W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 14, 2017 – 12:44 am GMT

PDB ID	:	4UI2
Title	:	Crystal structure of the ternary RGMB-BMP2-NEO1 complex
Authors	:	Healey, E.G.; Bishop, B.; Elegheert, J.; Bell, C.H.; Padilla-Parra, S.; Siebold, C.
Deposited on	:	2015-03-27
Resolution	:	3.15 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

|		
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28949

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 3.15 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.
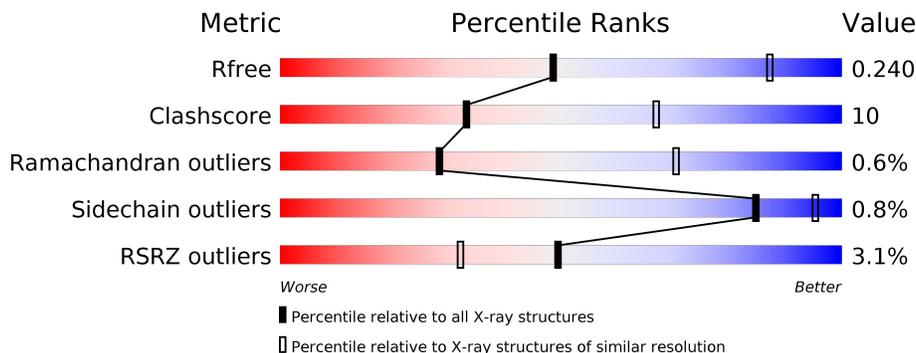
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1259 (3.20-3.12)
Clashscore	112137	1397 (3.20-3.12)
Ramachandran outliers	110173	1368 (3.20-3.12)
Sidechain outliers	110143	1367 (3.20-3.12)
RSRZ outliers	101464	1264 (3.20-3.12)

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. 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	264	2% / 66% / 9% · 24%
2	B	114	4% / 71% / 19% · 9%
3	C	122	2% / 56% / 9% · 34%
4	D	251	2% / 47% / 14% / 38%

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
6	SRT	B	1399	-	-	-	X
7	ACT	B	1400	-	-	-	X

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 4311 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 NEOGENIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	201	Total	C	N	O	S	0	0	0
			1596	1018	274	298	6			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	880	GLU	-	EXPRESSION TAG	UNP P97798
A	881	THR	-	EXPRESSION TAG	UNP P97798
A	882	GLY	-	EXPRESSION TAG	UNP P97798
A	1134	ASN	-	EXPRESSION TAG	UNP P97798
A	1135	GLY	-	EXPRESSION TAG	UNP P97798
A	1136	THR	-	EXPRESSION TAG	UNP P97798
A	1137	LYS	-	EXPRESSION TAG	UNP P97798
A	1138	HIS	-	EXPRESSION TAG	UNP P97798
A	1139	HIS	-	EXPRESSION TAG	UNP P97798
A	1140	HIS	-	EXPRESSION TAG	UNP P97798
A	1141	HIS	-	EXPRESSION TAG	UNP P97798
A	1142	HIS	-	EXPRESSION TAG	UNP P97798
A	1143	HIS	-	EXPRESSION TAG	UNP P97798

- Molecule 2 is a protein called BONE MORPHOGENETIC PROTEIN 2, BMP2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	104	Total	C	N	O	S	0	0	0
			814	516	137	152	9			

- Molecule 3 is a protein called REPULSIVE GUIDANCE MOLECULE C, RGMC, HEMO-JUVELIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	80	Total	C	N	O	S	0	0	0
			616	378	111	118	9			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	47	GLU	-	EXPRESSION TAG	UNP Q6NW40
C	48	THR	-	EXPRESSION TAG	UNP Q6NW40
C	49	GLY	-	EXPRESSION TAG	UNP Q6NW40

- Molecule 4 is a protein called REPULSIVE GUIDANCE MOLECULE C, RGMC, HEMO-JUVELIN.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	156	1212	757	210	238	7	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	411	GLY	-	EXPRESSION TAG	UNP Q6NW40
D	412	THR	-	EXPRESSION TAG	UNP Q6NW40
D	413	LYS	-	EXPRESSION TAG	UNP Q6NW40
D	414	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	415	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	416	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	417	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	418	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	419	HIS	-	EXPRESSION TAG	UNP Q6NW40
D	225	GLY	GLU	CONFLICT	UNP Q6NW40

- Molecule 5 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
5	A	3	39	22	2	15	0	0

- Molecule 6 is S,R MESO-TARTARIC ACID (three-letter code: SRT) (formula: $C_4H_6O_6$).

