wwPDB X-ray Structure Validation Summary Report (i)

Aug 21, 2020 – 10:55 AM BST

PDB ID	:	6EXH
Title	:	Crystal structure of the complex Fe(II)/alpha-ketoglutarate dependent dioxygenase KDO5 with Fe(II)/succinate/(4R)-4-hydroxy-L-lysine
Authors	:	Isabet, T.; Stura, E.; Legrand, P.; Zaparucha, A.; Bastard, K.
Deposited on	:	2017-11-08
Resolution	:	2.60 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 (i)) 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	:	2.13.1
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)	:	2.13.1

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 2.60 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.
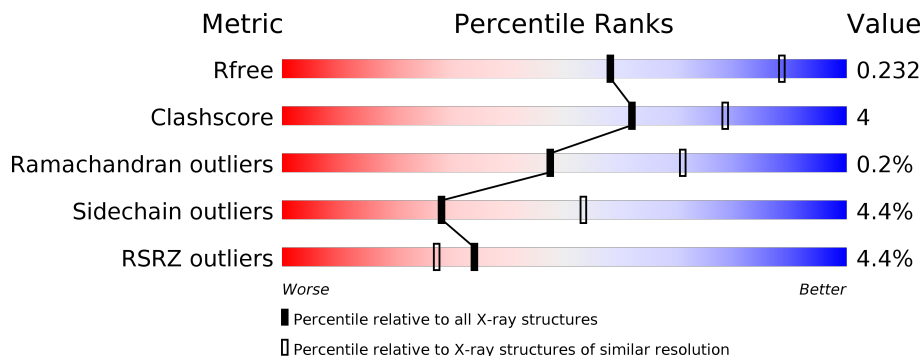
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	372	3% / 81% / 12% / · 6%
1	B	372	4% / 76% / 13% / · 10%
1	C	372	3% / 81% / 13% / 5%
1	D	372	6% / 83% / 12% / 6%

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	LYO	D	404	-	-	-	X

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 11611 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.
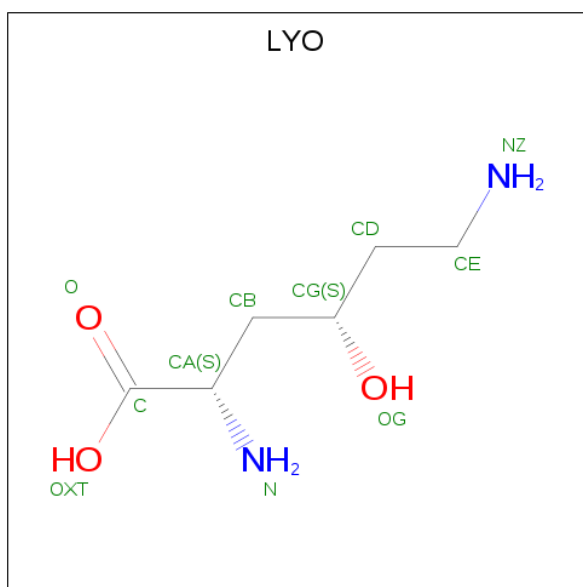
- Molecule 1 is a protein called L-lysine 4-hydroxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	348	Total	C	N	O	S	0	5	0
			2799	1776	479	531	13			
1	B	336	Total	C	N	O	S	0	5	0
			2703	1721	462	507	13			
1	C	353	Total	C	N	O	S	0	3	0
			2817	1788	481	535	13			
1	D	350	Total	C	N	O	S	0	0	0
			2768	1759	472	525	12			

- Molecule 2 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Fe	0	0
			1	1		
2	A	1	Total	Fe	0	0
			1	1		
2	D	1	Total	Fe	0	0
			1	1		
2	C	1	Total	Fe	0	0
			1	1		

- Molecule 3 is 4-HYDROXY-LYSINE (three-letter code: LYO) (formula: $C_6H_{14}N_2O_3$).

LYO

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	1
			11	6	2	3		
3	B	1	Total	C	N	O	0	1
			11	6	2	3		
3	C	1	Total	C	N	O	0	0
			11	6	2	3		
3	D	1	Total	C	N	O	0	0
			11	6	2	3		

- Molecule 4 is SUCCINIC ACID (three-letter code: SIN) (formula: $C_4H_6O_4$).

