



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:25 AM GMT

PDB ID : 3ARB  
Title : Ternary crystal structure of the NKT TCR-CD1d-alpha-galactosylceramide analogue-OCH  
Authors : Wun, K.S.; Rossjohn, J.  
Deposited on : 2010-11-26  
Resolution : 2.70 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 (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



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
5	NAG	A	501	-	-	-	X
6	PEG	B	100	-	-	-	X
9	D12	A	611	-	-	-	X

## 2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 6706 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 Antigen-presenting glycoprotein CD1d1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	295	Total	C	N	O	S	0	0	0
			2333	1491	411	417	14			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	201	HIS	ASP	SEE REMARK 999	UNP P11609
A	280	GLY	-	EXPRESSION TAG	UNP P11609
A	281	SER	-	EXPRESSION TAG	UNP P11609
A	282	LEU	-	EXPRESSION TAG	UNP P11609
A	283	HIS	-	EXPRESSION TAG	UNP P11609
A	284	HIS	-	EXPRESSION TAG	UNP P11609
A	285	ILE	-	EXPRESSION TAG	UNP P11609
A	286	LEU	-	EXPRESSION TAG	UNP P11609
A	287	ASP	-	EXPRESSION TAG	UNP P11609
A	288	ALA	-	EXPRESSION TAG	UNP P11609
A	289	GLN	-	EXPRESSION TAG	UNP P11609
A	290	LYS	-	EXPRESSION TAG	UNP P11609
A	291	MET	-	EXPRESSION TAG	UNP P11609
A	292	VAL	-	EXPRESSION TAG	UNP P11609
A	293	TRP	-	EXPRESSION TAG	UNP P11609
A	294	ASN	-	EXPRESSION TAG	UNP P11609
A	295	HIS	-	EXPRESSION TAG	UNP P11609
A	296	ARG	-	EXPRESSION TAG	UNP P11609
A	297	HIS	-	EXPRESSION TAG	UNP P11609
A	298	HIS	-	EXPRESSION TAG	UNP P11609
A	299	HIS	-	EXPRESSION TAG	UNP P11609
A	300	HIS	-	EXPRESSION TAG	UNP P11609
A	301	HIS	-	EXPRESSION TAG	UNP P11609
A	302	HIS	-	EXPRESSION TAG	UNP P11609

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	98	Total	C	N	O	S	0	0	0
			777	497	131	142	7			

- Molecule 3 is a protein called NKT Valpha14-Jalpha18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	203	Total	C	N	O	S	0	1	0
			1550	964	260	319	7			

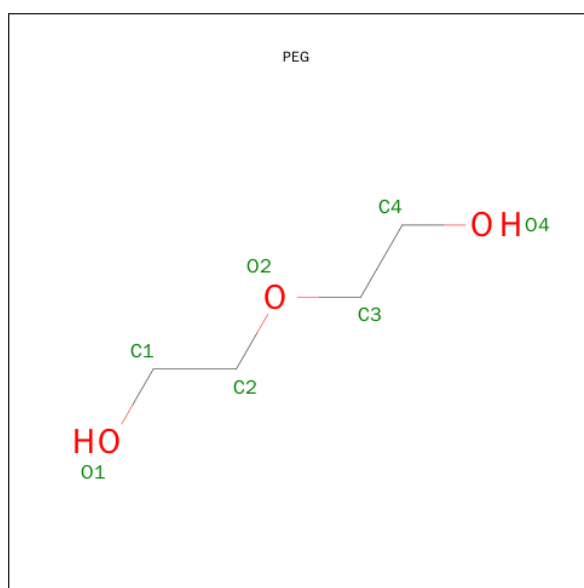
- Molecule 4 is a protein called NKT Vbeta8.2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	236	Total	C	N	O	S	0	0	0
			1815	1147	319	343	6			

- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

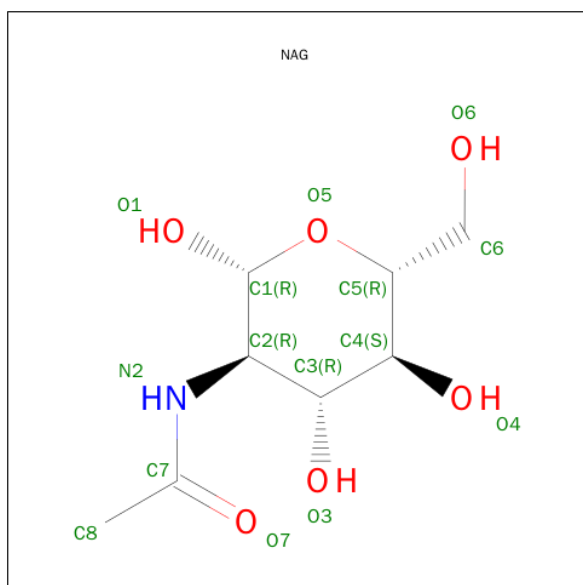
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	2	Total	C	N	O	0	0
			25	14	2	9		
5	A	2	Total	C	N	O	0	0
			24	14	2	8		
5	A	2	Total	C	N	O	0	0
			25	14	2	9		

