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

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 29, 2024 – 09:23 PM EDT

PDB ID	:	4HNY
Title	:	Apo N-terminal acetyltransferase complex A
Authors	:	Neubauer, J.L.; Immormino, R.M.; Dollins, D.E.; Endo-Streeter, S.T.; Pemble IV, C.W.; York, J.D.
Deposited on	:	2012-10-21
Resolution	:	2.25 Å(reported)

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

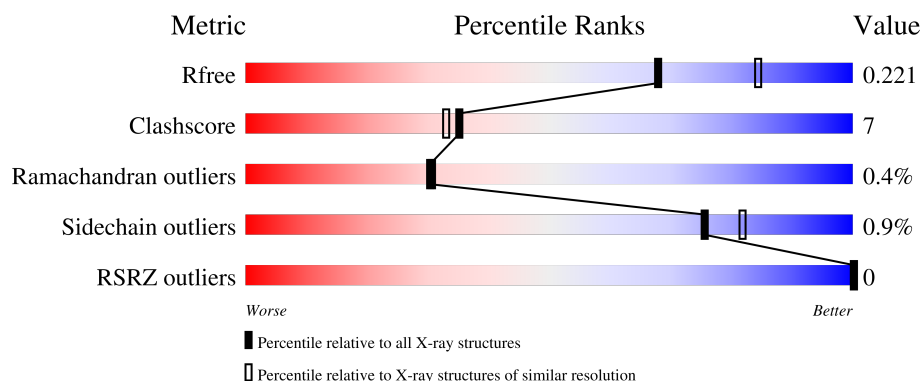|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.25 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		0.221
Clashscore		7
Ramachandran outliers		0.4%
Sidechain outliers		0.9%
RSRZ outliers		0

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}	164625	3139 (2.26-2.22)
Clashscore	180529	3381 (2.26-2.22)
Ramachandran outliers	177936	3334 (2.26-2.22)
Sidechain outliers	177891	3335 (2.26-2.22)
RSRZ outliers	164620	3138 (2.26-2.22)

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 of 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	863	73% 12% 15%
1	C	863	76% 11% • 12%
2	B	248	76% 11% 13%
2	D	248	78% 9% 13%

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
3	PG4	A	902	-	-	X	-
3	PG4	B	301	-	-	X	-
3	PG4	B	302	-	-	X	-
4	GOL	A	903	-	-	X	-

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 16061 atoms, of which 8 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 N-terminal acetyltransferase A complex subunit NAT1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	733	Total 5753	C 3703	N 961	O 1067	S 22	0	2	0
1	C	759	Total 5910	C 3807	N 977	O 1104	S 22	2	1	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	855	TYR	-	expression tag	UNP P12945
A	856	PRO	-	expression tag	UNP P12945
A	857	TYR	-	expression tag	UNP P12945
A	858	ASP	-	expression tag	UNP P12945
A	859	VAL	-	expression tag	UNP P12945
A	860	PRO	-	expression tag	UNP P12945
A	861	ASP	-	expression tag	UNP P12945
A	862	TYR	-	expression tag	UNP P12945
A	863	ALA	-	expression tag	UNP P12945
C	855	TYR	-	expression tag	UNP P12945
C	856	PRO	-	expression tag	UNP P12945
C	857	TYR	-	expression tag	UNP P12945
C	858	ASP	-	expression tag	UNP P12945
C	859	VAL	-	expression tag	UNP P12945
C	860	PRO	-	expression tag	UNP P12945
C	861	ASP	-	expression tag	UNP P12945
C	862	TYR	-	expression tag	UNP P12945
C	863	ALA	-	expression tag	UNP P12945

- Molecule 2 is a protein called N-terminal acetyltransferase A complex catalytic subunit ARD1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	216	Total 1666	C 1043	N 291	O 322	S 10	0	2	0

