wwPDB X-ray Structure Validation Summary Report (i)

Apr 27, 2021 – 12:09 pm BST

PDB ID	:	6YGU
Title	:	Crystal structure of the minimal Mtr4-Red1 complex (single chain) from Chaetomium thermophilum
Authors	:	Dobrev, N.; Ahmed, Y.L.; Sinning, I.
Deposited on	:	2020-03-27
Resolution	:	1.99 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 (1)) were used in the production of this report:

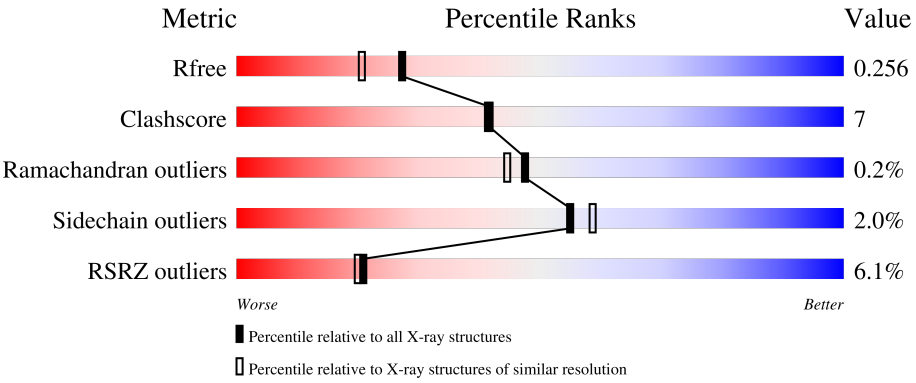
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.18
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.18

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 1.99 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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% 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	219	5% 81% 14%
1	C	219	8% 81% 17%
2	B	86	3% 65% 12% 23%
2	D	86	3% 67% 7% 24%

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	ACT	A	901	-	-	X	-
4	EDO	A	905	-	-	X	-

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 4711 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 ATP dependent RNA helicase (Dob1)-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	211	Total	C	N	O	S	0	2	0
			1702	1065	308	318	11			
1	C	215	Total	C	N	O	S	0	0	0
			1714	1070	311	321	12			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	653	MET	-	initiating methionine	UNP G0RZ64
A	866	GLY	-	expression tag	UNP G0RZ64
A	867	GLY	-	expression tag	UNP G0RZ64
A	868	SER	-	expression tag	UNP G0RZ64
A	869	GLY	-	expression tag	UNP G0RZ64
A	870	GLY	-	expression tag	UNP G0RZ64
A	871	SER	-	expression tag	UNP G0RZ64
C	653	MET	-	initiating methionine	UNP G0RZ64
C	866	GLY	-	expression tag	UNP G0RZ64
C	867	GLY	-	expression tag	UNP G0RZ64
C	868	SER	-	expression tag	UNP G0RZ64
C	869	GLY	-	expression tag	UNP G0RZ64
C	870	GLY	-	expression tag	UNP G0RZ64
C	871	SER	-	expression tag	UNP G0RZ64

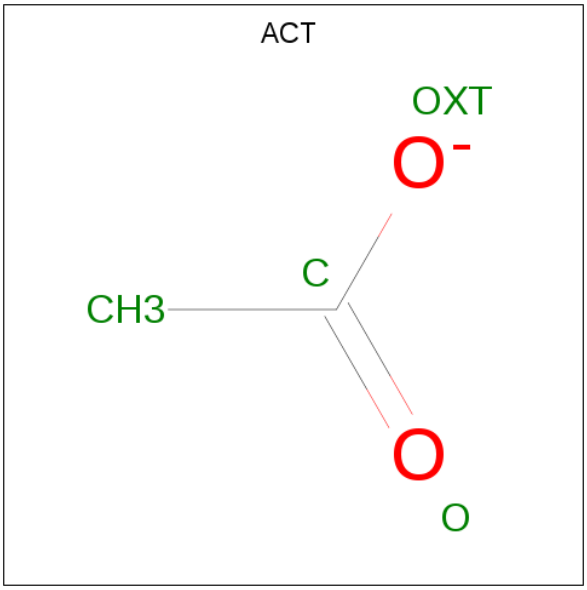
- Molecule 2 is a protein called Red1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	66	Total	C	N	O	S	0	0	0
			539	345	93	97	4			
2	D	65	Total	C	N	O	S	0	0	0
			534	342	92	96	4			

