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 ⓘ

Nov 25, 2024 – 05:59 PM EST

PDB ID	:	3D6K
Title	:	The crystal structure of a putative aminotransferase from Corynebacterium diphtheriae
Authors	:	Tan, K.; Zhang, R.; Duggan, E.; Clancy, S.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2008-05-19
Resolution	:	2.00 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

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 2.00 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		0.242
Clashscore		11
Ramachandran outliers		0.1%
Sidechain outliers		9.0%
RSRZ outliers		5.6%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
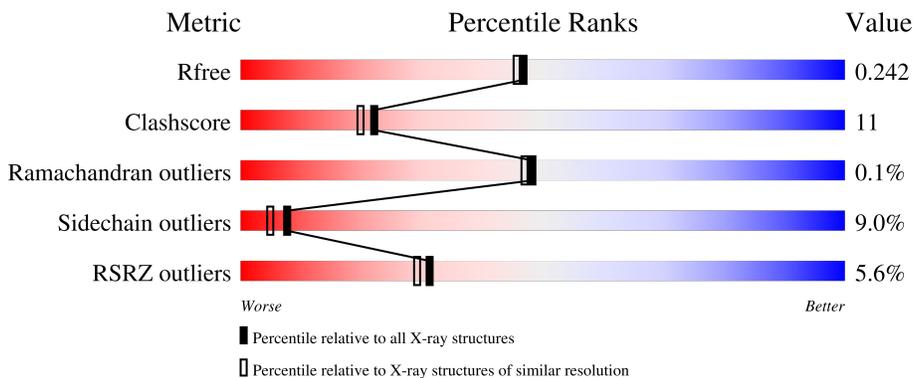
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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%

Mol	Chain	Length	Quality of chain
1	A	422	3% · 71% · 22% · ·
1	B	422	9% · 72% · 21% · ·
1	C	422	8% · 76% · 18% · 5% ·
1	D	422	% · 77% · 18% · ·

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 13332 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 Putative aminotransferase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	405	Total 3118	C 1983	N 527	O 594	S 4	Se 10	0	0	0
1	B	409	Total 3147	C 2002	N 532	O 599	S 4	Se 10	0	0	0
1	C	413	Total 3176	C 2018	N 539	O 605	S 4	Se 10	0	0	0
1	D	412	Total 3168	C 2014	N 538	O 602	S 4	Se 10	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q6NJY4
A	-1	ASN	-	expression tag	UNP Q6NJY4
A	0	ALA	-	expression tag	UNP Q6NJY4
B	-2	SER	-	expression tag	UNP Q6NJY4
B	-1	ASN	-	expression tag	UNP Q6NJY4
B	0	ALA	-	expression tag	UNP Q6NJY4
C	-2	SER	-	expression tag	UNP Q6NJY4
C	-1	ASN	-	expression tag	UNP Q6NJY4
C	0	ALA	-	expression tag	UNP Q6NJY4
D	-2	SER	-	expression tag	UNP Q6NJY4
D	-1	ASN	-	expression tag	UNP Q6NJY4
D	0	ALA	-	expression tag	UNP Q6NJY4

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total 5	O 4	S 1	0	0
2	B	1	Total 5	O 4	S 1	0	0
2	C	1	Total 5	O 4	S 1	0	0
2	D	1	Total 5	O 4	S 1	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total 1	Cl 1	0	0
3	B	1	Total 1	Cl 1	0	0
3	C	1	Total 1	Cl 1	0	0
3	D	1	Total 1	Cl 1	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO

