



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 13, 2023 – 11:24 PM JST

PDB ID : 5YHV
Title : Crystal structure of an aminotransferase from Mycobacterium tuberculosis
Authors : Saroj, D.C.; Biswal, B.K.
Deposited on : 2017-09-30
Resolution : 2.70 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 ⓘ 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 ⓘ](#)) 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 : 2.36
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

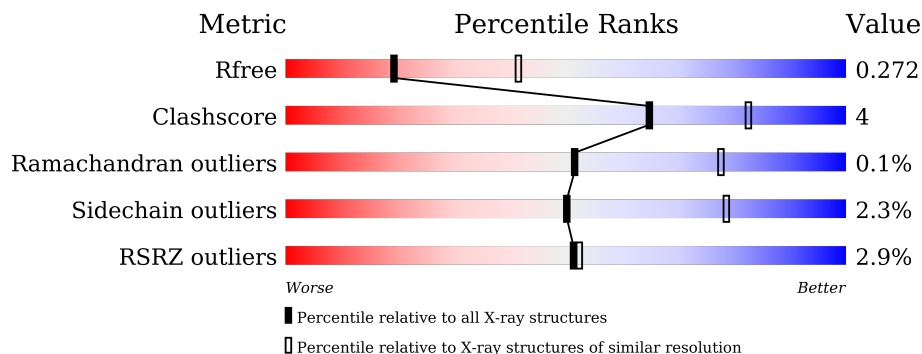
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 $\leq 5\%$. 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	394	
1	B	394	
1	C	394	
1	D	394	

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
3	GOL	B	402	-	-	-	X
4	PLP	A	403	-	-	X	-
4	PLP	D	401	-	-	X	-

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 11078 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 Aminotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	394	2894	1837	499	545	13	0	0	0
1	B	387	2826	1788	484	542	12	0	0	0
1	C	384	2707	1708	467	520	12	0	0	0
1	D	364	2342	1464	411	455	12	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

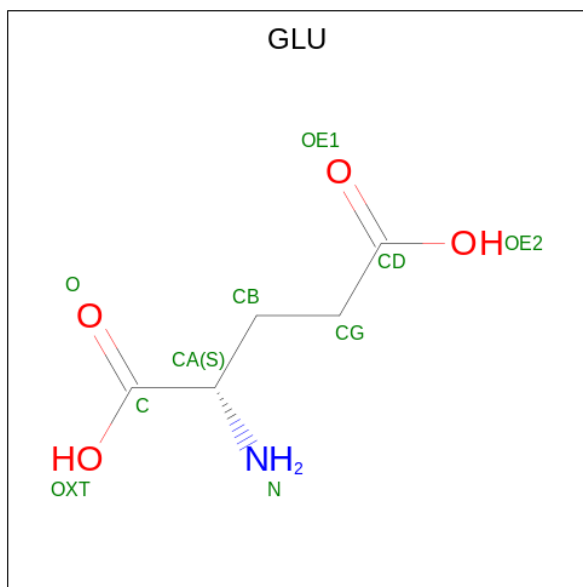
Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	expression tag	UNP P96847
A	-4	HIS	-	expression tag	UNP P96847
A	-3	HIS	-	expression tag	UNP P96847
A	-2	HIS	-	expression tag	UNP P96847
A	-1	HIS	-	expression tag	UNP P96847
A	0	HIS	-	expression tag	UNP P96847
B	-5	HIS	-	expression tag	UNP P96847
B	-4	HIS	-	expression tag	UNP P96847
B	-3	HIS	-	expression tag	UNP P96847
B	-2	HIS	-	expression tag	UNP P96847
B	-1	HIS	-	expression tag	UNP P96847
B	0	HIS	-	expression tag	UNP P96847
C	-5	HIS	-	expression tag	UNP P96847
C	-4	HIS	-	expression tag	UNP P96847
C	-3	HIS	-	expression tag	UNP P96847
C	-2	HIS	-	expression tag	UNP P96847
C	-1	HIS	-	expression tag	UNP P96847
C	0	HIS	-	expression tag	UNP P96847
D	-5	HIS	-	expression tag	UNP P96847
D	-4	HIS	-	expression tag	UNP P96847
D	-3	HIS	-	expression tag	UNP P96847

Continued on next page...

Continued from previous page...