There are 18 discrepancies between the modelled and reference sequences:

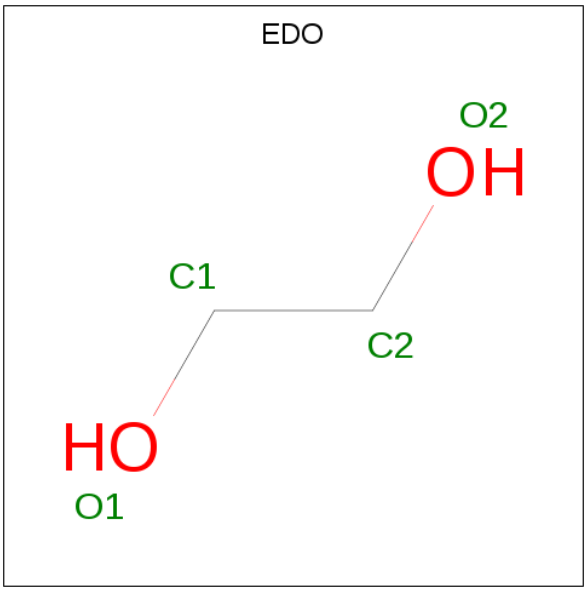
Chain	Residue	Modelled	Actual	Comment	Reference
B	1081	ALA	-	insertion	UNP G0S1V1
B	1092	GLY	-	expression tag	UNP G0S1V1
B	1093	SER	-	expression tag	UNP G0S1V1
B	1094	HIS	-	expression tag	UNP G0S1V1
B	1095	HIS	-	expression tag	UNP G0S1V1
B	1096	HIS	-	expression tag	UNP G0S1V1
B	1097	HIS	-	expression tag	UNP G0S1V1
B	1098	HIS	-	expression tag	UNP G0S1V1
B	1099	HIS	-	expression tag	UNP G0S1V1
D	1081	ALA	-	insertion	UNP G0S1V1
D	1092	GLY	-	expression tag	UNP G0S1V1
D	1093	SER	-	expression tag	UNP G0S1V1
D	1094	HIS	-	expression tag	UNP G0S1V1
D	1095	HIS	-	expression tag	UNP G0S1V1
D	1096	HIS	-	expression tag	UNP G0S1V1
D	1097	HIS	-	expression tag	UNP G0S1V1
D	1098	HIS	-	expression tag	UNP G0S1V1
D	1099	HIS	-	expression tag	UNP G0S1V1

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



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

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	B	1	Total Zn 1 1	0	0
5	D	1	Total Zn 1 1	0	0

- Molecule 6 is water.

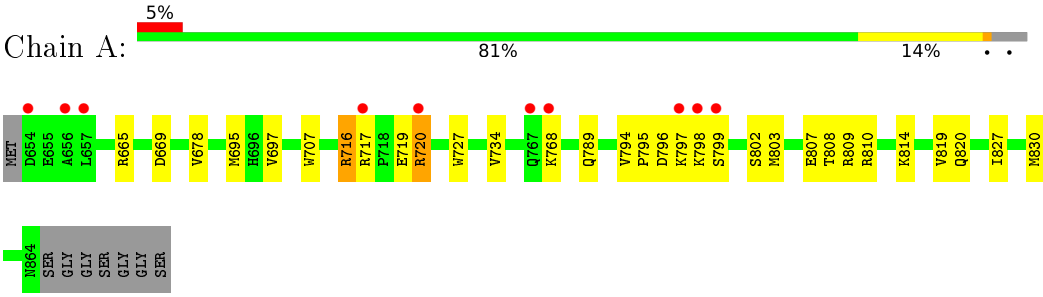
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	72	Total O 72 72	0	0
6	B	35	Total O 35 35	0	0
6	C	49	Total O 49 49	0	0
6	D	24	Total O 24 24	0	0

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.

