wwPDB X-ray Structure Validation Summary Report (i)

Sep 10, 2023 – 12:19 AM EDT

PDB ID	:	4HMZ
Title	:	Crystal Structure of ChmJ, a 3'-monoepimerase from Streptomyces bikiniensis in complex with dTDP-quinovose
Authors	:	Holden, H.M.; Kubiak, R.L.
Deposited on	:	2012-10-18
Resolution	:	2.00 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:

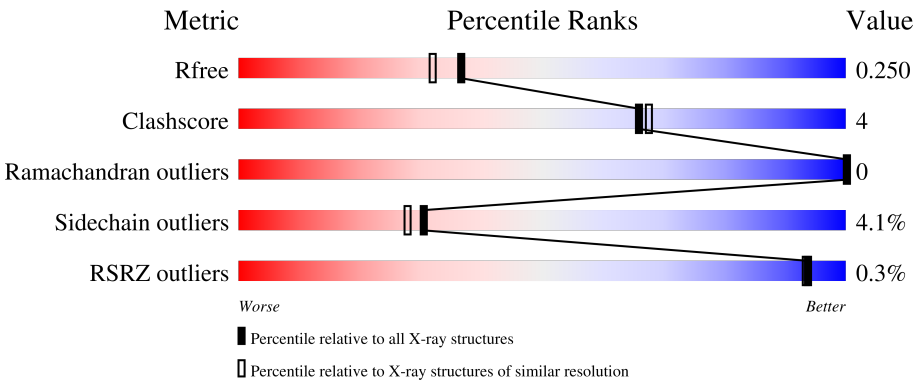
MolProbity	:	4.02b-467
Mogul	:	FAILED
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
buster-report	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.00 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	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	200	88% 10% • •
1	B	200	86% 12%
1	C	200	82% 16% • •
1	D	200	87% 10% •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6861 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.

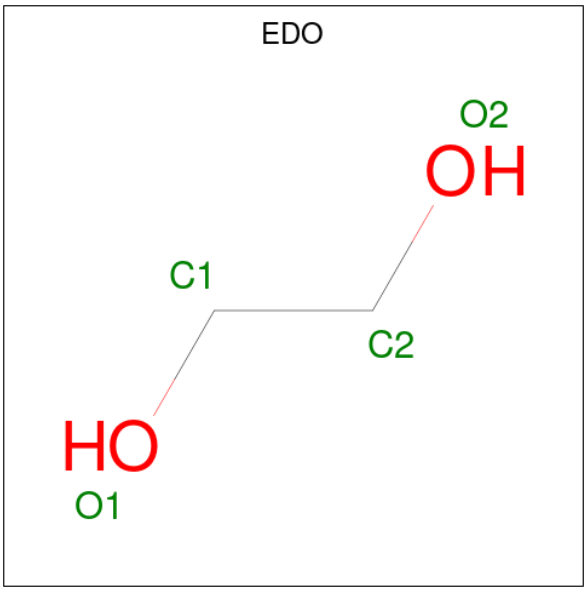
- Molecule 1 is a protein called Putative 3-epimerase in D-allose pathway.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	198	1560	980	283	293	4	0	1	0
1	B	199	1559	980	282	293	4	0	0	0
1	C	200	1593	1002	290	297	4	0	4	0
1	D	200	1596	1004	296	292	4	0	3	0

There are 16 discrepancies between the modelled and reference sequences:

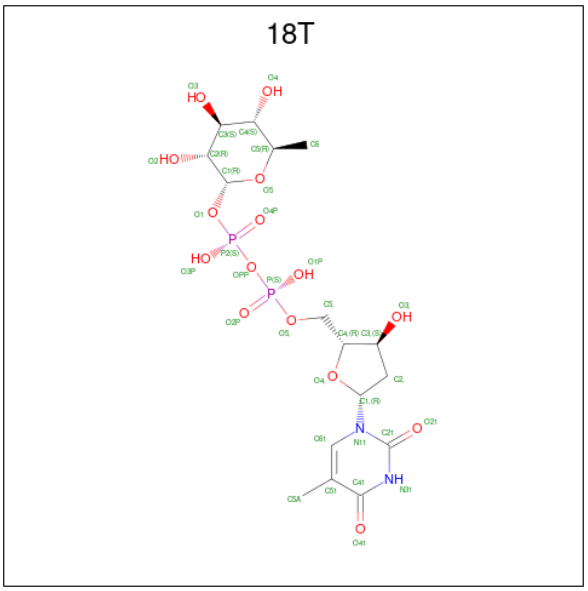
Chain	Residue	Modelled	Actual	Comment	Reference
A	197	LEU	-	expression tag	UNP Q5SFD1
A	198	GLU	-	expression tag	UNP Q5SFD1
A	199	HIS	-	expression tag	UNP Q5SFD1
A	200	HIS	-	expression tag	UNP Q5SFD1
B	197	LEU	-	expression tag	UNP Q5SFD1
B	198	GLU	-	expression tag	UNP Q5SFD1
B	199	HIS	-	expression tag	UNP Q5SFD1
B	200	HIS	-	expression tag	UNP Q5SFD1
C	197	LEU	-	expression tag	UNP Q5SFD1
C	198	GLU	-	expression tag	UNP Q5SFD1
C	199	HIS	-	expression tag	UNP Q5SFD1
C	200	HIS	-	expression tag	UNP Q5SFD1
D	197	LEU	-	expression tag	UNP Q5SFD1
D	198	GLU	-	expression tag	UNP Q5SFD1
D	199	HIS	-	expression tag	UNP Q5SFD1
D	200	HIS	-	expression tag	UNP Q5SFD1

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



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

- Molecule 3 is [(2R,3S,5R)-3-hydroxy-5-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-2-yl]methyl (2R,3R,4S,5S,6R)-3,4,5-trihydroxy-6-methyltetrahydro-2H-pyran-2-yl dihydrogen diphosphate (three-letter code: 18T) (formula: $C_{16}H_{26}N_2O_{15}P_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
3	A	1	35	16	2	15	2	0	0
3	B	1	35	16	2	15	2	0	0
3	C	1	35	16	2	15	2	0	0
3	D	1	35	16	2	15	2	0	0

- Molecule 4 is water.

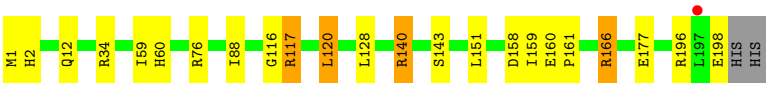
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	123	123	123	0	0
4	B	99	99	99	0	0
4	C	79	79	79	0	0
4	D	84	84	84	0	0

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: Putative 3-epimerase in D-allose pathway

Chain A:

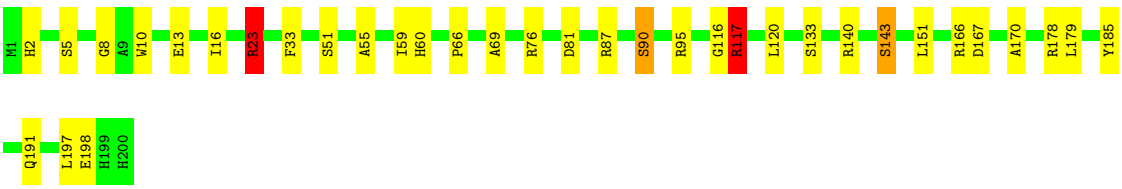


• Molecule 1: Putative 3-epimerase in D-allose pathway

Chain B:

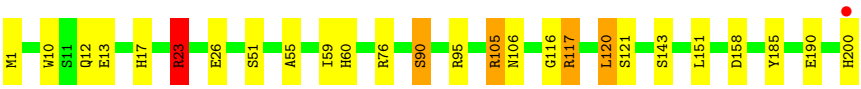
• Molecule 1: Putative 3-epimerase in D-allose pathway

Chain C:



• Molecule 1: Putative 3-epimerase in D-allose pathway

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α, β, γ	140.72Å 140.72Å 117.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.00 37.05 – 2.00	Depositor EDS
% Data completeness (in resolution range)	94.9 (30.00-2.00) 94.9 (37.05-2.00)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.71 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.199 , 0.249 0.201 , 0.250	Depositor DCC
R_{free} test set	3693 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	18.9	Xtriage
Anisotropy	0.109	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 17.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.488 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6861	wwPDB-VP
Average B, all atoms (Å^2)	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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: EDO, 18T

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	0.49	0/1603	1.17	9/2188 (0.4%)
1	B	0.48	0/1603	1.15	10/2189 (0.5%)
1	C	0.45	0/1651	1.11	8/2254 (0.4%)
1	D	0.45	0/1654	1.08	7/2257 (0.3%)
All	All	0.47	0/6511	1.13	34/8888 (0.4%)

There are no bond length outliers.