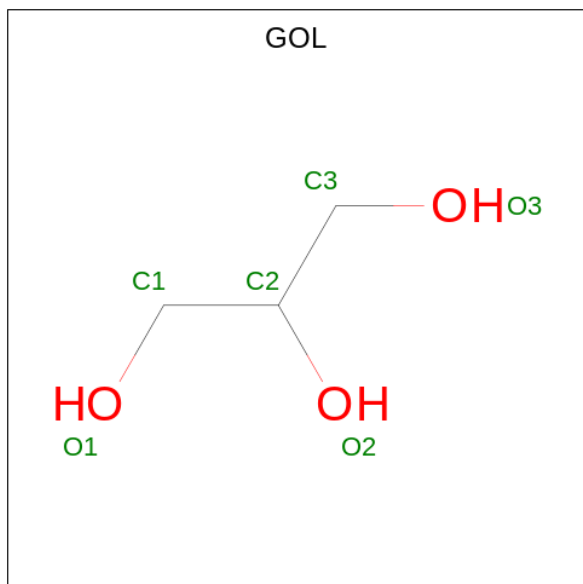
Chain	Residue	Modelled	Actual	Comment	Reference
D	-2	HIS	-	expression tag	UNP P96847
D	-1	HIS	-	expression tag	UNP P96847
D	0	HIS	-	expression tag	UNP P96847

- Molecule 2 is GLUTAMIC ACID (three-letter code: GLU) (formula: C₅H₉NO₄).



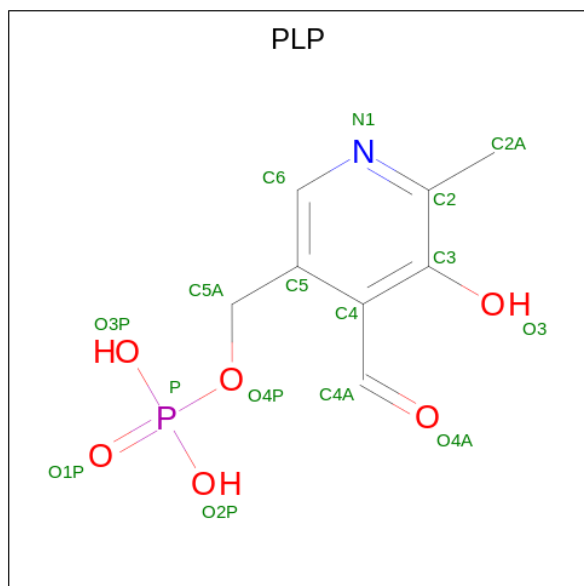
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	10	5	1	4	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



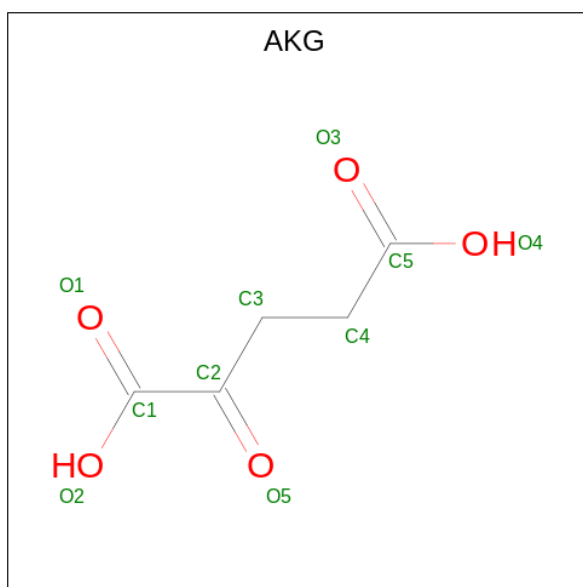
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		
3	C	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is PYRIDOXAL-5'-PHOSPHATE (three-letter code: PLP) (formula: C₈H₁₀NO₆P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			15	8	1	5	1		
4	B	1	Total	C	N	O	P	0	0
			15	8	1	5	1		
4	C	1	Total	C	N	O	P	0	0
			15	8	1	5	1		
4	D	1	Total	C	N	O	P	0	0
			15	8	1	5	1		

- Molecule 5 is 2-OXOGLUTARIC ACID (three-letter code: AKG) (formula: C₅H₆O₅).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total C O 10 5 5	0	0
5	C	1	Total C O 10 5 5	0	0