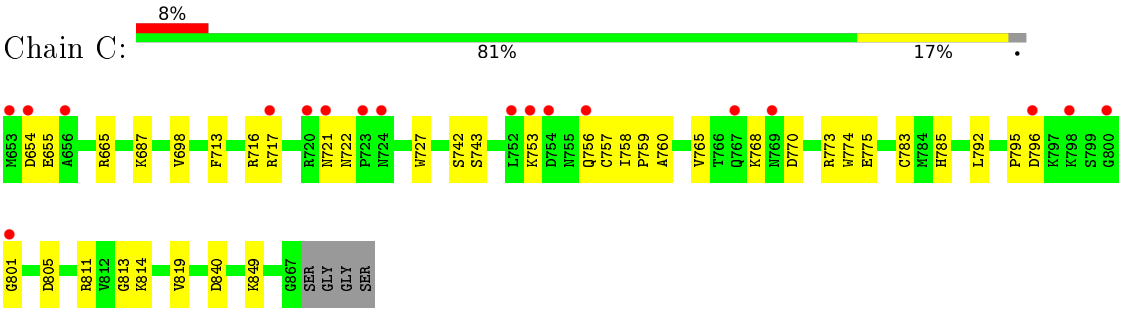
- Molecule 1: ATP dependent RNA helicase (Dob1)-like protein

Chain A:



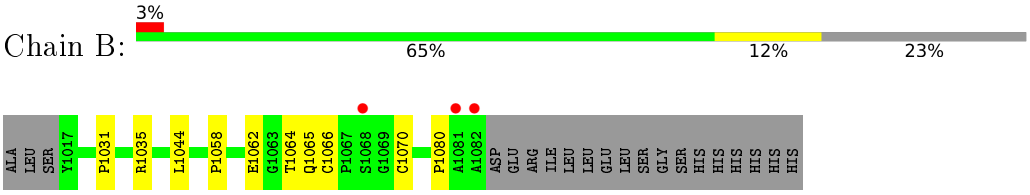
- Molecule 1: ATP dependent RNA helicase (Dob1)-like protein

Chain C:



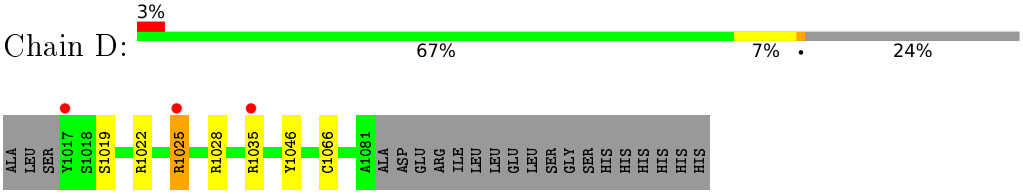
- Molecule 2: Red1

Chain B:



- Molecule 2: Red1

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	44.98Å 88.91Å 168.37Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.46 – 1.99 47.46 – 1.99	Depositor EDS
% Data completeness (in resolution range)	99.9 (47.46-1.99) 100.0 (47.46-1.99)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.97 (at 1.98Å)	Xtriage
Refinement program	PHENIX (1.14_3260: ???)	Depositor
R, R_{free}	0.204 , 0.252 0.207 , 0.256	Depositor DCC
R_{free} test set	2395 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	38.8	Xtriage
Anisotropy	0.737	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 45.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4711	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.60% 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: ZN, EDO, ACT

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	0.62	1/1742 (0.1%)	0.84	6/2358 (0.3%)
1	C	0.56	0/1751	0.67	2/2368 (0.1%)
2	B	0.62	0/558	0.69	0/756
2	D	0.81	2/553 (0.4%)	0.94	5/749 (0.7%)
All	All	0.63	3/4604 (0.1%)	0.78	13/6231 (0.2%)

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	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1025	ARG	CG-CD	-9.20	1.28	1.51
2	D	1066	CYS	CB-SG	-7.99	1.68	1.82
1	A	678	VAL	CB-CG2	5.81	1.65	1.52

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	720	ARG	NE-CZ-NH2	14.50	127.55	120.30
2	D	1025	ARG	CA-CB-CG	-10.57	90.15	113.40
1	A	720	ARG	CG-CD-NE	-9.63	91.58	111.80
2	D	1025	ARG	NE-CZ-NH1	-8.79	115.91	120.30
2	D	1025	ARG	CG-CD-NE	-8.06	94.86	111.80

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	719	GLU	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	1702	0	1697	25	0
1	C	1714	0	1704	24	0
2	B	539	0	499	7	2
2	D	534	0	494	7	0
3	A	4	0	3	3	0
3	B	4	0	3	0	0
4	A	24	0	36	6	0
4	B	4	0	6	0	0
4	C	4	0	6	0	0
5	B	1	0	0	0	0
5	D	1	0	0	0	0
6	A	72	0	0	5	0
6	B	35	0	0	3	1
6	C	49	0	0	2	0
6	D	24	0	0	0	0
All	All	4711	0	4448	63	2

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 7.

