WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 21, 2020 – 09:38 PM BST

PDB ID	:	4KNG
Title	:	Crystal structure of human LGR5-RSPO1-RNF43
Authors	:	Chen, P.H.; He, X.
Deposited on	:	2013-05-09
Resolution	:	2.50 Å(reported)

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

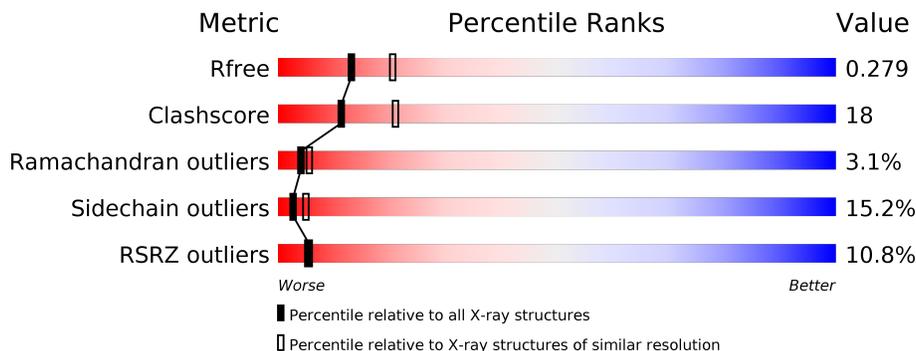|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	Percentile Ranks	Value
Rfree		0.279
Clashscore		18
Ramachandran outliers		3.1%
Sidechain outliers		15.2%
RSRZ outliers		10.8%

Worse *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 <=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	531	7% / 54% 26% 6% · 13%
1	B	531	5% / 55% 24% 7% · 13%
2	M	115	9% / 64% 17% · 15%
2	P	115	10% / 53% 23% 6% 18%
3	E	160	6% / 54% 31% 8% · 7%
3	F	160	38% / 54% 29% 7% · 8%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 11252 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 Leucine-rich repeat-containing G-protein coupled receptor 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	461	Total 3622	C 2307	N 626	O 673	S 16	0	1	0
1	B	463	Total 3629	C 2312	N 627	O 674	S 16	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	30	GLY	-	expression tag	UNP O75473
A	31	PRO	-	expression tag	UNP O75473
A	558	ALA	-	expression tag	UNP O75473
A	559	ALA	-	expression tag	UNP O75473
A	560	ALA	-	expression tag	UNP O75473
B	30	GLY	-	expression tag	UNP O75473
B	31	PRO	-	expression tag	UNP O75473
B	558	ALA	-	expression tag	UNP O75473
B	559	ALA	-	expression tag	UNP O75473
B	560	ALA	-	expression tag	UNP O75473

- Molecule 2 is a protein called R-spondin-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	M	98	Total 750	C 465	N 131	O 136	S 18	0	0	0
2	P	94	Total 720	C 448	N 127	O 128	S 17	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	33	GLY	-	expression tag	UNP Q2MKA7
M	34	PRO	-	expression tag	UNP Q2MKA7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	145	ALA	-	expression tag	UNP Q2MKA7
M	146	ALA	-	expression tag	UNP Q2MKA7
M	147	ALA	-	expression tag	UNP Q2MKA7
P	33	GLY	-	expression tag	UNP Q2MKA7
P	34	PRO	-	expression tag	UNP Q2MKA7
P	145	ALA	-	expression tag	UNP Q2MKA7
P	146	ALA	-	expression tag	UNP Q2MKA7
P	147	ALA	-	expression tag	UNP Q2MKA7