SIN
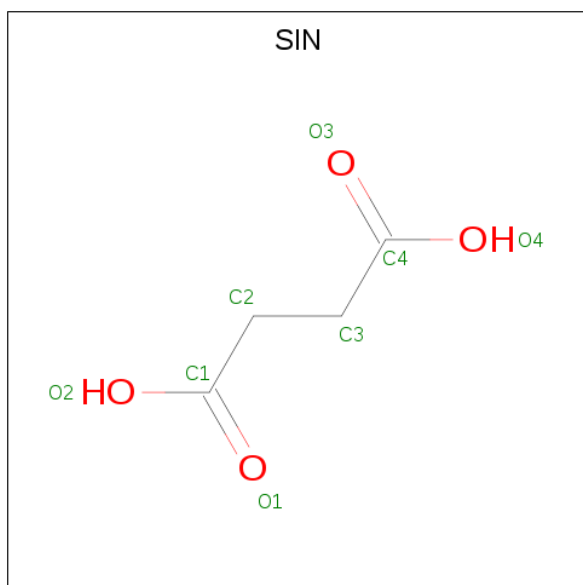
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total C O 8 4 4			0	0
4	B	1	Total C O 8 4 4			0	0
4	D	1	Total C O 8 4 4			0	1

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
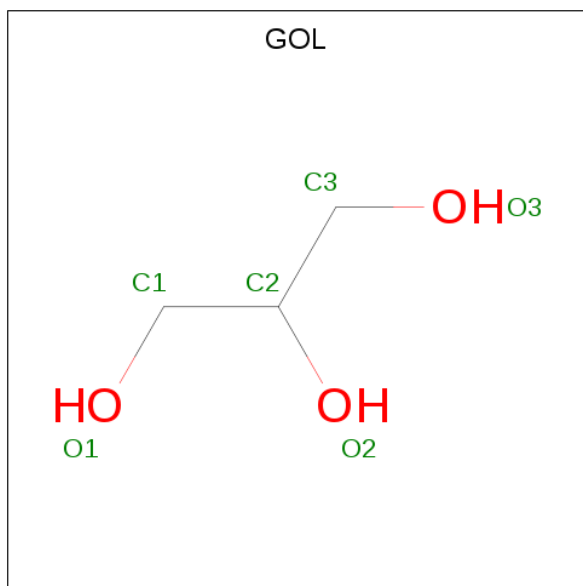
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total C O 6 3 3			0	0
5	C	1	Total C O 6 3 3			0	0
5	D	1	Total C O 6 3 3			0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	159	Total O 159 159		0	2
6	B	96	Total O 96 96		0	2
6	C	78	Total O 78 78		0	0
6	D	101	Total O 101 101		0	2

3 Residue-property plots ⓘ

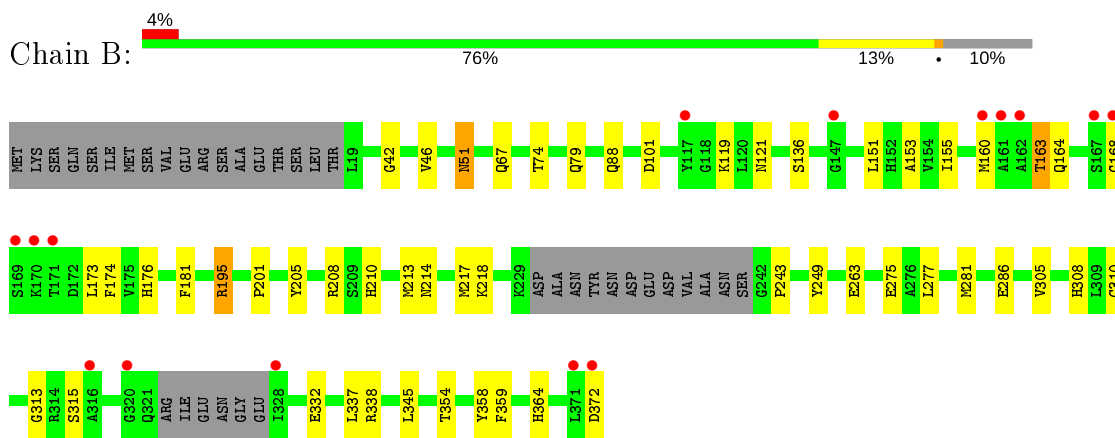These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: L-lysine 4-hydroxylase
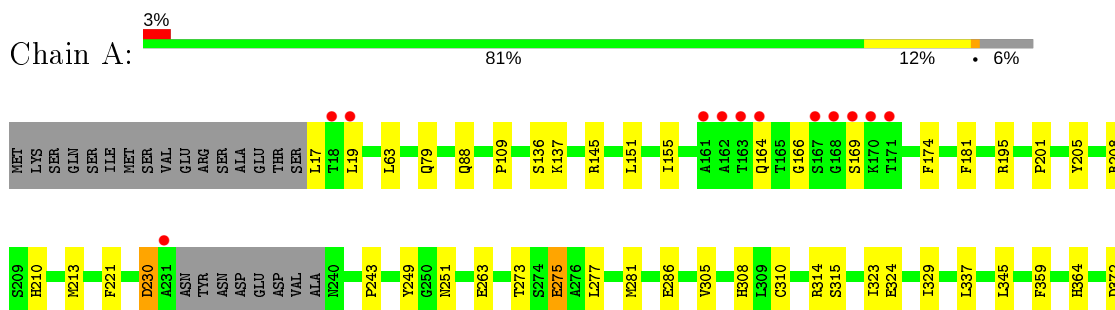
- Molecule 1: L-lysine 4-hydroxylase