The worst 5 of 63 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:814:LYS:NZ	6:A:1001:HOH:O	1.75	1.19
1:A:707:TRP:H	3:A:901:ACT:H2	1.17	1.05
2:D:1025:ARG:NH2	2:D:1046:TYR:O	2.14	0.80
1:C:795:PRO:HG2	1:C:811:ARG:HH12	1.46	0.80
1:A:814:LYS:HD2	6:A:1060:HOH:O	1.83	0.78

All (2) 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)
2:B:1066:CYS:SG	6:B:1201:HOH:O[4_556]	2.15	0.05
2:B:1035:ARG:NH2	2:B:1064:THR:O[4_456]	2.17	0.03

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	211/219 (96%)	204 (97%)	6 (3%)	1 (0%)	29 23
1	C	213/219 (97%)	208 (98%)	5 (2%)	0	100 100
2	B	64/86 (74%)	64 (100%)	0	0	100 100
2	D	63/86 (73%)	63 (100%)	0	0	100 100
All	All	551/610 (90%)	539 (98%)	11 (2%)	1 (0%)	47 44

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	720	ARG

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	196/198 (99%)	191 (97%)	5 (3%)	46 48

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	196/198 (99%)	190 (97%)	6 (3%)	40 40
2	B	59/77 (77%)	59 (100%)	0	100 100
2	D	59/77 (77%)	59 (100%)	0	100 100
All	All	510/550 (93%)	499 (98%)	11 (2%)	55 55

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	721	ASN
1	C	722	ASN
1	C	796	ASP
1	C	768	LYS
1	A	848	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	722	ASN
1	C	756	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)

Of 12 ligands modelled in this entry, 2 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	$\#\|Z\| > 2$
4	EDO	A	902	-	3,3,3	0.53	0	2,2,2	0.30	0
3	ACT	A	901	-	1,3,3	11.38	1 (100%)	0,3,3	0.00	-
4	EDO	C	901	-	3,3,3	0.52	0	2,2,2	0.20	0
4	EDO	A	904	-	3,3,3	0.48	0	2,2,2	0.39	0
4	EDO	A	905	-	3,3,3	0.38	0	2,2,2	0.32	0
4	EDO	B	1102	-	3,3,3	0.62	0	2,2,2	0.30	0
4	EDO	A	906	-	3,3,3	0.55	0	2,2,2	0.10	0
3	ACT	B	1101	-	1,3,3	7.60	1 (100%)	0,3,3	0.00	-
4	EDO	A	907	-	3,3,3	0.63	0	2,2,2	0.16	0
4	EDO	A	903	-	3,3,3	0.52	0	2,2,2	0.26	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	A	902	-	-	0/1/1/1	-
4	EDO	C	901	-	-	0/1/1/1	-
4	EDO	A	905	-	-	1/1/1/1	-
4	EDO	A	904	-	-	0/1/1/1	-
4	EDO	B	1102	-	-	0/1/1/1	-
4	EDO	A	906	-	-	1/1/1/1	-
4	EDO	A	907	-	-	1/1/1/1	-
4	EDO	A	903	-	-	0/1/1/1	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	901	ACT	CH3-C	11.38	1.63	1.48
3	B	1101	ACT	CH3-C	7.60	1.58	1.48

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	906	EDO	O1-C1-C2-O2
4	A	905	EDO	O1-C1-C2-O2
4	A	907	EDO	O1-C1-C2-O2

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	901	ACT	3	0
4	A	904	EDO	1	0
4	A	905	EDO	4	0
4	A	906	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

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	211/219 (96%)	0.54	10 (4%)	31	30	29, 49, 92, 118	0
1	C	215/219 (98%)	0.59	18 (8%)	11	10	34, 53, 104, 117	0
2	B	66/86 (76%)	0.49	3 (4%)	33	32	33, 43, 67, 104	0
2	D	65/86 (75%)	0.34	3 (4%)	32	31	39, 51, 69, 78	0
All	All	557/610 (91%)	0.53	34 (6%)	21	20	29, 50, 94, 118	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	720	ARG	6.8
1	C	800	GLY	5.2
1	C	801	GLY	5.2
2	B	1081	ALA	5.0
1	C	756	GLN	4.9

