wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 10:23 PM GMT

PDB ID	:	5D4S
Title	:	Crystal Structure of AraR(DBD) in complex with operator ORX1
Authors	:	Jain, D.; Narayanan, N.; Nair, D.T.
Deposited on	:	2015-08-08
Resolution	:	1.97 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:

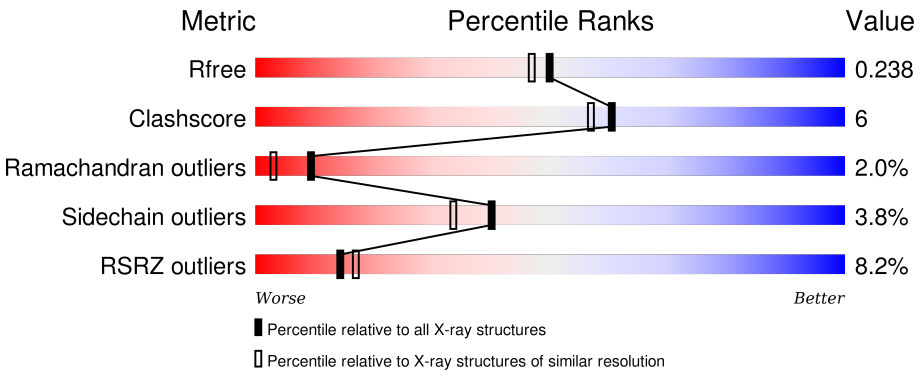
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 1.97 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	8664 (2.00-1.96)
Clashscore	102246	9905 (2.00-1.96)
Ramachandran outliers	100387	9792 (2.00-1.96)
Sidechain outliers	100360	9791 (2.00-1.96)
RSRZ outliers	91569	8679 (2.00-1.96)

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	88	2%, 73%, 5%, 20%
1	B	88	16%, 80%, 11%, 6%
2	T	21	76%, 24%
3	U	21	81%, 19%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2220 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 Arabinose metabolism transcriptional repressor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	70	551	352	93	104	2	0	0	0
1	B	83	653	419	111	121	2	0	0	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	expression tag	UNP P96711
A	-18	HIS	-	expression tag	UNP P96711
A	-17	HIS	-	expression tag	UNP P96711
A	-16	HIS	-	expression tag	UNP P96711
A	-15	HIS	-	expression tag	UNP P96711
A	-14	HIS	-	expression tag	UNP P96711
A	-13	HIS	-	expression tag	UNP P96711
A	-12	LEU	-	expression tag	UNP P96711
A	-11	GLU	-	expression tag	UNP P96711
A	-10	VAL	-	expression tag	UNP P96711
A	-9	LEU	-	expression tag	UNP P96711
A	-8	PHE	-	expression tag	UNP P96711
A	-7	GLN	-	expression tag	UNP P96711
A	-6	GLY	-	expression tag	UNP P96711
A	-5	PRO	-	expression tag	UNP P96711
A	-4	LEU	-	expression tag	UNP P96711
A	-3	GLY	-	expression tag	UNP P96711
A	-2	SER	-	expression tag	UNP P96711
A	-1	GLU	-	expression tag	UNP P96711
A	0	PHE	-	expression tag	UNP P96711
B	-19	MET	-	expression tag	UNP P96711
B	-18	HIS	-	expression tag	UNP P96711
B	-17	HIS	-	expression tag	UNP P96711
B	-16	HIS	-	expression tag	UNP P96711
B	-15	HIS	-	expression tag	UNP P96711

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-14	HIS	-	expression tag	UNP P96711
B	-13	HIS	-	expression tag	UNP P96711
B	-12	LEU	-	expression tag	UNP P96711
B	-11	GLU	-	expression tag	UNP P96711
B	-10	VAL	-	expression tag	UNP P96711
B	-9	LEU	-	expression tag	UNP P96711
B	-8	PHE	-	expression tag	UNP P96711
B	-7	GLN	-	expression tag	UNP P96711
B	-6	GLY	-	expression tag	UNP P96711
B	-5	PRO	-	expression tag	UNP P96711
B	-4	LEU	-	expression tag	UNP P96711
B	-3	GLY	-	expression tag	UNP P96711
B	-2	SER	-	expression tag	UNP P96711
B	-1	GLU	-	expression tag	UNP P96711
B	0	PHE	-	expression tag	UNP P96711

- Molecule 2 is a DNA chain called DNA (5'-D(*TP*AP*AP*TP*AP*TP*TP*TP*GP*TP*AP*CP*GP*TP*AP*TP*GP*TP*AP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	21	428	209	70	129	20	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*AP*AP*TP*AP*CP*AP*TP*AP*CP*GP*TP*AP*CP*AP*AP*AP*TP*AP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	U	21	427	207	81	119	20	0	0	0

- Molecule 4 is water.

