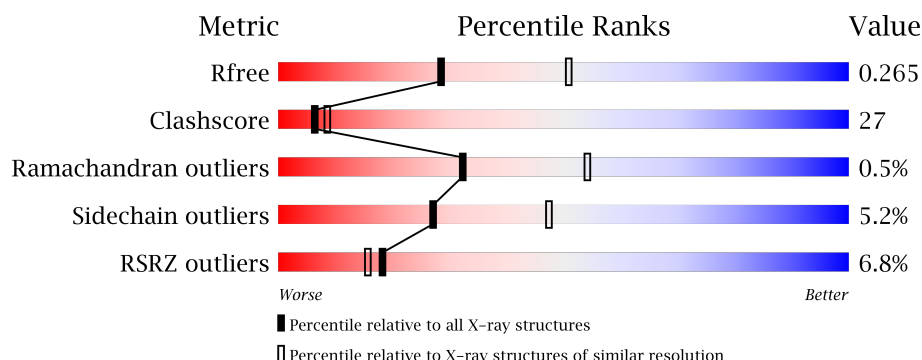
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.57 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))
$R_{free}$	100719	3689 (2.60-2.52)
Clashscore	112137	4096 (2.60-2.52)
Ramachandran outliers	110173	4037 (2.60-2.52)
Sidechain outliers	110143	4037 (2.60-2.52)
RSRZ outliers	101464	3700 (2.60-2.52)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	C	284	<div> <div>5%</div> <div>56% 33% 10%</div> </div>
2	A	220	<div> <div>5%</div> <div>47% 33% 15%</div> </div>
3	B	259	<div> <div>8%</div> <div>58% 33% 7%</div> </div>
4	D	100	<div> <div>8%</div> <div>52% 43%</div> </div>

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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	NAG	A	302	-	-	X	-
5	NAG	C	301	-	-	X	-
8	GOL	B	303	-	-	X	-

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 6389 atoms, of which 8 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 Antigen-presenting glycoprotein CD1d.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	257	Total	C	N	O	S	0	0	0
			2056	1323	351	375	7			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	ALA	-	EXPRESSION TAG	UNP P15813
C	1	ASP	-	EXPRESSION TAG	UNP P15813
C	2	PRO	-	EXPRESSION TAG	UNP P15813
C	278	HIS	-	EXPRESSION TAG	UNP P15813
C	279	HIS	-	EXPRESSION TAG	UNP P15813
C	280	HIS	-	EXPRESSION TAG	UNP P15813
C	281	HIS	-	EXPRESSION TAG	UNP P15813
C	282	HIS	-	EXPRESSION TAG	UNP P15813
C	283	HIS	-	EXPRESSION TAG	UNP P15813

- Molecule 2 is a protein called HUMAN NKT TCR ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	186	Total	C	N	O	S	0	0	0
			1369	858	231	273	7			

- Molecule 3 is a protein called HUMAN NKT TCR BETA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	240	Total	C	N	O	S	0	0	0
			1892	1193	326	365	8			

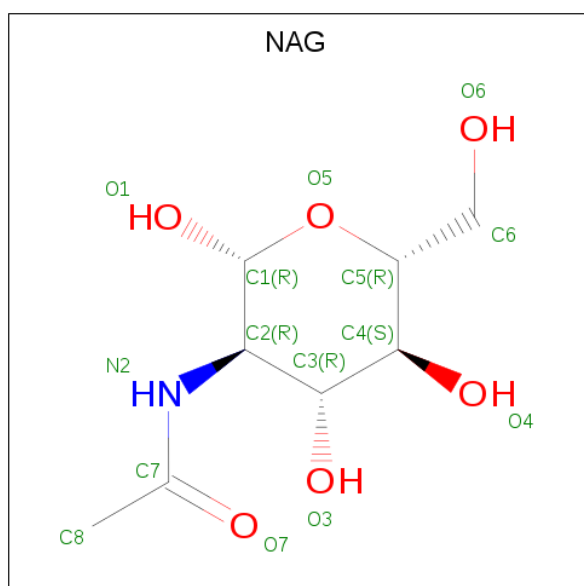
- Molecule 4 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	99	Total	C	N	O	S	0	0	0
			811	517	135	157	2			

There is a discrepancy between the modelled and reference sequences:

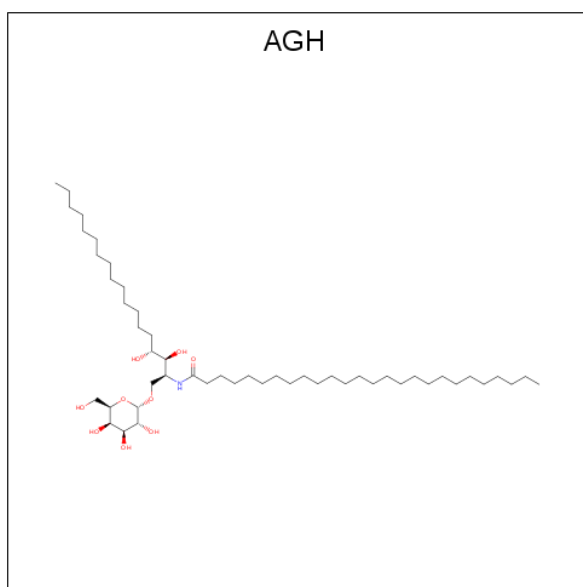
Chain	Residue	Modelled	Actual	Comment	Reference
D	0	PRO	-	EXPRESSION TAG	UNP P61769

- Molecule 5 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



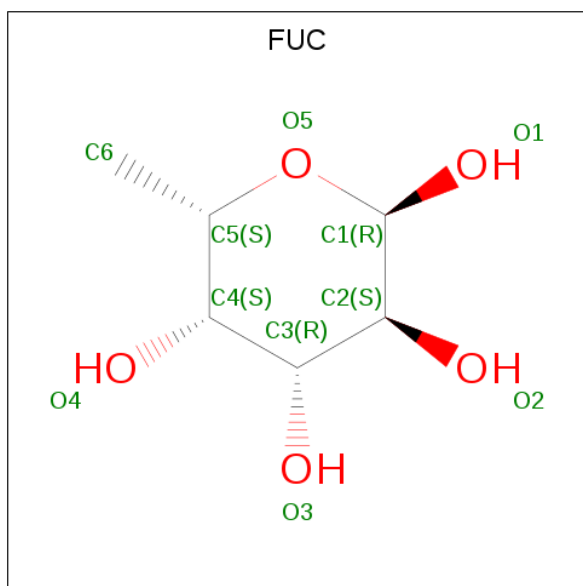
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	C	1	Total	C	N	O	0	0
			14	8	1	5		
5	C	1	Total	C	N	O	0	0
			14	8	1	5		
5	C	1	Total	C	N	O	0	0
			14	8	1	5		
5	A	1	Total	C	N	O	0	0
			14	8	1	5		
5	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 6 is N-{(1S,2R,3S)-1-[(ALPHA-D-GALACTOPYRANOSYLOXY)METHYL]-2,3-DIHYDROXYHEPTADECYL}HEXACOSANAMIDE (three-letter code: AGH) (formula:  $C_{50}H_{99}NO_9$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	C	1	Total	C	N	O	0	0
			60	50	1	9		

- Molecule 7 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula:  $C_6H_{12}O_5$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			10	6	4		

- Molecule 8 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	B	1	Total	C	O	0	0
			6	3	3		
8	B	1	Total	C	O	0	0
			6	3	3		

