



# wwPDB X-ray Structure Validation Summary Report

Feb 13, 2017 – 10:15 pm GMT

PDB ID : 4KJR  
Title : Crystal structure of selenium substituted Ca<sup>2+</sup>/H<sup>+</sup> antiporter proteinYfkE  
Authors : Wu, M.; Tong, S.; Zheng, L.  
Deposited on : 2013-05-03  
Resolution : 3.00 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

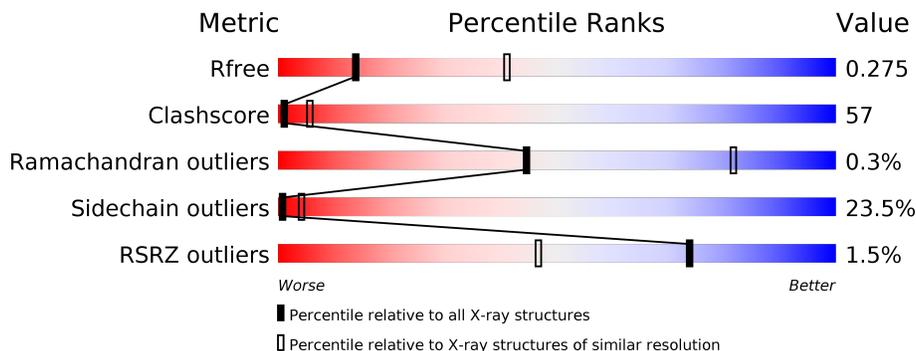
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	351	 % <span style="margin-left: 20px;">39%</span> <span style="margin-left: 20px;">39%</span> <span style="margin-left: 20px;">14%</span> <span style="margin-left: 20px;">• 7%</span>
1	B	351	 % <span style="margin-left: 20px;">37%</span> <span style="margin-left: 20px;">39%</span> <span style="margin-left: 20px;">12%</span> <span style="margin-left: 20px;">11%</span>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4785 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 cation exchanger YfkE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	325	2441	1628	379	420	2	12	0	0	0
1	B	311	2332	1560	360	398	2	12	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	77	MSE	LEU	ENGINEERED MUTATION	UNP O34840
A	116	ALA	LYS	ENGINEERED MUTATION	UNP O34840
B	77	MSE	LEU	ENGINEERED MUTATION	UNP O34840
B	116	ALA	LYS	ENGINEERED MUTATION	UNP O34840

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	8	Total	O	0	0
			8	8		
2	B	4	Total	O	0	0
			4	4		



## 4 Data and refinement statistics i

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	170.30Å 170.30Å 95.27Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	85.16 – 3.00 48.11 – 2.99	Depositor EDS
% Data completeness (in resolution range)	99.6 (85.16-3.00) 98.9 (48.11-2.99)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.96 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.239 , 0.268 0.250 , 0.275	Depositor DCC
$R_{free}$ test set	1058 reflections (5.40%)	DCC
Wilson B-factor (Å <sup>2</sup> )	89.6	Xtriage
Anisotropy	0.125	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.19 , 11.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.42$ , $\langle L^2 \rangle = 0.24$	Xtriage
Estimated twinning fraction	0.448 for h,-h-k,-l	Xtriage
Reported twinning fraction	0.531 for H, K, L 0.469 for K, H, -L	Depositor
Outliers	0 of 20634 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	4785	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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](#)

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	0.46	4/2478 (0.2%)	0.58	1/3352 (0.0%)
1	B	0.45	1/2367 (0.0%)	0.59	0/3202
All	All	0.46	5/4845 (0.1%)	0.58	1/6554 (0.0%)

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	A	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	237	TRP	CD2-CE2	5.23	1.47	1.41
1	A	201	TRP	CD2-CE2	5.15	1.47	1.41
1	A	331	TRP	CD2-CE2	5.13	1.47	1.41
1	A	24	TRP	CD2-CE2	5.08	1.47	1.41
1	B	331	TRP	CD2-CE2	5.02	1.47	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	101	LEU	N-CA-C	-5.37	96.51	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	100	LEU	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	2441	0	2554	304	0
1	B	2332	0	2449	256	0
2	A	8	0	0	0	0
2	B	4	0	0	0	0
All	All	4785	0	5003	559	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 57.

The worst 5 of 559 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:78:PHE:CE2	1:A:225:VAL:HG13	1.20	1.63
1:A:30:PHE:HE1	1:A:228:PHE:CE2	1.25	1.53
1:A:30:PHE:CE1	1:A:228:PHE:CE2	1.95	1.52
1:A:30:PHE:CE1	1:A:228:PHE:HE2	1.24	1.51
1:A:78:PHE:CE2	1:A:225:VAL:CG1	1.95	1.49

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	319/351 (91%)	295 (92%)	24 (8%)	0	100 100
1	B	305/351 (87%)	286 (94%)	17 (6%)	2 (1%)	25 67

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	624/702 (89%)	581 (93%)	41 (7%)	2 (0%)	44 81

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	24	TRP
1	B	251	GLY

### 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	259/267 (97%)	200 (77%)	59 (23%)	1 5
1	B	247/267 (92%)	187 (76%)	60 (24%)	1 4
All	All	506/534 (95%)	387 (76%)	119 (24%)	1 4

5 of 119 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	326	ASP
1	B	77	MSE
1	B	314	VAL
1	A	332	PHE
1	B	9	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	266	ASN
1	B	69	ASN
1	B	256	HIS
1	A	256	HIS
1	B	252	ASN

### 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](#)

There are no ligands in this entry.

### 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, 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	313/351 (89%)	-0.45	5 (1%) 72 44	50, 90, 131, 157	0
1	B	299/351 (85%)	-0.47	4 (1%) 77 51	47, 87, 123, 143	0
All	All	612/702 (87%)	-0.46	9 (1%) 74 47	47, 89, 126, 157	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	211	PHE	4.4
1	A	266	ASN	3.2
1	B	7	ILE	3.1
1	A	127	HIS	2.7
1	B	5	PHE	2.6

### 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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.