wwPDB X-ray Structure Validation Summary Report (i)

Jun 26, 2024 – 06:28 AM EDT

PDB ID	:	6YFV
Title	:	Crystal structure of Mtr4-Red1 minimal complex from Chaetomium thermophilum
Authors	:	Dobrev, N.; Ahmed, Y.L.; Sinning, I.
Deposited on	:	2020-03-26
Resolution	:	2.75 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

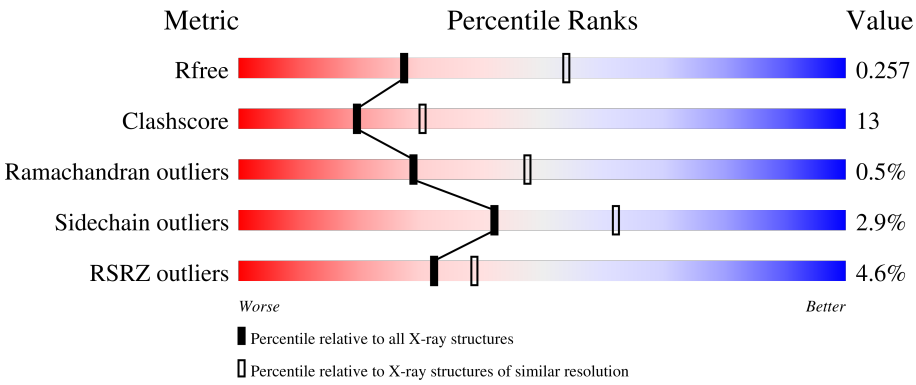
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	213	82% 16% .
1	C	213	4% 64% 31% . .
2	B	89	% 82% 7% 11%
2	D	89	18% 45% 27% 6% 22%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4599 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
			Total	C	N	O	S			
1	A	212	1699	1061	308	319	11	0	0	0
1	C	209	1685	1052	307	315	11	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	653	MET	-	initiating methionine	UNP G0RZ64
C	653	MET	-	initiating methionine	UNP G0RZ64

- Molecule 2 is a protein called Red1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	79	637	406	109	118	4	0	0	0
2	D	69	561	359	96	101	5	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1011	GLY	-	expression tag	UNP G0S1V1
B	1012	ALA	-	expression tag	UNP G0S1V1
B	1013	MET	-	expression tag	UNP G0S1V1
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

Continued on next page...

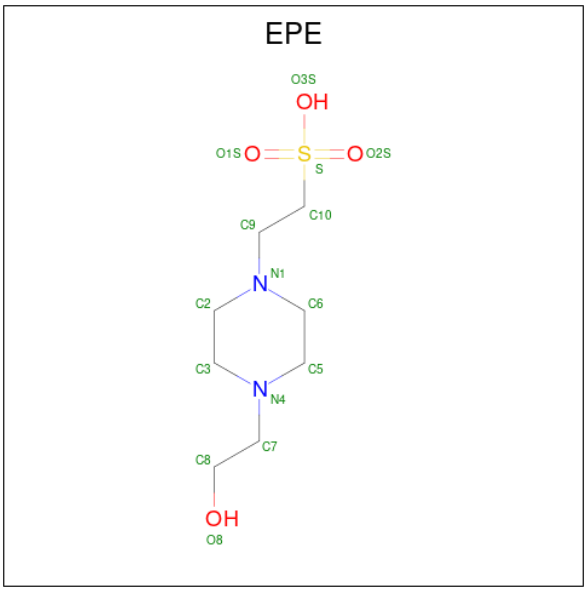
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	1098	HIS	-	expression tag	UNP G0S1V1
B	1099	HIS	-	expression tag	UNP G0S1V1
D	1011	GLY	-	expression tag	UNP G0S1V1
D	1012	ALA	-	expression tag	UNP G0S1V1
D	1013	MET	-	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 ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 4 is 4-(2-HYDROXYETHYL)-1-PIPERAZINE ETHANESULFONIC ACID (three-letter code: EPE) (formula: $C_8H_{18}N_2O_4S$).



