



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 24, 2024 – 09:54 AM EST

PDB ID : 4JA4  
Title : Inward open conformation of the xylose transporter XylE from E. coli  
Authors : Quistgaard, E.M.; Low, C.; Moberg, P.; Tresaugues, L.; Nordlund, P.  
Deposited on : 2013-02-18  
Resolution : 4.20 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](mailto: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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
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) : 2.36

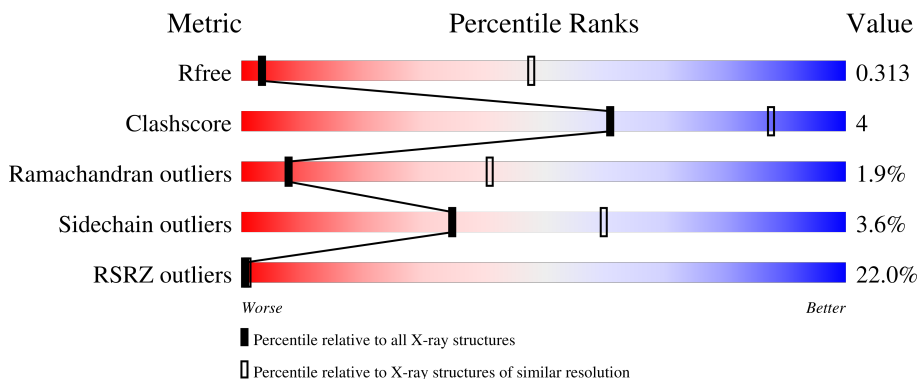
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.20 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	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)
RSRZ outliers	127900	1063 (4.70-3.70)

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  $\geq 3$ , 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	485	 20% 75% 15% • 9%
1	B	485	 18% 76% 13% • 9%
1	C	485	 22% 76% 12% • 10%

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10130 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 D-xylose-proton symporter.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	442	3401	2257	533	587	24	0	0	0
1	B	439	3379	2242	530	583	24	0	0	0
1	C	436	3349	2225	524	576	24	0	0	0

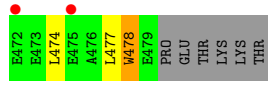
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P0AGF4
B	1	GLY	-	expression tag	UNP P0AGF4
C	1	GLY	-	expression tag	UNP P0AGF4

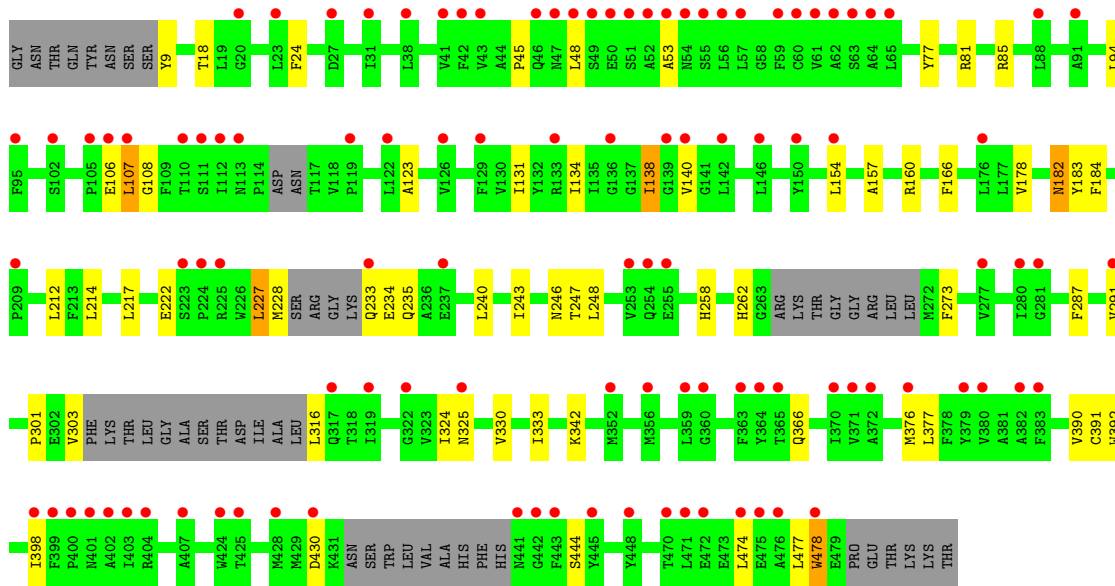
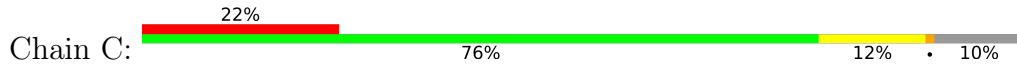
- Molecule 2 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cd	0	0
			1	1		





