



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:15 PM GMT

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)

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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

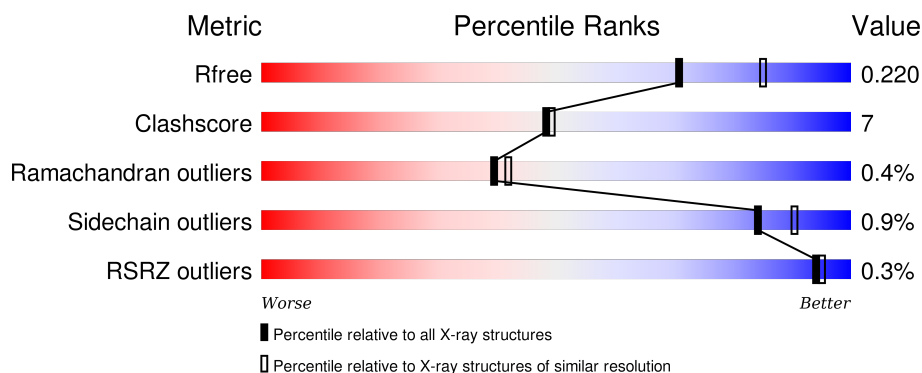
# 1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1611 (2.26-2.22)
Clashscore	102246	1764 (2.26-2.22)
Ramachandran outliers	100387	1724 (2.26-2.22)
Sidechain outliers	100360	1724 (2.26-2.22)
RSRZ outliers	91569	1616 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ . 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	863	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, orange 1%, yellow 12%, green 73%, grey 15%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>73%</span> <span>12%</span> <span>15%</span> </div> </div>
1	C	863	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 76%, yellow 11%, orange 1%, red 12%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>76%</span> <span>11%</span> <span>12%</span> </div> </div>
2	B	248	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 76%, yellow 11%, orange 1%, red 13%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>76%</span> <span>11%</span> <span>13%</span> </div> </div>
2	D	248	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 78%, yellow 9%, orange 1%, red 13%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>78%</span> <span>9%</span> <span>13%</span> </div> </div>

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	X
3	PG4	B	301	-	-	X	-
3	PG4	B	302	-	-	X	X
3	PG4	C	901	-	-	-	X
3	PG4	C	902	-	-	-	X
4	GOL	A	903	-	-	X	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	C	N	O	S	0	2	0
			5753	3703	961	1067	22			
1	C	759	Total	C	N	O	S	2	1	0
			5910	3807	977	1104	22			

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	C	N	O	S	0	2	0
			1666	1043	291	322	10			

*Continued on next page...*

*Continued from previous page...*

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

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<sub>8</sub>H<sub>18</sub>O<sub>5</sub>).



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

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

- Molecule 5 is water.

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

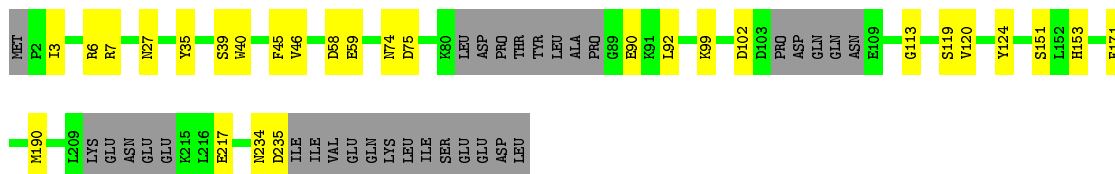






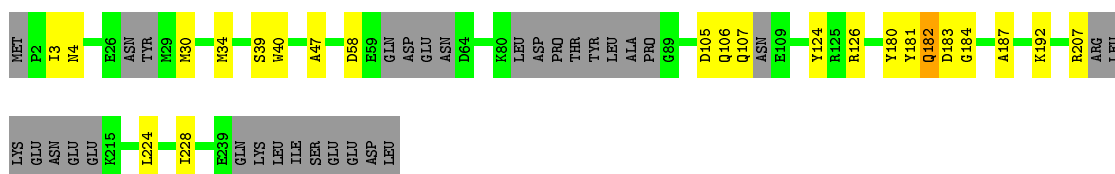
- Molecule 2: N-terminal acetyltransferase A complex catalytic subunit ARD1