SRT

O1
OH

O11 O — C1

C2(S) — OH O2

O3 HO — C3(R)

C4 — O O4

HO
O41
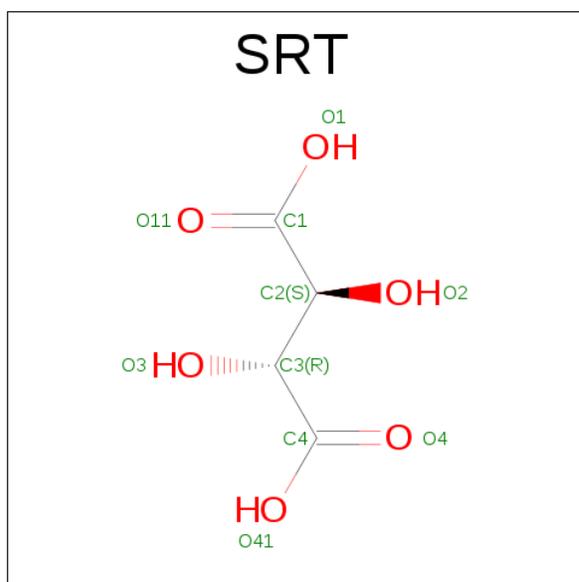
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	B	1	Total 10	C 4	O 6	0	0
6	B	1	Total 10	C 4	O 6	0	0
6	B	1	Total 10	C 4	O 6	0	0

• Molecule 7 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

ACT

OXT
O^-

C

CH3 —

O

O
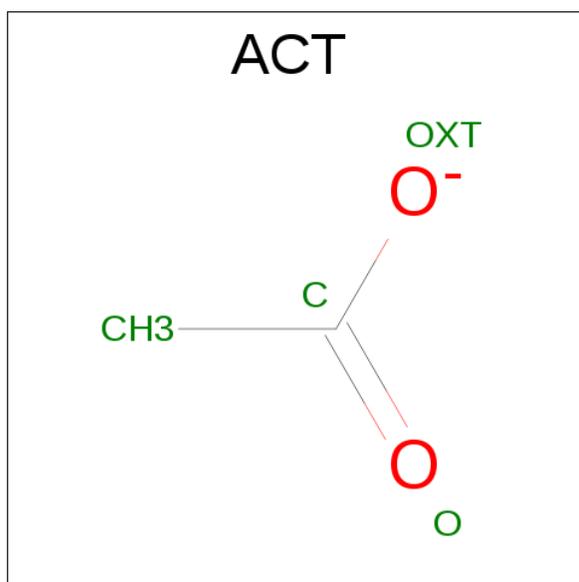
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	B	1	Total 4	C 2	O 2	0	0

