W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Dec 3, 2023 – 08:38 pm GMT

PDB ID	:	2BHT
Title	:	Crystal structure of O-acetylserine sulfhydrylase B
Authors	:	Claus, M.T.; Zocher, G.E.; Maier, T.H.P.; Schulz, G.E.
Deposited on	:	2005-01-18
Resolution	:	2.10 Å(reported)

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

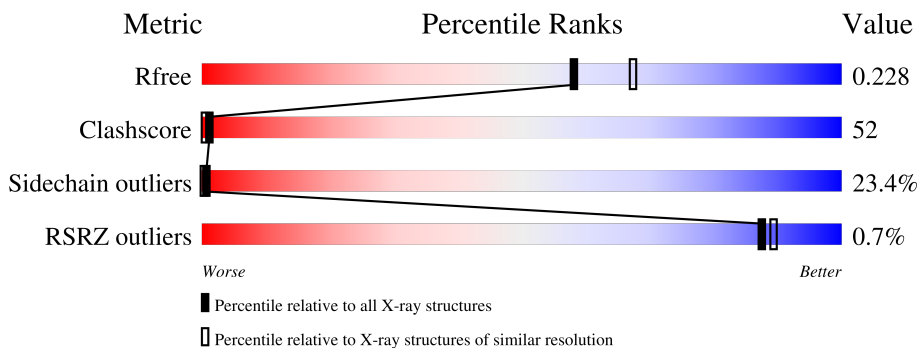|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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		0.228
Clashscore		52
Sidechain outliers		23.4%
RSRZ outliers		0.7%

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$_{free}$	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	31% 51% 15% •
1	B	303	% 25% 52% 14% • 7%
1	C	303	30% 55% 11% •
1	D	303	% 37% 42% 15% • 6%

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 8978 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 CYSTEINE SYNTHASE B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	294	Total	C	N	O	S	0	0	0
			2217	1375	398	429	15			
1	B	281	Total	C	N	O	S	0	0	0
			2122	1315	384	410	13			
1	C	292	Total	C	N	O	S	0	0	0
			2204	1368	396	425	15			
1	D	285	Total	C	N	O	S	0	0	0
			2152	1336	387	415	14			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	57	ARG	GLU	engineered mutation	UNP P16703
A	148	LYS	TYR	engineered mutation	UNP P16703
A	184	GLU	ARG	engineered mutation	UNP P16703
B	57	ARG	GLU	engineered mutation	UNP P16703
B	148	LYS	TYR	engineered mutation	UNP P16703
B	184	GLU	ARG	engineered mutation	UNP P16703
C	57	ARG	GLU	engineered mutation	UNP P16703
C	148	LYS	TYR	engineered mutation	UNP P16703
C	184	GLU	ARG	engineered mutation	UNP P16703
D	57	ARG	GLU	engineered mutation	UNP P16703
D	148	LYS	TYR	engineered mutation	UNP P16703
D	184	GLU	ARG	engineered mutation	UNP P16703

- Molecule 2 is PYRIDOXAL-5'-PHOSPHATE (three-letter code: PLP) (formula: $C_8H_{10}NO_6P$).

PLP

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

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	56	Total	O	0	0
			56	56		
3	B	49	Total	O	0	0
			49	49		
3	C	56	Total	O	0	0
			56	56		
3	D	62	Total	O	0	0
			62	62		