• Molecule 1: D-xylose-proton symporter



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	125.82Å 125.82Å 380.68Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	41.18 – 4.20 71.68 – 4.20	Depositor EDS
% Data completeness (in resolution range)	72.2 (41.18-4.20) 72.4 (71.68-4.20)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.14 (at 4.15Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.288 , 0.318 0.288 , 0.313	Depositor DCC
$R_{free}$ test set	1374 reflections (7.66%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	161.4	Xtrriage
Anisotropy	0.806	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 139.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.25$	Xtrriage
Estimated twinning fraction	0.085 for h,-h-k,-l	Xtrriage
$F_o, F_c$ correlation	0.61	EDS
Total number of atoms	10130	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	116.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.94% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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: CD

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	0.22	0/3488	0.37	0/4745
1	B	0.22	0/3465	0.37	0/4710
1	C	0.22	0/3433	0.37	0/4667
All	All	0.22	0/10386	0.37	0/14122

There are no bond length outliers.

There are no bond angle outliers.

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	3401	0	3445	36	0
1	B	3379	0	3418	30	0
1	C	3349	0	3402	26	0
2	A	1	0	0	0	0
All	All	10130	0	10265	91	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 91 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:258:HIS:CE1	1:C:262:HIS:NE2	2.57	0.72
1:B:258:HIS:CE1	1:B:262:HIS:NE2	2.59	0.70
1:A:258:HIS:CE1	1:A:262:HIS:NE2	2.62	0.66
1:B:227:LEU:N	1:B:228:MET:HA	2.13	0.64
1:A:227:LEU:N	1:A:228:MET:HA	2.13	0.62

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	432/485 (89%)	399 (92%)	24 (6%)	9 (2%)	7	39
1	B	427/485 (88%)	393 (92%)	26 (6%)	8 (2%)	8	41
1	C	424/485 (87%)	394 (93%)	23 (5%)	7 (2%)	9	44
All	All	1283/1455 (88%)	1186 (92%)	73 (6%)	24 (2%)	8	41

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	107	LEU
1	A	117	THR
1	A	391	CYS
1	B	107	LEU
1	B	391	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	355/390 (91%)	344 (97%)	11 (3%)	40	62
1	B	352/390 (90%)	339 (96%)	13 (4%)	34	59
1	C	349/390 (90%)	335 (96%)	14 (4%)	31	57
All	All	1056/1170 (90%)	1018 (96%)	38 (4%)	35	60

5 of 38 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	183	TYR
1	C	376	MET
1	C	184	PHE
1	C	303	VAL
1	C	478	TRP

Sometimes 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 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	442/485 (91%)	1.13	95 (21%) 0 1	32, 111, 172, 239	0
1	B	439/485 (90%)	1.03	88 (20%) 1 1	29, 112, 170, 237	0
1	C	436/485 (89%)	1.08	107 (24%) 0 1	33, 113, 170, 243	0
All	All	1317/1455 (90%)	1.08	290 (22%) 0 1	29, 112, 171, 243	0

The worst 5 of 290 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	50	GLU	11.4
1	A	53	ALA	9.3
1	A	52	ALA	9.3
1	A	54	ASN	9.1
1	C	471	LEU	8.8

### 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 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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
2	CD	A	501	1/1	0.92	0.23	297,297,297,297	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.