wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 10:00 PM GMT

PDB ID	:	1RKQ
Title	:	Crystal structure of HAD-like phosphatase yidA from E. coli
Authors	:	Ramagopal, U.A.; Almo, S.C.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2003-11-23
Resolution	:	1.40 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

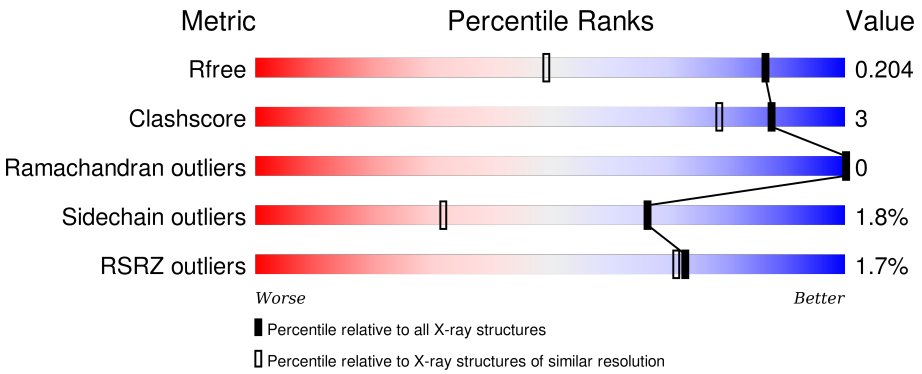
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.40 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	1199 (1.40-1.40)
Clashscore	102246	1295 (1.40-1.40)
Ramachandran outliers	100387	1259 (1.40-1.40)
Sidechain outliers	100360	1258 (1.40-1.40)
RSRZ outliers	91569	1198 (1.40-1.40)

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 <=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	282	2% 84% 12% •
1	B	282	% 80% 13% • •

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4830 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 Hypothetical protein yidA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	271	2102	1341	347	405	9	0	3	0
1	B	271	2106	1342	350	405	9	0	4	0

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	INITIATING METHIONINE	UNP P0A8Y5
A	1	SER	-	CLONING ARTIFACT	UNP P0A8Y5
A	2	LEU	-	CLONING ARTIFACT	UNP P0A8Y5
A	238	VAL	MET	CONFLICT	UNP P0A8Y5
A	272	GLU	-	CLONING ARTIFACT	UNP P0A8Y5
A	273	GLY	-	CLONING ARTIFACT	UNP P0A8Y5
A	274	GLY	-	CLONING ARTIFACT	UNP P0A8Y5
A	275	SER	-	CLONING ARTIFACT	UNP P0A8Y5
A	276	HIS	-	EXPRESSION TAG	UNP P0A8Y5
A	277	HIS	-	EXPRESSION TAG	UNP P0A8Y5
A	278	HIS	-	EXPRESSION TAG	UNP P0A8Y5
A	279	HIS	-	EXPRESSION TAG	UNP P0A8Y5
A	280	HIS	-	EXPRESSION TAG	UNP P0A8Y5
A	281	HIS	-	EXPRESSION TAG	UNP P0A8Y5
B	0	MET	-	INITIATING METHIONINE	UNP P0A8Y5
B	1	SER	-	CLONING ARTIFACT	UNP P0A8Y5
B	2	LEU	-	CLONING ARTIFACT	UNP P0A8Y5
B	238	VAL	MET	CONFLICT	UNP P0A8Y5
B	272	GLU	-	CLONING ARTIFACT	UNP P0A8Y5
B	273	GLY	-	CLONING ARTIFACT	UNP P0A8Y5
B	274	GLY	-	CLONING ARTIFACT	UNP P0A8Y5
B	275	SER	-	CLONING ARTIFACT	UNP P0A8Y5
B	276	HIS	-	EXPRESSION TAG	UNP P0A8Y5
B	277	HIS	-	EXPRESSION TAG	UNP P0A8Y5
B	278	HIS	-	EXPRESSION TAG	UNP P0A8Y5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	279	HIS	-	EXPRESSION TAG	UNP P0A8Y5
B	280	HIS	-	EXPRESSION TAG	UNP P0A8Y5
B	281	HIS	-	EXPRESSION TAG	UNP P0A8Y5

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Cl 1 1	0	0
2	A	1	Total Cl 1 1	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Mg 1 1	0	0
3	A	1	Total Mg 1 1	0	0

- Molecule 4 is water.

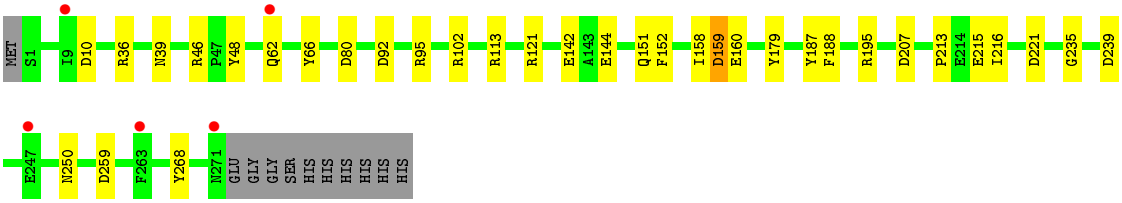
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	312	Total O 312 312	0	0
4	B	306	Total O 306 306	0	0

3 Residue-property plots (i)

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 (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.