3 Residue-property plots ⓘ

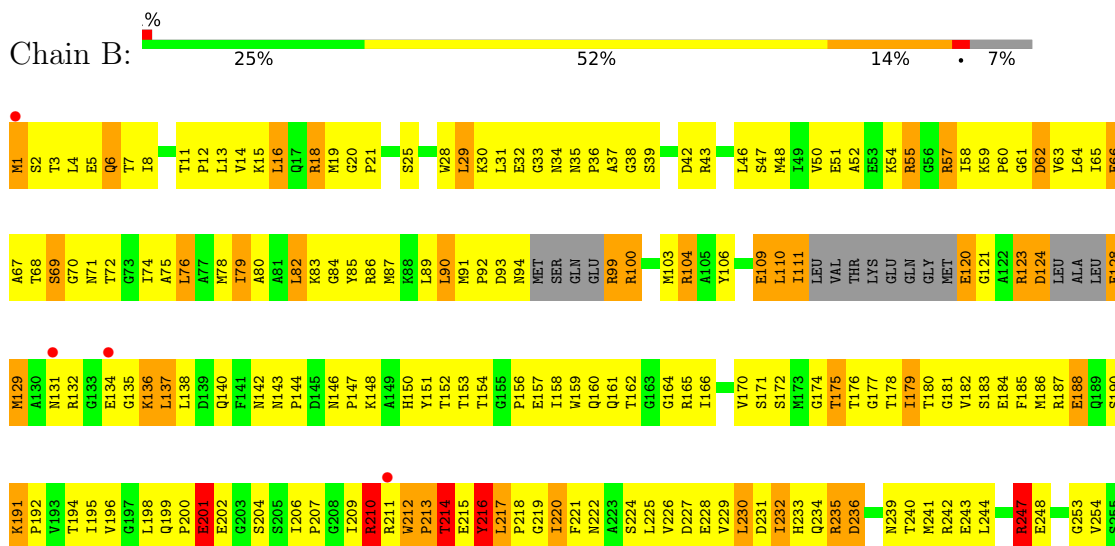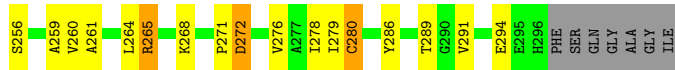These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.
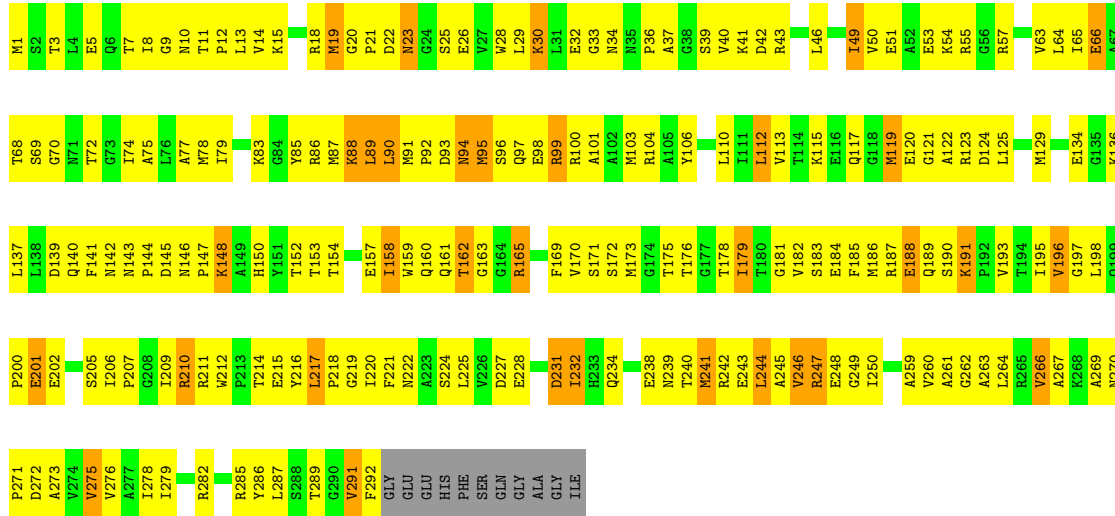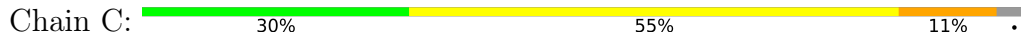
- Molecule 1: CYSTEINE SYNTHASE B

Chain A:

31%   51%   15%

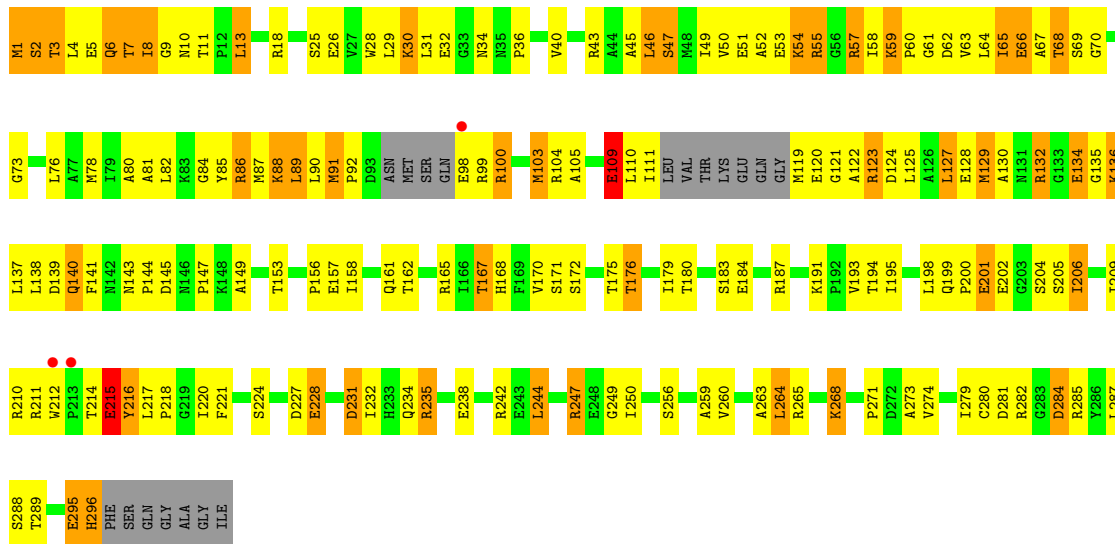- Molecule 1: CYSTEINE SYNTHASE B

Chain B:

25%   52%   14%   7%

S256 A259 V260 A261 L264 R265 K268 P271 D272 V276 A277 I278 I279 C280 Y286 T289 G290 V291 E294 E295 H296 PHE SER GLN GLY ALA GLY ILE

• Molecule 1: CYSTEINE SYNTHASE B

Chain C:

30%   55%   11%   •

M1 S2 T3 L4 E5 Q6 T7 I8 G9 N10 T11 P12 L13 V14 K15 R18 M19 G20 P21 D22 N23 G24 S25 E26 V27 W28 L29 K30 L31 E32 G33 N34 N35 P36 A37 G38 S39 V40 K41 D42 R43 L46 I49 V50 E51 A52 E53 K54 R55 G56 R57 V63 L64 I65 E66 A67

T68 S69 G70 N71 T72 G73 I74 A75 L76 A77 M78 I79 K83 G84 Y85 R86 M87 K88 L89 L90 M91 P92 D93 N94 M95 S96 Q97 E98 R99 R100 A101 A102 M103 R104 A105 Y106 L110 I111 L112 V113 T114 K115 E116 Q117 G118 M119 E120 G121 A122 R123 D124 L125 M129 E134 G135 K136

L137 L138 D139 Q140 F141 N142 N143 P144 D145 N146 P147 K148 A149 H150 Y151 T152 T153 T154 E157 I158 W159 Q160 Q161 T162 G163 G164 R165 F169 V170 S171 S172 M173 G174 T175 T176 G177 T178 I179 T180 V182 S183 E184 F185 M186 R187 E188 Q189 S190 K191 P192 V193 T194 I195 V196 G197 L198 Q199

P200 E201 E202 S205 I206 P207 G208 I209 R210 R211 W212 P213 T214 E215 Y216 L217 P218 G219 I220 F221 N222 A223 S224 L225 V226 D227 E228 D231 I232 H233 Q234 E238 N239 T240 M241 R242 E243 L244 A245 V246 R247 E248 G249 I250 A259 V260 A261 G262 A263 L264 R265 V266 A267 K268 A269 N270

P271 D272 A273 V274 V275 V276 A277 I278 I279 R282 R285 Y286 L287 S288 T289 G290 V291 F292 GLY GLU GLU HIS PHE SER GLN GLY ALA GLY ILE

• Molecule 1: CYSTEINE SYNTHASE B

Chain D:

%

37%   42%   15%   • 6%

M1 S2 T3 L4 E5 Q6 T7 I8 G9 N10 T11 P12 L13 R18 S25 E26 V27 W28 L29 K30 L31 E32 G33 N34 N35 P36 V40 R43 A44 A45 L46 S47 M48 I49 V50 E51 A52 E53 K54 R55 G56 R57 I58 K59 P60 G61 D62 V63 L64 I65 E66 A67 T68 S69 G70

G73 L76 A77 M78 I79 A80 A81 L82 K83 G84 Y85 R86 M87 K88 L89 L90 M91 P92 D93 ASN MET SER GLN E98 R99 R100 M103 R104 A105 E109 L110 I111 LEU VAL THR LYS GLU GLN GLY M119 E120 G121 A122 R123 D124 L125 A126 L127 E128 M129 A130 N131 R132 G133 E134 G135 K136

L137 L138 D139 Q140 F141 N142 N143 P144 D145 N146 P147 K148 A149 T153 P156 E157 I158 Q161 T162 R165 I166 T167 H168 F169 V170 S171 S172 T175 T176 I179 T180 S183 E184 R187 K191 P192 V193 T194 I195 L198 Q199 P200 E201 E202 G203 S204 S205 I206 I209

R210 R211 W212 P213 T214 E215 Y216 L217 P218 G219 I220 F221 S224 D227 E228 D231 I232 H233 Q234 R235 E238 R242 E243 L244 R247 E248 G249 I250 S256 A259 V260 A263 L264 R265 K268 P271 D272 A273 V274 I279 C280 D281 R282 G283 D284 R285 Y286 L287

S288 T289 E295 H296 PHE SER GLN GLY ALA GLY ILE

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	I 41	Depositor
Cell constants a, b, c, α, β, γ	149.94Å 149.94Å 194.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.10 74.97 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (20.00-2.10) 99.6 (74.97-2.10)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 2.10Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.225 , 0.244 0.235 , 0.228	Depositor DCC
R_{free} test set	4937 reflections (1.03%)	wwPDB-VP
Wilson B-factor (Å^2)	30.4	Xtriage
Anisotropy	0.474	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 74.0	EDS
L-test for twinning[2]	$< |L| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.419 for -h,k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	8978	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.93% 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: PLP

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	0.92	1/2250 (0.0%)	0.68	2/3039 (0.1%)
1	B	0.47	2/2153 (0.1%)	0.67	4/2905 (0.1%)
1	C	0.23	0/2237	0.55	0/3022
1	D	0.24	0/2184	0.58	0/2948
All	All	0.54	3/8824 (0.0%)	0.62	6/11914 (0.1%)

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	3
1	B	0	4
1	C	0	2
1	D	0	4
All	All	0	13

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	292	PHE	C-O	41.94	2.03	1.23
1	B	210	ARG	CZ-NH1	13.48	1.50	1.33
1	B	210	ARG	NE-CZ	12.76	1.49	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	292	PHE	O-C-N	-18.34	92.02	123.20
1	B	210	ARG	NE-CZ-NH1	13.94	127.27	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	292	PHE	CA-C-O	-9.78	99.56	120.10
1	B	247	ARG	CD-NE-CZ	7.70	134.38	123.60
1	B	210	ARG	NE-CZ-NH2	-7.42	116.59	120.30

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1	MET	Peptide
1	A	292	PHE	Mainchain
1	A	41	LYS	Mainchain
1	B	201	GLU	Peptide
1	B	213	PRO	Peptide

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	2217	0	2228	262	0
1	B	2122	0	2116	256	0
1	C	2204	0	2220	191	0
1	D	2152	0	2154	220	0
2	A	15	0	6	2	0
2	B	15	0	6	4	0
2	C	15	0	6	2	0
2	D	15	0	6	1	0
3	A	56	0	0	1	0
3	B	49	0	0	4	0
3	C	56	0	0	10	0
3	D	62	0	0	6	0
All	All	8978	0	8742	909	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 52.