- Molecule 6 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



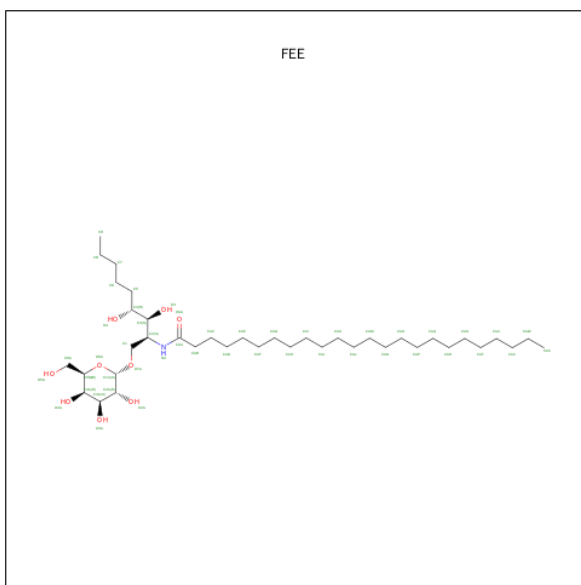
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			7	4	3		
6	B	1	Total	C	O	0	0
			7	4	3		

- Molecule 7 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



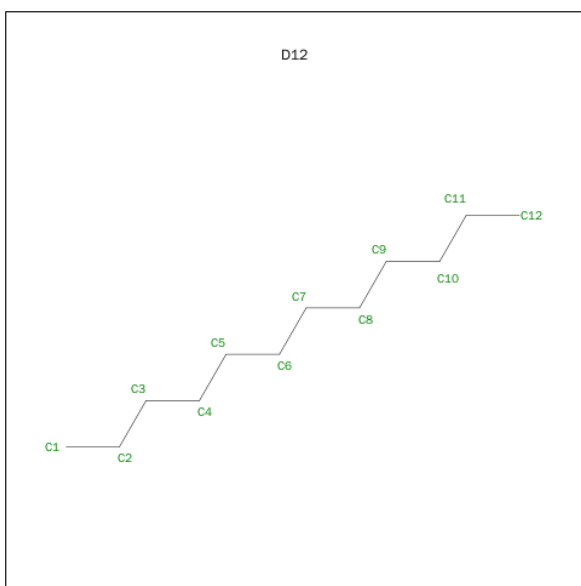
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 8 is N-{(1S,2S,3R)-1-[(ALPHA-D-GALACTOPYRANOSYLOXY)METHYL]-2,3-DIHYDROXYOCTYL}TETRACOSANAMIDE (three-letter code: FEE) (formula:  $C_{39}H_{77}NO_9$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	A	1	Total	C	N	O	0	0
			49	39	1	9		

- Molecule 9 is DODECANE (three-letter code: D12) (formula:  $C_{12}H_{26}$ ).



Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	1	Total	C	0	0
			12	12		

- Molecule 10 is water.

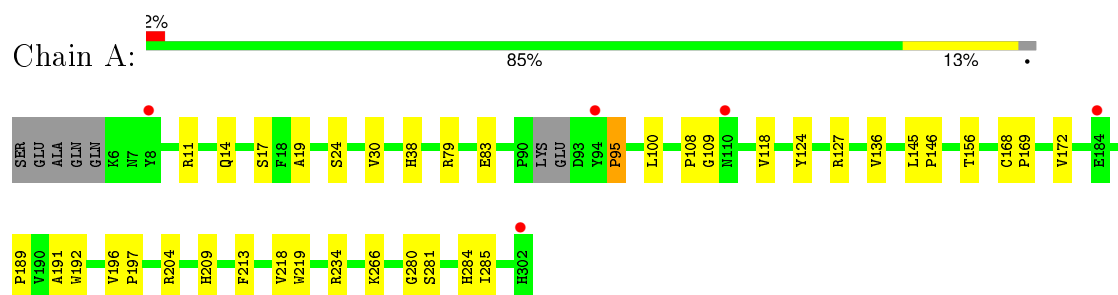
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	20	Total 20	O 20	0	0
10	B	10	Total 10	O 10	0	0
10	C	24	Total 24	O 24	0	0
10	D	14	Total 14	O 14	0	0