- Molecule 3 is a protein called E3 ubiquitin-protein ligase RNF43.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	149	Total 1152	C 734	N 200	O 212	S 6	0	0	0
3	F	148	Total 1145	C 729	N 199	O 211	S 6	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	42	GLY	-	expression tag	UNP Q68DV7
E	43	PRO	-	expression tag	UNP Q68DV7
E	199	ALA	-	expression tag	UNP Q68DV7
E	200	ALA	-	expression tag	UNP Q68DV7
E	201	ALA	-	expression tag	UNP Q68DV7
F	42	GLY	-	expression tag	UNP Q68DV7
F	43	PRO	-	expression tag	UNP Q68DV7
F	199	ALA	-	expression tag	UNP Q68DV7
F	200	ALA	-	expression tag	UNP Q68DV7
F	201	ALA	-	expression tag	UNP Q68DV7

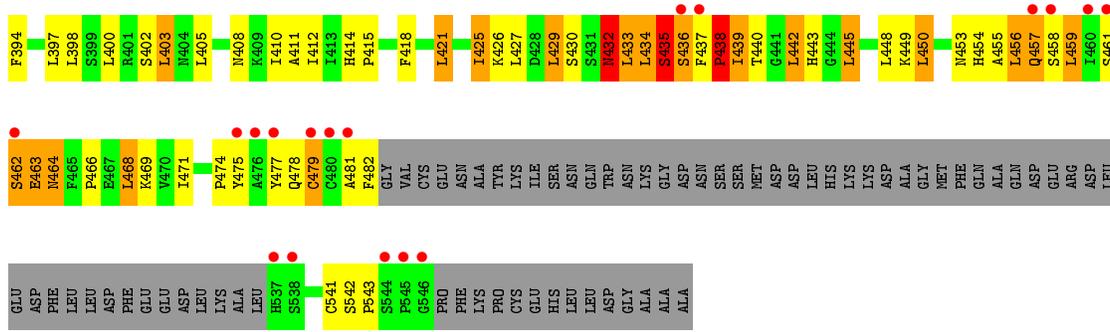- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).
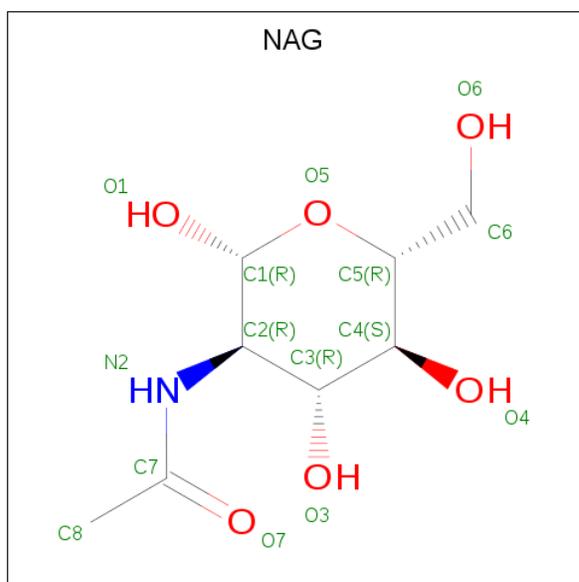
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Ni	0	0
			1	1		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	70	Total	O	0	0
			70	70		
6	B	85	Total	O	0	0
			85	85		
6	M	21	Total	O	0	0
			21	21		
6	P	7	Total	O	0	0
			7	7		
6	E	18	Total	O	0	0
			18	18		
6	F	4	Total	O	0	0
			4	4		

3 Residue-property plots ⓘ

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: Leucine-rich repeat-containing G-protein coupled receptor 5