O2

OH

C1

C2

HO

O1
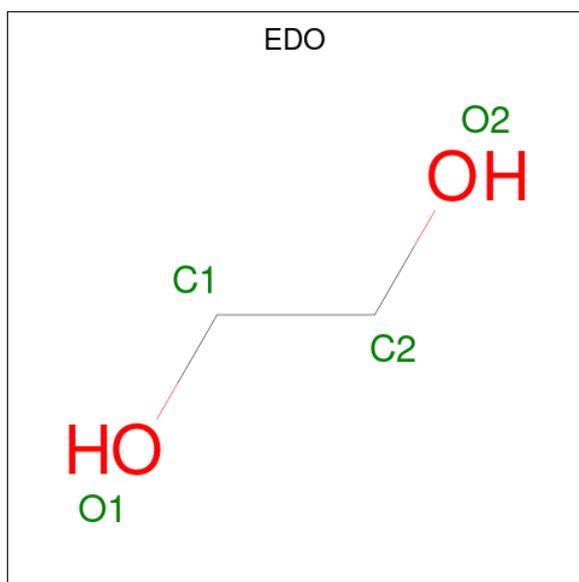
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total 4	C 2	O 2	0	0
4	B	1	Total 4	C 2	O 2	0	0
4	C	1	Total 4	C 2	O 2	0	0
4	D	1	Total 4	C 2	O 2	0	0

- Molecule 5 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	211	Total 211	O 211	0	0
5	B	182	Total 182	O 182	0	0
5	C	164	Total 164	O 164	0	0
5	D	126	Total 126	O 126	0	0

W O R L D W I D E

PDB
PROTEIN DATA BANK

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. 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. 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: Putative aminotransferase
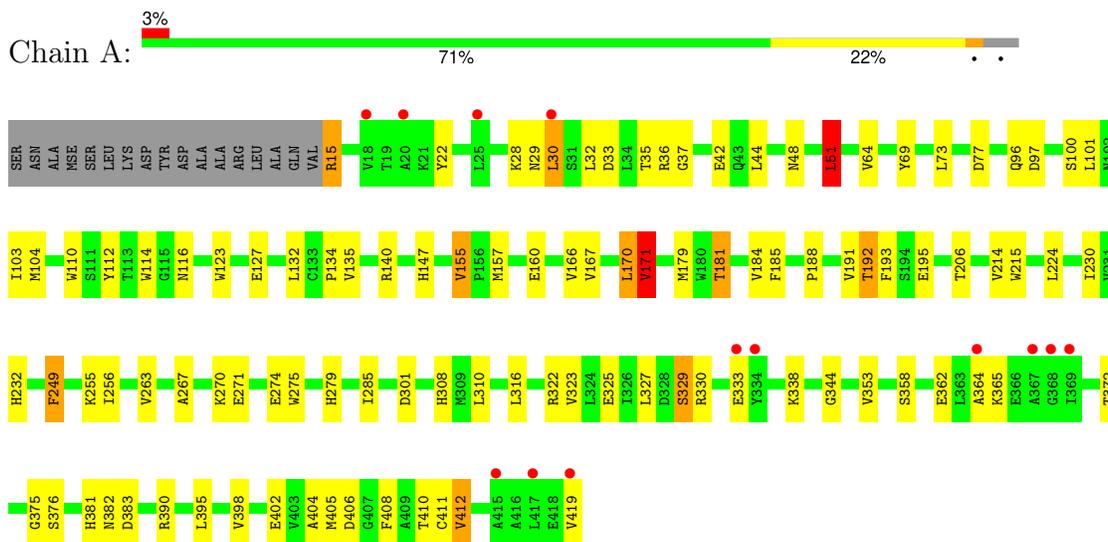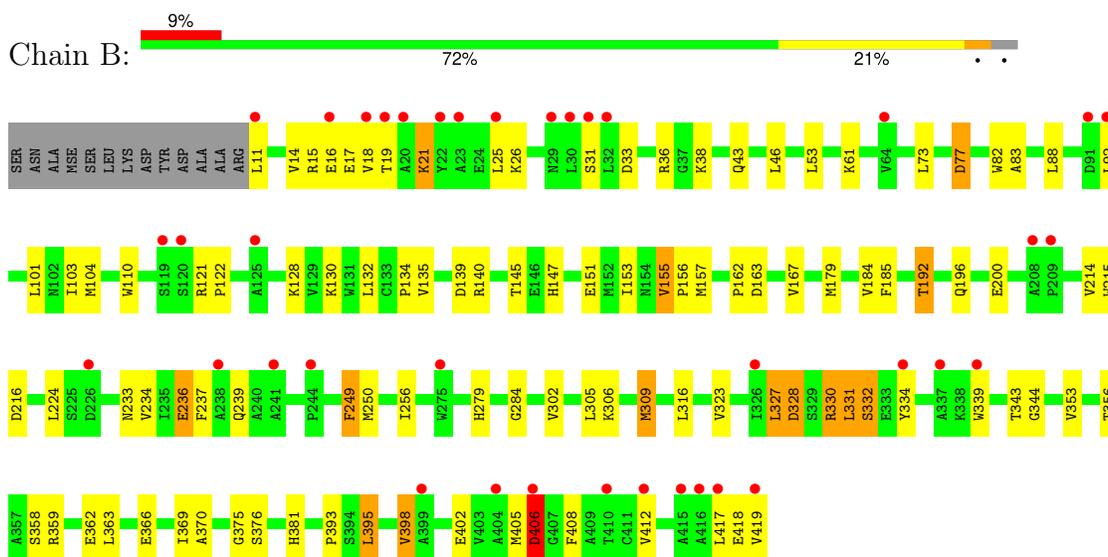