The worst 5 of 34 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	23[A]	ARG	NE-CZ-NH1	8.20	124.40	120.30
1	C	23[B]	ARG	NE-CZ-NH1	8.20	124.40	120.30
1	C	87	ARG	NE-CZ-NH1	-7.70	116.45	120.30
1	A	34	ARG	NE-CZ-NH1	7.40	124.00	120.30
1	B	34	ARG	NE-CZ-NH1	6.82	123.71	120.30

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	1560	0	1493	10	0
1	B	1559	0	1488	11	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1593	0	1526	17	0
1	D	1596	0	1537	15	0
2	A	8	0	12	0	0
2	B	4	0	6	0	0
2	C	8	0	12	1	0
2	D	8	0	12	0	0
3	A	35	0	24	0	0
3	B	35	0	24	0	0
3	C	35	0	24	0	0
3	D	35	0	24	0	0
4	A	123	0	0	2	0
4	B	99	0	0	2	0
4	C	79	0	0	2	0
4	D	84	0	0	1	0
All	All	6861	0	6182	51	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 51 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:D:23[C]:ARG:NH2	1:D:106:ASN:OD1	2.14	0.80
1:B:70:LYS:HE3	4:B:404:HOH:O	1.87	0.73
1:A:117:ARG:HG2	1:A:117:ARG:O	1.90	0.63
1:B:111:LEU:CD2	1:B:117:ARG:HB3	2.29	0.61
1:C:76:ARG:HD2	1:D:76:ARG:HD2	1.84	0.58

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	197/200 (98%)	194 (98%)	3 (2%)	0	100	100
1	B	197/200 (98%)	195 (99%)	2 (1%)	0	100	100
1	C	202/200 (101%)	198 (98%)	4 (2%)	0	100	100
1	D	202/200 (101%)	197 (98%)	5 (2%)	0	100	100
All	All	798/800 (100%)	784 (98%)	14 (2%)	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	163/164 (99%)	158 (97%)	5 (3%)	40	40
1	B	163/164 (99%)	159 (98%)	4 (2%)	47	49
1	C	168/164 (102%)	156 (93%)	12 (7%)	14	10
1	D	167/164 (102%)	158 (95%)	9 (5%)	22	18
All	All	661/656 (101%)	631 (96%)	30 (4%)	30	24

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	120	LEU
1	D	151	LEU
1	C	151	LEU
1	D	200	HIS
1	D	90	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	12	GLN
1	B	12	GLN
1	C	25	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	191	GLN
1	D	12	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)

Mogul failed to run properly - this section is therefore empty.

5.5 Carbohydrates (i)

Mogul failed to run properly - this section is therefore empty.

5.6 Ligand geometry (i)

Mogul failed to run properly - this section is therefore empty.

5.7 Other polymers (i)

Mogul failed to run properly - this section is therefore empty.

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	198/200 (99%)	-0.77	1 (0%) 91 90	7, 19, 36, 52	0
1	B	199/200 (99%)	-0.76	0 100 100	9, 19, 35, 56	0
1	C	200/200 (100%)	-0.69	0 100 100	11, 23, 43, 65	0
1	D	200/200 (100%)	-0.67	1 (0%) 91 90	11, 24, 42, 66	0
All	All	797/800 (99%)	-0.72	2 (0%) 94 93	7, 21, 40, 66	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	LEU	2.7
1	D	200	HIS	2.1

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
2	EDO	B	302	4/4	0.95	0.08	16,17,20,21	0
2	EDO	A	302	4/4	0.96	0.07	18,20,21,21	0
2	EDO	A	301	4/4	0.98	0.07	16,18,21,23	0
2	EDO	C	303	4/4	0.98	0.06	20,21,22,23	0
2	EDO	D	301	4/4	0.98	0.06	20,22,22,23	0
3	18T	C	302	35/35	0.98	0.08	18,24,29,29	0
3	18T	D	303	35/35	0.98	0.07	18,23,26,30	0
3	18T	A	303	35/35	0.99	0.07	17,21,25,27	0
3	18T	B	301	35/35	0.99	0.08	12,16,21,22	0
2	EDO	C	301	4/4	0.99	0.08	12,16,20,25	0
2	EDO	D	302	4/4	0.99	0.07	11,15,17,23	0

6.5 Other polymers (i)

There are no such residues in this entry.