- Molecule 1: Leucine-rich repeat-containing G-protein coupled receptor 5

• Molecule 2: R-spondin-1

Chain M:
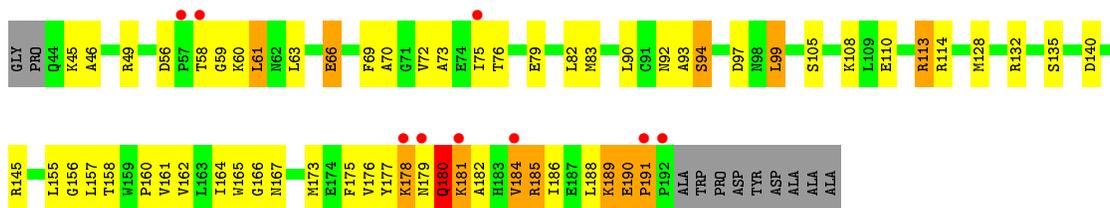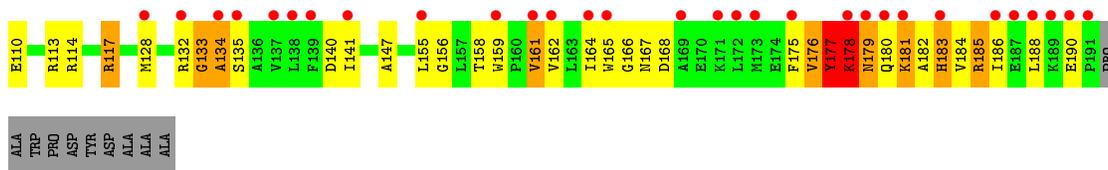

• Molecule 2: R-spondin-1

Chain P:

• Molecule 3: E3 ubiquitin-protein ligase RNF43

Chain E:

• Molecule 3: E3 ubiquitin-protein ligase RNF43

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	104.58Å 120.97Å 181.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.51 – 2.50	Depositor
	39.51 – 2.50	EDS
% Data completeness	97.1 (39.51-2.50)	Depositor
(in resolution range)	97.1 (39.51-2.50)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.63 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.232 , 0.276	Depositor
	0.232 , 0.279	DCC
R_{free} test set	3894 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	43.4	Xtriage
Anisotropy	0.321	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 54.1	EDS
L-test for twinning[2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	11252	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 16.19% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ 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: NI, 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	0.61	0/3704	0.87	9/5043 (0.2%)
1	B	0.72	1/3712 (0.0%)	0.93	8/5055 (0.2%)
2	M	0.73	0/765	0.90	0/1025
2	P	0.64	0/735	0.87	0/985
3	E	0.54	0/1175	0.80	0/1593
3	F	0.48	0/1167	0.88	2/1581 (0.1%)
All	All	0.64	1/11258 (0.0%)	0.89	19/15282 (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	B	0	8
2	P	0	1
3	E	0	3
3	F	0	3
All	All	0	16

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	168	TRP	CG-CD1	5.66	1.44	1.36

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	86	LEU	CA-CB-CG	-7.22	98.69	115.30

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	442	LEU	CA-CB-CG	6.30	129.79	115.30
1	B	96	ARG	NE-CZ-NH1	6.22	123.41	120.30
1	A	214	VAL	CB-CA-C	-6.15	99.71	111.40
1	A	170	ASP	N-CA-CB	-6.15	99.54	110.60

There are no chirality outliers.

5 of 16 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	409	LYS	Peptide
1	B	432	ASN	Peptide
1	B	57	LEU	Peptide
1	B	59	GLU	Peptide
1	B	61	PRO	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	3622	0	3626	129	0
1	B	3629	0	3634	146	0
2	M	750	0	719	17	0
2	P	720	0	692	26	0
3	E	1152	0	1182	43	0
3	F	1145	0	1175	47	0
4	A	14	0	13	0	0
4	B	14	0	13	0	0
5	B	1	0	0	0	0
6	A	70	0	0	8	0
6	B	85	0	0	7	0
6	E	18	0	0	4	0
6	F	4	0	0	1	0
6	M	21	0	0	2	0
6	P	7	0	0	8	0
All	All	11252	0	11054	394	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 18.

