



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 6, 2020 – 08:39 AM BST

PDB ID : 6EJ8
Title : Human Xylosyltransferase 1 in complex with peptide QEEEGSGGGQGG
Authors : Briggs, D.C.; Hohenester, E.
Deposited on : 2017-09-20
Resolution : 2.09 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 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 : 2.13.1
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) : 2.13.1

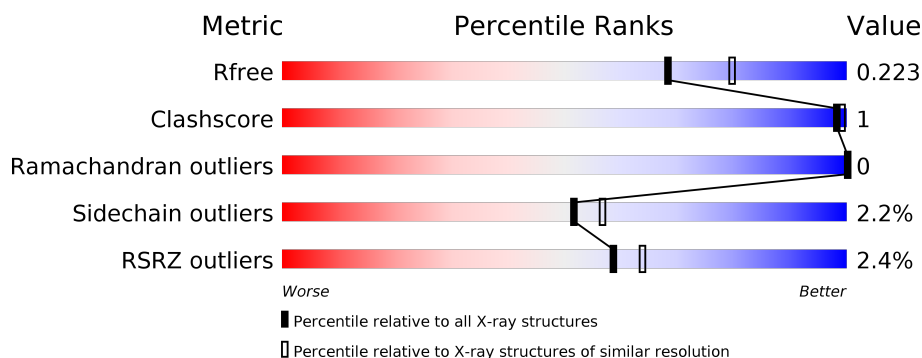
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.09 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	755	<div> <div>2%</div> <div> <div></div> <div>90%</div> <div>7%</div> </div> </div>
2	B	12	<div> <div>75%</div> <div>25%</div> </div>

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	NAG	A	1001	-	-	-	X

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11282 atoms, of which 5392 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 Xylosyltransferase 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	703	Total	C	H	N	O	S	0	0	0
			10936	3580	5353	972	999	32			

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	205	ARG	-	expression tag	UNP Q86Y38
A	206	ALA	-	expression tag	UNP Q86Y38
A	207	LEU	-	expression tag	UNP Q86Y38
A	208	ALA	-	expression tag	UNP Q86Y38
A	209	ALA	-	expression tag	UNP Q86Y38
A	210	PRO	-	expression tag	UNP Q86Y38
A	211	LEU	-	expression tag	UNP Q86Y38
A	212	VAL	-	expression tag	UNP Q86Y38
A	213	HIS	-	expression tag	UNP Q86Y38
A	214	HIS	-	expression tag	UNP Q86Y38
A	215	HIS	-	expression tag	UNP Q86Y38
A	216	HIS	-	expression tag	UNP Q86Y38
A	217	HIS	-	expression tag	UNP Q86Y38
A	218	HIS	-	expression tag	UNP Q86Y38
A	219	ALA	-	expression tag	UNP Q86Y38
A	220	LEU	-	expression tag	UNP Q86Y38
A	221	ASP	-	expression tag	UNP Q86Y38
A	222	GLU	-	expression tag	UNP Q86Y38
A	223	ASN	-	expression tag	UNP Q86Y38
A	224	LEU	-	expression tag	UNP Q86Y38
A	225	TYR	-	expression tag	UNP Q86Y38
A	226	PHE	-	expression tag	UNP Q86Y38
A	227	GLN	-	expression tag	UNP Q86Y38
A	228	GLY	-	expression tag	UNP Q86Y38
A	229	ALA	-	expression tag	UNP Q86Y38
A	230	LEU	-	expression tag	UNP Q86Y38
A	231	ALA	-	expression tag	UNP Q86Y38

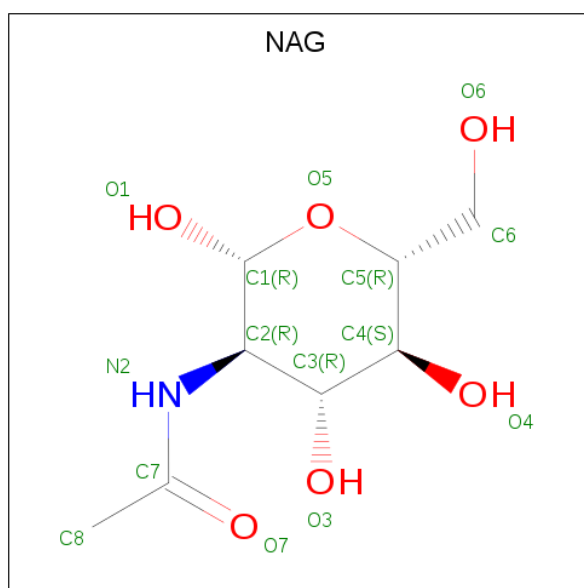
- Molecule 2 is a protein called Protein AMBP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	9	Total	C	H	N	O	0	0	0
			76	27	26	10	13			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	220	GLY	LEU	engineered mutation	UNP P02760
B	221	GLY	VAL	engineered mutation	UNP P02760