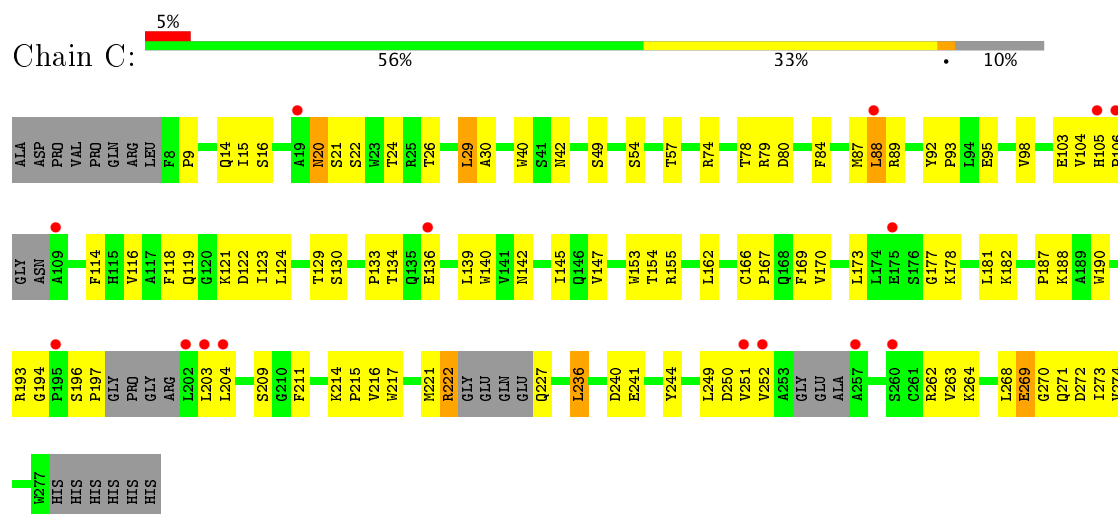
- Molecule 9 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
9	C	33	Total	H	O	0	0
			35	2	33		
9	A	20	Total	H	O	0	0
			22	2	20		
9	B	26	Total	H	O	0	0
			30	4	26		
9	D	22	Total	O		0	0
			22	22			

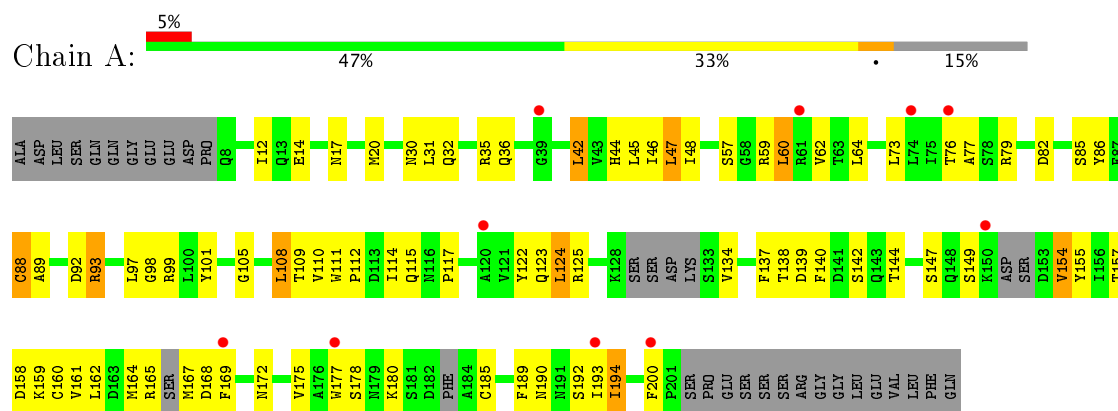
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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.

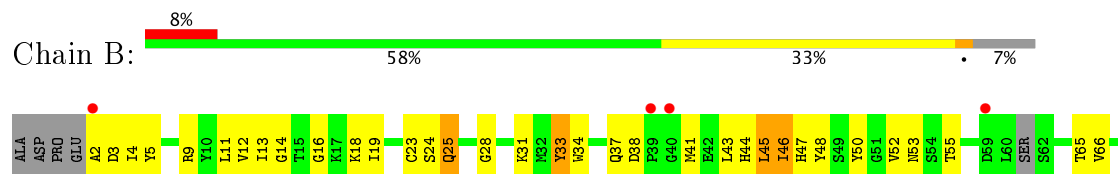
#### • Molecule 1: Antigen-presenting glycoprotein CD1d



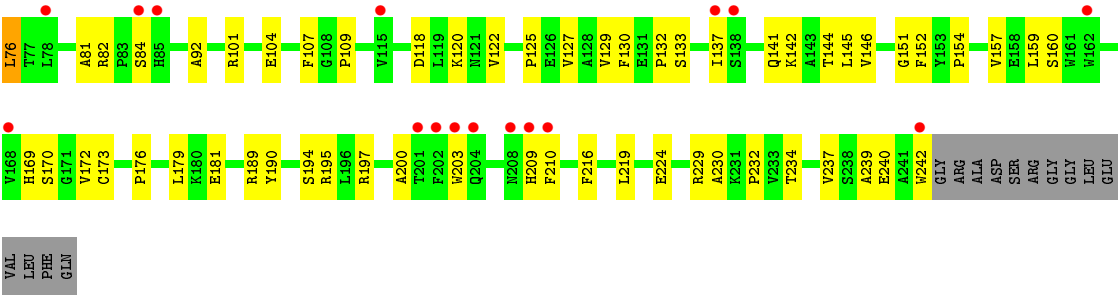
#### • Molecule 2: HUMAN NKT TCR ALPHA CHAIN



