Full wwPDB X-ray Structure Validation Report ⓘ

Jan 7, 2024 – 10:36 pm GMT

PDB ID	:	6GBL
Title	:	Repertoires of functionally diverse enzymes through computational design at epistatic active-site positions
Authors	:	Khersonsky, O.; Lipsh, R.; Avizemer, Z.; Goldsmith, M.; Ashani, Y.; Leader, H.; Dym, O.; Rogotner, S.; Trudeau, D.; Tawfik, D.S.; Fleishman, S.J.
Deposited on	:	2018-04-15
Resolution	:	1.95 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

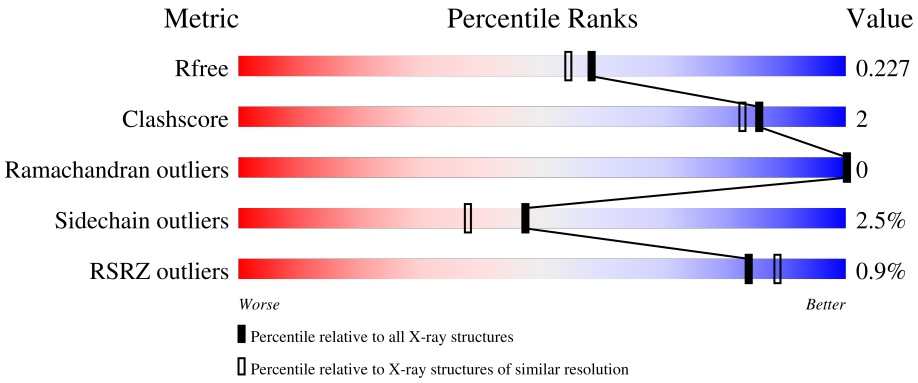
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	336	90% 7% .
1	B	336	% 91% 5% . .

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 5304 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 Parathion hydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	328	Total	C	N	O	S	0	2	0
			2534	1596	443	487	8			
1	B	327	Total	C	N	O	S	0	3	0
			2529	1592	441	488	8			

There are 54 discrepancies between the modelled and reference sequences:

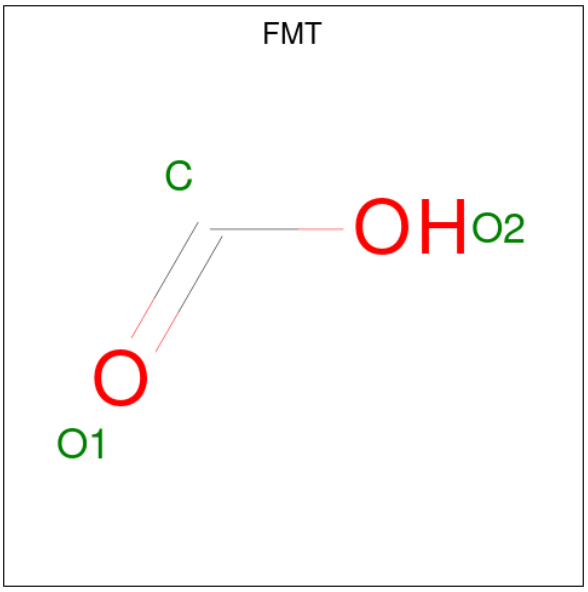
Chain	Residue	Modelled	Actual	Comment	Reference
A	30	ILE	-	expression tag	UNP P0A434
A	31	THR	-	expression tag	UNP P0A434
A	32	ASN	-	expression tag	UNP P0A434
A	33	SER	-	expression tag	UNP P0A434
A	54	MET	THR	engineered mutation	UNP P0A434
A	77	ASP	LYS	engineered mutation	UNP P0A434
A	106	LEU	ILE	engineered mutation	UNP P0A434
A	111	GLU	SER	engineered mutation	UNP P0A434
A	118	GLU	ARG	engineered mutation	UNP P0A434
A	182	ARG	LEU	engineered mutation	UNP P0A434
A	185	ARG	LYS	engineered mutation	UNP P0A434
A	203	ASP	ALA	engineered mutation	UNP P0A434
A	214	ASP	ALA	engineered mutation	UNP P0A434
A	222	ASP	SER	engineered mutation	UNP P0A434
A	238	ASP	SER	engineered mutation	UNP P0A434
A	254	GLY	HIS	engineered mutation	UNP P0A434
A	257	TRP	HIS	engineered mutation	UNP P0A434
A	269	ALA	SER	engineered mutation	UNP P0A434
A	274	LEU	ILE	engineered mutation	UNP P0A434
A	293	ALA	MET	engineered mutation	UNP P0A434
A	294	ASP	LYS	engineered mutation	UNP P0A434
A	303	THR	LEU	engineered mutation	UNP P0A434
A	343	ASP	GLN	engineered mutation	UNP P0A434
A	347	GLU	ALA	engineered mutation	UNP P0A434
A	348	THR	GLY	engineered mutation	UNP P0A434