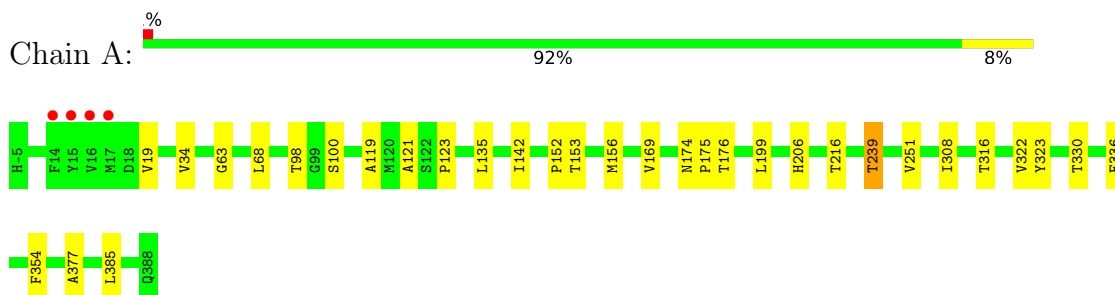
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	79	Total O 79 79	0	0
6	B	73	Total O 73 73	0	0
6	C	34	Total O 34 34	0	0
6	D	15	Total O 15 15	0	0

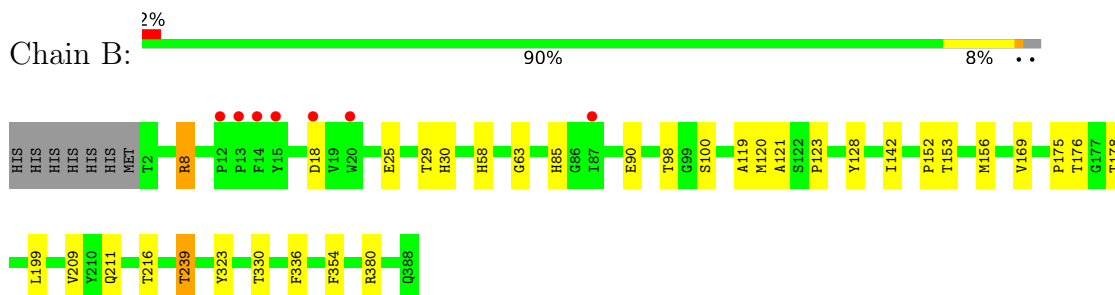
3 Residue-property plots [i](#)

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.