- Molecule 1: L-lysine 4-hydroxylase

L337 F359 D372

● Molecule 1: L-lysine 4-hydroxylase

Chain D:

6%

83% 12% 6%

MET LYS SER GLN SER ILE MET SER VAL GLU ARG SER ALA GLU THR SER L17 L63 Q67 Q79 Q88 S136 L151 M160 A161 A162 T163 Q164 G168 S169 K170 T171 D172 L173 F174 R195 E198 P201 Y205 R208 S209 H210 M213 K218

D230 A231 N232 Y233 N234 ASP GLU ASP VAL ALA ASN S241 G242 P243 Y249 E263 E275 E286 V305 H308 L309 C310 A311 H312 G313 R314 S315 I318 A319 G320 Q321 R322 I323 E324 N325 G326 E327 I328 I329 L337 F359 H364 D372
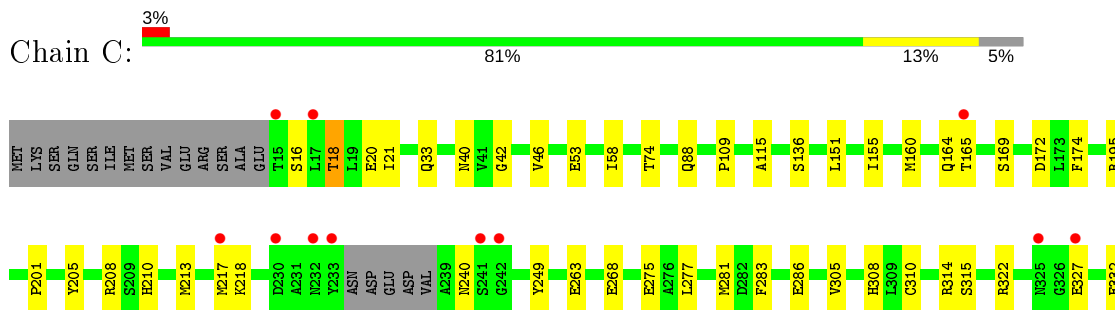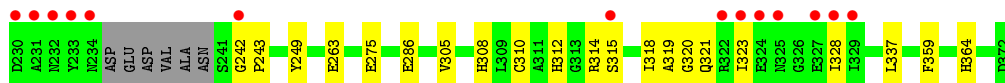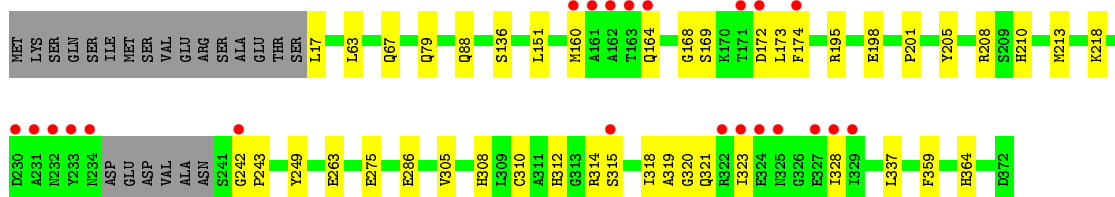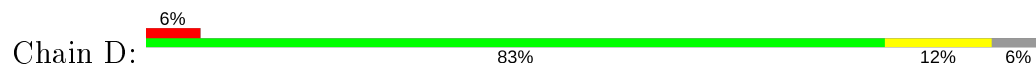
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	91.74Å 99.23Å 165.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.07 − 2.60 63.57 − 2.60	Depositor EDS
% Data completeness (in resolution range)	100.0 (38.07-2.60) 99.8 (63.57-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.86 (at 2.61Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.177 , 0.226 0.183 , 0.232	Depositor DCC
R_{free} test set	2363 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	62.0	Xtriage
Anisotropy	0.616	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 64.5	EDS
L-test for twinning[2]	$<|L|>=0.48$, $<L^2>=0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11611	wwPDB-VP
Average B, all atoms (Å^2)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.99% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, LYO, FE, SIN