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

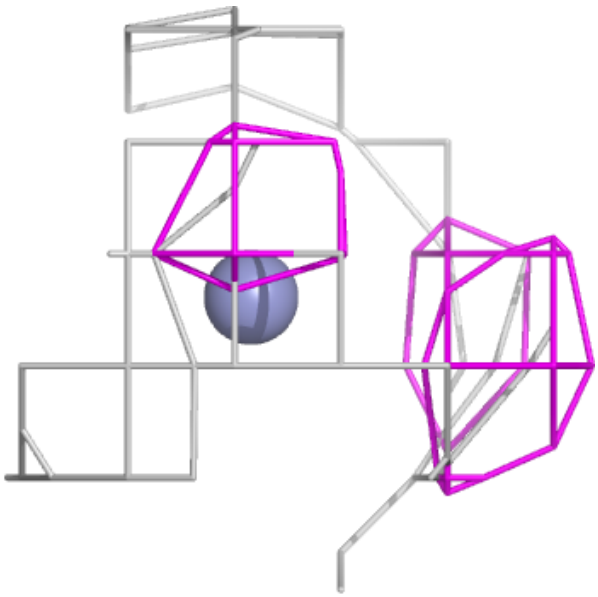
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

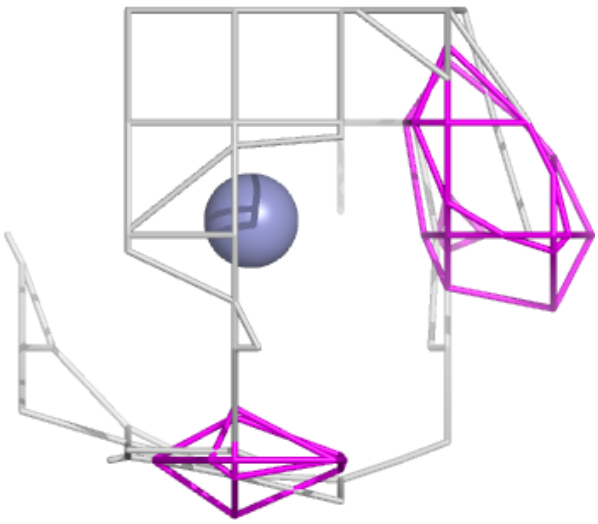
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	EDO	A	904	4/4	0.57	0.34	86,87,89,90	0
3	ACT	A	901	4/4	0.59	0.18	44,57,61,64	0
4	EDO	B	1102	4/4	0.70	0.16	55,62,68,72	0
4	EDO	C	901	4/4	0.78	0.19	74,77,77,78	0
4	EDO	A	907	4/4	0.79	0.14	63,65,67,69	0
4	EDO	A	902	4/4	0.79	0.19	70,71,74,80	0
3	ACT	B	1101	4/4	0.79	0.20	65,67,67,72	0
4	EDO	A	906	4/4	0.81	0.40	63,66,71,75	0
4	EDO	A	905	4/4	0.84	0.31	57,58,60,62	0
4	EDO	A	903	4/4	0.87	0.15	68,72,77,83	0
5	ZN	B	1103	1/1	0.99	0.08	50,50,50,50	0
5	ZN	D	1101	1/1	0.99	0.09	48,48,48,48	0

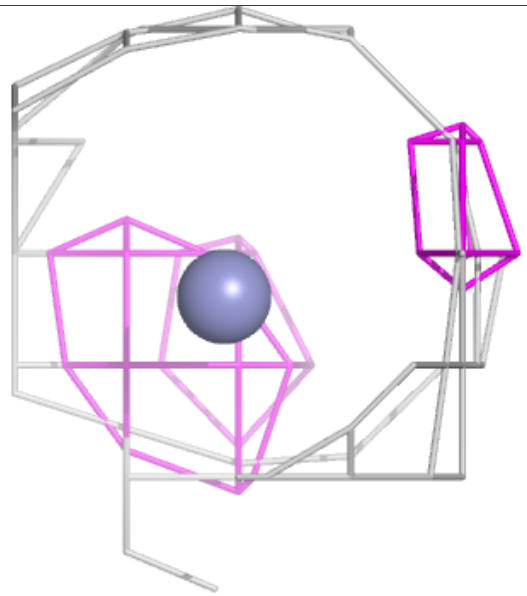
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around ZN B 1103:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