Chain B: 76% 11% 13%



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

Chain D: 78% 9% 13%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	129.42Å 129.42Å 171.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.16 – 2.25 46.93 – 2.25	Depositor EDS
% Data completeness (in resolution range)	95.3 (30.16-2.25) 95.3 (46.93-2.25)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.73 (at 2.24Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.178 , 0.220 0.185 , 0.220	Depositor DCC
$R_{free}$ test set	1957 reflections (1.36%)	DCC
Wilson B-factor (Å <sup>2</sup> )	28.4	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 35.4	EDS
Estimated twinning fraction	0.019 for -h,-k,l 0.093 for h,-h-k,-l 0.028 for -k,-h,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 152763 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	16061	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: GOL, PG4

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 [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	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:H32	3:A:902:PG4:H81	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%)	46	51
1	C	748/863 (87%)	730 (98%)	17 (2%)	1 (0%)	56	65
2	B	210/248 (85%)	204 (97%)	5 (2%)	1 (0%)	34	34

*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%)	13	7
All	All	1886/2222 (85%)	1840 (98%)	39 (2%)	7 (0%)	39	42

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%)	92	95
1	C	583/767 (76%)	574 (98%)	9 (2%)	72	82
2	B	169/226 (75%)	168 (99%)	1 (1%)	90	94
2	D	168/226 (74%)	167 (99%)	1 (1%)	90	94
All	All	1494/1986 (75%)	1480 (99%)	14 (1%)	84	90

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

Mol	Chain	Res	Type
1	C	241	PHE
1	C	291	ASN
1	C	745	SER
1	C	84	ILE
1	C	617	GLN

Some 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 [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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

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	A	901	-	10,10,12	0.59	0	9,9,11	0.88	0
3	PG4	A	902	-	9,9,12	0.62	0	8,8,11	0.83	0
4	GOL	A	903	-	5,5,5	0.65	0	5,5,5	0.87	0
3	PG4	B	301	-	12,12,12	0.71	0	11,11,11	0.69	0
3	PG4	B	302	-	9,9,12	0.60	0	8,8,11	0.85	0
3	PG4	B	303	-	9,9,12	0.57	0	8,8,11	0.91	0
3	PG4	C	901	-	10,10,12	0.61	0	9,9,11	0.92	0
3	PG4	C	902	-	9,9,12	0.64	0	8,8,11	0.81	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	A	901	-	-	0/8/8/10	0/0/0/0
3	PG4	A	902	-	-	0/7/7/10	0/0/0/0
4	GOL	A	903	-	-	0/4/4/4	0/0/0/0
3	PG4	B	301	-	-	0/10/10/10	0/0/0/0
3	PG4	B	302	-	-	0/7/7/10	0/0/0/0
3	PG4	B	303	-	-	0/7/7/10	0/0/0/0
3	PG4	C	901	-	-	0/8/8/10	0/0/0/0
3	PG4	C	902	-	-	0/7/7/10	0/0/0/0

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.

8 monomers are involved in 46 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	901	PG4	4	0
3	A	902	PG4	8	0
4	A	903	GOL	7	0
3	B	301	PG4	8	0
3	B	302	PG4	7	0
3	B	303	PG4	4	0
3	C	901	PG4	3	0
3	C	902	PG4	5	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	733/863 (84%)	-0.41	5 (0%) 89 89	16, 37, 74, 93	0
1	C	759/863 (87%)	-0.45	0 100 100	17, 38, 64, 83	0
2	B	216/248 (87%)	-0.43	0 100 100	15, 30, 70, 94	1 (0%)
2	D	216/248 (87%)	-0.38	0 100 100	17, 37, 65, 83	0
All	All	1924/2222 (86%)	-0.43	5 (0%) 94 95	15, 36, 69, 94	1 (0%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	57	LEU	4.2
1	A	537	ALA	3.6
1	A	68	GLY	3.5
1	A	545	TRP	2.8
1	A	58	ALA	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 carbohydrates 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. 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
3	PG4	B	302	10/13	0.85	0.39	14.63	51,62,64,65	0
4	GOL	A	903	6/6	0.90	0.24	8.47	28,44,53,53	0
3	PG4	C	902	10/13	0.86	0.25	7.23	71,73,78,81	0
3	PG4	A	902	10/13	0.84	0.25	5.92	64,68,76,80	0
3	PG4	C	901	11/13	0.80	0.16	3.23	42,64,73,74	0
3	PG4	B	303	10/13	0.88	0.15	1.74	50,59,63,70	0
3	PG4	B	301	13/13	0.87	0.18	1.59	36,55,59,60	0
3	PG4	A	901	11/13	0.84	0.14	1.35	49,63,69,70	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.