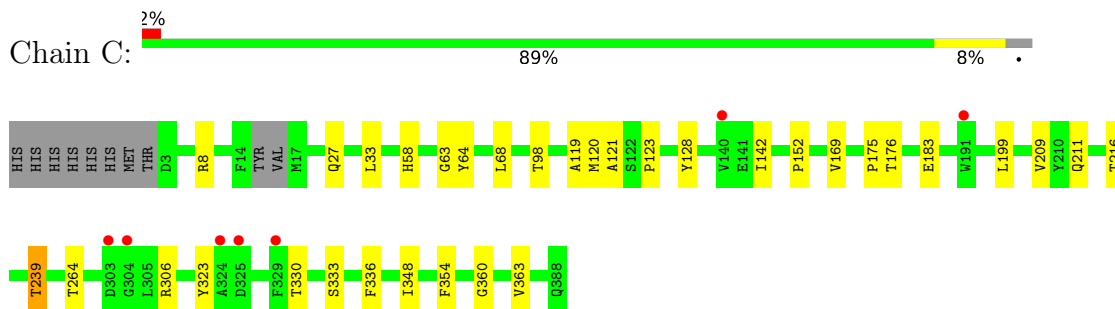
- Molecule 1: Aminotransferase



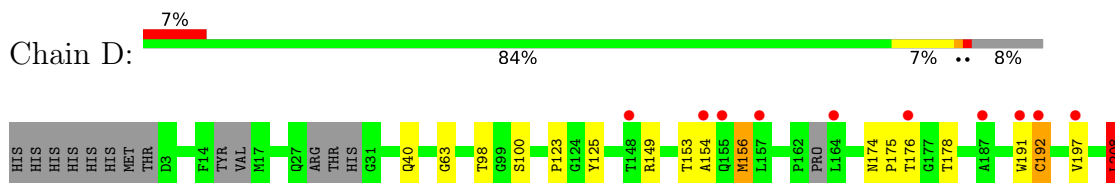
- Molecule 1: Aminotransferase

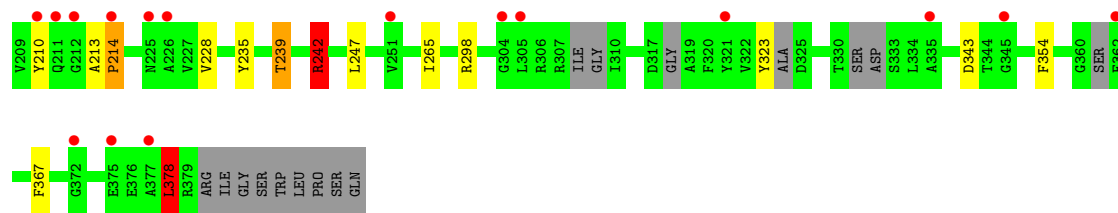


- Molecule 1: Aminotransferase



- Molecule 1: Aminotransferase





4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	93.28Å 98.22Å 198.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	99.39 – 2.70 47.33 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.3 (99.39-2.70) 99.3 (47.33-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.59 (at 2.69Å)	Xtrriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.236 , 0.272 0.235 , 0.272	Depositor DCC
R_{free} test set	2563 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	49.4	Xtrriage
Anisotropy	0.639	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 36.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.013 for k,h,-l	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11078	wwPDB-VP
Average B, all atoms (Å ²)	56.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.88% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: PLP, GOL, AKG

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	0.41	0/2968	0.64	0/4070
1	B	0.41	0/2894	0.65	1/3964 (0.0%)
1	C	0.39	0/2769	0.63	0/3797
1	D	0.45	0/2393	0.68	5/3264 (0.2%)
All	All	0.41	0/11024	0.65	6/15095 (0.0%)

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	D	0	2

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	242	ARG	NE-CZ-NH1	-8.02	116.29	120.30
1	D	208	LEU	CB-CG-CD1	6.60	122.22	111.00
1	D	214	PRO	CA-N-CD	-5.76	103.43	111.50
1	D	242	ARG	NE-CZ-NH2	5.75	123.18	120.30
1	B	380	ARG	NE-CZ-NH2	5.54	123.07	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	149	ARG	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	D	343	ASP	Mainchain

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	2894	0	2786	25	0
1	B	2826	0	2717	20	0
1	C	2707	0	2506	22	0
1	D	2342	0	1894	31	0
2	A	10	0	5	3	0
3	A	6	0	8	0	0
3	B	6	0	8	0	0
3	C	6	0	8	0	0
4	A	15	0	6	6	0
4	B	15	0	6	1	0
4	C	15	0	6	1	0
4	D	15	0	6	10	0
5	B	10	0	4	1	0
5	C	10	0	4	0	0
6	A	79	0	0	1	0
6	B	73	0	0	0	0
6	C	34	0	0	0	0
6	D	15	0	0	0	0
All	All	11078	0	9964	86	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 4.

The worst 5 of 86 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:D:192:CYS:SG	1:D:197:VAL:O	1.95	1.23
1:D:100:SER:OG	4:D:401:PLP:H5A1	1.59	1.02
1:A:206:HIS:HB3	2:A:401:GLU:HA	1.41	1.00
1:D:100:SER:OG	4:D:401:PLP:C5A	2.16	0.93

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:100:SER:HG	4:D:401:PLP:H5A1	1.33	0.90

There are no symmetry-related clashes.

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	392/394 (100%)	384 (98%)	8 (2%)	0	100	100
1	B	385/394 (98%)	378 (98%)	7 (2%)	0	100	100
1	C	380/394 (96%)	372 (98%)	8 (2%)	0	100	100
1	D	346/394 (88%)	334 (96%)	10 (3%)	2 (1%)	25	50
All	All	1503/1576 (95%)	1468 (98%)	33 (2%)	2 (0%)	51	78

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	214	PRO
1	D	213	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	291/306 (95%)	287 (99%)	4 (1%)	67	86

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	286/306 (94%)	280 (98%)	6 (2%)	53	80
1	C	252/306 (82%)	247 (98%)	5 (2%)	55	81
1	D	178/306 (58%)	171 (96%)	7 (4%)	32	61
All	All	1007/1224 (82%)	985 (98%)	22 (2%)	50	79

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	306	ARG
1	D	192	CYS
1	D	178	THR
1	D	208	LEU
1	B	90	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	30	HIS
1	B	58	HIS
1	B	155	GLN
1	D	40	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