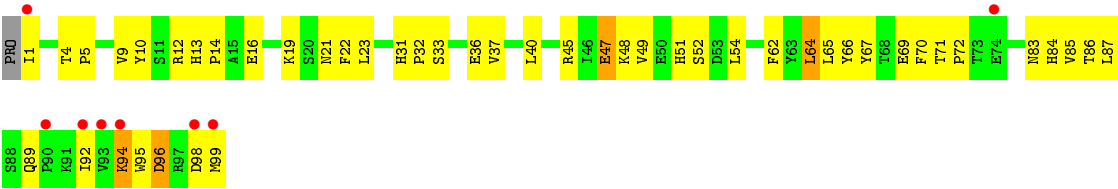
#### • Molecule 3: HUMAN NKT TCR BETA CHAIN







● Molecule 4: Beta-2-microglobulin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	57.11Å 72.57Å 113.75Å 90.00° 103.31° 90.00°	Depositor
Resolution (Å)	44.01 – 2.57 44.01 – 2.57	Depositor EDS
% Data completeness (in resolution range)	99.1 (44.01-2.57) 93.0 (44.01-2.57)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.84 (at 2.58Å)	Xtriage
Refinement program	PHENIX 1.8_1069	Depositor
R, $R_{free}$	0.208 , 0.266 0.206 , 0.265	Depositor DCC
$R_{free}$ test set	1378 reflections (5.09%)	DCC
Wilson B-factor (Å <sup>2</sup> )	51.1	Xtriage
Anisotropy	0.685	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 60.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.021 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6389	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.48% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, AGH, FUC, 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	C	0.21	0/2116	0.37	0/2882
2	A	0.21	0/1392	0.37	0/1896
3	B	0.21	0/1944	0.39	0/2651
4	D	0.21	0/834	0.37	0/1134
All	All	0.21	0/6286	0.38	0/8563

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 [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	C	2056	0	1972	104	0
2	A	1369	0	1218	100	0
3	B	1892	0	1759	104	0
4	D	811	0	754	41	0
5	A	14	0	13	9	0
5	B	14	0	13	3	0
5	C	42	0	39	12	0
6	C	60	0	99	8	0
7	A	10	0	10	4	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	B	12	0	16	6	0
9	A	20	2	0	1	0
9	B	26	4	0	1	0
9	C	33	2	0	4	0
9	D	22	0	0	1	0
All	All	6381	8	5893	325	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 27.

The worst 5 of 325 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)
3:B:137:ILE:HG23	3:B:200:ALA:HB1	1.41	1.03
1:C:42:ASN:HD21	5:C:301:NAG:C1	1.77	0.97
3:B:41:MET:HE2	3:B:41:MET:HA	1.51	0.93
1:C:154:THR:HG23	6:C:304:AGH:HAB2	1.50	0.92
2:A:20:MET:HE1	2:A:108:LEU:HB2	1.48	0.91

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	C	247/284 (87%)	240 (97%)	6 (2%)	1 (0%)	38	59
2	A	176/220 (80%)	165 (94%)	10 (6%)	1 (1%)	28	49
3	B	236/259 (91%)	220 (93%)	16 (7%)	0	100	100
4	D	97/100 (97%)	91 (94%)	4 (4%)	2 (2%)	8	14
All	All	756/863 (88%)	716 (95%)	36 (5%)	4 (0%)	32	54

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	48	LYS
1	C	269	GLU
2	A	164	MET
4	D	47	GLU

### 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	C	224/249 (90%)	215 (96%)	9 (4%)	36	60
2	A	137/195 (70%)	128 (93%)	9 (7%)	19	35
3	B	203/227 (89%)	193 (95%)	10 (5%)	29	50
4	D	90/95 (95%)	84 (93%)	6 (7%)	19	34
All	All	654/766 (85%)	620 (95%)	34 (5%)	27	47