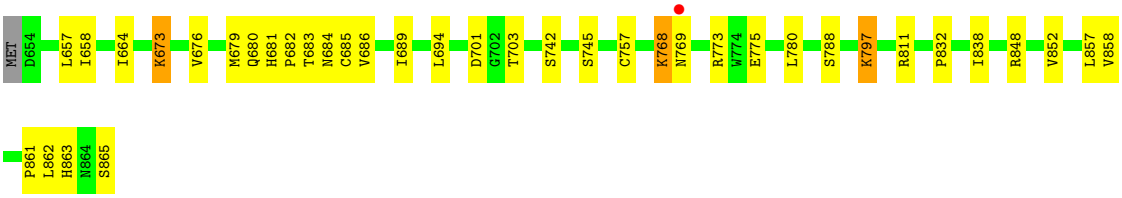
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
4	C	1	15	8	2	4	1	0	0

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.

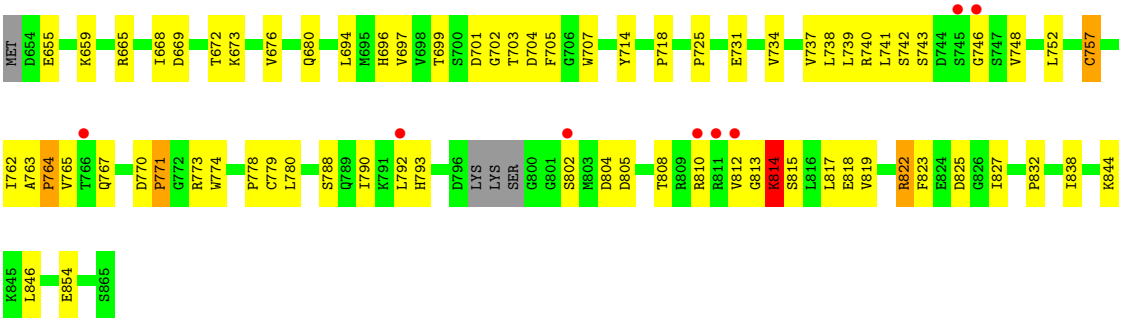
• Molecule 1: ATP dependent RNA helicase (Dob1)-like protein

Chain A:



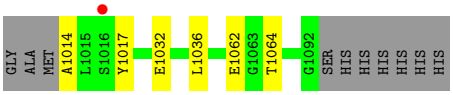
• Molecule 1: ATP dependent RNA helicase (Dob1)-like protein

Chain C:



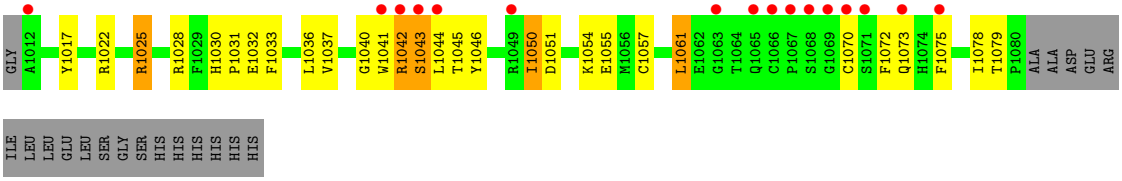
• Molecule 2: Red1

Chain B:



• Molecule 2: Red1

Chain D:


GLY
A1012
Y1017
R1022
R1025
R1028
F1029
H1030
P1031
E1032
F1033
L1036
V1037
G1040
W1041
R1042
S1043
L1044
T1045
Y1046
R1049
I1050
D1051
K1054
E1055
M1056
C1057
L1061
E1062
G1063
T1064
Q1065
C1066
P1067
S1068
G1069
C1070
S1071
F1072
Q1073
H1074
F1075
I1078
T1079
P1080
ALA
ALA
ASP
GLU
ARG
ILE
LEU
LEU
GLU
LEU
SER
GLY
SER
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	172.19Å 172.19Å 145.08Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.71 – 2.75 49.71 – 2.75	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.71-2.75) 100.0 (49.71-2.75)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.65 (at 2.77Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.215 , 0.257 0.215 , 0.257	Depositor DCC
R_{free} test set	1705 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	86.7	Xtriage
Anisotropy	0.206	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 67.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	4599	wwPDB-VP
Average B, all atoms (Å^2)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.08% 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: ZN, EPE

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	0.48	0/1736	0.75	2/2348 (0.1%)
1	C	0.48	0/1721	0.81	4/2329 (0.2%)
2	B	0.49	0/656	0.59	0/888
2	D	0.41	0/580	0.72	0/785
All	All	0.48	0/4693	0.75	6/6350 (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	C	0	1