Chain A:

- Molecule 1: Putative aminotransferase

Chain B:

- Molecule 1: Putative aminotransferase

Chain C:

• Molecule 1: Putative aminotransferase

Chain D:
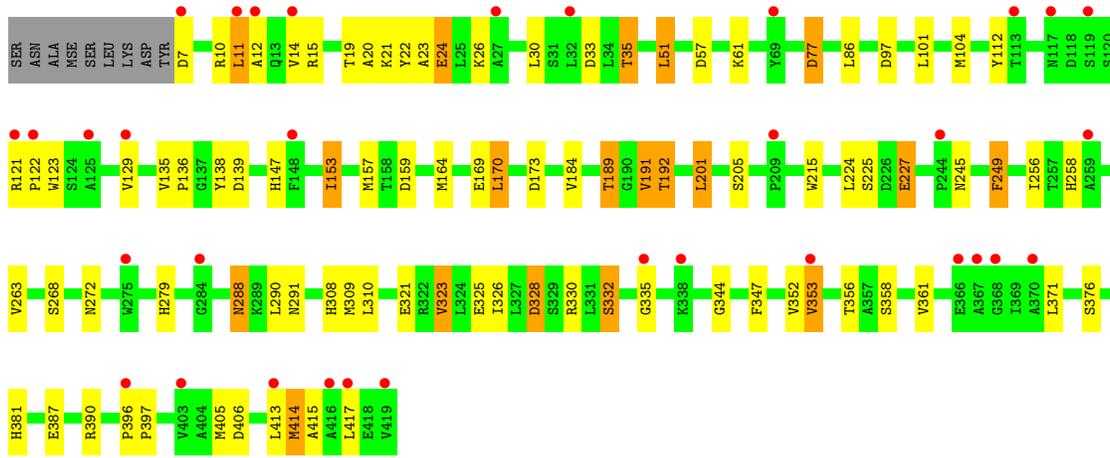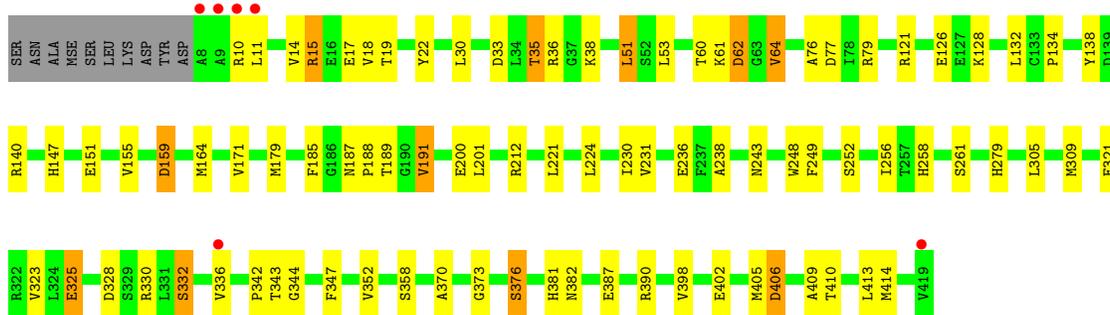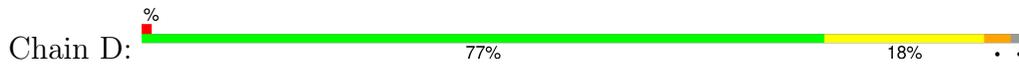
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.25Å 56.08Å 173.59Å 90.00° 102.47° 90.00°	Depositor
Resolution (Å)	42.68 – 2.00 42.68 – 2.00	Depositor EDS
% Data completeness (in resolution range)	90.4 (42.68-2.00) 90.4 (42.68-2.00)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.03 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.190 , 0.241 0.191 , 0.242	Depositor DCC
R_{free} test set	5228 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	32.3	Xtriage
Anisotropy	0.197	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 33.0	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.028 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	13332	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.28% 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: CL, SO4, EDO

