



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:22 AM GMT

PDB ID : 2QG7
Title : Plasmodium vivax ethanolamine kinase Pv091845
Authors : Lunin, V.V.; Wernimont, A.K.; Mulichak, A.; Lew, J.; Wasney, G.; Senisterra, G.; Kozieradzki, I.; Vedadi, M.; Bochkarev, A.; Arrowsmith, C.H.; Sundstrom, M.; Weigelt, J.; Edwards, A.E.; Hui, R.; Hills, T.; Artz, J.; Xiao, T.; Structural Genomics Consortium (SGC)
Deposited on : 2007-06-28
Resolution : 2.41 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

<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
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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)	:	trunk26865

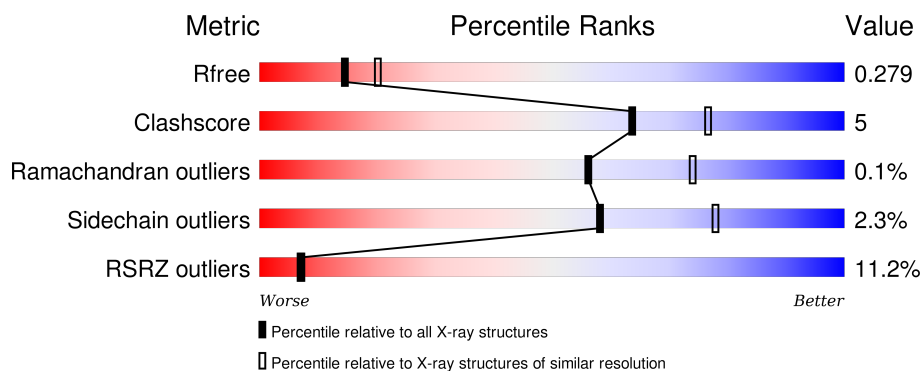
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.41 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}	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

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. 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	458	<div> <div>5%</div> <div>69% 11% 20%</div> </div>
1	B	458	<div> <div>8%</div> <div>65% 14% 21%</div> </div>
1	D	458	<div> <div>15%</div> <div>64% 16% 20%</div> </div>
1	E	458	<div> <div>9%</div> <div>70% 10% 19%</div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 12638 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 ethanolamine kinase Pv091845.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	368	Total	C	N	O	S	0	2	0
			3057	1969	504	567	17			
1	B	363	Total	C	N	O	S	0	1	0
			3021	1944	500	560	17			
1	D	367	Total	C	N	O	S	0	0	0
			3041	1958	503	563	17			
1	E	369	Total	C	N	O	S	0	1	0
			3058	1973	501	566	18			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	Cloning Artifact	UNP A5K4Q6
A	2	SER	-	Cloning Artifact	UNP A5K4Q6
B	1	GLY	-	Cloning Artifact	UNP A5K4Q6
B	2	SER	-	Cloning Artifact	UNP A5K4Q6
D	1	GLY	-	Cloning Artifact	UNP A5K4Q6
D	2	SER	-	Cloning Artifact	UNP A5K4Q6
E	1	GLY	-	Cloning Artifact	UNP A5K4Q6
E	2	SER	-	Cloning Artifact	UNP A5K4Q6

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	162	Total	O	0	0
			162	162		
3	B	101	Total	O	0	0
			101	101		
3	D	87	Total	O	0	0
			87	87		
3	E	106	Total	O	0	0
			106	106		

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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 ($\text{RSRZ} > 2$). 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.

- [illegible]

- Chain B:
-
- 8% 65% 14% 21%
- GLY SER GLU GLN LYS ARG LEU GLU GLY THR ARG ALA ASN SER VAL ALA GLU SER PRO PRO PHE LEU ARG SER LYS THR SER VAL GLY SER THR ASN GLN ILE THR GLU THR LYS ASN LYS GLN GLY VAL THR TYR PRO ILE THR E54 L57 R58 L59 L60 L61
- G62 E63 D64 R65 S66 E67 K68 A69 K70 L73 K74 N79 W80 F81 E82 N83 E84 T86 Y90 H97 N106 E112 I115 I116 N117 G119 I123 L124 V127 L137 I138 R139 L140 Y141 GLY PRO LYS THR D146 E147 I148 E154 I157 S158 L159 L160 L161
- N163 I166 A167 K168 N176 G177 R178 E181 D184 R190 E191 P196 K197 F198 Q199 K200 T212 Y219 K220 E221 L222 T225 S234 R251 L254 C255 T263 L264 K266 L267 V271 C283 L290 L298 V306 G307 GLY GLY GLY GLY
- F436 N443 T444 R445 S449 R455 S456 K457 V458 F352 E353 R366 I391 P392 R393 L406 Q411 F412 F413 V414 I415 H418 H431 S432 S433 I435 P436

