wwPDB X-ray Structure Validation Summary Report (i)

Nov 21, 2023 – 12:37 pm GMT

PDB ID	:	2WKJ
Title	:	Crystal structure of the E192N mutant of E. Coli N-acetylneuraminic acid lyase in complex with pyruvate at 1.45A resolution in space group P212121
Authors	:	Campeotto, I.; Carr, S.B.; Trinh, C.H.; Nelson, A.S.; Berry, A.; Phillips, S.E.V.; Pearson, A.R.
Deposited on	:	2009-06-11
Resolution	:	1.45 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

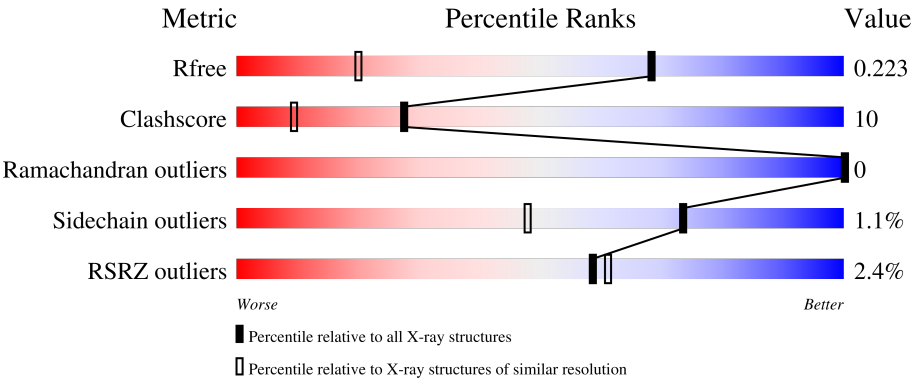
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.45 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}	130704	1156 (1.46-1.46)
Clashscore	141614	1202 (1.46-1.46)
Ramachandran outliers	138981	1178 (1.46-1.46)
Sidechain outliers	138945	1178 (1.46-1.46)
RSRZ outliers	127900	1139 (1.46-1.46)

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	303	3% 88% 10%
1	B	303	% 85% 12%
1	C	303	% 85% 13%
1	D	303	4% 85% 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
2	1PE	A	1297	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 10677 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.

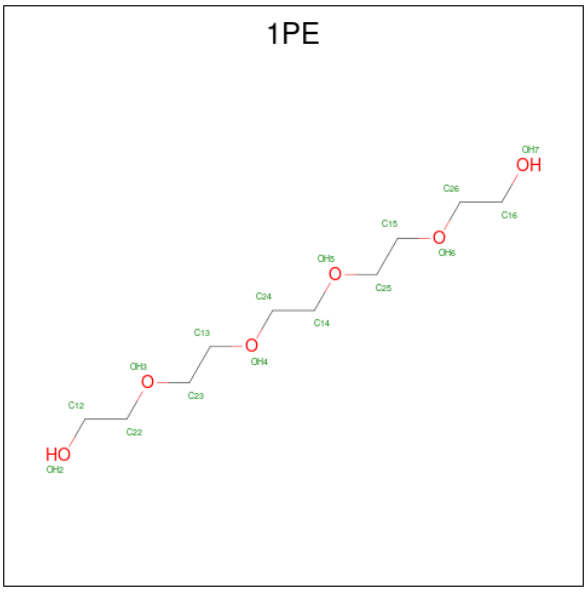
- Molecule 1 is a protein called N-ACETYLNEURAMINATE LYASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	298	Total	C	N	O	S	0	8	0
			2342	1497	394	440	11			
1	B	298	Total	C	N	O	S	0	7	0
			2341	1492	398	440	11			
1	C	298	Total	C	N	O	S	0	5	0
			2329	1485	396	436	12			
1	D	298	Total	C	N	O	S	0	4	0
			2324	1482	394	437	11			

There are 4 discrepancies between the modelled and reference sequences:

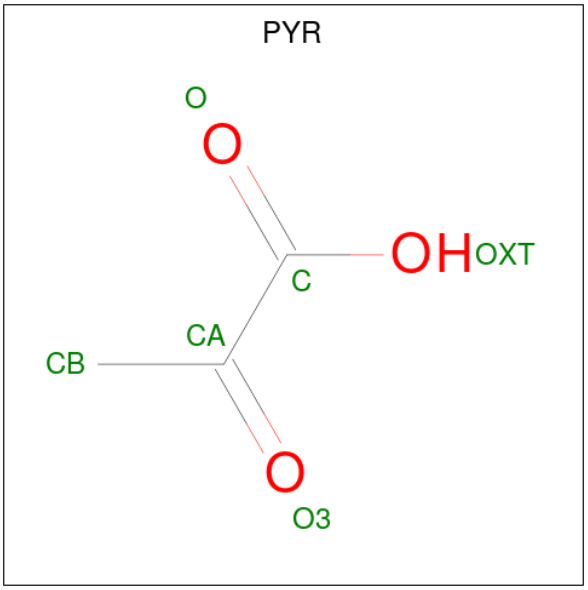
Chain	Residue	Modelled	Actual	Comment	Reference
A	192	ASN	GLU	engineered mutation	UNP P0A6L4
B	192	ASN	GLU	engineered mutation	UNP P0A6L4
C	192	ASN	GLU	engineered mutation	UNP P0A6L4
D	192	ASN	GLU	engineered mutation	UNP P0A6L4

- Molecule 2 is PENTAETHYLENE GLYCOL (three-letter code: 1PE) (formula: $C_{10}H_{22}O_6$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 8 5 3	0	0
2	C	1	Total C O 8 5 3	0	0

- Molecule 3 is PYRUVIC ACID (three-letter code: PYR) (formula: $C_3H_4O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total C O 12 6 6	0	1
3	D	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

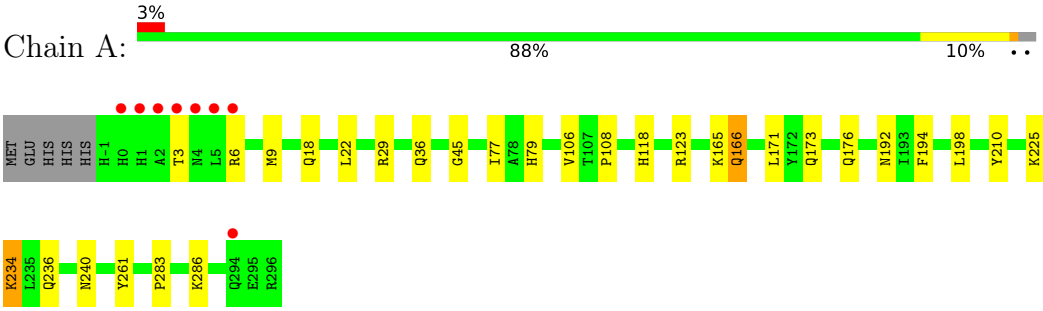
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	345	Total O 345 345	0	0
4	B	347	Total O 347 347	0	0
4	C	340	Total O 340 340	0	0
4	D	275	Total O 275 275	0	0

3 Residue-property plots (i)

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.

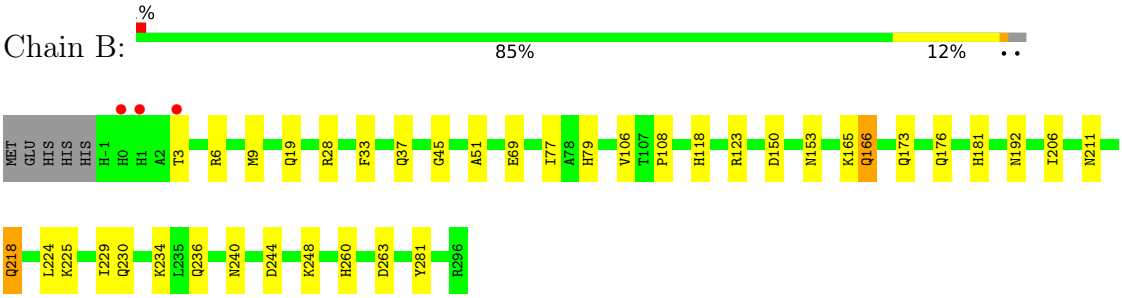
• Molecule 1: N-ACETYLNEURAMINATE LYASE

Chain A:



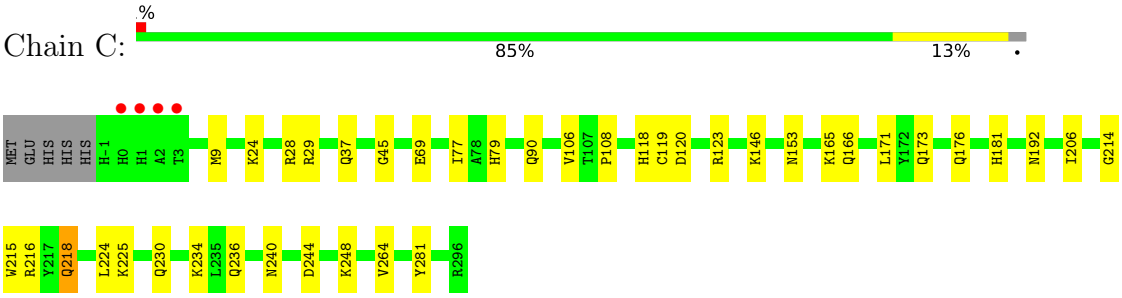
• Molecule 1: N-ACETYLNEURAMINATE LYASE

Chain B:



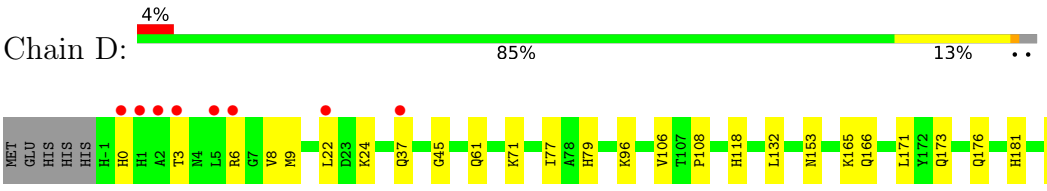
• Molecule 1: N-ACETYLNEURAMINATE LYASE

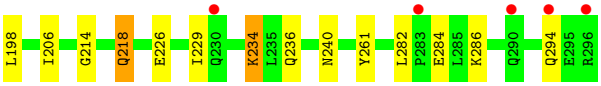
Chain C:



• Molecule 1: N-ACETYLNEURAMINATE LYASE

Chain D:




L198
I206
G214
Q218
E226
I229
Q230
K234
L235
Q236
N240
Y261
L282
P283
E284
L285
K286
Q290
Q294
E295
R296

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.33Å 108.05Å 148.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	87.33 – 1.45 61.14 – 1.45	Depositor EDS
% Data completeness (in resolution range)	95.9 (87.33-1.45) 95.9 (61.14-1.45)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.42 (at 1.45Å)	Xtriage
Refinement program	REFMAC 5.5.0066	Depositor
R, R_{free}	0.167 , 0.189 0.209 , 0.223	Depositor DCC
R_{free} test set	10717 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	13.4	Xtriage
Anisotropy	0.503	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 40.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10677	wwPDB-VP
Average B, all atoms (Å^2)	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 76.71 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.8480e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: PYR, 1PE, KPI

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	0.56	0/2392	0.70	2/3236 (0.1%)
1	B	0.50	0/2388	0.65	3/3231 (0.1%)
1	C	0.54	0/2370	0.70	2/3206 (0.1%)
1	D	0.52	0/2359	0.67	0/3193
All	All	0.53	0/9509	0.68	7/12866 (0.1%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	244	ASP	CB-CG-OD1	6.98	124.58	118.30
1	B	244	ASP	CB-CG-OD1	6.26	123.94	118.30
1	C	244	ASP	CB-CG-OD2	-6.14	112.77	118.30
1	B	244	ASP	CB-CG-OD2	-5.91	112.98	118.30
1	A	210	TYR	CB-CG-CD1	5.58	124.35	121.00

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	2342	0	2376	60	0
1	B	2341	0	2365	34	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2329	0	2354	44	0
1	D	2324	0	2342	56	1
2	A	8	0	9	9	0
2	C	8	0	9	0	0
3	C	12	0	0	0	0
3	D	6	0	0	0	0
4	A	345	0	0	15	0
4	B	347	0	0	8	0
4	C	340	0	0	10	0
4	D	275	0	0	17	0
All	All	10677	0	9455	189	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 10.

The worst 5 of 189 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:C:9[B]:MET:CE	1:C:77:ILE:HD12	1.44	1.46
1:C:9[B]:MET:CE	1:C:77:ILE:CD1	1.95	1.43
1:C:9[B]:MET:HE1	1:C:77:ILE:CD1	1.44	1.43
1:D:9[B]:MET:CE	1:D:77:ILE:HD12	1.63	1.28
1:D:9[B]:MET:HE1	1:D:77:ILE:CD1	1.69	1.22

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)
1:B:123:ARG:NH2	1:D:226:GLU:OE2[4_555]	2.06	0.14

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	303/303 (100%)	298 (98%)	5 (2%)	0	100	100
1	B	302/303 (100%)	296 (98%)	6 (2%)	0	100	100
1	C	300/303 (99%)	294 (98%)	6 (2%)	0	100	100
1	D	299/303 (99%)	293 (98%)	6 (2%)	0	100	100
All	All	1204/1212 (99%)	1181 (98%)	23 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	250/250 (100%)	246 (98%)	4 (2%)	62	31
1	B	249/250 (100%)	246 (99%)	3 (1%)	71	43
1	C	247/250 (99%)	245 (99%)	2 (1%)	81	62
1	D	246/250 (98%)	242 (98%)	4 (2%)	62	31
All	All	992/1000 (99%)	979 (99%)	13 (1%)	73	40

