



wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 07:30 pm BST

PDB ID : 5UHL
Title : Crystal structure of the core catalytic domain of human O-GlcNAcase complexed with Thiamet G
Authors : Klein, D.J.; Elsen, N.L.
Deposited on : 2017-01-11
Resolution : 3.14 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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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 : 2.11
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) : 2.11

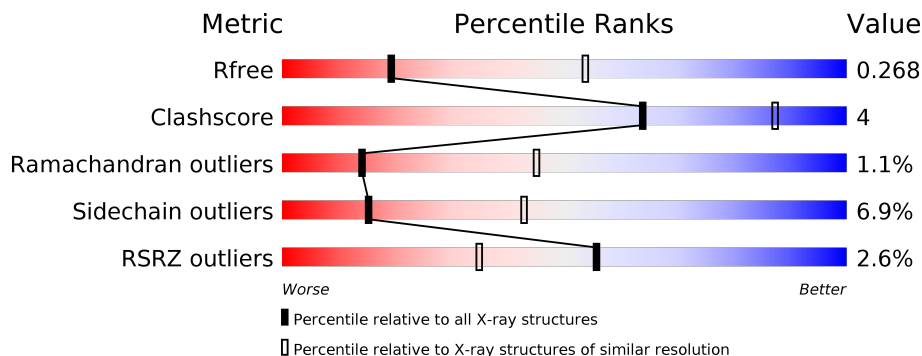
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.14 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	1626 (3.18-3.10)
Clashscore	141614	1735 (3.18-3.10)
Ramachandran outliers	138981	1677 (3.18-3.10)
Sidechain outliers	138945	1677 (3.18-3.10)
RSRZ outliers	127900	1588 (3.18-3.10)

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 on 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 $\leq 5\%$. 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	345	
1	C	345	
2	B	163	
2	D	163	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6951 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 O-GlcNAcase TIM-barrel domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	304	2498	1616	413	455	14	0	0	0
1	C	302	2471	1601	405	451	14	0	0	0

- Molecule 2 is a protein called O-GlcNAcase stalk domain.

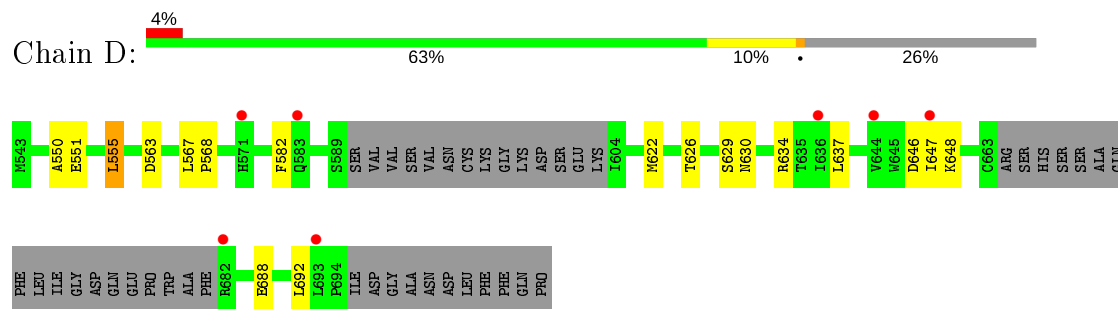
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	120	970	627	160	172	11	0	0	0
2	D	120	973	629	160	173	11	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	543	MET	-	initiating methionine	UNP O60502
D	543	MET	-	initiating methionine	UNP O60502

- Molecule 3 is (2Z,3aR,5R,6S,7R,7aR)-2-(ethylimino)-5-(hydroxymethyl)hexahydro-3aH-pyrano[3,2-d][1,3]thiazole-6,7-diol (three-letter code: 8BJ) (formula: C₉H₁₆N₂O₄S).