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	216	1657	1038	285	323	11	0	1	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	239	GLU	-	expression tag	UNP P07347
B	240	GLN	-	expression tag	UNP P07347
B	241	LYS	-	expression tag	UNP P07347
B	242	LEU	-	expression tag	UNP P07347
B	243	ILE	-	expression tag	UNP P07347
B	244	SER	-	expression tag	UNP P07347
B	245	GLU	-	expression tag	UNP P07347
B	246	GLU	-	expression tag	UNP P07347
B	247	ASP	-	expression tag	UNP P07347
B	248	LEU	-	expression tag	UNP P07347
D	239	GLU	-	expression tag	UNP P07347
D	240	GLN	-	expression tag	UNP P07347
D	241	LYS	-	expression tag	UNP P07347
D	242	LEU	-	expression tag	UNP P07347
D	243	ILE	-	expression tag	UNP P07347
D	244	SER	-	expression tag	UNP P07347
D	245	GLU	-	expression tag	UNP P07347
D	246	GLU	-	expression tag	UNP P07347
D	247	ASP	-	expression tag	UNP P07347
D	248	LEU	-	expression tag	UNP P07347

- Molecule 3 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).

PG4

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 11	C 7	O 4	0	0
3	A	1	Total 10	C 6	O 4	0	0
3	B	1	Total 13	C 8	O 5	0	0
3	B	1	Total 10	C 6	O 4	0	0
3	B	1	Total 10	C 6	O 4	0	0
3	C	1	Total 11	C 7	O 4	0	0
3	C	1	Total 10	C 6	O 4	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 14	C 3	H 8	O 3	0	0

- Molecule 5 is water.

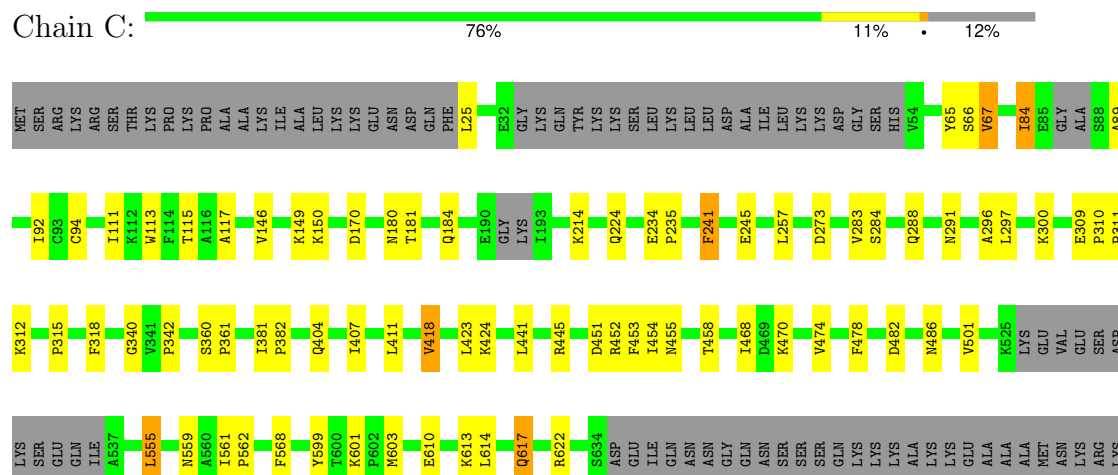| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	379	Total 379	O 379	0	0
5	B	138	Total 138	O 138	0	0
5	C	342	Total 342	O 342	0	0
5	D	127	Total 127	O 127	0	0

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.

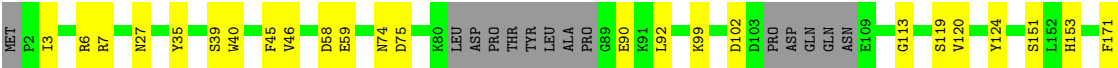• Molecule 1: N-terminal acetyltransferase A complex subunit NAT1

Chain A:

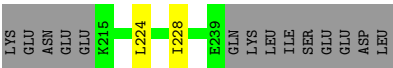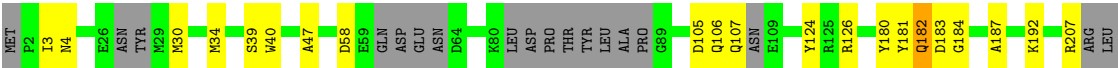• Molecule 1: N-terminal acetyltransferase A complex subunit NAT1

