W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2020 – 01:12 AM BST

PDB ID	:	4AG4
Title	:	Crystal structure of a DDR1-Fab complex
Authors	:	Carafoli, F.; Mayer, M.C.; Shiraishi, K.; Pecheva, M.A.; Chan, L.Y.; Nan, R.; Leitinger, B.; Hohenester, E.
Deposited on	:	2012-01-24
Resolution	:	2.80 Å(reported)

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

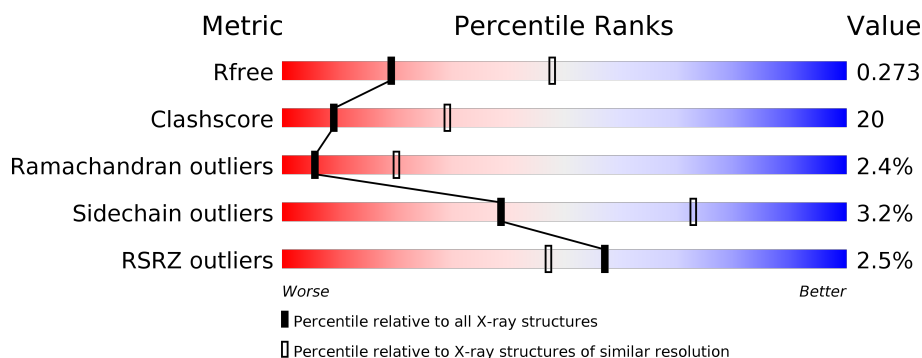|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 2.80 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		0.273
Clashscore		20
Ramachandran outliers		2.4%
Sidechain outliers		3.2%
RSRZ outliers		2.5%

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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	351	4% — 64% green, 32% yellow · ·
2	H	215	% — 63% green, 31% yellow · ·
3	L	213	% — 53% green, 44% yellow · ·
4	B	2	50% green, 50% yellow
4	C	2	50% green, 50% yellow

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5885 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 EPITHELIAL DISCOIDIN DOMAIN-CONTAINING RECEPTOR 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	341	2644	1661	481	485	17	0	0	1

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	26	ALA	-	expression tag	UNP Q08345
A	27	PRO	-	expression tag	UNP Q08345
A	28	LEU	-	expression tag	UNP Q08345
A	368	ALA	-	expression tag	UNP Q08345
A	369	ALA	-	expression tag	UNP Q08345
A	370	ALA	-	expression tag	UNP Q08345
A	371	HIS	-	expression tag	UNP Q08345
A	372	HIS	-	expression tag	UNP Q08345
A	373	HIS	-	expression tag	UNP Q08345
A	374	HIS	-	expression tag	UNP Q08345
A	375	HIS	-	expression tag	UNP Q08345
A	376	HIS	-	expression tag	UNP Q08345

- Molecule 2 is a protein called MONOCLONAL ANTIBODY 3E3 HEAVY CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	209	1567	995	254	313	5	0	0	0

- Molecule 3 is a protein called MONOCLONAL ANTIBODY 3E3 LIGHT CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	L	211	1602	1000	263	330	9	0	0	1

- Molecule 4 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	B	2	Total	C	N	O	0	0	0
			28	16	2	10			
4	C	2	Total	C	N	O	0	0	0
			28	16	2	10			

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	2	Total	Ca	0	0
			2	2		

- Molecule 6 is water.

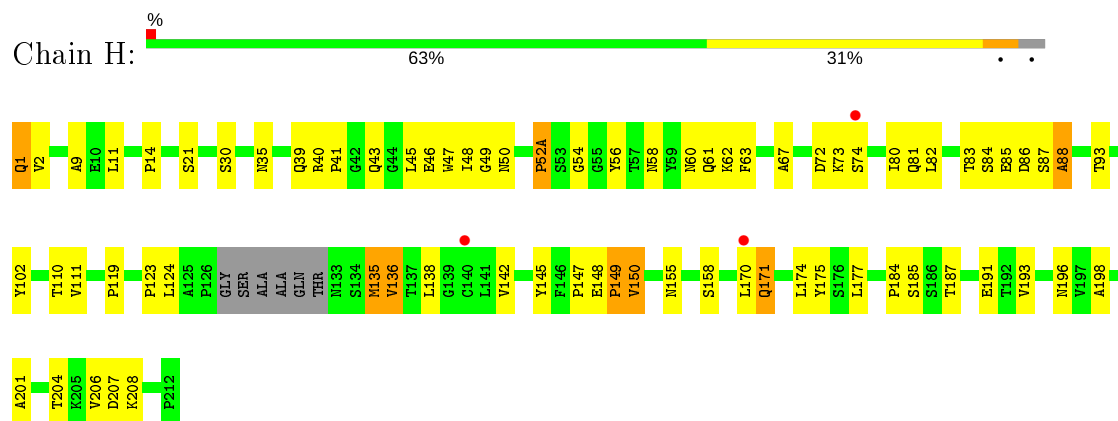| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	6	Total	O	0	0
			6	6		
6	H	5	Total	O	0	0
			5	5		
6	L	3	Total	O	0	0
			3	3		

