



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 9, 2020 – 01:31 AM BST

PDB ID : 2CH5
Title : Crystal structure of human N-acetylglucosamine kinase in complex with N-acetylglucosamine
Authors : Weihofen, W.A.; Berger, M.; Chen, H.; Saenger, W.; Hinderlich, S.
Deposited on : 2006-03-13
Resolution : 1.90 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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 : 2.13.1
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) : 2.13.1

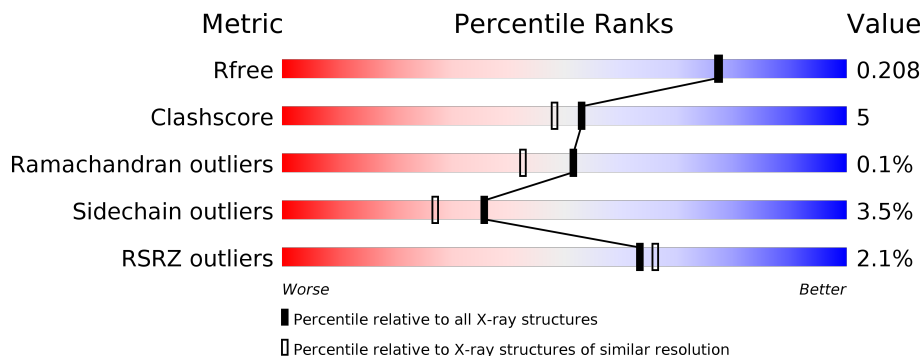
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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 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	347	
1	B	347	
1	C	347	
1	D	347	

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
2	NAG	A	500	X	-	-	-
2	NAG	B	500	X	-	-	-
2	NAG	C	500	X	-	-	-

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11843 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.

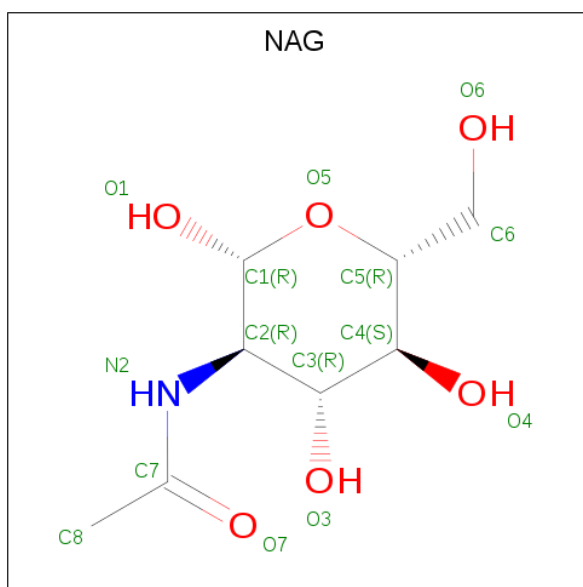
- Molecule 1 is a protein called NAGK PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	344	2632	1673	457	488	14	0	0	0
1	B	342	2615	1663	453	485	14	0	0	0
1	C	342	2619	1666	453	486	14	0	0	0
1	D	342	2616	1663	453	486	14	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