Chain C:

GLU THR GLU ALA LYS SER VAL ALA ALA TYR PRO SER ASP GLN ASP N678 N702 M705 Q706 E709 Y714 F718 F731 S745 M751 L756 T759 K770 K771 V772 S776 L777 L787 I790 K831 S839 P840 L850 L854 Y862 ALA

• Molecule 2: N-terminal acetyltransferase A complex catalytic subunit ARD1

Chain B:

76% 11% 13%

MET P2 I3 R6 R7 N27 Y35 S39 W40 F45 V46 D58 E59 N74 D75 K80 LEU ASP PRO THR TYR LEU ALA PRO G89 E90 K91 L92 K99 D102 D103 PRO ASP GLN GLN ASN E109 G113 S119 V120 Y124 S151 L152 H153 F171

M190 L209 LYS GLU ASN GLU GLU K215 L216 E217 N234 D235 ILE ILE VAL GLU GLN LYS LEU ILE SER GLU GLU ASP LEU

• Molecule 2: N-terminal acetyltransferase A complex catalytic subunit ARD1

Chain D:

78% 9% 13%

MET P2 I3 N4 E26 ASN TYR M29 M30 M34 S39 W40 A47 D58 E59 GLN ASP GLU ASN D64 K80 LEU ASP PRO THR TYR LEU ALA PRO G89 D105 Q106 Q107 ASN E109 Y124 R125 R126 Y180 Y181 Q182 D183 G184 A187 K192 R207 ARG LEU

LYS GLU ASN GLU GLU K215 L224 I228 E239 GLN LYS LEU ILE SER GLU GLU ASP LEU

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, α, β, γ	129.42Å 129.42Å 171.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.16 – 2.25 30.16 – 2.25	Depositor EDS
% Data completeness (in resolution range)	95.3 (30.16-2.25) 95.3 (30.16-2.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.08	Depositor
$< I/\sigma(I) >$ [1]	4.73 (at 2.24Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869)	Depositor
R, R_{free}	0.178 , 0.220 0.182 , 0.221	Depositor DCC
R_{free} test set	1848 reflections (1.27%)	wwPDB-VP
Wilson B-factor (Å²)	28.4	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent k_{sol}(e/Å³), B_{sol}(Å²)	0.33 , 33.0	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.019 for -h,-k,l 0.093 for h,-h-k,-l 0.028 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	16061	wwPDB-VP
Average B, all atoms (Å²)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.20% 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: PG4, GOL

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	0.39	0/5882	0.49	0/7967
1	C	0.38	0/6044	0.50	0/8209
2	B	0.42	0/1698	0.54	0/2302
2	D	0.43	0/1684	0.53	0/2282
All	All	0.40	0/15308	0.50	0/20760

There are no bond length outliers.

There are no bond angle outliers.

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	5753	0	5450	75	0
1	C	5910	0	5504	84	0
2	B	1666	0	1546	42	0
2	D	1657	0	1528	19	0
3	A	21	0	26	12	0
3	B	33	0	44	19	0
3	C	21	0	26	8	0
4	A	6	8	8	7	0
5	A	379	0	0	7	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	138	0	0	3	0
5	C	342	0	0	10	0
5	D	127	0	0	5	1
All	All	16053	8	14132	214	1

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 7.

The worst 5 of 214 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:D:182:GLN:O	2:D:184:GLY:N	1.83	1.09
1:C:84:ILE:HD11	1:C:89:ALA:HB2	1.28	1.09
2:B:153[A]:HIS:H	3:B:301:PG4:H11	1.24	0.98
2:B:153[B]:HIS:H	3:B:301:PG4:H11	1.24	0.98
3:A:902:PG4:H81	3:A:902:PG4:H32	1.47	0.97

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:1314:HOH:O	5:D:358:HOH:O[3_554]	2.13	0.07

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	723/863 (84%)	709 (98%)	12 (2%)	2 (0%)	37	39
1	C	748/863 (87%)	730 (98%)	17 (2%)	1 (0%)	48	55
2	B	210/248 (85%)	204 (97%)	5 (2%)	1 (0%)	25	24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	D	205/248 (83%)	197 (96%)	5 (2%)	3 (2%)	8 4
All	All	1886/2222 (85%)	1840 (98%)	39 (2%)	7 (0%)	30 31