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.
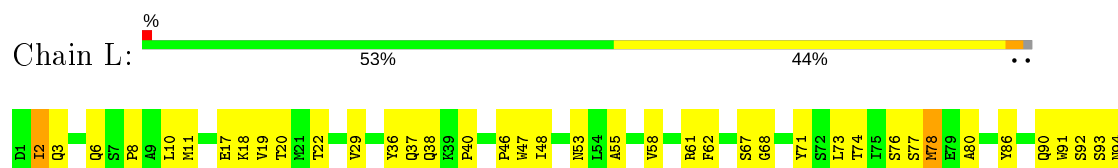
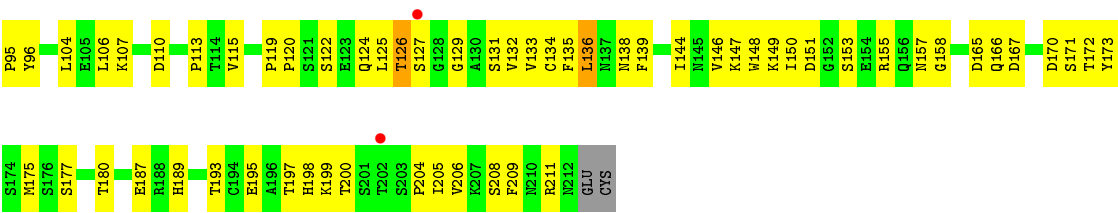• Molecule 1: EPITHELIAL DISCOIDIN DOMAIN-CONTAINING RECEPTOR 1
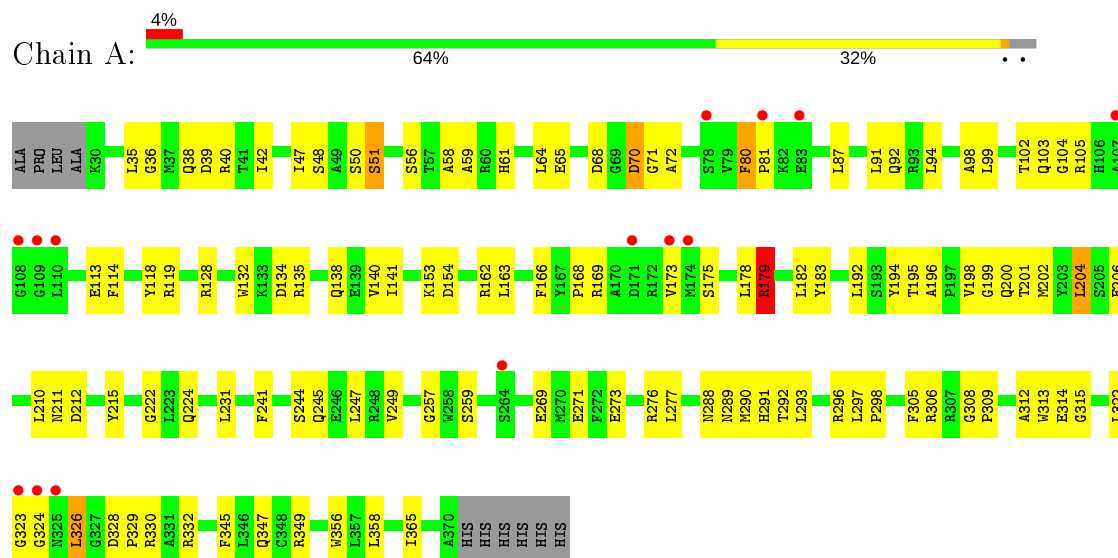
Chain A:

• Molecule 2: MONOCLONAL ANTIBODY 3E3 HEAVY CHAIN

Chain H:

• Molecule 3: MONOCLONAL ANTIBODY 3E3 LIGHT CHAIN

Chain L:

P95 Y96 L104 E105 L106 K107 D110 P113 T114 V115 P119 P120 S121 S122 E123 Q124 L125 T126 S127 G128 G129 A130 S131 V132 V133 C134 F135 L136 N137 N138 F139 I144 N145 V146 K147 W148 K149 I150 D151 G152 S153 E154 R155 Q156 N157 G158 D165 Q166 D167 D170 S171 T172 Y173

S174 M175 S176 S177 T180 E187 R188 H189 T193 C194 E195 A196 T197 H198 K199 T200 S201 T202 S203 P204 I205 V206 K207 S208 F209 N210 R211 N212 GLU CYS

• Molecule 4: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-gluc opyranose

Chain B:

50%   50%

NAG1 NAG2

• Molecule 4: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-gluc opyranose

Chain C:

