wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 10:18 am BST

PDB ID	:	4W1Y
Title	:	Crystal structure of Escherichia coli Tryptophanase in 'semi-holo' form
Authors	:	Goldgur, Y.
Deposited on	:	2014-08-13
Resolution	:	3.20 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 following versions of software and data (see references (1)) were used in the production of this report:

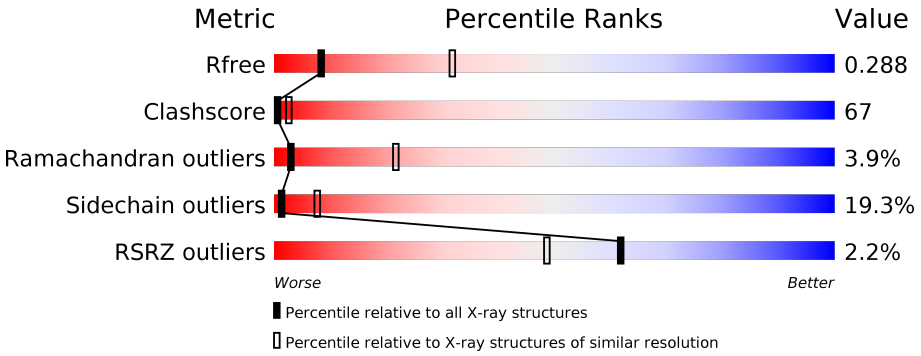
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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	467	2% 30% 49% 14% • 7%
2	B	467	2% 24% 54% 14% • 7%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6957 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.

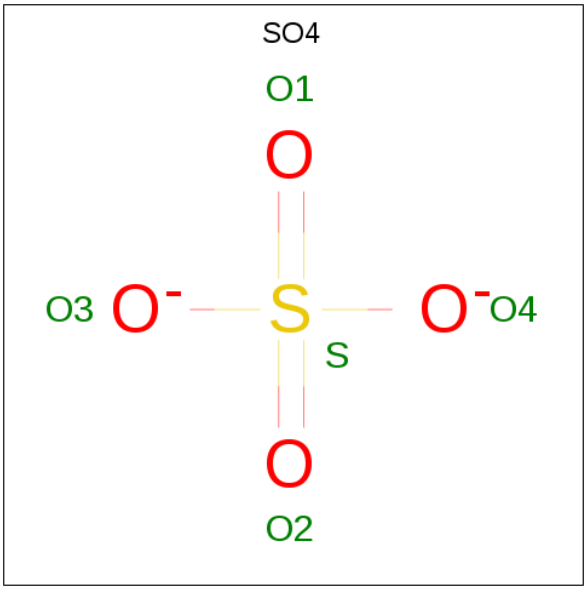
- Molecule 1 is a protein called Tryptophanase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	435	Total	C	N	O	P	S	0	0	0
			3428	2181	575	650	1	21			

- Molecule 2 is a protein called Tryptophanase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	433	Total	C	N	O	S	0	0	0
			3412	2175	575	641	21			