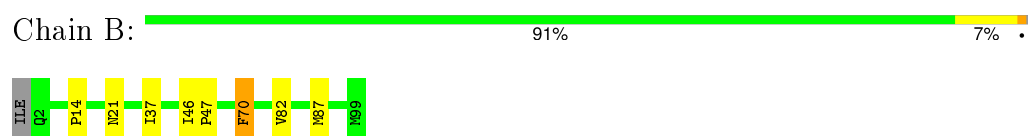
### 3 Residue-property plots [i](#)

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.

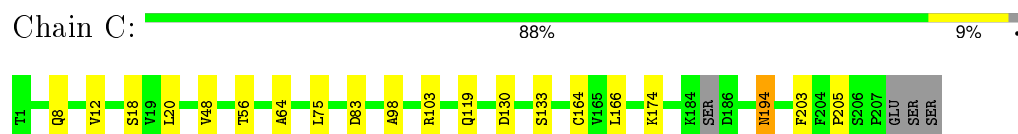
- Molecule 1: Antigen-presenting glycoprotein CD1d1



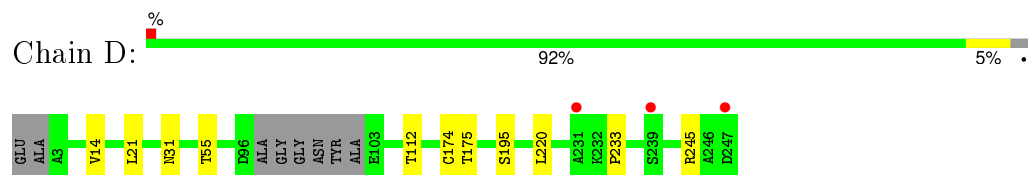
- Molecule 2: Beta-2-microglobulin



- Molecule 3: NKT Valpha14-Jalpha18



- Molecule 4: NKT Vbeta8.2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	59.04Å 87.31Å 235.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.84 – 2.70 48.84 – 2.70	Depositor EDS
% Data completeness (in resolution range)	95.0 (48.84-2.70) 95.2 (48.84-2.70)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.01 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.214 , 0.257 0.210 , 0.250	Depositor DCC
$R_{free}$ test set	1635 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.3	Xtriage
Anisotropy	0.653	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 26.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 32741 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6706	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: FEE, PEG, NAG, D12

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	0.31	0/2406	0.47	0/3277
2	B	0.32	0/803	0.47	0/1098
3	C	0.32	0/1581	0.48	0/2153
4	D	0.31	0/1865	0.46	0/2547
All	All	0.32	0/6655	0.47	0/9075

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	A	2333	0	2168	21	0
2	B	777	0	712	4	0
3	C	1550	0	1453	8	0
4	D	1815	0	1676	5	0
5	A	74	0	61	0	0
6	A	7	0	10	0	0
6	B	7	0	10	0	0
7	A	14	0	13	0	0
8	A	49	0	77	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	A	12	0	26	0	0
10	A	20	0	0	0	0
10	B	10	0	0	0	0
10	C	24	0	0	0	0
10	D	14	0	0	0	0
All	All	6706	0	6206	35	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 3.

The worst 5 of 35 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:280:GLY:H	1:A:284:HIS:HD2	1.38	0.72
1:A:189:PRO:HB3	1:A:213:PHE:HB3	1.73	0.68
1:A:218:VAL:O	1:A:234:ARG:NH2	2.33	0.61
1:A:196:VAL:HG22	1:A:197:PRO:HD2	1.87	0.56
1:A:280:GLY:H	1:A:284:HIS:CD2	2.24	0.53

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	291/302 (96%)	286 (98%)	5 (2%)	0	100	100
2	B	96/99 (97%)	95 (99%)	1 (1%)	0	100	100
3	C	200/207 (97%)	197 (98%)	3 (2%)	0	100	100
4	D	232/244 (95%)	225 (97%)	7 (3%)	0	100	100
All	All	819/852 (96%)	803 (98%)	16 (2%)	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	239/264 (90%)	235 (98%)	4 (2%)	68	90
2	B	82/93 (88%)	80 (98%)	2 (2%)	57	85
3	C	174/186 (94%)	166 (95%)	8 (5%)	33	64
4	D	186/206 (90%)	183 (98%)	3 (2%)	70	91
All	All	681/749 (91%)	664 (98%)	17 (2%)	55	84