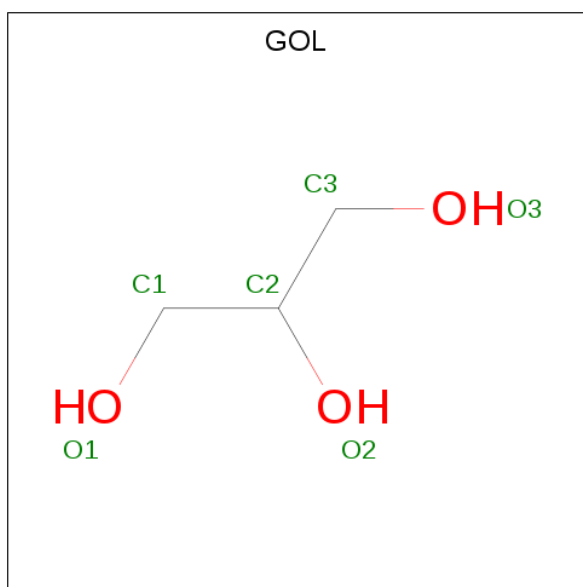
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	PRO	-	expression tag	UNP Q6IA84
A	-1	GLN	-	expression tag	UNP Q6IA84
A	0	PHE	-	expression tag	UNP Q6IA84
B	-2	PRO	-	expression tag	UNP Q6IA84
B	-1	GLN	-	expression tag	UNP Q6IA84
B	0	PHE	-	expression tag	UNP Q6IA84
C	-2	PRO	-	expression tag	UNP Q6IA84
C	-1	GLN	-	expression tag	UNP Q6IA84
C	0	PHE	-	expression tag	UNP Q6IA84
D	-2	PRO	-	expression tag	UNP Q6IA84
D	-1	GLN	-	expression tag	UNP Q6IA84
D	0	PHE	-	expression tag	UNP Q6IA84

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C₈H₁₅NO₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	15	8	1	6	0	0
2	B	1	15	8	1	6	0	0
2	C	1	15	8	1	6	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



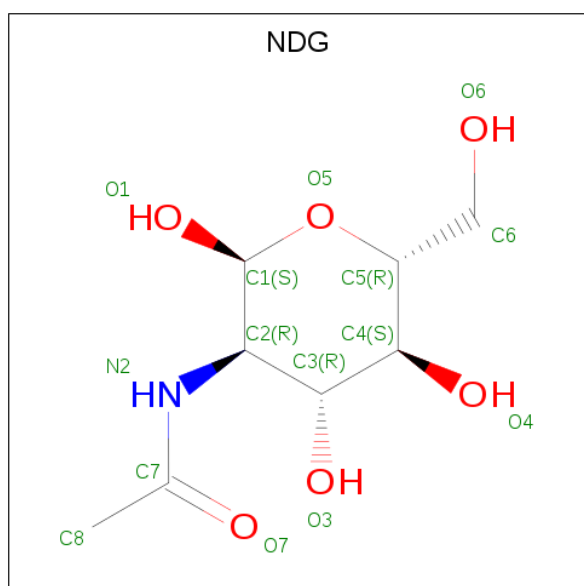
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	6	3	3	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 4 is 2-acetamido-2-deoxy-alpha-D-glucopyranose (three-letter code: NDG) (formula: C₈H₁₅NO₆).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	D	1	Total	C	N	O	0	0
			14	8	1	5		

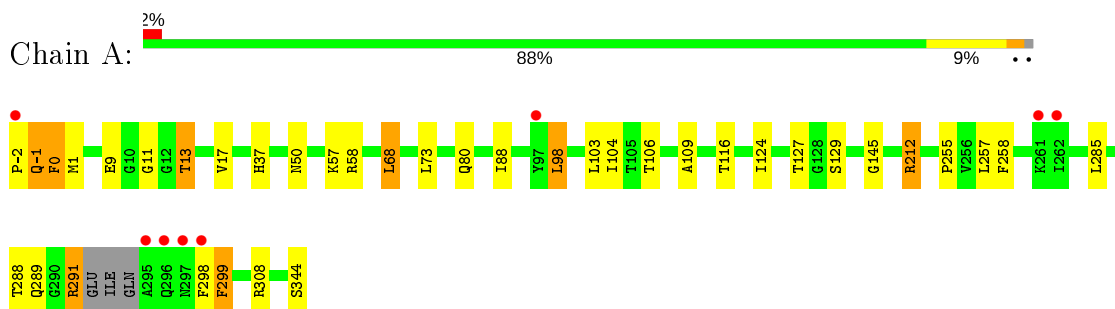
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	315	Total	O	0	0
			315	315		
5	B	306	Total	O	0	0
			306	306		
5	C	317	Total	O	0	0
			317	317		
5	D	334	Total	O	0	0
			334	334		