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	0.67	0/3188	0.76	3/4322 (0.1%)
1	B	0.79	4/3217 (0.1%)	0.76	0/4362
1	C	0.76	3/3246 (0.1%)	0.74	4/4401 (0.1%)
1	D	0.64	0/3238	0.70	0/4390
All	All	0.72	7/12889 (0.1%)	0.74	7/17475 (0.0%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	406	ASP	C-N	7.62	1.46	1.33
1	C	415	ALA	C-O	5.92	1.34	1.23
1	C	24	GLU	CD-OE2	5.56	1.31	1.25
1	B	16	GLU	C-O	5.40	1.33	1.23
1	B	16	GLU	CD-OE1	5.26	1.31	1.25

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	171	VAL	CG1-CB-CG2	5.97	120.45	110.90
1	C	22	TYR	O-C-N	5.21	131.03	122.70
1	C	328	ASP	CB-CG-OD2	-5.16	113.66	118.30
1	A	383	ASP	CB-CG-OD1	5.06	122.86	118.30
1	C	173	ASP	CB-CG-OD1	5.04	122.84	118.30

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	3118	0	3026	59	0
1	B	3147	0	3059	89	0
1	C	3176	0	3086	59	0
1	D	3168	0	3082	66	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
2	D	5	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	4	0	6	0	0
4	B	4	0	6	0	0
4	C	4	0	6	1	0
4	D	4	0	6	1	0
5	A	211	0	0	3	0
5	B	182	0	0	5	0
5	C	164	0	0	1	0
5	D	126	0	0	0	0
All	All	13332	0	12277	262	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 11.

The worst 5 of 262 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:104:MSE:HE2	1:B:214:VAL:CG1	1.50	1.39
1:B:104:MSE:HE3	1:B:250:MSE:HG3	1.21	1.18
1:A:15:ARG:HB3	1:A:15:ARG:NH2	1.59	1.15
1:B:104:MSE:HE2	1:B:214:VAL:HG11	1.16	1.12
1:D:22:TYR:HB2	1:D:414:MSE:HE3	1.30	1.12

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	403/422 (96%)	389 (96%)	13 (3%)	1 (0%)	44	42
1	B	407/422 (96%)	388 (95%)	19 (5%)	0	100	100
1	C	411/422 (97%)	395 (96%)	15 (4%)	1 (0%)	44	42
1	D	410/422 (97%)	394 (96%)	16 (4%)	0	100	100
All	All	1631/1688 (97%)	1566 (96%)	63 (4%)	2 (0%)	48	47