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	124	LEU
3	B	11	LEU
4	D	92	ILE
2	A	194	ILE
1	C	227	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	20	ASN
1	C	42	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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

9 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	$\# Z  > 2$
7	FUC	A	301	-	9,10,11	1.52	1 (11%)	13,14,16	1.20	1 (7%)
5	NAG	A	302	-	14,14,15	0.58	0	15,19,21	1.80	3 (20%)
8	GOL	B	301	-	5,5,5	0.33	0	5,5,5	0.31	0
5	NAG	B	302	-	14,14,15	0.58	0	15,19,21	0.77	1 (6%)
8	GOL	B	303	-	5,5,5	0.34	0	5,5,5	0.31	0
5	NAG	C	301	-	14,14,15	0.57	0	15,19,21	1.37	1 (6%)
5	NAG	C	302	-	14,14,15	0.44	0	15,19,21	0.79	0
5	NAG	C	303	-	14,14,15	0.59	0	15,19,21	1.07	1 (6%)
6	AGH	C	304	-	60,60,60	0.33	0	64,69,69	0.80	1 (1%)

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
7	FUC	A	301	-	-	0/0/17/20	0/1/1/1
5	NAG	A	302	-	-	0/6/23/26	0/1/1/1
8	GOL	B	301	-	-	0/4/4/4	0/0/0/0
5	NAG	B	302	-	-	0/6/23/26	0/1/1/1
8	GOL	B	303	-	-	0/4/4/4	0/0/0/0
5	NAG	C	301	-	-	0/6/23/26	0/1/1/1
5	NAG	C	302	-	-	0/6/23/26	0/1/1/1
5	NAG	C	303	-	-	0/6/23/26	0/1/1/1
6	AGH	C	304	-	-	0/58/78/78	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	301	FUC	O5-C1	-3.56	1.37	1.43

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	301	NAG	O5-C1-C2	-4.06	105.83	111.47
5	A	302	NAG	O5-C1-C2	-3.93	106.01	111.47
5	C	303	NAG	O5-C1-C2	-3.51	106.59	111.47
6	C	304	AGH	C1-C2-N2	-3.18	105.36	109.76
5	B	302	NAG	O5-C1-C2	-2.05	108.62	111.47

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

9 monomers are involved in 38 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	A	301	FUC	4	0
5	A	302	NAG	9	0
8	B	301	GOL	2	0
5	B	302	NAG	3	0
8	B	303	GOL	4	0
5	C	301	NAG	7	0
5	C	302	NAG	3	0
5	C	303	NAG	5	0
6	C	304	AGH	8	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	C	257/284 (90%)	0.48	15 (5%) 24 21	35, 55, 89, 97	0
2	A	186/220 (84%)	0.52	10 (5%) 26 23	40, 67, 108, 124	0
3	B	240/259 (92%)	0.62	20 (8%) 12 10	35, 61, 106, 118	0
4	D	99/100 (99%)	0.54	8 (8%) 13 11	33, 59, 92, 113	0
All	All	782/863 (90%)	0.54	53 (6%) 18 16	33, 59, 102, 124	0

The worst 5 of 53 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	1	ILE	8.6
1	C	195	PRO	4.7
1	C	252	VAL	4.7
3	B	242	TRP	4.5
1	C	202	LEU	4.4

### 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 carbohydrates 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. 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<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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
6	AGH	C	304	60/60	0.94	0.23	1.74	32,48,61,63	0
8	GOL	B	303	6/6	0.93	0.28	1.41	41,57,59,61	0
8	GOL	B	301	6/6	0.92	0.22	1.13	55,60,62,62	0
5	NAG	A	302	14/15	0.70	0.26	0.34	98,105,112,114	0
7	FUC	A	301	10/11	0.87	0.21	-0.45	112,116,121,124	0
5	NAG	C	301	14/15	0.90	0.19	-	54,70,82,83	0
5	NAG	C	303	14/15	0.71	0.23	-	75,92,95,96	0
5	NAG	C	302	14/15	0.79	0.19	-	102,107,111,112	0
5	NAG	B	302	14/15	0.84	0.23	-	104,105,108,109	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.