



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 9, 2020 – 12:50 PM BST

PDB ID : 2PHL
Title : THE STRUCTURE OF PHASEOLIN AT 2.2 ANGSTROMS RESOLUTION:
IMPLICATIONS FOR A COMMON VICILIN(SLASH)LEGUMIN STRUC-
TURE AND THE GENETIC ENGINEERING OF SEED STORAGE PRO-
TEINS
Authors : Lawrence, M.C.; Izard, T.; Beuchat, M.; Blagrove, R.J.; Colman, P.M.
Deposited on : 1994-07-07
Resolution : 2.20 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

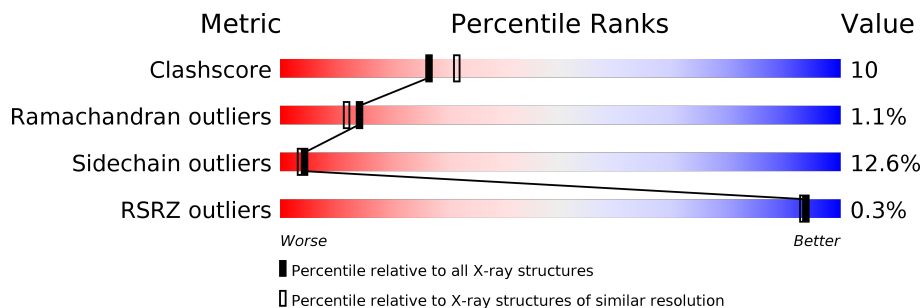
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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))
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	397	 66% 20% 5% 9%
1	B	397	 59% 27% 1 9%
1	C	397	 62% 24% 5% 9%

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	NAG	C	902	X	-	-	-

2 Entry composition [i](#)

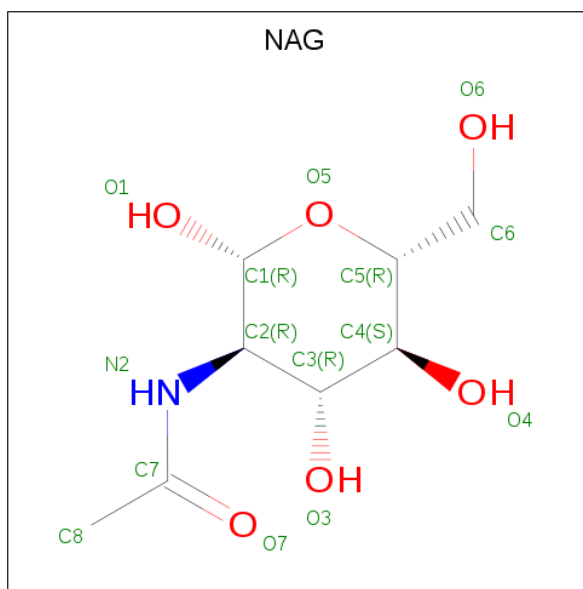
There are 4 unique types of molecules in this entry. The entry contains 8746 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 PHASEOLIN.

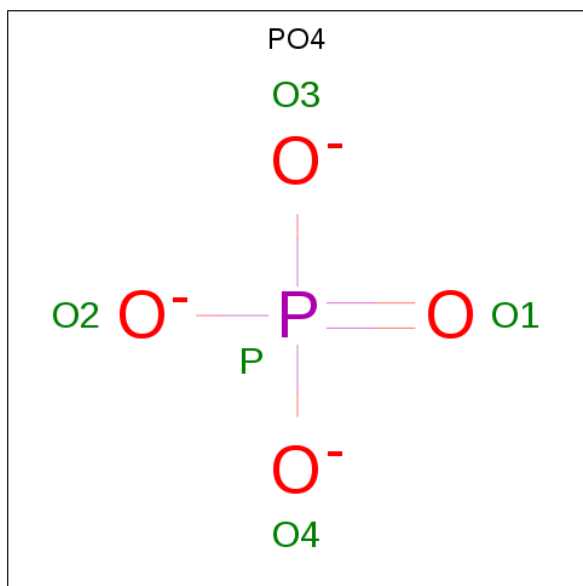
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	360	Total 2870	C 1819	N 484	O 564	S 3	0	0	0
1	B	360	Total 2870	C 1819	N 484	O 564	S 3	0	0	0
1	C	361	Total 2879	C 1824	N 486	O 566	S 3	0	0	0