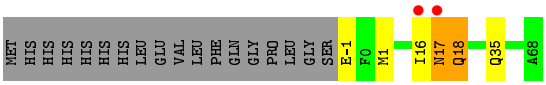
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	31	31	31	0	0
4	B	40	40	40	0	0
4	T	49	49	49	0	0
4	U	41	41	41	0	0

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.

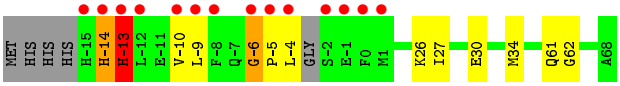
• Molecule 1: Arabinose metabolism transcriptional repressor

Chain A:



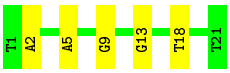
• Molecule 1: Arabinose metabolism transcriptional repressor

Chain B:



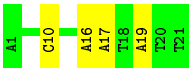
• Molecule 2: DNA (5'-D(*TP*AP*AP*TP*AP*TP*TP*TP*GP*TP*AP*CP*GP*TP*AP*TP*GP*TP*AP*TP*T)-3')

Chain T:



• Molecule 3: DNA (5'-D(*AP*AP*AP*TP*AP*CP*AP*TP*AP*CP*GP*TP*AP*CP*AP*AP*AP*TP*AP*TP*T)-3')

Chain U:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	138.23Å 42.45Å 67.21Å 90.00° 114.99° 90.00°	Depositor
Resolution (Å)	35.87 – 1.97 36.62 – 1.97	Depositor EDS
% Data completeness (in resolution range)	99.4 (35.87-1.97) 99.7 (36.62-1.97)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.88 (at 1.97Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, R_{free}	0.195 , 0.241 0.195 , 0.238	Depositor DCC
R_{free} test set	1278 reflections (5.09%)	DCC
Wilson B-factor (Å^2)	32.5	Xtriage
Anisotropy	0.517	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 57.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	7 of 25204 reflections (0.028%)	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2220	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.45% 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)

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	0.40	0/561	0.67	1/754 (0.1%)
1	B	0.43	0/666	0.87	4/896 (0.4%)
2	T	0.87	0/478	1.05	0/737
3	U	0.87	0/480	1.00	1/738 (0.1%)
All	All	0.65	0/2185	0.91	6/3125 (0.2%)

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
1	B	0	2
All	All	0	3

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	-13	HIS	CB-CA-C	10.73	131.87	110.40
1	B	-14	HIS	CB-CA-C	-9.13	92.14	110.40
1	B	-14	HIS	N-CA-C	7.35	130.86	111.00
1	B	-13	HIS	N-CA-C	-7.07	91.91	111.00
1	A	18	GLN	N-CA-CB	6.75	122.75	110.60

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	17	ASN	Mainchain
1	B	-13	HIS	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	B	-6	GLY	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	551	0	559	4	0
1	B	653	0	657	12	0
2	T	428	0	244	7	0
3	U	427	0	239	3	0
4	A	31	0	0	0	0
4	B	40	0	0	0	0
4	T	49	0	0	3	0
4	U	41	0	0	1	0
All	All	2220	0	1699	21	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 6.

The worst 5 of 21 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:B:-14:HIS:O	1:B:-13:HIS:CG	2.00	1.13
1:B:-14:HIS:O	1:B:-13:HIS:ND1	1.89	1.06
1:B:-14:HIS:O	1:B:-13:HIS:CE1	2.37	0.78
1:A:-1:GLU:HG3	1:A:1:MET:H	1.57	0.69
2:T:13:DG:OP2	4:T:101:HOH:O	2.11	0.68

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	68/88 (77%)	65 (96%)	3 (4%)	0	100	100
1	B	79/88 (90%)	72 (91%)	4 (5%)	3 (4%)	4	1
All	All	147/176 (84%)	137 (93%)	7 (5%)	3 (2%)	9	3

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	-13	HIS
1	B	-5	PRO
1	B	-6	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	60/76 (79%)	59 (98%)	1 (2%)	68	69
1	B	71/76 (93%)	67 (94%)	4 (6%)	26	18
All	All	131/152 (86%)	126 (96%)	5 (4%)	40	33

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	35	GLN
1	B	-13	HIS
1	B	-10	VAL
1	B	-9	LEU
1	B	-4	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	35	GLN
1	B	-14	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)

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, 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	70/88 (79%)	0.17	2 (2%) 55 58	24, 37, 55, 69	0
1	B	83/88 (94%)	1.41	14 (16%) 2 3	24, 38, 116, 131	0
2	T	21/21 (100%)	-0.29	0 100 100	28, 43, 54, 63	0
3	U	21/21 (100%)	-0.31	0 100 100	29, 40, 53, 57	0
All	All	195/218 (89%)	0.60	16 (8%) 14 17	24, 40, 87, 131	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	-5	PRO	14.6
1	B	-13	HIS	13.3
1	B	-4	LEU	13.0
1	B	-6	GLY	10.8
1	B	-2	SER	7.0

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.