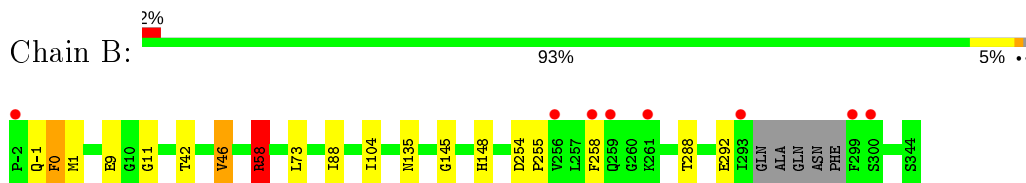
3 Residue-property plots [i](#)

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.

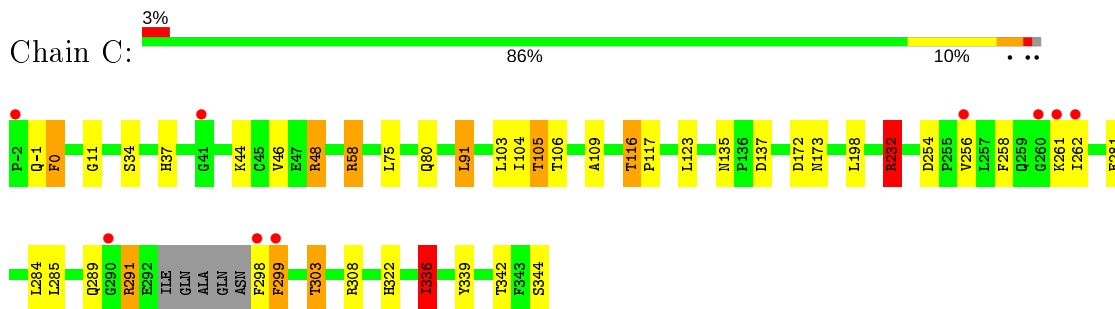
- Molecule 1: NAGK PROTEIN



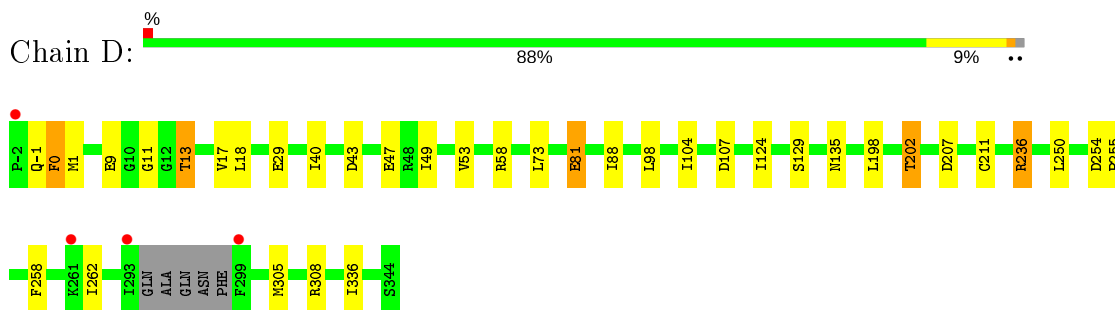
- Molecule 1: NAGK PROTEIN



- Molecule 1: NAGK PROTEIN



- Molecule 1: NAGK PROTEIN



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	57.54Å 98.48Å 101.98Å 63.43° 75.66° 75.06°	Depositor
Resolution (Å)	90.17 – 1.90 43.22 – 1.90	Depositor EDS
% Data completeness (in resolution range)	93.3 (90.17-1.90) 93.3 (43.22-1.90)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.43 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.167 , 0.199 0.177 , 0.208	Depositor DCC
R_{free} test set	2817 reflections (2.00%)	wwPDB-VP
Wilson B-factor (Å ²)	24.0	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 49.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.105 for -h,-l,-k	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	11843	wwPDB-VP
Average B, all atoms (Å ²)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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, NAG, NDG