- 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	C 8	N 1	O 5	0	0
2	B	1	Total 14	C 8	N 1	O 5	0	0
2	C	1	Total 14	C 8	N 1	O 5	0	0

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O P 5 4 1	0	0
3	B	1	Total O P 5 4 1	0	0
3	C	1	Total O P 5 4 1	0	0

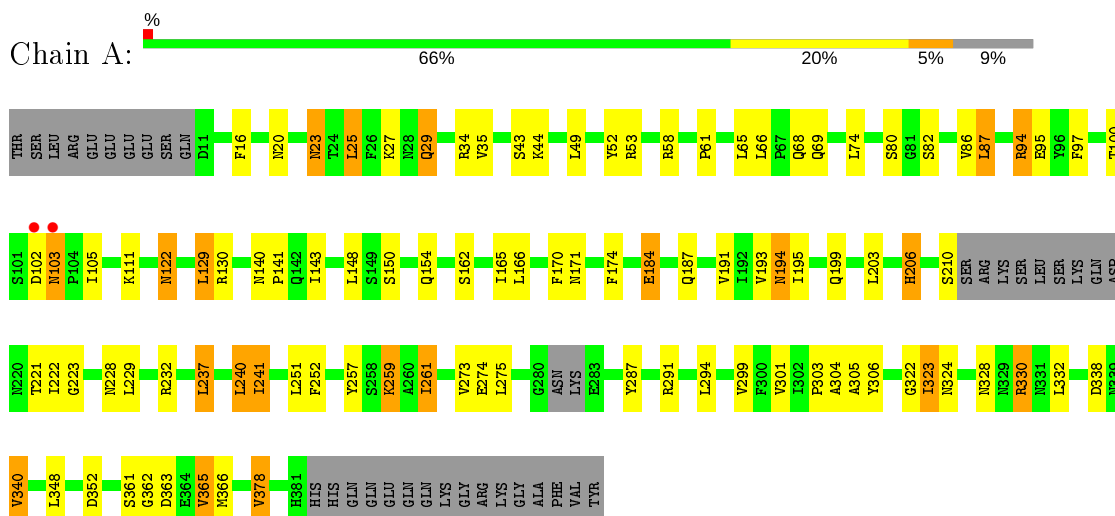
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	24	Total O 24 24	0	0
4	B	24	Total O 24 24	0	0
4	C	22	Total O 22 22	0	0

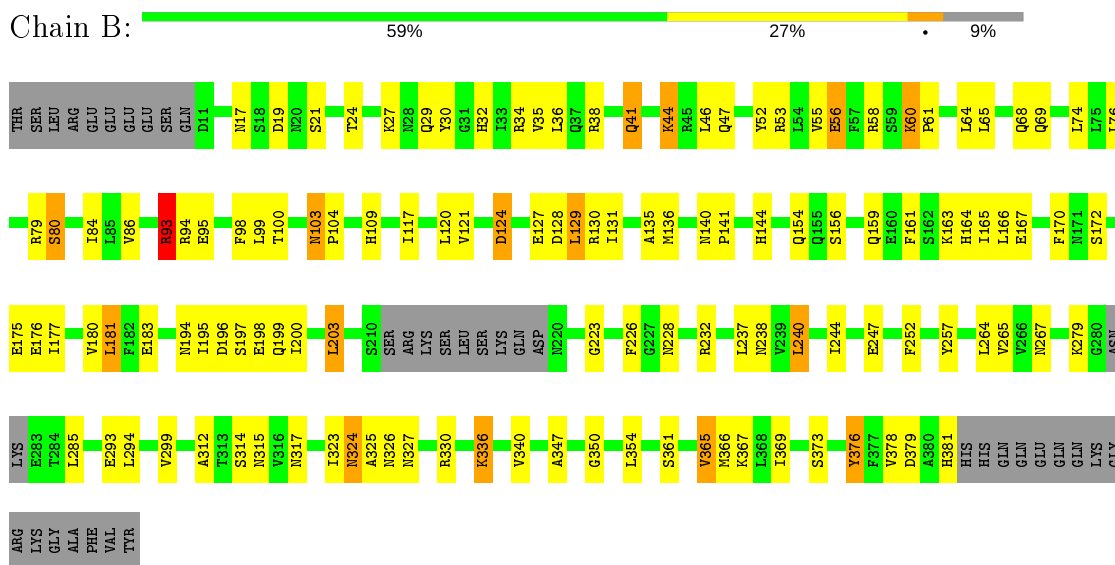
3 Residue-property plots [i](#)