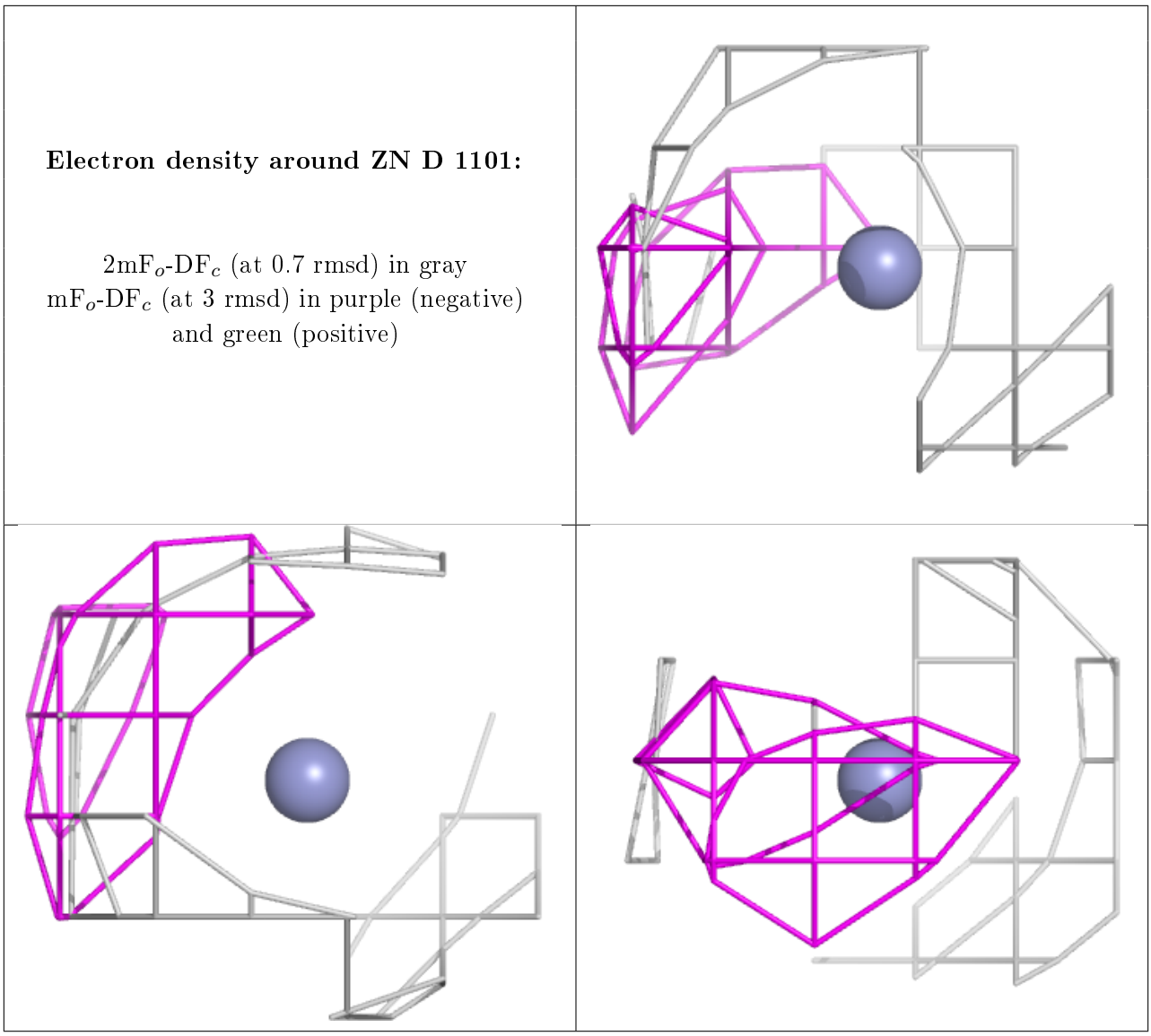






Electron density around ZN D 1101:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.