WORLD WIDE
PDB
PROTEIN DATA BANK

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA 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: NEOGENIN
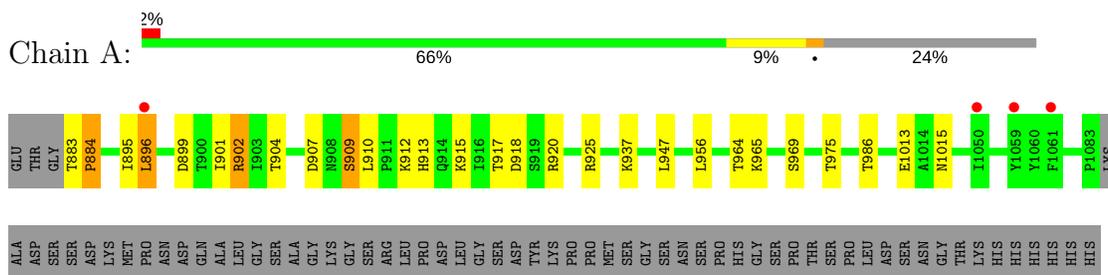
• Molecule 2: BONE MORPHOGENETIC PROTEIN 2, BMP2
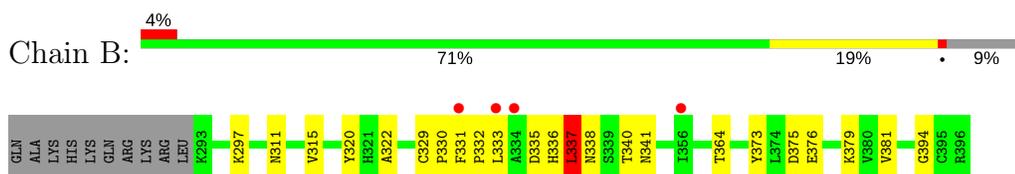
• Molecule 3: REPULSIVE GUIDANCE MOLECULE C, RGMC, HEMOJUVELIN
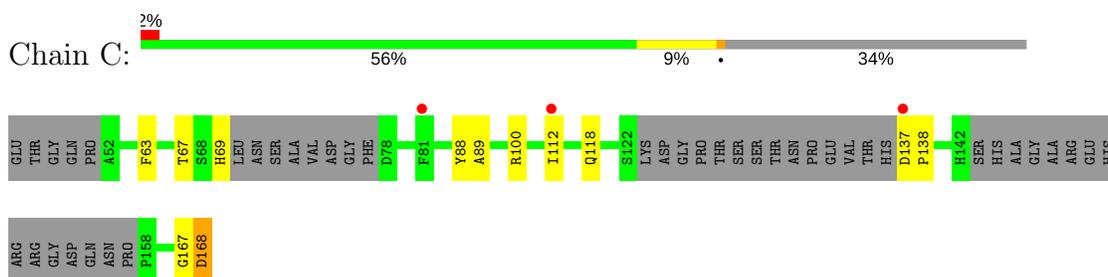
• Molecule 4: REPULSIVE GUIDANCE MOLECULE C, RGMC, HEMOJUVELIN
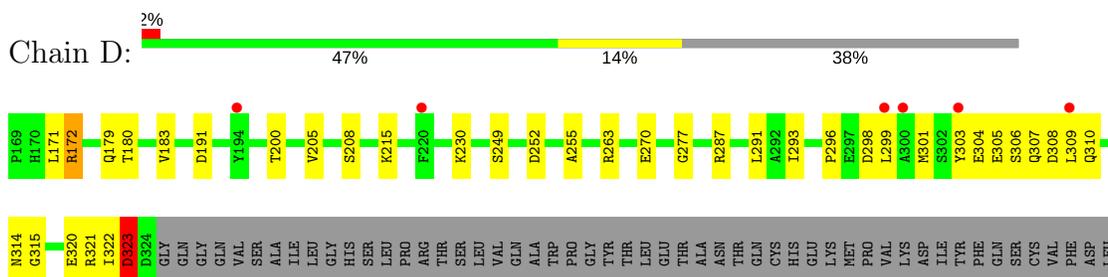
LEU THR THR GLY ASP ALA ASN PHE THR ALA ALA ALA HIS SER ALA LEU GLU ASP VAL GLU ALA LEU HIS PRO ARG LYS GLU ARG TRP HIS ILE PHE PRO SER GLY THR LYS HIS HIS HIS HIS HIS

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	120.08Å 120.08Å 204.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.01 – 3.15 103.50 – 3.10	Depositor EDS
% Data completeness (in resolution range)	98.6 (50.01-3.15) 98.5 (103.50-3.10)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.70 (at 3.13Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, R_{free}	0.196 , 0.236 0.203 , 0.240	Depositor DCC
R_{free} test set	1250 reflections (5.01%)	DCC
Wilson B-factor (Å^2)	107.9	Xtriage
Anisotropy	0.695	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 109.9	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	0.95	EDS
Total number of atoms	4311	wwPDB-VP
Average B, all atoms (Å^2)	135.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.09% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: BMA, NAG, SRT, ACT

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	0.62	0/1640	0.84	3/2241 (0.1%)
2	B	0.69	0/837	0.86	2/1140 (0.2%)
3	C	0.58	0/625	0.86	2/835 (0.2%)
4	D	0.64	0/1236	0.89	4/1680 (0.2%)
All	All	0.63	0/4338	0.86	11/5896 (0.2%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	168	ASP	CB-CG-OD2	-9.23	109.99	118.30
1	A	896	LEU	CB-CG-CD1	8.96	126.22	111.00
4	D	323	ASP	CB-CG-OD2	6.66	124.29	118.30
4	D	172	ARG	NE-CZ-NH1	6.63	123.61	120.30
2	B	337	LEU	CA-CB-CG	5.43	127.79	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

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	1596	0	1594	19	0
2	B	814	0	775	27	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	616	0	576	8	0
4	D	1212	0	1176	32	0
5	A	39	0	32	0	0
6	B	30	0	12	0	0
7	B	4	0	3	0	0
All	All	4311	0	4168	83	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 83 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)
2:B:332:PRO:HA	2:B:335:ASP:OD2	1.16	1.26
2:B:332:PRO:CA	2:B:335:ASP:OD2	1.92	1.15
4:D:305:GLU:OE1	4:D:306:SER:N	2.14	0.81
2:B:337:LEU:CG	2:B:338:ASN:HA	2.11	0.81
2:B:337:LEU:HG	2:B:338:ASN:HA	1.61	0.80