The worst 5 of 909 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:110:LEU:O	1:D:111:ILE:HD13	1.44	1.15
1:B:152:THR:HG23	1:B:153:THR:HG23	1.38	1.04
1:D:99:ARG:NH2	1:D:288:SER:OG	1.94	1.00
1:A:168:HIS:HB2	1:A:275:VAL:HG22	1.46	0.97
1:A:292:PHE:C	1:A:292:PHE:O	2.03	0.97

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

There are no protein backbone outliers to report in this entry.

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	234/240 (98%)	184 (79%)	50 (21%)	1	0
1	B	223/240 (93%)	168 (75%)	55 (25%)	0	0
1	C	233/240 (97%)	181 (78%)	52 (22%)	1	0
1	D	226/240 (94%)	169 (75%)	57 (25%)	0	0
All	All	916/960 (95%)	702 (77%)	214 (23%)	1	0

5 of 214 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	69	SER
1	C	205	SER
1	D	215	GLU
1	C	90	LEU
1	C	137	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20

such sidechains are listed below:

Mol	Chain	Res	Type
1	C	97	GLN
1	C	270	ASN
1	D	296	HIS
1	D	6	GLN
1	B	94	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

4 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	$\#|Z| > 2$
2	PLP	A	320	1	15,15,16	2.05	4 (26%)	20,22,23	1.86	4 (20%)
2	PLP	D	320	1	15,15,16	2.09	4 (26%)	20,22,23	1.85	4 (20%)
2	PLP	C	320	1	15,15,16	2.10	4 (26%)	20,22,23	1.72	3 (15%)
2	PLP	B	320	1	15,15,16	2.07	4 (26%)	20,22,23	1.75	3 (15%)

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
2	PLP	A	320	1	-	0/6/6/8	0/1/1/1
2	PLP	D	320	1	-	0/6/6/8	0/1/1/1
2	PLP	C	320	1	-	2/6/6/8	0/1/1/1
2	PLP	B	320	1	-	0/6/6/8	0/1/1/1

The worst 5 of 16 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	320	PLP	O3-C3	-5.87	1.23	1.37
2	D	320	PLP	O3-C3	-5.87	1.23	1.37
2	B	320	PLP	O3-C3	-5.82	1.23	1.37
2	A	320	PLP	O3-C3	-5.79	1.23	1.37
2	C	320	PLP	P-O1P	2.98	1.60	1.50

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	320	PLP	O4P-C5A-C5	6.12	121.00	109.35
2	A	320	PLP	O4P-C5A-C5	5.81	120.43	109.35
2	C	320	PLP	O4P-C5A-C5	5.62	120.06	109.35
2	B	320	PLP	O4P-C5A-C5	5.56	119.95	109.35
2	A	320	PLP	C6-C5-C4	3.15	120.63	118.16

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	320	PLP	C4-C5-C5A-O4P
2	C	320	PLP	C6-C5-C5A-O4P

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	320	PLP	2	0
2	D	320	PLP	1	0
2	C	320	PLP	2	0
2	B	320	PLP	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(Å²)	Q<0.9
1	A	294/303 (97%)	-0.47	1 (0%)	94	94	19, 38, 66, 97	0
1	B	281/303 (92%)	-0.27	4 (1%)	75	78	17, 42, 80, 144	0
1	C	292/303 (96%)	-0.50	0	100	100	18, 37, 58, 91	0
1	D	285/303 (94%)	-0.33	3 (1%)	80	84	15, 41, 78, 112	0
All	All	1152/1212 (95%)	-0.39	8 (0%)	87	89	15, 39, 72, 144	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	98	GLU	5.8
1	D	212	TRP	3.9
1	B	1	MET	3.1
1	B	131	ASN	2.9
1	B	134	GLU	2.7

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
2	PLP	A	320	15/16	0.98	0.09	19,28,38,44	0
2	PLP	B	320	15/16	0.98	0.10	25,39,48,50	0
2	PLP	C	320	15/16	0.98	0.10	16,32,41,46	0
2	PLP	D	320	15/16	0.99	0.10	14,38,55,56	0

6.5 Other polymers ⓘ

There are no such residues in this entry.