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	350	MET	THR	engineered mutation	UNP P0A434
A	352	ASP	THR	engineered mutation	UNP P0A434
B	30	ILE	-	expression tag	UNP P0A434
B	31	THR	-	expression tag	UNP P0A434
B	32	ASN	-	expression tag	UNP P0A434
B	33	SER	-	expression tag	UNP P0A434
B	54	MET	THR	engineered mutation	UNP P0A434
B	77	ASP	LYS	engineered mutation	UNP P0A434
B	106	LEU	ILE	engineered mutation	UNP P0A434
B	111	GLU	SER	engineered mutation	UNP P0A434
B	118	GLU	ARG	engineered mutation	UNP P0A434
B	182	ARG	LEU	engineered mutation	UNP P0A434
B	185	ARG	LYS	engineered mutation	UNP P0A434
B	203	ASP	ALA	engineered mutation	UNP P0A434
B	214	ASP	ALA	engineered mutation	UNP P0A434
B	222	ASP	SER	engineered mutation	UNP P0A434
B	238	ASP	SER	engineered mutation	UNP P0A434
B	254	GLY	HIS	engineered mutation	UNP P0A434
B	257	TRP	HIS	engineered mutation	UNP P0A434
B	269	ALA	SER	engineered mutation	UNP P0A434
B	274	LEU	ILE	engineered mutation	UNP P0A434
B	293	ALA	MET	engineered mutation	UNP P0A434
B	294	ASP	LYS	engineered mutation	UNP P0A434
B	303	THR	LEU	engineered mutation	UNP P0A434
B	343	ASP	GLN	engineered mutation	UNP P0A434
B	347	GLU	ALA	engineered mutation	UNP P0A434
B	348	THR	GLY	engineered mutation	UNP P0A434
B	350	MET	THR	engineered mutation	UNP P0A434
B	352	ASP	THR	engineered mutation	UNP P0A434

- Molecule 2 is FORMIC ACID (three-letter code: FMT) (formula: CH_2O_2).

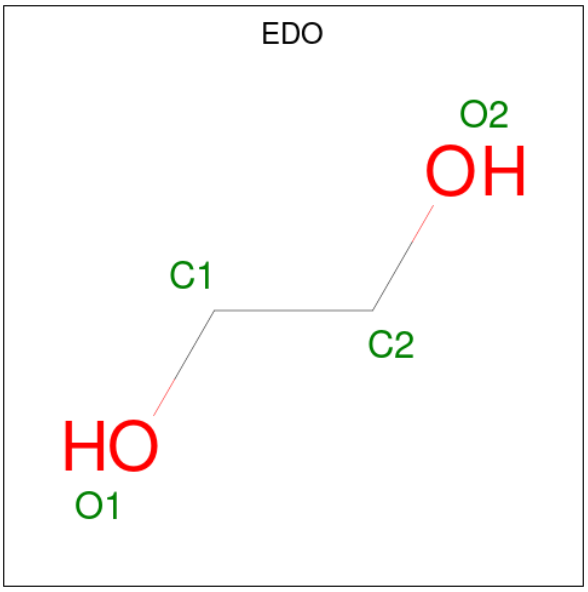


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 3 1 2	0	0
2	B	1	Total C O 3 1 2	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

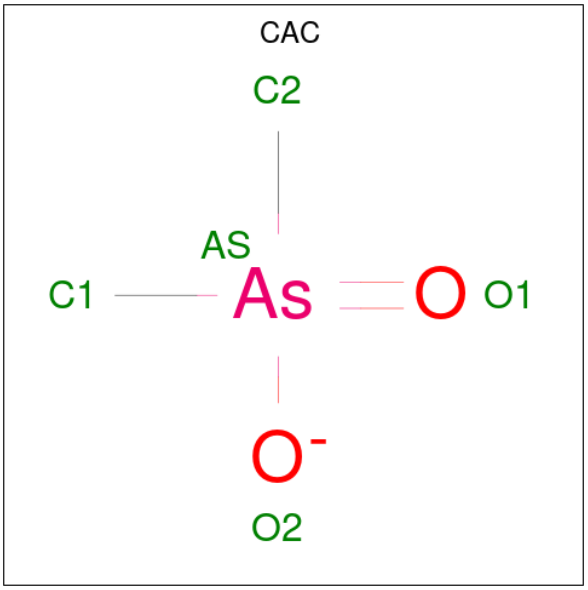
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Zn 2 2	0	0
3	B	2	Total Zn 2 2	0	0