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	48	VAL
3	C	83	ASP
3	C	194	ASN
3	C	8	GLN
4	D	14	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	273	GLN
1	A	284	HIS
4	D	24	ASN
1	A	186	GLN
3	C	194	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 ⓘ

6 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	$\# Z  > 2$
5	NAG	A	314	1,5	14,14,15	0.46	0	15,19,21	2.18	4 (26%)
5	NAG	A	315	5	10,10,15	0.67	0	10,13,21	2.01	2 (20%)
5	NAG	A	400	1,5	14,14,15	0.53	0	15,19,21	0.78	0
5	NAG	A	401	5	11,11,15	0.62	0	13,15,21	0.84	0
5	NAG	A	501	1,5	14,14,15	0.47	0	15,19,21	1.71	4 (26%)
5	NAG	A	502	5	11,11,15	0.60	0	13,15,21	1.55	2 (15%)

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
5	NAG	A	314	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	315	5	-	0/2/15/26	0/1/1/1
5	NAG	A	400	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	401	5	-	0/2/19/26	0/1/1/1
5	NAG	A	501	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	502	5	-	0/2/19/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	501	NAG	C4-C3-C2	-3.76	105.39	111.23

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	314	NAG	C4-C3-C2	-3.36	106.00	111.23
5	A	501	NAG	O4-C4-C3	2.27	115.45	110.34
5	A	314	NAG	O4-C4-C5	2.77	116.59	109.24
5	A	314	NAG	O4-C4-C3	2.81	116.66	110.34

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.6 Ligand geometry

5 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	$\# Z  > 2$
6	PEG	A	303	-	6,6,6	0.47	0	5,5,5	0.23	0
7	NAG	A	551	1	14,14,15	0.48	0	15,19,21	0.93	1 (6%)
8	FEE	A	601	-	48,49,49	0.39	0	54,58,58	0.70	1 (1%)
9	D12	A	611	-	11,11,11	0.27	0	10,10,10	0.49	0
6	PEG	B	100	-	6,6,6	0.46	0	5,5,5	0.29	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
6	PEG	A	303	-	-	0/4/4/4	0/0/0/0
7	NAG	A	551	1	-	0/6/23/26	0/1/1/1
8	FEE	A	601	-	-	0/47/67/67	0/1/1/1
9	D12	A	611	-	-	0/9/9/9	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	PEG	B	100	-	-	0/4/4/4	0/0/0/0

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	A	601	FEE	C1-C2-N2	-3.18	105.39	109.60
7	A	551	NAG	C1-O5-C5	2.14	114.96	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	A	601	FEE	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	295/302 (97%)	0.02	5 (1%) 73 74	38, 56, 76, 102	13 (4%)
2	B	98/99 (98%)	-0.17	0 100 100	43, 54, 68, 73	1 (1%)
3	C	203/207 (98%)	-0.12	0 100 100	36, 55, 72, 83	13 (6%)
4	D	236/244 (96%)	-0.04	3 (1%) 79 79	39, 60, 87, 106	6 (2%)
All	All	832/852 (97%)	-0.05	8 (0%) 84 85	36, 56, 79, 106	33 (3%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	302	HIS	4.3
1	A	94	TYR	2.8
1	A	8	TYR	2.8
4	D	247	ASP	2.6
1	A	110	ASN	2.2

### 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](#)

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
5	NAG	A	501	14/15	0.92	0.18	3.16	71,75,78,82	0
5	NAG	A	314	14/15	0.90	0.13	-1.17	64,68,73,73	0
5	NAG	A	400	14/15	0.87	0.31	-	84,91,97,98	0
5	NAG	A	502	11/15	0.85	0.20	-	80,82,86,90	0
5	NAG	A	401	11/15	0.75	0.35	-	89,95,103,108	0
5	NAG	A	315	10/15	0.89	0.17	-	68,74,80,81	0

## 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	PEG	B	100	7/7	0.85	0.26	4.16	72,72,72,72	0
9	D12	A	611	12/12	0.90	0.28	2.69	50,55,64,66	0
8	FEE	A	601	49/49	0.96	0.21	1.49	45,48,52,52	0
6	PEG	A	303	7/7	0.86	0.12	-3.02	68,68,69,69	0
7	NAG	A	551	14/15	0.71	0.37	-	96,104,110,113	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.