50%   50%

NAG1 NAG2

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	102.51Å 251.48Å 75.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.80 19.92 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.5 (20.00-2.80) 98.5 (19.92-2.80)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.20 (at 2.79Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.216 , 0.285 0.206 , 0.273	Depositor DCC
R_{free} test set	2388 reflections (9.94%)	wwPDB-VP
Wilson B-factor (Å^2)	57.5	Xtriage
Anisotropy	0.260	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 40.7	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	5885	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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: CA, 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	0.41	0/2714	0.69	1/3679 (0.0%)
2	H	0.40	0/1611	0.68	0/2211
3	L	0.38	0/1643	0.62	0/2239
All	All	0.40	0/5968	0.67	1/8129 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	179	ARG	N-CA-C	-5.62	95.82	111.00

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	2644	0	2486	95	0
2	H	1567	0	1497	60	0
3	L	1602	0	1506	74	0
4	B	28	0	25	0	0
4	C	28	0	25	0	0
5	A	2	0	0	0	0
6	A	6	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	H	5	0	0	0	0
6	L	3	0	0	0	0
All	All	5885	0	5539	228	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 20.

The worst 5 of 228 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:80:PHE:HB3	1:A:81:PRO:HD2	1.29	1.13
1:A:138:GLN:HE21	1:A:140:VAL:H	1.12	0.95
1:A:72:ALA:HB2	1:A:179:ARG:HD3	1.54	0.89
3:L:195:GLU:HG2	3:L:206:VAL:HG22	1.57	0.84
2:H:1:GLN:NE2	2:H:1:GLN:N	2.28	0.82

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	339/351 (97%)	303 (89%)	29 (9%)	7 (2%)	7	23
2	H	205/215 (95%)	185 (90%)	15 (7%)	5 (2%)	6	20
3	L	209/213 (98%)	185 (88%)	18 (9%)	6 (3%)	4	15
All	All	753/779 (97%)	673 (89%)	62 (8%)	18 (2%)	6	20

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	323	GLY
2	H	84	SER
1	A	51	SER
2	H	171	GLN

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	268/284 (94%)	262 (98%)	6 (2%)	52	83
2	H	176/184 (96%)	168 (96%)	8 (4%)	27	60
3	L	181/188 (96%)	175 (97%)	6 (3%)	38	72
All	All	625/656 (95%)	605 (97%)	20 (3%)	39	73

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	136	VAL
2	H	149	PRO
3	L	78	MET
2	H	21	SER
2	H	135	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	281	GLN
2	H	1	GLN
2	H	81	GLN
1	A	224	GLN
2	H	5	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)

4 monosaccharides 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
4	NAG	B	1	1,4	14,14,15	0.56	0	17,19,21	0.98	1 (5%)
4	NAG	B	2	4	14,14,15	0.53	0	17,19,21	0.66	0
4	NAG	C	1	1,4	14,14,15	0.53	0	17,19,21	0.64	0
4	NAG	C	2	4	14,14,15	0.78	1 (7%)	17,19,21	0.68	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
4	NAG	B	1	1,4	-	4/6/23/26	0/1/1/1
4	NAG	B	2	4	-	6/6/23/26	0/1/1/1
4	NAG	C	1	1,4	-	2/6/23/26	0/1/1/1
4	NAG	C	2	4	-	4/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	2	NAG	C1-C2	2.14	1.55	1.52

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	1	NAG	C2-N2-C7	-3.17	118.39	122.90

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	2	NAG	C8-C7-N2-C2
4	C	2	NAG	O7-C7-N2-C2
4	B	1	NAG	C8-C7-N2-C2
4	B	1	NAG	O7-C7-N2-C2
4	B	2	NAG	C1-C2-N2-C7

There are no ring outliers.

No monomer is involved in short contacts.

5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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	341/351 (97%)	-0.07	14 (4%)	37	27	30, 47, 75, 95	0
2	H	209/215 (97%)	-0.13	3 (1%)	75	70	30, 50, 70, 82	0
3	L	211/213 (99%)	0.05	2 (0%)	84	80	33, 61, 88, 92	0
All	All	761/779 (97%)	-0.06	19 (2%)	57	47	30, 51, 81, 95	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	171	ASP	6.1
1	A	110	LEU	4.1
1	A	109	GLY	3.8
1	A	174	MET	3.2
1	A	81	PRO	3.1

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)

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(Å²)	Q<0.9
4	NAG	C	1	14/15	0.84	0.23	60,63,65,67	0
4	NAG	C	2	14/15	0.88	0.43	70,72,75,75	0
4	NAG	B	2	14/15	0.92	0.26	64,68,70,70	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	NAG	B	1	14/15	0.94	0.15	55,60,64,66	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. 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
5	CA	A	5002	1/1	0.96	0.16	69,69,69,69	0
5	CA	A	5001	1/1	0.97	0.10	39,39,39,39	0

6.5 Other polymers (i)

There are no such residues in this entry.