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	0.76	1/2690 (0.0%)	0.73	4/3632 (0.1%)
1	B	0.79	1/2672 (0.0%)	0.72	4/3609 (0.1%)
1	C	0.74	1/2677 (0.0%)	0.78	6/3614 (0.2%)
1	D	0.77	1/2673 (0.0%)	0.71	5/3609 (0.1%)
All	All	0.77	4/10712 (0.0%)	0.73	19/14464 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	1	1
1	C	1	0
All	All	2	1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	0	PHE	C-N	21.80	1.84	1.34
1	A	0	PHE	C-N	18.49	1.76	1.34
1	D	0	PHE	C-N	16.04	1.71	1.34
1	C	0	PHE	C-N	14.12	1.66	1.34

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	48	ARG	NE-CZ-NH1	9.48	125.04	120.30
1	B	0	PHE	C-N-CA	-8.78	99.76	121.70
1	C	232	ARG	NE-CZ-NH1	7.82	124.21	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	17	VAL	CB-CA-C	-7.34	97.45	111.40
1	A	212	ARG	NE-CZ-NH1	7.28	123.94	120.30

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	106	THR	CB
1	C	106	THR	CB

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	13	THR	Peptide

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	2632	0	2600	33	0
1	B	2615	0	2589	20	0
1	C	2619	0	2585	37	0
1	D	2616	0	2588	28	0
2	A	15	0	15	1	0
2	B	15	0	15	1	0
2	C	15	0	15	0	0
3	A	6	0	8	2	0
3	B	6	0	8	2	0
3	C	6	0	8	0	0
3	D	12	0	16	0	0
4	D	14	0	9	0	0
5	A	315	0	0	9	0
5	B	306	0	0	2	0
5	C	317	0	0	10	0
5	D	334	0	0	2	0
All	All	11843	0	10456	114	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 114 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:D:0:PHE:C	1:D:1:MET:N	1.70	1.39
1:A:0:PHE:C	1:A:1:MET:N	1.76	1.37
1:B:0:PHE:C	1:B:1:MET:N	1.84	1.31
1:A:0:PHE:C	1:A:1:MET:H2	1.39	1.20
1:B:0:PHE:C	1:B:1:MET:H2	1.37	1.19

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	340/347 (98%)	327 (96%)	12 (4%)	1 (0%)	41	31
1	B	338/347 (97%)	330 (98%)	8 (2%)	0	100	100
1	C	338/347 (97%)	331 (98%)	7 (2%)	0	100	100
1	D	338/347 (97%)	330 (98%)	8 (2%)	0	100	100
All	All	1354/1388 (98%)	1318 (97%)	35 (3%)	1 (0%)	51	43

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	298	PHE

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	273/276 (99%)	264 (97%)	9 (3%)	38	29
1	B	272/276 (99%)	269 (99%)	3 (1%)	73	73
1	C	272/276 (99%)	254 (93%)	18 (7%)	16	8
1	D	272/276 (99%)	264 (97%)	8 (3%)	42	35
All	All	1089/1104 (99%)	1051 (96%)	38 (4%)	36	27

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	105	THR
1	C	232	ARG
1	D	198	LEU
1	C	123	LEU
1	C	261	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	259	GLN
1	B	309	HIS
1	C	225	GLN
1	B	193	GLN
1	C	289	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 monosaccharides 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
3	GOL	D	501	-	5,5,5	0.45	0	5,5,5	0.56	0
2	NAG	B	500	-	15,15,15	1.10	1 (6%)	21,21,21	1.26	2 (9%)
4	NDG	D	500	-	14,14,15	1.37	3 (21%)	18,19,21	1.62	6 (33%)
3	GOL	C	501	-	5,5,5	0.34	0	5,5,5	0.42	0
3	GOL	A	501	-	5,5,5	0.39	0	5,5,5	0.47	0
3	GOL	B	501	-	5,5,5	0.47	0	5,5,5	0.42	0
2	NAG	A	500	-	15,15,15	0.92	1 (6%)	21,21,21	1.01	0
3	GOL	D	502	-	5,5,5	0.31	0	5,5,5	0.30	0
2	NAG	C	500	-	15,15,15	0.73	0	21,21,21	1.35	2 (9%)