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	0.52	0/2862	0.71	0/3872
1	B	0.51	0/2765	0.70	0/3740
1	C	0.50	0/2880	0.70	0/3897
1	D	0.50	0/2831	0.68	0/3834
All	All	0.51	0/11338	0.70	0/15343

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	2799	0	2731	27	0
1	B	2703	0	2649	31	0
1	C	2817	0	2755	25	0
1	D	2768	0	2697	19	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	11	0	13	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	11	0	13	2	0
3	C	11	0	13	1	0
3	D	11	0	13	1	0
4	A	8	0	4	0	0
4	B	8	0	4	0	0
4	D	8	0	4	1	0
5	A	6	0	8	0	0
5	C	6	0	8	1	0
5	D	6	0	8	0	0
6	A	159	0	0	0	0
6	B	96	0	0	0	0
6	C	78	0	0	0	0
6	D	101	0	0	0	0
All	All	11611	0	10920	97	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 97 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:B:213[A]:MET:HG3	1:B:218:LYS:HE3	1.73	0.71
1:A:324:GLU:HB2	1:A:329:ILE:HD11	1.80	0.64
1:A:166:GLY:H	3:A:402[A]:LYO:HB1	1.63	0.64
1:A:145:ARG:HE	3:A:402[A]:LYO:H	1.45	0.64
1:B:151:LEU:HD21	1:B:337:LEU:HD13	1.80	0.63

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	349/372 (94%)	333 (95%)	15 (4%)	1 (0%)	41	64
1	B	335/372 (90%)	323 (96%)	12 (4%)	0	100	100
1	C	352/372 (95%)	341 (97%)	11 (3%)	0	100	100
1	D	346/372 (93%)	332 (96%)	12 (4%)	2 (1%)	25	47
All	All	1382/1488 (93%)	1329 (96%)	50 (4%)	3 (0%)	47	71

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	230	ASP
1	D	319	ALA
1	D	242	GLY

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	305/322 (95%)	292 (96%)	13 (4%)	29	54
1	B	295/322 (92%)	282 (96%)	13 (4%)	28	53
1	C	307/322 (95%)	291 (95%)	16 (5%)	23	46
1	D	300/322 (93%)	288 (96%)	12 (4%)	31	57
All	All	1207/1288 (94%)	1153 (96%)	54 (4%)	28	52

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	359	PHE
1	C	169	SER
1	D	275	GLU
1	B	372	ASP
1	C	18	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	298	ASN
1	C	164	GLN
1	D	308	HIS
1	B	308	HIS
1	C	40	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 ⓘ

Of 14 ligands modelled in this entry, 4 are monoatomic - leaving 10 for Mogul analysis.