10 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
4	PLP	B	401	1	15,15,16	2.91	3 (20%)	20,22,23	1.87	7 (35%)
3	GOL	A	402	-	5,5,5	0.42	0	5,5,5	0.41	0
4	PLP	C	403	1	15,15,16	3.28	3 (20%)	20,22,23	1.69	5 (25%)
5	AKG	C	402	-	9,9,9	1.94	1 (11%)	11,11,11	1.59	2 (18%)
4	PLP	D	401	1	15,15,16	3.14	3 (20%)	20,22,23	1.74	5 (25%)
3	GOL	C	401	-	5,5,5	0.34	0	5,5,5	0.17	0
4	PLP	A	403	1	15,15,16	3.53	3 (20%)	20,22,23	2.53	7 (35%)
2	GLU	A	401	-	8,9,9	1.06	1 (12%)	10,11,11	1.26	2 (20%)
5	AKG	B	403	-	9,9,9	2.05	1 (11%)	11,11,11	1.26	1 (9%)
3	GOL	B	402	-	5,5,5	0.35	0	5,5,5	0.23	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
4	PLP	B	401	1	-	0/6/6/8	0/1/1/1
3	GOL	A	402	-	-	1/4/4/4	-
4	PLP	C	403	1	-	1/6/6/8	0/1/1/1
5	AKG	C	402	-	-	4/9/9/9	-
4	PLP	D	401	1	-	5/6/6/8	0/1/1/1
3	GOL	C	401	-	-	2/4/4/4	-
4	PLP	A	403	1	-	2/6/6/8	0/1/1/1
2	GLU	A	401	-	-	5/9/9/9	-
5	AKG	B	403	-	-	5/9/9/9	-
3	GOL	B	402	-	-	2/4/4/4	-

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	403	PLP	C3-C2	9.21	1.50	1.40
4	A	403	PLP	C5-C4	8.78	1.50	1.40
4	C	403	PLP	C3-C2	8.69	1.49	1.40
4	D	401	PLP	C3-C2	8.36	1.49	1.40
4	C	403	PLP	C5-C4	7.95	1.49	1.40

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	403	PLP	O4P-C5A-C5	6.95	122.59	109.35
4	A	403	PLP	C5A-C5-C6	-5.22	110.79	119.37
4	D	401	PLP	O4P-C5A-C5	3.92	116.81	109.35
4	B	401	PLP	C5A-C5-C6	3.62	125.33	119.37
4	A	403	PLP	C4A-C4-C5	3.49	124.53	120.94

There are no chirality outliers.

5 of 27 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	401	GLU	C-CA-CB-CG
3	B	402	GOL	O1-C1-C2-O2
3	B	402	GOL	O1-C1-C2-C3
3	C	401	GOL	O1-C1-C2-C3
4	A	403	PLP	C4-C5-C5A-O4P

There are no ring outliers.

6 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	401	PLP	1	0
4	C	403	PLP	1	0
4	D	401	PLP	10	0
4	A	403	PLP	6	0
2	A	401	GLU	3	0
5	B	403	AKG	1	0

5.7 Other polymers [i](#)

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 [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, 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	394/394 (100%)	-0.08	4 (1%) 82 83	28, 43, 70, 97	0
1	B	387/394 (98%)	0.06	7 (1%) 68 70	29, 42, 72, 118	0
1	C	384/394 (97%)	0.15	7 (1%) 68 70	33, 63, 95, 116	0
1	D	364/394 (92%)	0.58	26 (7%) 16 14	36, 84, 112, 125	0
All	All	1529/1576 (97%)	0.17	44 (2%) 51 52	28, 49, 103, 125	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	13	PRO	5.2
1	D	375	GLU	3.9
1	C	329	PHE	3.9
1	A	16	VAL	3.9
1	D	225	ASN	3.8

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 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, 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(\AA^2)	Q<0.9
3	GOL	C	401	6/6	0.73	0.34	61,63,65,66	0
3	GOL	B	402	6/6	0.78	0.41	63,65,65,66	0
5	AKG	B	403	10/10	0.78	0.33	70,72,73,74	0
3	GOL	A	402	6/6	0.85	0.21	51,56,58,58	0
2	GLU	A	401	10/10	0.85	0.21	60,64,71,72	0
5	AKG	C	402	10/10	0.88	0.22	67,73,78,81	0
4	PLP	D	401	15/16	0.91	0.27	59,76,79,82	0
4	PLP	C	403	15/16	0.95	0.21	50,59,63,65	0
4	PLP	B	401	15/16	0.95	0.21	43,54,57,58	1
4	PLP	A	403	15/16	0.96	0.18	40,50,55,56	0

6.5 Other polymers [i](#)

There are no such residues in this entry.