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



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

- Molecule 5 is CACODYLATE ION (three-letter code: CAC) (formula: $C_2H_6AsO_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	As	C	O	0	0
			5	1	2	2		
5	B	1	Total	As	C	O	0	0
			5	1	2	2		

- Molecule 6 is water.

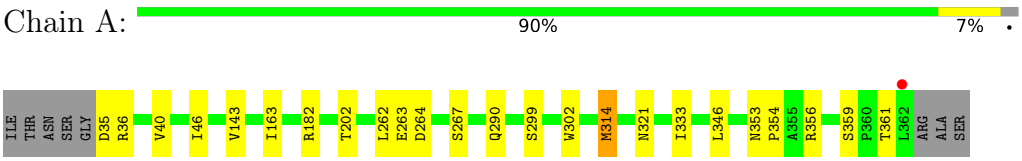
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	117	Total O 117 117	0	0
6	B	96	Total O 96 96	0	0

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.

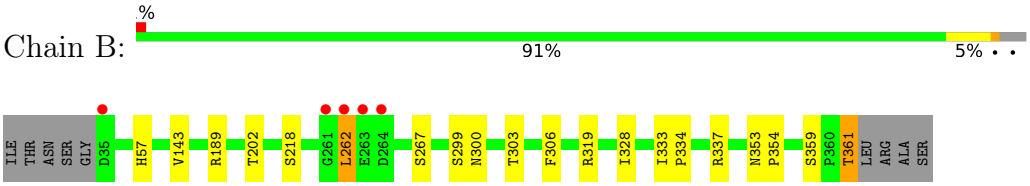
• Molecule 1: Parathion hydrolase

Chain A:



• Molecule 1: Parathion hydrolase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	155.80Å 53.56Å 89.34Å 90.00° 107.21° 90.00°	Depositor
Resolution (Å)	41.61 – 1.95 41.61 – 1.95	Depositor EDS
% Data completeness (in resolution range)	87.8 (41.61-1.95) 87.8 (41.61-1.95)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.77 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.8.0222	Depositor
R, R_{free}	0.177 , 0.220 0.186 , 0.227	Depositor DCC
R_{free} test set	2272 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	18.7	Xtriage
Anisotropy	0.782	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 34.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	5304	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.54% 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: FMT, EDO, CAC, ZN

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	0.60	0/2582	0.77	2/3511 (0.1%)
1	B	0.60	0/2577	0.73	0/3504
All	All	0.60	0/5159	0.75	2/7015 (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	A	0	4
1	B	0	1
All	All	0	5

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	263	GLU	CB-CA-C	-5.84	98.71	110.40
1	A	264	ASP	CB-CA-C	5.69	121.78	110.40

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	182	ARG	Sidechain
1	A	202	THR	Peptide
1	A	356	ARG	Sidechain
1	A	36	ARG	Sidechain

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	202	THR	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	2534	0	2503	10	0
1	B	2529	0	2491	11	3
2	A	3	0	0	0	0
2	B	3	0	0	0	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	4	0	6	0	0
4	B	4	0	6	2	0
5	A	5	0	0	0	0
5	B	5	0	0	1	0
6	A	117	0	0	2	0
6	B	96	0	0	1	0
All	All	5304	0	5006	22	3

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 2.

All (22) 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:A:361:THR:HG21	6:A:511:HOH:O	1.82	0.79
1:A:359[A]:SER:O	1:A:361:THR:HG23	1.93	0.68
1:A:359[B]:SER:O	1:A:361:THR:HG23	1.95	0.66
1:B:262:LEU:N	1:B:262:LEU:HD13	2.17	0.59
1:A:333:ILE:HG23	1:A:346:LEU:HD13	1.89	0.54
1:B:262:LEU:N	1:B:262:LEU:CD1	2.74	0.50
1:A:353:ASN:HB2	1:A:354:PRO:HD3	1.94	0.50
1:B:353:ASN:HB2	1:B:354:PRO:HD3	1.96	0.48
1:B:359[B]:SER:O	1:B:361:THR:HG22	2.14	0.47
1:A:143:VAL:HG23	6:A:564:HOH:O	2.13	0.47