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
3	GOL	D	501	-	-	2/4/4/4	-
2	NAG	B	500	-	1/1/6/7	2/6/26/26	0/1/1/1
2	NAG	A	500	-	1/1/6/7	0/6/26/26	0/1/1/1
3	GOL	C	501	-	-	4/4/4/4	-
3	GOL	A	501	-	-	4/4/4/4	-
3	GOL	B	501	-	-	4/4/4/4	-
4	NDG	D	500	-	-	1/5/25/26	0/1/1/1
3	GOL	D	502	-	-	3/4/4/4	-
2	NAG	C	500	-	1/1/6/7	2/6/26/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	500	NDG	O1-C1	3.21	1.49	1.39
2	B	500	NAG	O1-C1	2.79	1.48	1.39
2	A	500	NAG	O1-C1	2.36	1.47	1.39
4	D	500	NDG	O5-C1	2.15	1.48	1.42
4	D	500	NDG	C1-C2	-2.15	1.50	1.52

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	D	500	NDG	O5-C1-C2	-2.95	106.55	109.52
2	C	500	NAG	C6-C5-C4	-2.77	106.51	113.00
4	D	500	NDG	O5-C5-C6	2.69	113.12	106.44
4	D	500	NDG	C6-C5-C4	-2.66	106.77	113.00
2	C	500	NAG	O5-C5-C6	2.59	112.88	106.44

All (3) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	500	NAG	C1
2	A	500	NAG	C1
2	C	500	NAG	C1

5 of 22 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	501	GOL	C1-C2-C3-O3
3	B	501	GOL	C1-C2-C3-O3
3	D	502	GOL	C1-C2-C3-O3
3	B	501	GOL	O2-C2-C3-O3
3	D	501	GOL	O1-C1-C2-C3

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	500	NAG	1	0
3	A	501	GOL	2	0
3	B	501	GOL	2	0
2	A	500	NAG	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1
1	A	1
1	D	1
1	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	0:PHE	C	1:MET	N	1.84
1	A	0:PHE	C	1:MET	N	1.76
1	D	0:PHE	C	1:MET	N	1.70
1	C	0:PHE	C	1:MET	N	1.66

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	344/347 (99%)	0.11	8 (2%) 60 63	20, 26, 37, 64	0
1	B	342/347 (98%)	0.00	8 (2%) 60 63	21, 26, 36, 51	0
1	C	342/347 (98%)	0.18	9 (2%) 56 58	19, 26, 38, 59	0
1	D	342/347 (98%)	0.08	4 (1%) 79 81	19, 24, 34, 45	0
All	All	1370/1388 (98%)	0.09	29 (2%) 63 66	19, 26, 37, 64	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	298	PHE	7.8
1	A	-2	PRO	6.9
1	A	298	PHE	6.9
1	C	-2	PRO	5.7
1	C	261	LYS	5.1

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, 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(\AA^2)	Q<0.9
3	GOL	B	501	6/6	0.83	0.17	42,45,46,48	0
3	GOL	D	502	6/6	0.86	0.14	50,52,52,54	0
3	GOL	A	501	6/6	0.88	0.17	41,45,46,47	0
3	GOL	C	501	6/6	0.89	0.13	34,44,47,47	0
4	NDG	D	500	14/15	0.94	0.09	17,19,21,24	0
3	GOL	D	501	6/6	0.96	0.08	37,40,41,45	0
2	NAG	A	500	15/15	0.97	0.06	18,22,24,26	0
2	NAG	B	500	15/15	0.97	0.06	18,19,22,23	0
2	NAG	C	500	15/15	0.97	0.07	21,22,24,25	0

6.5 Other polymers [i](#)

There are no such residues in this entry.