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

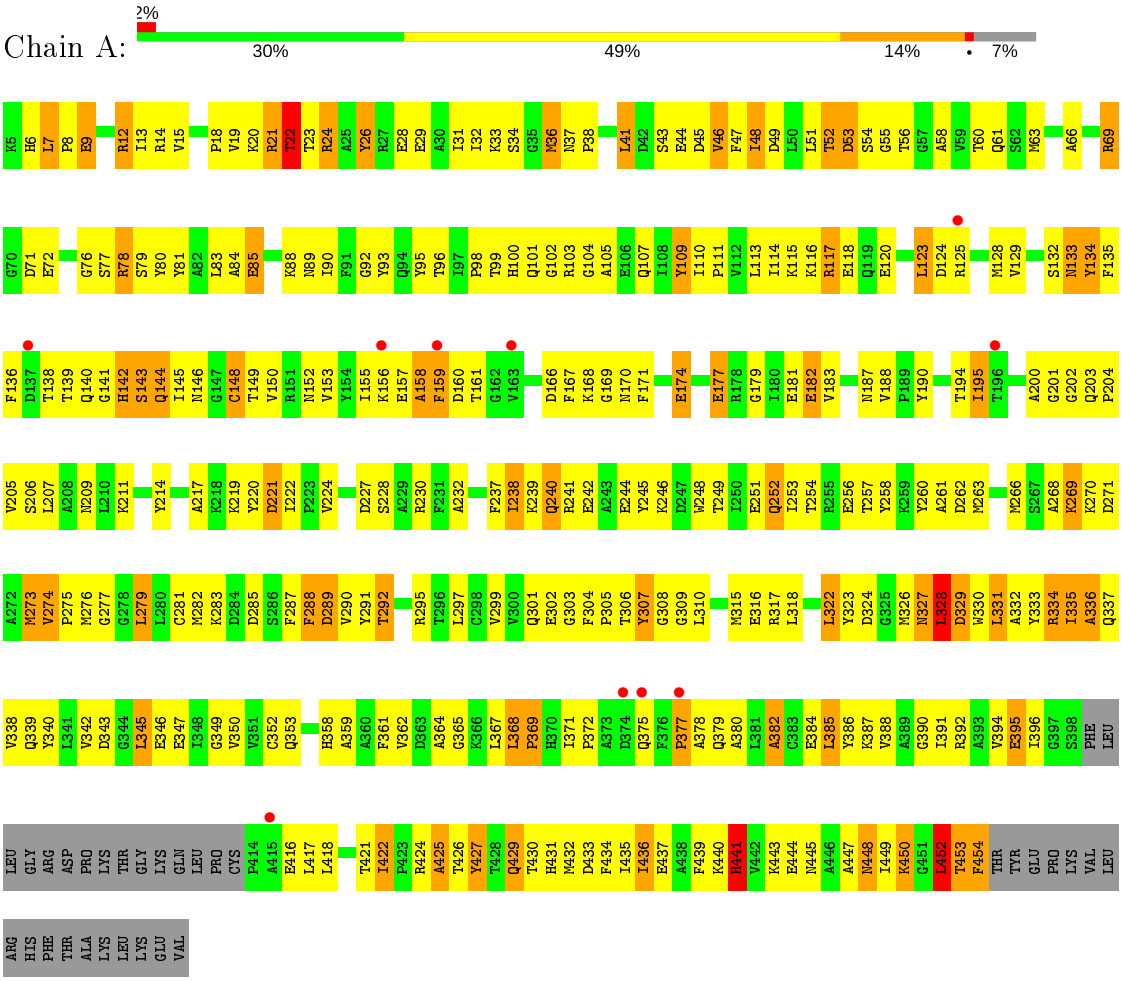
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	42	Total O 42 42	0	0
4	B	65	Total O 65 65	0	0

3 Residue-property plots (i)

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.

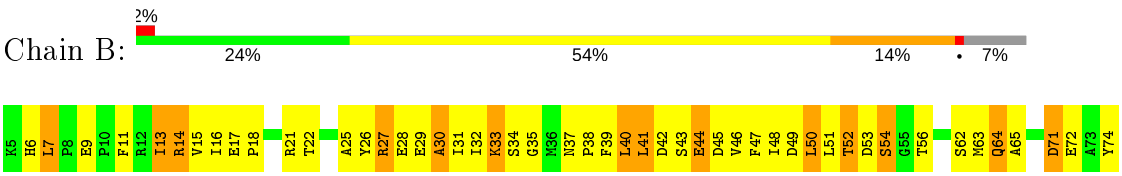
- Molecule 1: Tryptophanase

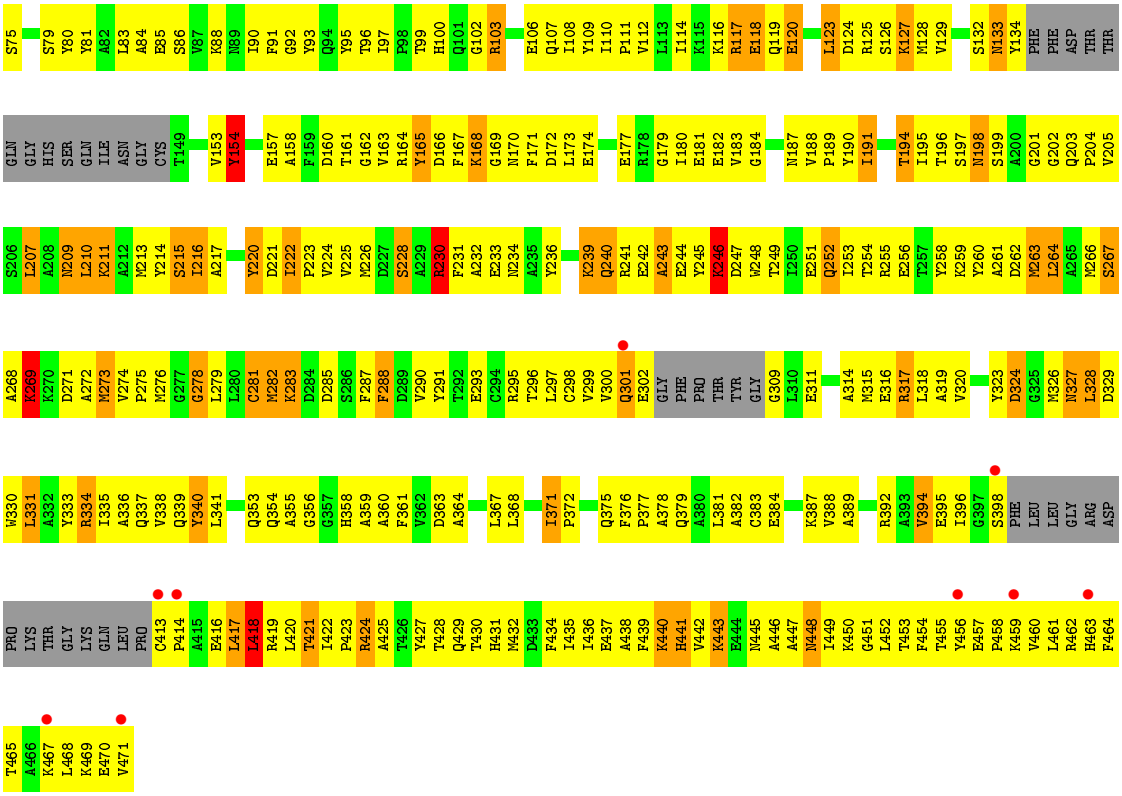
Chain A:



- Molecule 2: Tryptophanase

Chain B:




S75 S79 Y80 Y81 A82 L83 A84 E85 S86 V87 K88 N89 I90 F91 G92 Y93 Q94 Y95 T96 I97 P98 T99 H100 Q101 G102 R103 E106 Q107 I108 Y109 I110 P111 V112 L113 I114 K115 K116 R117 E118 Q119 E120 L123 D124 R125 S126 K127 M128 V129 S132 N133 Y134 PHE PHE ASP THR THR
GLN GLY HIS SER GLN ILE ASN GLY CYS T149 V153 Y154 E157 A158 F159 D160 T161 G162 V163 R164 Y165 D166 F167 K168 G169 N170 F171 D172 L173 E174 E177 R178 G179 I180 E181 E182 V183 G184 N187 V188 P189 Y190 I191 T194 I195 T196 S197 N198 S199 A200 G201 G202 Q203 P204 V205
S206 L207 A208 N209 L210 K211 A212 M213 Y214 S215 I216 A217 Y220 D221 T222 P223 V224 V225 M226 D227 S228 A229 R230 F231 A232 E233 N234 A235 Y236 K239 Q240 R241 E242 A243 E244 Y245 K246 D247 W248 T249 I250 E251 Q252 I253 T254 R255 E256 T257 Y258 K259 Y260 A261 D262 M263 L264 A265 M266 S267
A268 K269 K270 D271 A272 M273 V274 P275 M276 G277 G278 L279 L280 C281 M282 K283 D284 D285 S286 F287 F288 D289 V290 Y291 T292 E293 C294 R295 T296 L297 C298 V299 V300 Q301 E302 GLY PHE PRO THR TYR GLY G309 L310 E311 A314 M315 E316 R317 L318 A319 V320 Y323 D324 G325 M326 N327 L328 D329
W330 L331 A332 Y333 R334 I335 A336 Q337 V338 Q339 Y340 L341 Q353 Q354 A355 G356 G357 H358 A359 A360 F361 V362 D363 A364 L367 L368 I371 P372 Q375 F376 P377 A378 Q379 A380 L381 A382 C383 E384 K387 V388 A389 R392 A393 V394 E395 I396 G397 S398 PHE LEU LEU GLY ARG ASP
PRO LYS THR GLY LYS GLN LEU PRO C413 P414 A415 E416 I417 I418 R419 I420 T421 I422 P423 R424 A425 T426 Y427 T428 Q429 T430 H431 M432 D433 F434 I435 I436 E437 A438 F439 K440 H441 V442 K443 E444 N445 A446 A447 N448 I449 K450 G451 L452 T453 F454 T455 Y456 E457 P458 K459 V460 L461 R462 H463 F464
T465 A466 K467 L468 K469 E470 V471

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	109.97Å 109.97Å 238.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 3.20 29.06 – 3.20	Depositor EDS
% Data completeness (in resolution range)	94.7 (15.00-3.20) 94.9 (29.06-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	2.51 (at 3.18Å)	Xtriage
Refinement program	REFMAC 5.4.0077	Depositor
R, R_{free}	0.214 , 0.290 0.212 , 0.288	Depositor DCC
R_{free} test set	1207 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	86.8	Xtriage
Anisotropy	0.071	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 81.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.53$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6957	wwPDB-VP
Average B, all atoms (Å^2)	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.13% 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: LLP, SO4

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	1.02	2/3473 (0.1%)	1.12	11/4692 (0.2%)
2	B	1.09	4/3478 (0.1%)	1.20	16/4694 (0.3%)
All	All	1.05	6/6951 (0.1%)	1.16	27/9386 (0.3%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
2	B	0	2
All	All	0	3