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:333:ILE:HB	1:B:334:PRO:HD3	1.96	0.46
1:B:143:VAL:HG11	1:B:189:ARG:CZ	2.46	0.45
1:B:359[A]:SER:O	1:B:361:THR:HG22	2.17	0.44
1:B:218[A]:SER:HA	6:B:590:HOH:O	2.16	0.44
5:B:404:CAC:C2	4:B:405:EDO:O1	2.67	0.43
1:B:57:HIS:CD2	4:B:405:EDO:H11	2.54	0.42
1:B:57:HIS:O	1:B:303:THR:HA	2.19	0.42
1:A:46:ILE:N	1:A:46:ILE:HD13	2.35	0.42
1:A:302:TRP:CH2	1:A:321:ASN:HB3	2.55	0.41
1:B:300:ASN:OD1	1:B:328:ILE:HG12	2.21	0.40
1:A:314:MET:HA	1:A:314:MET:CE	2.52	0.40
1:A:40:VAL:HA	1:A:163:ILE:HG23	2.04	0.40

All (3) 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)
1:B:189:ARG:NH1	1:B:189:ARG:NH1[2_555]	1.23	0.97
1:B:189:ARG:CZ	1:B:189:ARG:NH1[2_555]	2.08	0.12
1:B:189:ARG:CZ	1:B:189:ARG:CZ[2_555]	2.10	0.10

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	328/336 (98%)	318 (97%)	10 (3%)	0	100	100
1	B	328/336 (98%)	318 (97%)	10 (3%)	0	100	100
All	All	656/672 (98%)	636 (97%)	20 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	267/271 (98%)	260 (97%)	7 (3%)	46 36
1	B	266/271 (98%)	258 (97%)	8 (3%)	41 30
All	All	533/542 (98%)	518 (97%)	15 (3%)	47 33

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	35	ASP
1	A	262	LEU
1	A	267[A]	SER
1	A	267[B]	SER
1	A	290	GLN
1	A	299	SER
1	A	314	MET
1	B	262	LEU
1	B	267[A]	SER
1	B	267[B]	SER
1	B	299	SER
1	B	306	PHE
1	B	319	ARG
1	B	337	ARG
1	B	361	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

Of 10 ligands modelled in this entry, 4 are monoatomic - leaving 6 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	405	-	3,3,3	0.37	0	2,2,2	1.01	0
2	FMT	A	401	3,1	2,2,2	0.65	0	1,1,1	0.53	0
5	CAC	B	404	3	0,4,4	-	-	0,6,6	-	-
5	CAC	A	405	3	0,4,4	-	-	0,6,6	-	-
2	FMT	B	401	3,1	2,2,2	0.95	0	1,1,1	0.37	0
4	EDO	A	404	-	3,3,3	0.47	0	2,2,2	1.42	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	B	405	-	-	1/1/1/1	-
4	EDO	A	404	-	-	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	405	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	B	405	EDO	2	0
5	B	404	CAC	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	328/336 (97%)	-0.44	1 (0%)	94	96	11, 17, 28, 39	0
1	B	327/336 (97%)	-0.19	5 (1%)	73	81	12, 19, 32, 61	0
All	All	655/672 (97%)	-0.32	6 (0%)	84	89	11, 18, 30, 61	0

All (6) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	262	LEU	3.4
1	B	261	GLY	3.1
1	B	264	ASP	2.8
1	B	263	GLU	2.3
1	B	35	ASP	2.1
1	A	362	LEU	2.1

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, 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
4	EDO	A	404	4/4	0.83	0.19	20,22,27,29	0
4	EDO	B	405	4/4	0.89	0.15	24,24,27,28	0
2	FMT	B	401	3/3	0.97	0.11	12,12,13,13	0
2	FMT	A	401	3/3	0.99	0.04	12,12,12,13	0
3	ZN	B	402	1/1	1.00	0.06	13,13,13,13	0
3	ZN	B	403	1/1	1.00	0.04	15,15,15,15	0
3	ZN	A	402	1/1	1.00	0.04	12,12,12,12	0
3	ZN	A	403	1/1	1.00	0.04	12,12,12,12	0
5	CAC	A	405	5/5	1.00	0.06	12,12,13,14	0
5	CAC	B	404	5/5	1.00	0.07	14,14,14,15	0

6.5 Other polymers (i)

There are no such residues in this entry.