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: PHASEOLIN

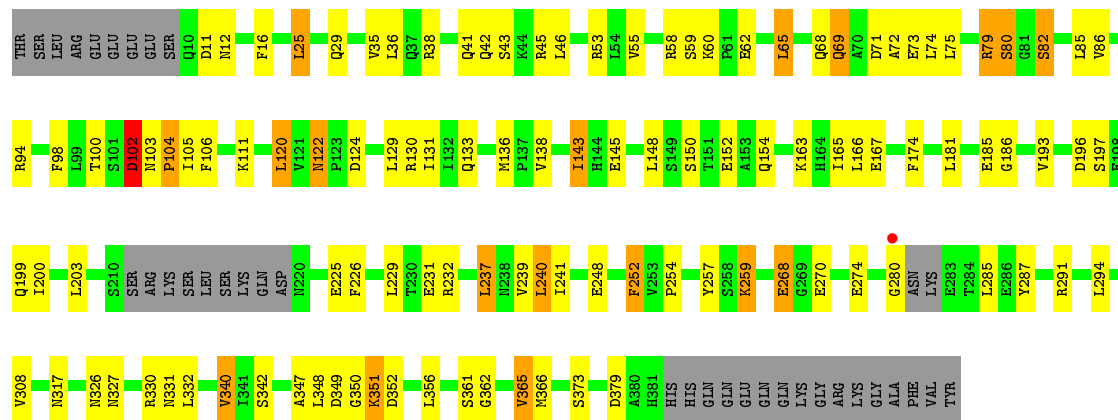


- Molecule 1: PHASEOLIN



- Molecule 1: PHASEOLIN





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	89.81Å 114.08Å 137.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	6.00 – 2.20 6.00 – 2.20	Depositor EDS
% Data completeness (in resolution range)	(Not available) (6.00-2.20) 51.7 (6.00-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$	-	Xtrriage
Refinement program	X-PLOR	Depositor
R, R_{free}	0.178 , (Not available) 0.173 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å ²)	30.7	Xtrriage
Anisotropy	0.339	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 109.9	EDS
L-test for twinning ¹	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	8746	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.42% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Theoretical 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: PO4, 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	0.54	0/2925	0.82	1/3958 (0.0%)
1	B	0.55	1/2925 (0.0%)	0.82	4/3958 (0.1%)
1	C	0.54	0/2934	0.84	5/3970 (0.1%)
All	All	0.54	1/8784 (0.0%)	0.83	10/11886 (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	0	1
1	C	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	93	ARG	CZ-NH1	5.56	1.40	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	C	120	LEU	CA-CB-CG	5.87	128.81	115.30
1	B	120	LEU	CA-CB-CG	5.79	128.63	115.30
1	B	240	LEU	CA-CB-CG	5.61	128.19	115.30
1	C	240	LEU	CA-CB-CG	5.56	128.09	115.30
1	B	350	GLY	N-CA-C	5.37	126.52	113.10

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	287	TYR	Sidechain
1	C	287	TYR	Sidechain

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	2870	0	2797	55	0
1	B	2870	0	2797	73	0
1	C	2879	0	2805	59	0
2	A	14	0	13	1	0
2	B	14	0	13	0	0
2	C	14	0	13	0	0
3	A	5	0	0	0	0
3	B	5	0	0	0	0
3	C	5	0	0	1	0
4	A	24	0	0	0	0
4	B	24	0	0	1	0
4	C	22	0	0	1	0
All	All	8746	0	8438	176	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 10.