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	256	GLU	CG-CD	6.80	1.62	1.51
2	B	281	CYS	CB-SG	-6.40	1.71	1.82
1	A	427	TYR	CD1-CE1	6.06	1.48	1.39
2	B	154	TYR	CD2-CE2	-5.92	1.30	1.39
2	B	74	TYR	CE1-CZ	5.05	1.45	1.38

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	266	MET	CB-CG-SD	-7.79	89.05	112.40
2	B	14	ARG	NE-CZ-NH1	-7.49	116.56	120.30
2	B	418	LEU	CA-CB-CG	7.07	131.56	115.30
2	B	7	LEU	CB-CG-CD2	-6.96	99.17	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	282	MET	CG-SD-CE	6.80	111.08	100.20

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	453	THR	Peptide
2	B	162	GLY	Peptide
2	B	309	GLY	Peptide

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	3428	0	3371	418	0
2	B	3412	0	3395	515	1
3	B	10	0	0	1	0
4	A	42	0	0	28	0
4	B	65	0	0	40	0
All	All	6957	0	6766	916	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 67.

The worst 5 of 916 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:22:THR:CG2	2:B:27:ARG:HH11	1.02	1.58
2:B:22:THR:CG2	2:B:27:ARG:HG3	1.39	1.52
2:B:450:LYS:HB2	2:B:470:GLU:CD	1.31	1.46
2:B:375:GLN:HE21	2:B:471:VAL:CG1	1.26	1.46
1:A:170:ASN:CB	1:A:209:ASN:ND2	1.79	1.45

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)
2:B:103:ARG:NH1	2:B:301:GLN:O[8_665]	2.16	0.04

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	430/467 (92%)	331 (77%)	81 (19%)	18 (4%)	3 20
2	B	425/467 (91%)	327 (77%)	83 (20%)	15 (4%)	3 24
All	All	855/934 (92%)	658 (77%)	164 (19%)	33 (4%)	3 22

5 of 33 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	328	LEU
2	B	246	LYS
1	A	157	GLU
1	A	238	ILE
2	B	243	ALA

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	356/385 (92%)	286 (80%)	70 (20%)	1 7
2	B	358/386 (93%)	290 (81%)	68 (19%)	1 8
All	All	714/771 (93%)	576 (81%)	138 (19%)	1 8

5 of 138 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	430	THR
2	B	50	LEU
2	B	371	ILE
1	A	441	HIS
2	B	13	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	327	ASN
1	A	445	ASN
2	B	133	ASN
1	A	209	ASN
2	B	170	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is 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	LLP	A	270	1	23,24,25	2.17	6 (26%)	25,32,34	2.43	7 (28%)

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	LLP	A	270	1	-	4/16/17/19	0/1/1/1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	270	LLP	O3-C3	-6.59	1.21	1.37
1	A	270	LLP	C4-C4'	3.71	1.53	1.46
1	A	270	LLP	C2-N1	2.99	1.39	1.33
1	A	270	LLP	C6-N1	2.81	1.40	1.34
1	A	270	LLP	C4'-NZ	2.63	1.36	1.27

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	270	LLP	OP4-C5'-C5	7.15	122.98	109.35
1	A	270	LLP	C4-C3-C2	6.41	124.16	120.19
1	A	270	LLP	CE-NZ-C4'	-3.67	107.62	118.90
1	A	270	LLP	OP3-P-OP1	3.06	122.68	110.68
1	A	270	LLP	C3-C2-N1	-2.44	117.62	120.77

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	270	LLP	C4-C5-C5'-OP4
1	A	270	LLP	C6-C5-C5'-OP4
1	A	270	LLP	C4-C4'-NZ-CE
1	A	270	LLP	CA-CB-CG-CD

There are no ring outliers.

1 monomer is involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	270	LLP	6	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

2 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	SO4	B	500	-	4,4,4	0.14	0	6,6,6	0.60	0
3	SO4	B	501	-	4,4,4	0.17	0	6,6,6	0.20	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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	501	SO4	1	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	434/467 (92%)	-0.32	10 (2%) 60 47	50, 82, 120, 126	0
2	B	433/467 (92%)	-0.27	9 (2%) 63 49	48, 82, 129, 170	0
All	All	867/934 (92%)	-0.29	19 (2%) 62 48	48, 82, 121, 170	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	471	VAL	5.9
2	B	414	PRO	4.6
2	B	413	CYS	3.8
1	A	415	ALA	3.2
2	B	398	SER	3.1

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	LLP	A	270	24/25	0.97	0.18	63,73,78,80	0

6.3 Carbohydrates

There are no carbohydrates 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, 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	SO4	B	501	5/5	0.81	0.26	184,184,184,185	0
3	SO4	B	500	5/5	0.92	0.15	110,110,111,112	0

6.5 Other polymers

There are no such residues in this entry.