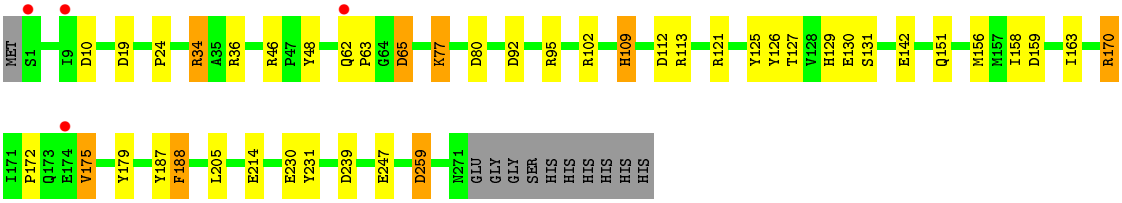
- Molecule 1: Hypothetical protein yidA

Chain A:



- Molecule 1: Hypothetical protein yidA

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	45.12Å 54.87Å 67.72Å 112.58° 96.37° 106.79°	Depositor
Resolution (Å)	24.50 – 1.40 24.54 – 1.40	Depositor EDS
% Data completeness (in resolution range)	95.5 (24.50-1.40) 81.4 (24.54-1.40)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	3.15 (at 1.40Å)	Xtriage
Refinement program	REFMAC 4.0.6	Depositor
R, R_{free}	0.177 , 0.213 0.168 , 0.204	Depositor DCC
R_{free} test set	5253 reflections (5.26%)	DCC
Wilson B-factor (Å^2)	15.4	Xtriage
Anisotropy	0.554	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 52.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Outliers	0 of 105081 reflections	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4830	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.54% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: MG, CL

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	0.97	0/2154	1.68	38/2925 (1.3%)
1	B	0.98	1/2162 (0.0%)	1.66	44/2936 (1.5%)
All	All	0.97	1/4316 (0.0%)	1.67	82/5861 (1.4%)

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	B	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	131	SER	CB-OG	5.18	1.49	1.42

The worst 5 of 82 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	102	ARG	NE-CZ-NH2	-18.08	111.26	120.30
1	B	121	ARG	NE-CZ-NH2	-16.43	112.09	120.30
1	B	34[A]	ARG	CD-NE-CZ	14.15	143.41	123.60
1	B	34[B]	ARG	CD-NE-CZ	14.15	143.41	123.60
1	A	259	ASP	CB-CG-OD1	12.89	129.90	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	259	ASP	Mainchain

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	2102	0	2122	9	0
1	B	2106	0	2125	13	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	312	0	0	4	0
4	B	306	0	0	6	0
All	All	4830	0	4247	22	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 3.

The worst 5 of 22 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:39:ASN:HB2	4:A:1461:HOH:O	1.71	0.89
1:A:235:GLY:H	1:A:250:ASN:HD22	1.27	0.82
1:B:34[A]:ARG:HH12	1:B:65:ASP:CG	1.85	0.79
1:B:63:PRO:O	1:B:77:LYS:HE3	1.97	0.65
1:B:172:PRO:O	1:B:175[A]:VAL:HG13	1.99	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	272/282 (96%)	266 (98%)	6 (2%)	0	100	100
1	B	273/282 (97%)	267 (98%)	6 (2%)	0	100	100
All	All	545/564 (97%)	533 (98%)	12 (2%)	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	228/234 (97%)	226 (99%)	2 (1%)	84	63
1	B	229/234 (98%)	222 (97%)	7 (3%)	47	12
All	All	457/468 (98%)	448 (98%)	9 (2%)	66	27

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	109	HIS
1	B	175[B]	VAL
1	B	159	ASP
1	B	62	GLN
1	B	158	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	39	ASN
1	A	149	ASN
1	A	250	ASN
1	B	109	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)

Of 4 ligands modelled in this entry, 4 are 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 (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	271/282 (96%)	-0.19	5 (1%)	71	70	12, 18, 28, 45	0
1	B	271/282 (96%)	-0.17	4 (1%)	76	75	12, 17, 30, 46	0
All	All	542/564 (96%)	-0.18	9 (1%)	73	71	12, 18, 29, 46	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1	SER	4.5
1	B	62	GLN	2.7
1	A	62	GLN	2.6
1	A	263	PHE	2.5
1	B	9	ILE	2.5

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(Å^2)	Q<0.9
3	MG	B	2273	1/1	1.00	0.04	-1.78	12,12,12,12	0
3	MG	A	1273	1/1	1.00	0.02	-2.82	14,14,14,14	0
2	CL	B	2272	1/1	1.00	0.03	-3.45	14,14,14,14	0
2	CL	A	1272	1/1	1.00	0.04	-4.31	14,14,14,14	0

6.5 Other polymers (i)

There are no such residues in this entry.