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	166	GLN
1	C	218	GLN
1	D	234	LYS
1	D	218[A]	GLN
1	D	218[B]	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 48 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	166	GLN
1	D	37	GLN
1	C	176	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	240	ASN
1	D	61	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)

4 non-standard protein/DNA/RNA residues 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	$\#\|Z\| > 2$
1	KPI	A	165	1	11,13,14	0.88	0	10,15,17	1.46	1 (10%)
1	KPI	B	165	1	11,13,14	1.01	1 (9%)	10,15,17	1.49	1 (10%)
1	KPI	D	165	1	11,13,14	1.34	1 (9%)	10,15,17	1.06	0
1	KPI	C	165	1	11,13,14	1.00	1 (9%)	10,15,17	1.71	2 (20%)

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
1	KPI	A	165	1	-	0/13/14/16	-
1	KPI	B	165	1	-	0/13/14/16	-
1	KPI	D	165	1	-	0/13/14/16	-
1	KPI	C	165	1	-	0/13/14/16	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	165	KPI	O1-CX2	-2.60	1.23	1.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	165	KPI	C1-CX1	2.06	1.53	1.49
1	B	165	KPI	O1-CX2	-2.02	1.24	1.30

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	165	KPI	O2-CX2-CX1	-3.88	116.43	121.38
1	A	165	KPI	O2-CX2-CX1	-3.85	116.46	121.38
1	C	165	KPI	C1-CX1-CX2	3.78	121.84	118.17
1	C	165	KPI	O2-CX2-CX1	-3.21	117.28	121.38

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

5 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	PYR	D	1297	-	5,5,5	1.09	0	3,6,6	1.01	0
2	1PE	C	1297	-	7,7,15	0.72	0	6,6,14	0.41	0
3	PYR	C	1298[A]	-	5,5,5	1.30	1 (20%)	3,6,6	0.96	0
3	PYR	C	1298[B]	-	5,5,5	1.10	0	3,6,6	0.88	0
2	1PE	A	1297	-	7,7,15	0.43	0	6,6,14	1.16	1 (16%)

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	PYR	D	1297	-	-	0/4/4/4	-
2	1PE	C	1297	-	-	2/5/5/13	-
3	PYR	C	1298[A]	-	-	0/4/4/4	-
3	PYR	C	1298[B]	-	-	0/4/4/4	-
2	1PE	A	1297	-	-	2/5/5/13	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	1298[A]	PYR	CB-CA	2.07	1.54	1.50

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1297	1PE	C23-OH3-C22	-2.21	103.71	113.29

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1297	1PE	OH4-C13-C23-OH3
2	C	1297	1PE	OH2-C12-C22-OH3
2	A	1297	1PE	C12-C22-OH3-C23
2	C	1297	1PE	OH4-C13-C23-OH3

There are no ring outliers.

1 monomer is involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1297	1PE	9	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

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, 95^{th} 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	297/303 (98%)	0.11	8 (2%) 54 56	8, 12, 29, 71	0
1	B	297/303 (98%)	0.05	3 (1%) 82 84	9, 14, 26, 42	0
1	C	297/303 (98%)	-0.08	4 (1%) 77 78	8, 12, 23, 41	0
1	D	297/303 (98%)	0.24	13 (4%) 34 37	9, 14, 32, 75	0
All	All	1188/1212 (98%)	0.08	28 (2%) 59 61	8, 13, 29, 75	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	3	THR	6.2
1	D	1	HIS	6.1
1	A	5	LEU	5.8
1	D	3	THR	5.2
1	D	5	LEU	4.2

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

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, 95^{th} 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
1	KPI	D	165	14/15	0.96	0.09	8,10,13,16	0
1	KPI	B	165	14/15	0.97	0.07	11,12,14,14	0
1	KPI	C	165	14/15	0.97	0.07	8,9,12,13	0
1	KPI	A	165	14/15	0.97	0.08	8,8,11,11	0

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, 95^{th} 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	PYR	C	1298[A]	6/6	0.80	0.27	12,14,15,15	6
3	PYR	C	1298[B]	6/6	0.80	0.27	36,37,37,38	6
2	1PE	A	1297	8/16	0.84	0.19	18,20,22,22	8
3	PYR	D	1297	6/6	0.84	0.21	23,28,28,31	0
2	1PE	C	1297	8/16	0.85	0.23	91,92,92,92	8

6.5 Other polymers (i)

There are no such residues in this entry.