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	192	LYS
2	D	182	GLN
2	D	183	ASP
2	B	59	GLU
2	D	106	GLN

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	574/767 (75%)	571 (100%)	3 (0%)	86 90
1	C	583/767 (76%)	574 (98%)	9 (2%)	60 69
2	B	169/226 (75%)	168 (99%)	1 (1%)	84 88
2	D	168/226 (74%)	167 (99%)	1 (1%)	84 88
All	All	1494/1986 (75%)	1480 (99%)	14 (1%)	75 82

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	291	ASN
1	C	418	VAL
2	D	34	MET
1	C	745	SER
1	C	787	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	180	ASN
1	A	184	GLN
2	B	27	ASN
1	C	180	ASN
1	C	184	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 ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

8 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	PG4	B	303	-	9,9,12	0.58	0	8,8,11	0.90	0
4	GOL	A	903	-	5,5,5	0.69	0	5,5,5	0.88	0
3	PG4	B	301	-	12,12,12	0.71	0	11,11,11	0.68	0
3	PG4	C	902	-	9,9,12	0.65	0	8,8,11	0.80	0
3	PG4	C	901	-	10,10,12	0.62	0	9,9,11	0.94	0
3	PG4	B	302	-	9,9,12	0.61	0	8,8,11	0.85	0
3	PG4	A	902	-	9,9,12	0.63	0	8,8,11	0.83	0
3	PG4	A	901	-	10,10,12	0.61	0	9,9,11	0.89	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
3	PG4	B	303	-	-	3/7/7/10	-
4	GOL	A	903	-	-	2/4/4/4	-
3	PG4	B	301	-	-	4/10/10/10	-
3	PG4	C	902	-	-	3/7/7/10	-
3	PG4	C	901	-	-	3/8/8/10	-
3	PG4	B	302	-	-	6/7/7/10	-
3	PG4	A	902	-	-	6/7/7/10	-
3	PG4	A	901	-	-	6/8/8/10	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 33 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	903	GOL	O1-C1-C2-C3
3	B	302	PG4	C4-C3-O2-C2
3	B	301	PG4	C4-C3-O2-C2
3	B	301	PG4	C8-C7-O4-C6
3	A	902	PG4	O3-C5-C6-O4

There are no ring outliers.

8 monomers are involved in 46 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	303	PG4	4	0
4	A	903	GOL	7	0
3	B	301	PG4	8	0
3	C	902	PG4	5	0
3	C	901	PG4	3	0
3	B	302	PG4	7	0
3	A	902	PG4	8	0
3	A	901	PG4	4	0

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	733/863 (84%)	-1.65	0	100	100	16, 37, 74, 93	2 (0%)
1	C	759/863 (87%)	-1.68	0	100	100	16, 38, 64, 83	1 (0%)
2	B	216/248 (87%)	-1.70	0	100	100	14, 30, 70, 94	3 (1%)
2	D	216/248 (87%)	-1.67	0	100	100	17, 37, 65, 83	1 (0%)
All	All	1924/2222 (86%)	-1.67	0	100	100	14, 36, 69, 94	7 (0%)

There are no RSRZ outliers to report.

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 ⓘ

There are no monosaccharides in this entry.

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. 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
3	PG4	C	901	11/13	0.98	0.05	42,64,73,74	0
3	PG4	A	902	10/13	0.99	0.06	64,68,76,80	0
3	PG4	B	301	13/13	0.99	0.05	36,55,59,60	0
3	PG4	B	302	10/13	0.99	0.04	51,62,64,65	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	PG4	B	303	10/13	0.99	0.05	50,59,63,70	0
3	PG4	A	901	11/13	0.99	0.05	49,63,69,70	0
3	PG4	C	902	10/13	0.99	0.04	71,73,78,81	0
4	GOL	A	903	6/6	0.99	0.09	28,44,53,53	0

6.5 Other polymers ⓘ

There are no such residues in this entry.