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})
1	C	822	ARG	NE-CZ-NH2	-11.76	114.42	120.30
1	A	673	LYS	CB-CG-CD	6.54	128.61	111.60
1	C	822	ARG	NE-CZ-NH1	6.35	123.48	120.30
1	C	757	CYS	CA-CB-SG	6.15	125.06	114.00
1	A	673	LYS	CD-CE-NZ	5.95	125.39	111.70

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	814	LYS	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	1699	0	1689	25	0
1	C	1685	0	1667	69	0
2	B	637	0	601	4	0
2	D	561	0	524	30	0
3	B	1	0	0	0	0
3	D	1	0	0	0	0
4	C	15	0	17	1	0
All	All	4599	0	4498	116	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 13.

The worst 5 of 116 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:C:774:TRP:HD1	1:C:810:ARG:HD3	1.36	0.89
1:C:774:TRP:CD1	1:C:810:ARG:HD3	2.08	0.88
1:C:822:ARG:O	2:D:1044:LEU:HD13	1.77	0.84
1:A:768:LYS:HA	1:A:768:LYS:HE2	1.65	0.79
1:C:668:ILE:O	1:C:672:THR:HG23	1.84	0.77

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	210/213 (99%)	208 (99%)	2 (1%)	0	100	100
1	C	206/213 (97%)	199 (97%)	6 (3%)	1 (0%)	29	47
2	B	77/89 (86%)	77 (100%)	0	0	100	100
2	D	67/89 (75%)	53 (79%)	12 (18%)	2 (3%)	4	6
All	All	560/604 (93%)	537 (96%)	20 (4%)	3 (0%)	29	47

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	764	PRO
2	D	1050	ILE
2	D	1070	CYS

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	195/196 (100%)	192 (98%)	3 (2%)	65	78
1	C	193/196 (98%)	187 (97%)	6 (3%)	40	60
2	B	70/78 (90%)	69 (99%)	1 (1%)	67	79
2	D	62/78 (80%)	56 (90%)	6 (10%)	8	14
All	All	520/548 (95%)	504 (97%)	16 (3%)	42	60

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	1043	SER
2	D	1042	ARG
1	C	779	CYS
2	D	1041	TRP
1	C	771	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	864	ASN
2	B	1065	GLN
1	C	721	ASN
2	D	1073	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 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 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	EPE	C	901	-	15,15,15	0.85	1 (6%)	18,20,20	1.98	4 (22%)

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	EPE	C	901	-	-	8/9/19/19	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	901	EPE	C10-S	2.67	1.81	1.77

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	C	901	EPE	C7-N4-C3	4.23	122.06	111.23
4	C	901	EPE	C5-N4-C3	4.15	118.16	108.83
4	C	901	EPE	C7-N4-C5	3.30	119.67	111.23
4	C	901	EPE	O1S-S-C10	2.97	110.49	106.92

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	901	EPE	C10-C9-N1-C6
4	C	901	EPE	C8-C7-N4-C5
4	C	901	EPE	N4-C7-C8-O8
4	C	901	EPE	S-C10-C9-N1
4	C	901	EPE	C9-C10-S-O1S

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	901	EPE	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(Å^2)	Q<0.9
1	A	212/213 (99%)	0.13	1 (0%)	91	94	56, 80, 131, 152	0
1	C	209/213 (98%)	0.38	8 (3%)	40	48	70, 112, 176, 195	0
2	B	79/89 (88%)	0.06	1 (1%)	77	84	58, 79, 127, 147	0
2	D	69/89 (77%)	1.18	16 (23%)	0	0	94, 157, 196, 204	0
All	All	569/604 (94%)	0.34	26 (4%)	32	39	56, 97, 175, 204	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	1069	GLY	6.4
2	D	1044	LEU	5.9
2	D	1068	SER	5.1
2	D	1042	ARG	4.6
2	D	1071	SER	4.4

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
4	EPE	C	901	15/15	0.71	0.31	90,139,162,166	0
3	ZN	D	1101	1/1	0.96	0.05	190,190,190,190	0
3	ZN	B	1101	1/1	0.98	0.19	84,84,84,84	0

6.5 Other polymers (i)

There are no such residues in this entry.