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

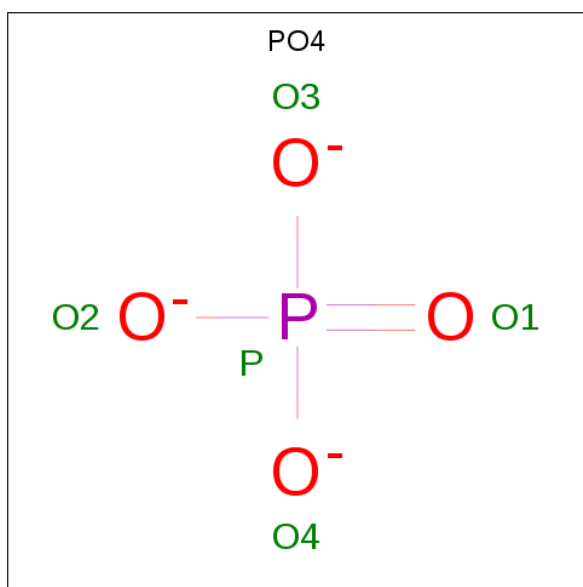


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	H	N	O	0	0
			27	8	13	1	5		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Na	0	0
			1	1		

- Molecule 5 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	P	0	0
			5	4	1		
5	A	1	Total	O	P	0	0
			5	4	1		

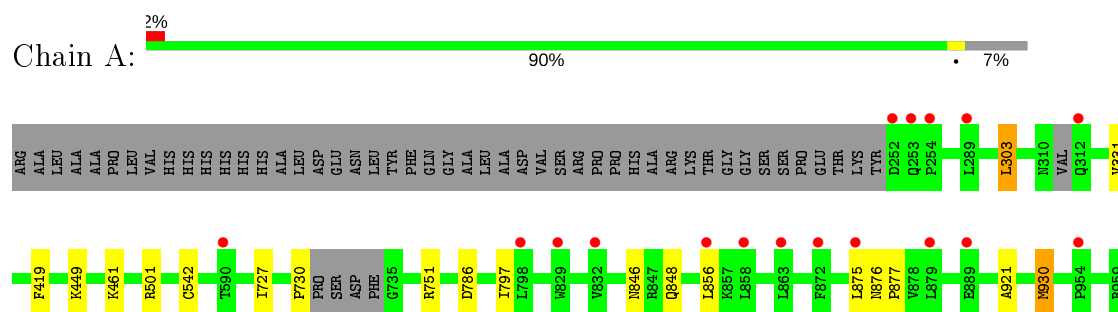
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	231	Total	O	0	0
			231	231		
6	B	1	Total	O	0	0
			1	1		

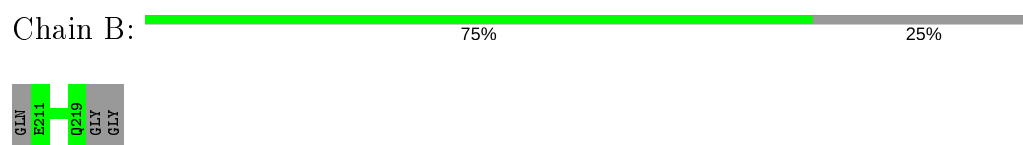
3 Residue-property plots [i](#)

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: Xylosyltransferase 1



- Molecule 2: Protein AMBP



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	67.31Å 86.68Å 152.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	57.33 – 2.09 57.33 – 2.09	Depositor EDS
% Data completeness (in resolution range)	99.6 (57.33-2.09) 99.6 (57.33-2.09)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.54 (at 2.08Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.195 , 0.223 0.195 , 0.223	Depositor DCC
R_{free} test set	2612 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å ²)	36.2	Xtriage
Anisotropy	0.613	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 47.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	11282	wwPDB-VP
Average B, all atoms (Å ²)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.83% 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: NA, PO4, NAG