- 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	77.03 Å 172.70 Å 93.35 Å 90.00° 110.72° 90.00°	Depositor
Resolution (Å)	24.93 – 2.41 24.93 – 2.41	Depositor EDS
% Data completeness (in resolution range)	96.1 (24.93-2.41) 96.2 (24.93-2.41)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.14 (at 2.41 Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.230 , 0.290 0.223 , 0.279	Depositor DCC
R_{free} test set	4237 reflections (5.27%)	DCC
Wilson B-factor (Å ²)	54.7	Xtriage
Anisotropy	0.072	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 51.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 84689 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	12638	wwPDB-VP
Average B, all atoms (Å ²)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.12% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: SO4

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	0.83	2/3121 (0.1%)	0.72	0/4207
1	B	0.84	2/3082 (0.1%)	0.70	1/4149 (0.0%)
1	D	0.81	3/3105 (0.1%)	0.67	2/4182 (0.0%)
1	E	0.81	4/3121 (0.1%)	0.69	1/4203 (0.0%)
All	All	0.82	11/12429 (0.1%)	0.70	4/16741 (0.0%)

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	213	LYS	CD-CE	10.60	1.77	1.51
1	E	128	LYS	CE-NZ	7.18	1.67	1.49
1	B	85	LYS	CE-NZ	7.17	1.67	1.49
1	D	126	LYS	CE-NZ	6.51	1.65	1.49
1	A	74	LYS	CD-CE	6.49	1.67	1.51

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	E	65	ARG	NE-CZ-NH2	-5.49	117.55	120.30
1	D	213	LYS	CD-CE-NZ	-5.35	99.40	111.70
1	D	213	LYS	CG-CD-CE	-5.23	96.20	111.90
1	B	190	ARG	NE-CZ-NH1	5.11	122.85	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

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	3057	0	3025	30	0
1	B	3021	0	3004	34	0
1	D	3041	0	3024	42	0
1	E	3058	0	3043	23	0
2	A	5	0	0	1	0
3	A	162	0	0	3	1
3	B	101	0	0	0	1
3	D	87	0	0	2	0
3	E	106	0	0	1	0
All	All	12638	0	12096	127	1

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 5.

The worst 5 of 127 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:213:LYS:CE	1:D:213:LYS:CD	1.77	1.61
1:D:213:LYS:CE	1:D:213:LYS:CG	2.54	0.84
1:A:232:ARG:HH21	1:A:232:ARG:HG2	1.45	0.80
1:E:207:LYS:HG3	1:E:394:TYR:O	1.82	0.80
1:A:64:ASP:HB3	1:A:67:GLU:HG2	1.63	0.79

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:655:HOH:O	3:B:520:HOH:O[2_656]	2.13	0.07

5.3 Torsion angles

5.3.1 Protein backbone

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	360/458 (79%)	349 (97%)	11 (3%)	0	100	100
1	B	354/458 (77%)	343 (97%)	11 (3%)	0	100	100
1	D	357/458 (78%)	341 (96%)	15 (4%)	1 (0%)	46	63
1	E	360/458 (79%)	343 (95%)	16 (4%)	1 (0%)	46	63
All	All	1431/1832 (78%)	1376 (96%)	53 (4%)	2 (0%)	56	74

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	133	GLN
1	E	227	LYS

5.3.2 Protein sidechains

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	342/407 (84%)	333 (97%)	9 (3%)	54	74
1	B	340/407 (84%)	333 (98%)	7 (2%)	61	80
1	D	342/407 (84%)	331 (97%)	11 (3%)	46	68
1	E	343/407 (84%)	338 (98%)	5 (2%)	72	87
All	All	1367/1628 (84%)	1335 (98%)	32 (2%)	58	78

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	298	LEU
1	D	123	ILE
1	E	215	ASN
1	D	109	ASP
1	D	127	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	411	GLN
1	D	122	ASN
1	D	301	ASN
1	B	404	ASN
1	D	375	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](#)

1 ligand is 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
2	SO4	A	600	-	4,4,4	0.46	0	6,6,6	0.16	0

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
2	SO4	A	600	-	-	0/0/0/0	0/0/0/0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	600	SO4	1	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 [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	368/458 (80%)	0.28	22 (5%) 25 25	34, 48, 78, 92	0
1	B	363/458 (79%)	0.56	35 (9%) 10 10	37, 58, 91, 127	0
1	D	367/458 (80%)	0.87	67 (18%) 2 2	40, 66, 114, 120	0
1	E	369/458 (80%)	0.53	41 (11%) 7 7	36, 59, 97, 104	0
All	All	1467/1832 (80%)	0.56	165 (11%) 7 7	34, 58, 100, 127	0

The worst 5 of 165 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	134	ALA	8.0
1	D	433	SER	7.3
1	E	131	SER	7.2
1	D	231	THR	6.8
1	E	130[A]	MET	6.5

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
2	SO4	A	600	5/5	0.96	0.09	-1.17	86,87,88,88	0

6.5 Other polymers [i](#)

There are no such residues in this entry.