- Molecule 2: O-GlcNAcase stalk domain



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	98.20Å 98.20Å 261.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	91.93 – 3.14 91.93 – 3.14	Depositor EDS
% Data completeness (in resolution range)	99.9 (91.93-3.14) 100.0 (91.93-3.14)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.22 (at 3.13Å)	Xtrriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.192 , 0.257 0.207 , 0.268	Depositor DCC
R_{free} test set	1167 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	95.6	Xtrriage
Anisotropy	0.475	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 87.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	6951	wwPDB-VP
Average B, all atoms (Å ²)	103.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.17% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: 8BJ

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	0.55	0/2565	0.75	1/3475 (0.0%)
1	C	0.56	0/2538	0.73	2/3441 (0.1%)
2	B	0.48	0/993	0.65	0/1337
2	D	0.50	0/996	0.68	0/1341
All	All	0.54	0/7092	0.72	3/9594 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	278	TRP	N-CA-C	-5.39	96.45	111.00
1	C	286	TYR	C-N-CA	5.32	135.00	121.70
1	C	278	TRP	N-CA-C	-5.16	97.08	111.00

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	2498	0	2453	23	0
1	C	2471	0	2420	19	0
2	B	970	0	948	17	0
2	D	973	0	949	6	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	16	0	0	0	0
3	C	16	0	0	0	0
4	A	3	0	0	0	0
4	B	1	0	0	0	0
4	C	3	0	0	0	0
All	All	6951	0	6770	58	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 58 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:111:TYR:H	1:A:160:GLN:HE22	1.26	0.83
2:B:617:MET:HA	2:B:620:LEU:HD12	1.70	0.71
1:C:375:SER:HB3	1:C:378:MET:HB2	1.77	0.66
1:A:384:LEU:HD22	2:B:558:LEU:HB3	1.78	0.66
2:B:624:MET:HG2	2:B:628:LEU:HD12	1.79	0.64

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	300/345 (87%)	275 (92%)	22 (7%)	3 (1%)	15	47
1	C	298/345 (86%)	273 (92%)	20 (7%)	5 (2%)	9	34
2	B	114/163 (70%)	106 (93%)	7 (6%)	1 (1%)	17	50
2	D	114/163 (70%)	108 (95%)	6 (5%)	0	100	100
All	All	826/1016 (81%)	762 (92%)	55 (7%)	9 (1%)	14	45

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	140	GLY
1	C	287	ASP
1	C	140	GLY
1	C	189	SER
1	C	288	GLN

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	272/310 (88%)	255 (94%)	17 (6%)	18	46
1	C	268/310 (86%)	252 (94%)	16 (6%)	19	48
2	B	102/141 (72%)	94 (92%)	8 (8%)	12	38
2	D	102/141 (72%)	92 (90%)	10 (10%)	8	27
All	All	744/902 (82%)	693 (93%)	51 (7%)	15	43

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	654	VAL
1	C	137	ILE
2	D	646	ASP
1	C	65	VAL
1	C	178	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	583	GLN
1	C	288	GLN
1	C	163	GLN
1	A	193	HIS
1	C	193	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

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
3	8BJ	A	501	-	14,17,17	0.96	1 (7%)	15,24,24	1.74	2 (13%)
3	8BJ	C	501	-	14,17,17	0.94	1 (7%)	15,24,24	1.79	2 (13%)

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
3	8BJ	A	501	-	-	0/4/33/33	0/2/2/2
3	8BJ	C	501	-	-	0/4/33/33	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	501	8BJ	C5-N1	3.14	1.34	1.26

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	501	8BJ	C5-N1	3.12	1.34	1.26

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	501	8BJ	S-C5-N	5.76	117.95	111.26
3	A	501	8BJ	S-C5-N	4.89	116.93	111.26
3	A	501	8BJ	C8-O3-C1	3.75	119.50	112.58
3	C	501	8BJ	C8-O3-C1	2.75	117.64	112.58

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	304/345 (88%)	0.29	3 (0%) 82 70	67, 96, 125, 181	0
1	C	302/345 (87%)	0.23	5 (1%) 70 51	71, 101, 131, 144	0
2	B	120/163 (73%)	0.32	7 (5%) 23 10	73, 111, 163, 189	0
2	D	120/163 (73%)	0.42	7 (5%) 23 10	74, 105, 146, 162	0
All	All	846/1016 (83%)	0.29	22 (2%) 56 35	67, 101, 139, 189	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	691	ARG	3.8
1	C	289	LYS	3.4
2	B	685	LEU	3.4
2	B	692	LEU	3.3
1	C	171	LEU	2.8

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 carbohydrates 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(\AA^2)	Q<0.9
3	8BJ	A	501	16/16	0.97	0.23	80,80,80,80	0
3	8BJ	C	501	16/16	0.98	0.29	87,87,87,87	0

6.5 Other polymers [i](#)

There are no such residues in this entry.