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	LYO	D	404	-	7,10,10	0.20	0	8,12,12	1.12	1 (12%)
5	GOL	D	403	-	5,5,5	0.10	0	5,5,5	0.29	0
4	SIN	D	402[A]	2	1,7,7	0.13	0	2,8,8	0.98	0
4	SIN	B	403	2	1,7,7	0.08	0	2,8,8	0.96	0
3	LYO	C	402	-	7,10,10	0.18	0	8,12,12	1.14	1 (12%)
3	LYO	A	402[A]	-	7,10,10	0.13	0	8,12,12	2.08	3 (37%)
3	LYO	B	402[A]	-	7,10,10	0.20	0	8,12,12	1.09	1 (12%)
4	SIN	A	403	2	1,7,7	0.08	0	2,8,8	1.03	0
5	GOL	A	404	-	5,5,5	0.13	0	5,5,5	0.32	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z|$ > 2	Counts	RMSZ	#$|Z|$ > 2
5	GOL	C	403	-	5,5,5	0.09	0	5,5,5	0.31	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	LYO	D	404	-	-	<mark>3/7/11/11</mark>	-
5	GOL	D	403	-	-	0/4/4/4	-
4	SIN	D	402[A]	2	-	1/1/5/5	-
4	SIN	B	403	2	-	0/1/5/5	-
3	LYO	C	402	-	-	<mark>2/7/11/11</mark>	-
3	LYO	A	402[A]	-	-	0/7/11/11	-
3	LYO	B	402[A]	-	-	1/7/11/11	-
4	SIN	A	403	2	-	0/1/5/5	-
5	GOL	A	404	-	-	0/4/4/4	-
5	GOL	C	403	-	-	1/4/4/4	-

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	402[A]	LYO	CB-CA-C	-3.36	104.27	112.61
3	A	402[A]	LYO	CB-CA-N	3.35	117.79	110.32
3	A	402[A]	LYO	CG-CB-CA	3.11	119.98	114.39
3	C	402	LYO	CB-CA-N	2.79	116.53	110.32
3	D	404	LYO	CB-CA-N	2.73	116.40	110.32

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	404	LYO	CG-CD-CE-NZ
4	D	402[A]	SIN	C1-C2-C3-C4
3	B	402[A]	LYO	N-CA-CB-CG
3	D	404	LYO	CE-CD-CG-OG
3	D	404	LYO	CE-CD-CG-CB

There are no ring outliers.

6 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	404	LYO	1	0
4	D	402[A]	SIN	1	0
3	C	402	LYO	1	0
3	A	402[A]	LYO	2	0
3	B	402[A]	LYO	2	0
5	C	403	GOL	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 (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(Å²)	Q<0.9
1	A	348/372 (93%)	0.04	12 (3%)	45	38	48, 64, 111, 139	0
1	B	336/372 (90%)	0.15	15 (4%)	33	26	46, 74, 136, 163	0
1	C	353/372 (94%)	0.10	12 (3%)	45	38	52, 78, 121, 155	0
1	D	350/372 (94%)	0.25	22 (6%)	20	15	49, 76, 142, 206	0
All	All	1387/1488 (93%)	0.13	61 (4%)	34	27	46, 73, 127, 206	0

The worst 5 of 61 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	162	ALA	9.6
1	D	232	ASN	9.0
1	A	171	THR	6.6
1	C	372	ASP	5.8
1	A	168	GLY	5.6

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(Å^2)	Q<0.9
5	GOL	A	404	6/6	0.71	0.28	92,95,101,101	0
3	LYO	C	402	11/11	0.75	0.25	109,111,114,115	0
3	LYO	D	404	11/11	0.75	0.47	117,123,129,130	0
5	GOL	D	403	6/6	0.84	0.20	86,93,95,97	0
3	LYO	B	402[A]	11/11	0.87	0.36	75,78,81,82	11
5	GOL	C	403	6/6	0.87	0.24	99,100,103,103	0
3	LYO	A	402[A]	11/11	0.88	0.30	70,71,73,73	11
4	SIN	D	402[A]	8/8	0.91	0.25	57,62,64,66	8
4	SIN	B	403	8/8	0.93	0.19	101,107,112,113	0
4	SIN	A	403	8/8	0.96	0.16	96,102,104,105	0
2	FE	C	401	1/1	0.97	0.25	89,89,89,89	0
2	FE	B	401	1/1	0.98	0.21	79,79,79,79	0
2	FE	A	401	1/1	0.99	0.16	75,75,75,75	0
2	FE	D	401	1/1	0.99	0.12	84,84,84,84	0

6.5 Other polymers (i)

There are no such residues in this entry.