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	199/264 (75%)	190 (96%)	8 (4%)	1 (0%)	32	73
2	B	102/114 (90%)	94 (92%)	7 (7%)	1 (1%)	18	60
3	C	72/122 (59%)	67 (93%)	5 (7%)	0	100	100
4	D	154/251 (61%)	137 (89%)	16 (10%)	1 (1%)	28	70
All	All	527/751 (70%)	488 (93%)	36 (7%)	3 (1%)	28	70

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	337	LEU
4	D	277	GLY
1	A	884	PRO

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	180/232 (78%)	177 (98%)	3 (2%)	66	88
2	B	93/102 (91%)	93 (100%)	0	100	100
3	C	69/103 (67%)	69 (100%)	0	100	100
4	D	135/215 (63%)	134 (99%)	1 (1%)	87	96
All	All	477/652 (73%)	473 (99%)	4 (1%)	85	95

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	909	SER
1	A	956	LEU
1	A	975	THR
4	D	323	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	179	GLN
4	D	314	ASN
4	D	307	GLN
1	A	938	ASN
4	D	310	GLN

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 ⓘ

3 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	2084	1,5	14,14,15	0.64	0	15,19,21	2.09	4 (26%)
5	NAG	A	2085	5	14,14,15	0.40	0	15,19,21	1.39	2 (13%)
5	BMA	A	2086	5	11,11,12	0.55	0	13,15,17	2.99	7 (53%)

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	2084	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	2085	5	-	0/6/23/26	0/1/1/1
5	BMA	A	2086	5	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	2086	BMA	O2-C2-C3	-5.98	98.43	110.17
5	A	2084	NAG	O5-C1-C2	-3.01	107.29	111.47

Continued on next page...

WORLDWIDE wwPDB PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	2085	NAG	C4-C3-C2	-2.91	106.76	111.02
5	A	2086	BMA	C6-C5-C4	-2.20	107.84	113.00
5	A	2086	BMA	C1-C2-C3	2.06	112.26	109.65

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 ⓘ

4 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	SRT	B	1397	-	3,9,9	0.41	0	6,12,12	1.10	1 (16%)
6	SRT	B	1398	-	3,9,9	0.95	0	6,12,12	0.53	0
6	SRT	B	1399	-	3,9,9	0.96	0	6,12,12	0.83	0
7	ACT	B	1400	-	1,3,3	0.05	0	0,3,3	0.00	-

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	SRT	B	1397	-	-	0/4/12/12	0/0/0/0
6	SRT	B	1398	-	-	0/4/12/12	0/0/0/0
6	SRT	B	1399	-	-	0/4/12/12	0/0/0/0
7	ACT	B	1400	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	B	1397	SRT	C4-C3-C2	-2.32	108.12	113.11

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 ⓘ

There are no such residues in this entry.

5.8   Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	201/264 (76%)	0.46	4 (1%)	65	49	97, 131, 194, 234	0
2	B	104/114 (91%)	0.62	4 (3%)	41	25	88, 109, 185, 237	0
3	C	80/122 (65%)	0.47	3 (3%)	41	25	96, 151, 213, 234	0
4	D	156/251 (62%)	0.66	6 (3%)	41	25	86, 113, 186, 220	0
All	All	541/751 (72%)	0.55	17 (3%)	49	32	86, 125, 200, 237	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	81	PHE	4.1
2	B	334	ALA	3.9
3	C	137	ASP	3.1
4	D	299	LEU	2.5
2	B	333	LEU	2.5

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

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, 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	LLDF	B-factors(Å^2)	Q<0.9
5	NAG	A	2084	14/15	0.90	0.20	-1.13	130,157,168,180	0
5	NAG	A	2085	14/15	0.90	0.21	-	148,184,204,216	0
5	BMA	A	2086	11/12	0.53	0.19	-	204,224,237,238	0

6.4 Ligands ⓘ

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, 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	LLDF	B-factors(Å^2)	Q<0.9
6	SRT	B	1399	10/10	0.57	0.50	1.86	139,171,181,184	0
6	SRT	B	1398	10/10	0.70	0.35	1.83	119,150,174,176	0
7	ACT	B	1400	4/4	0.88	0.41	1.59	94,106,107,108	0
6	SRT	B	1397	10/10	0.65	0.27	-	130,167,173,175	0

6.5 Other polymers ⓘ

There are no such residues in this entry.