The worst 5 of 394 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:F:177:TYR:O	3:F:179:ASN:HA	1.44	1.17
1:B:450:LEU:HD12	1:B:453:ASN:HD22	0.98	1.15
1:B:459:LEU:HD21	1:B:478:GLN:CB	1.76	1.14
1:B:459:LEU:CD2	1:B:478:GLN:HB2	1.78	1.14
1:B:459:LEU:HD21	1:B:478:GLN:HB2	1.22	1.08

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	458/531 (86%)	389 (85%)	58 (13%)	11 (2%)	6	9
1	B	459/531 (86%)	392 (85%)	56 (12%)	11 (2%)	6	9
2	M	94/115 (82%)	82 (87%)	10 (11%)	2 (2%)	7	11
2	P	91/115 (79%)	80 (88%)	8 (9%)	3 (3%)	4	5
3	E	147/160 (92%)	120 (82%)	19 (13%)	8 (5%)	2	2
3	F	146/160 (91%)	119 (82%)	19 (13%)	8 (6%)	2	2
All	All	1395/1612 (86%)	1182 (85%)	170 (12%)	43 (3%)	4	5

5 of 43 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	410	ILE
1	A	430	SER
1	B	84	ASN
1	B	430	SER
1	B	438	PRO

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	418/476 (88%)	356 (85%)	62 (15%)	3	5
1	B	419/476 (88%)	354 (84%)	65 (16%)	2	4
2	M	87/95 (92%)	77 (88%)	10 (12%)	5	11
2	P	83/95 (87%)	72 (87%)	11 (13%)	4	7
3	E	124/130 (95%)	106 (86%)	18 (14%)	3	6
3	F	123/130 (95%)	98 (80%)	25 (20%)	1	2
All	All	1254/1402 (89%)	1063 (85%)	191 (15%)	3	5

5 of 191 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	256	LEU
1	B	421	LEU
3	F	99	LEU
1	B	278	VAL
1	B	330	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	141	GLN
1	B	218	HIS
2	P	88	ASN
1	B	166	HIS
1	B	199	HIS

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)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 for Mogul analysis.

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	#$|Z| > 2$
4	NAG	A	601	1	14,14,15	0.75	1 (7%)	17,19,21	2.08	5 (29%)
4	NAG	B	602	1	14,14,15	0.88	0	17,19,21	2.02	5 (29%)

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
4	NAG	A	601	1	-	3/6/23/26	0/1/1/1
4	NAG	B	602	1	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	601	NAG	C1-C2	2.20	1.55	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	601	NAG	C4-C3-C2	-5.33	103.20	111.02

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	602	NAG	C1-O5-C5	-4.66	105.87	112.19
4	B	602	NAG	O5-C5-C6	3.90	113.32	107.20
4	A	601	NAG	C1-C2-N2	3.77	116.93	110.49
4	A	601	NAG	C1-O5-C5	3.62	117.10	112.19

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	601	NAG	C4-C5-C6-O6
4	A	601	NAG	O5-C5-C6-O6
4	A	601	NAG	C1-C2-N2-C7

There are no ring outliers.

No monomer is involved in short contacts.

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(Å²)	Q<0.9
1	A	461/531 (86%)	0.32	38 (8%)	11	11	25, 66, 145, 192	0
1	B	463/531 (87%)	0.13	25 (5%)	25	27	23, 52, 120, 159	0
2	M	98/115 (85%)	0.21	10 (10%)	6	6	32, 59, 110, 123	0
2	P	94/115 (81%)	0.45	11 (11%)	4	4	31, 83, 113, 125	0
3	E	149/160 (93%)	0.27	9 (6%)	21	22	40, 79, 132, 151	0
3	F	148/160 (92%)	1.98	60 (40%)	0	0	78, 133, 169, 200	0
All	All	1413/1612 (87%)	0.43	153 (10%)	5	5	23, 68, 148, 200	0

The worst 5 of 153 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	480	CYS	12.3
3	F	191	PRO	10.8
1	A	481	ALA	9.0
1	A	539	VAL	8.6
3	F	61	LEU	8.0

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
4	NAG	A	601	14/15	0.90	0.31	66,82,88,95	0
4	NAG	B	602	14/15	0.93	0.28	54,60,66,67	0
5	NI	B	601	1/1	0.99	0.12	44,44,44,44	0

6.5 Other polymers (i)

There are no such residues in this entry.