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	0.25	0/5746	0.45	1/7821 (0.0%)
2	B	0.25	0/49	0.47	0/63
All	All	0.25	0/5795	0.45	1/7884 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	730	PRO	N-CA-CB	5.85	110.32	103.30

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	5583	5353	5342	4	0
2	B	50	26	34	0	0
3	A	14	13	13	3	0
4	A	1	0	0	0	0
5	A	10	0	0	0	0
6	A	231	0	0	0	0
6	B	1	0	0	0	0
All	All	5890	5392	5389	7	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 1.

All (7) 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)
3:A:1001:NAG:H83	3:A:1001:NAG:H3	1.86	0.56
3:A:1001:NAG:C1	3:A:1001:NAG:H82	2.42	0.50
1:A:921:ALA:HB2	1:A:930:MET:HE3	1.96	0.48
1:A:331:VAL:HG21	1:A:419:PHE:CE2	2.54	0.43
1:A:876:ASN:HB2	1:A:877:PRO:HD3	2.01	0.43
3:A:1001:NAG:C1	3:A:1001:NAG:C8	2.98	0.41
1:A:303:LEU:HD22	1:A:542:CYS:SG	2.60	0.40

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	697/755 (92%)	680 (98%)	17 (2%)	0	100	100
2	B	7/12 (58%)	6 (86%)	1 (14%)	0	100	100
All	All	704/767 (92%)	686 (97%)	18 (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	585/658 (89%)	572 (98%)	13 (2%)	52	57
2	B	3/6 (50%)	3 (100%)	0	100	100
All	All	588/664 (89%)	575 (98%)	13 (2%)	52	57

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	303	LEU
1	A	449	LYS
1	A	461	LYS
1	A	501	ARG
1	A	727	ILE
1	A	751	ARG
1	A	786	ASP
1	A	797	ILE
1	A	846	ASN
1	A	848	GLN
1	A	856	LEU
1	A	875	LEU
1	A	930	MET

Some 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 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	$\# Z > 2$
5	PO4	A	1004	-	4,4,4	0.93	0	6,6,6	0.40	0
5	PO4	A	1003	-	4,4,4	0.92	0	6,6,6	0.44	0
3	NAG	A	1001	-	14,14,15	0.21	0	17,19,21	0.90	1 (5%)

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	NAG	A	1001	-	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
3	A	1001	NAG	C2-N2-C7	2.76	126.83	122.90

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1001	NAG	C8-C7-N2-C2
3	A	1001	NAG	O7-C7-N2-C2
3	A	1001	NAG	C1-C2-N2-C7
3	A	1001	NAG	C3-C2-N2-C7

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1001	NAG	3	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, 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	703/755 (93%)	0.04	17 (2%) 59 64	28, 46, 80, 223	0
2	B	9/12 (75%)	-0.11	0 100 100	41, 50, 72, 74	0
All	All	712/767 (92%)	0.04	17 (2%) 59 64	28, 46, 80, 223	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	829	TRP	3.9
1	A	872	PHE	2.8
1	A	856	LEU	2.7
1	A	252	ASP	2.6
1	A	798	LEU	2.3
1	A	879	LEU	2.3
1	A	254	PRO	2.3
1	A	875	LEU	2.3
1	A	590	THR	2.3
1	A	858	LEU	2.2
1	A	863	LEU	2.2
1	A	253	GLN	2.1
1	A	289	LEU	2.1
1	A	312	GLN	2.1
1	A	889	GLU	2.1
1	A	954	PRO	2.0
1	A	832	VAL	2.0

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 [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(\AA^2)	Q<0.9
3	NAG	A	1001	14/15	0.40	0.47	442,476,571,572	0
5	PO4	A	1003	5/5	0.75	0.32	100,102,103,103	0
5	PO4	A	1004	5/5	0.82	0.29	115,116,117,117	0
4	NA	A	1002	1/1	0.97	0.10	40,40,40,40	0

6.5 Other polymers [i](#)

There are no such residues in this entry.