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	97	ASP
1	C	97	ASP

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	332/334 (99%)	295 (89%)	37 (11%)	5	3
1	B	335/334 (100%)	307 (92%)	28 (8%)	9	6
1	C	337/334 (101%)	305 (90%)	32 (10%)	7	4
1	D	336/334 (101%)	313 (93%)	23 (7%)	13	10
All	All	1340/1336 (100%)	1220 (91%)	120 (9%)	8	5

5 of 120 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	331	LEU
1	D	236	GLU
1	C	51	LEU
1	D	221	LEU
1	D	406	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	308	HIS
1	D	102	ASN
1	D	96	GLN
1	D	239	GLN
1	B	102	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 oligosaccharides in this entry.

5.6   Ligand geometry ⓘ

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 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
4	EDO	B	422	-	3,3,3	0.59	0	2,2,2	0.33	0
4	EDO	C	422	-	3,3,3	0.60	0	2,2,2	0.26	0
2	SO4	C	420	-	4,4,4	0.47	0	6,6,6	0.56	0
4	EDO	A	422	-	3,3,3	0.75	0	2,2,2	0.36	0
2	SO4	A	420	-	4,4,4	0.40	0	6,6,6	0.35	0
2	SO4	D	420	-	4,4,4	0.54	0	6,6,6	0.22	0
2	SO4	B	420	-	4,4,4	0.31	0	6,6,6	0.30	0
4	EDO	D	422	-	3,3,3	0.77	0	2,2,2	0.46	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	EDO	C	422	-	-	0/1/1/1	-
4	EDO	A	422	-	-	0/1/1/1	-
4	EDO	B	422	-	-	1/1/1/1	-
4	EDO	D	422	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	422	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	422	EDO	1	0
4	D	422	EDO	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 ⓘ

<div style="background-color: yellow;">

Warning: The R factor obtained from EDS is 0.2507, which does not match the depositor's R factor of 0.18992. Please interpret the results in this section carefully.

</div>

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	395/422 (93%)	0.40	13 (3%)	49	47	27, 34, 49, 54	0
1	B	399/422 (94%)	0.91	38 (9%)	15	14	27, 35, 51, 55	0
1	C	403/422 (95%)	0.82	33 (8%)	19	17	28, 35, 46, 55	0
1	D	402/422 (95%)	0.39	6 (1%)	71	70	28, 35, 42, 62	0
All	All	1599/1688 (94%)	0.63	90 (5%)	31	29	27, 35, 48, 62	0

The worst 5 of 90 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	404	ALA	5.2
1	B	415	ALA	5.2
1	B	20	ALA	4.7
1	B	32	LEU	4.7
1	C	417	LEU	4.6

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($\mathring{A}^2$)	Q<0.9
2	SO4	A	420	5/5	0.80	0.16	25,30,31,36	5
4	EDO	D	422	4/4	0.83	0.13	27,28,29,32	0
4	EDO	C	422	4/4	0.88	0.14	38,42,43,44	0
2	SO4	D	420	5/5	0.88	0.15	37,39,39,41	5
4	EDO	A	422	4/4	0.89	0.18	25,28,29,29	0
2	SO4	B	420	5/5	0.90	0.15	41,42,45,47	5
2	SO4	C	420	5/5	0.92	0.13	24,27,29,31	5
4	EDO	B	422	4/4	0.93	0.12	37,38,38,39	0
3	CL	C	421	1/1	0.96	0.14	39,39,39,39	0
3	CL	D	421	1/1	0.96	0.17	42,42,42,42	0
3	CL	A	421	1/1	0.96	0.13	36,36,36,36	0
3	CL	B	421	1/1	0.97	0.12	43,43,43,43	0

6.5 Other polymers ⓘ

There are no such residues in this entry.