The worst 5 of 176 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:340:VAL:HG11	1:B:117:ILE:HD12	1.63	0.81
1:B:17:ASN:HD21	1:B:19:ASP:HB2	1.50	0.76
1:A:229:LEU:HD23	1:A:330:ARG:HH21	1.51	0.75
1:B:140:ASN:HB2	1:B:141:PRO:HD2	1.70	0.73
1:C:43:SER:HB3	1:C:46:LEU:HD13	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	354/397 (89%)	326 (92%)	27 (8%)	1 (0%)	41	46
1	B	354/397 (89%)	323 (91%)	25 (7%)	6 (2%)	9	6
1	C	355/397 (89%)	325 (92%)	25 (7%)	5 (1%)	11	8
All	All	1063/1191 (89%)	974 (92%)	77 (7%)	12 (1%)	14	12

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	350	GLY
1	A	29	GLN
1	C	80	SER
1	C	104	PRO
1	B	80	SER

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	320/354 (90%)	279 (87%)	41 (13%)	4	3
1	B	320/354 (90%)	283 (88%)	37 (12%)	5	5
1	C	321/354 (91%)	278 (87%)	43 (13%)	4	3
All	All	961/1062 (90%)	840 (87%)	121 (13%)	4	3

5 of 121 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	93	ARG
1	B	232	ARG
1	C	259	LYS
1	B	99	LEU
1	B	154	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	187	GLN
1	B	315	ASN
1	C	329	ASN
1	B	238	ASN
1	B	326	ASN

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 [i](#)

6 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	B	901	1	14,14,15	1.07	1 (7%)	17,19,21	1.78	5 (29%)
2	NAG	A	900	1	14,14,15	1.74	2 (14%)	17,19,21	2.50	5 (29%)
3	PO4	A	950	-	4,4,4	1.30	0	6,6,6	1.05	0
3	PO4	B	951	-	4,4,4	1.97	2 (50%)	6,6,6	1.28	1 (16%)
2	NAG	C	902	1	14,14,15	1.46	3 (21%)	17,19,21	3.36	11 (64%)
3	PO4	C	952	-	4,4,4	2.09	2 (50%)	6,6,6	1.13	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	C	902	1	1/1/5/7	2/6/23/26	0/1/1/1
2	NAG	B	901	1	-	1/6/23/26	0/1/1/1
2	NAG	A	900	1	-	1/6/23/26	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	900	NAG	C1-C2	4.39	1.58	1.52
2	A	900	NAG	C4-C5	-4.08	1.44	1.53
2	C	902	NAG	C4-C5	3.14	1.59	1.53
3	C	952	PO4	P-O2	-2.86	1.46	1.54
3	B	951	PO4	P-O2	-2.76	1.46	1.54

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	902	NAG	C1-C2-N2	8.00	124.15	110.49
2	A	900	NAG	C1-C2-N2	6.45	121.51	110.49
2	C	902	NAG	O5-C1-C2	6.37	121.35	111.29
2	C	902	NAG	C6-C5-C4	6.02	127.11	113.00
2	A	900	NAG	C1-O5-C5	5.31	119.38	112.19

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	C	902	NAG	C1

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	900	NAG	C1-C2-N2-C7
2	B	901	NAG	C3-C2-N2-C7
2	C	902	NAG	C1-C2-N2-C7
2	C	902	NAG	C3-C2-N2-C7

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	900	NAG	1	0
3	C	952	PO4	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, 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	360/397 (90%)	-1.12	2 (0%) 89 88	8, 28, 54, 84	0
1	B	360/397 (90%)	-0.98	0 100 100	13, 34, 61, 89	0
1	C	361/397 (90%)	-1.08	1 (0%) 94 93	12, 29, 57, 80	0
All	All	1081/1191 (90%)	-1.06	3 (0%) 94 93	8, 30, 58, 89	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	103	ASN	2.7
1	A	102	ASP	2.5
1	C	280	GLY	2.4

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 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(Å ²)	Q<0.9
2	NAG	B	901	14/15	0.80	0.21	71,77,90,94	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	NAG	A	900	14/15	0.89	0.17	53,57,64,67	0
2	NAG	C	902	14/15	0.90	0.15	49,63,78,78	0
3	PO4	C	952	5/5	0.98	0.11	57,60,65,65	0
3	PO4	A	950	5/5	0.99	0.05	38,39,45,45	0
3	PO4	B	951	5/5	0.99	0.06	33,39,43,43	0

6.5 Other polymers [i](#)

There are no such residues in this entry.