



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 08:43 PM GMT

PDB ID : 4YRV
Title : Crystal structure of Anabaena transcription factor HetR complexed with 21-bp DNA from hetP promoter
Authors : Hu, H.X.; Jiang, Y.L.; Zhao, M.X.; Zhang, C.C.; Chen, Y.; Zhou, C.Z.
Deposited on : 2015-03-16
Resolution : 2.80 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 : 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

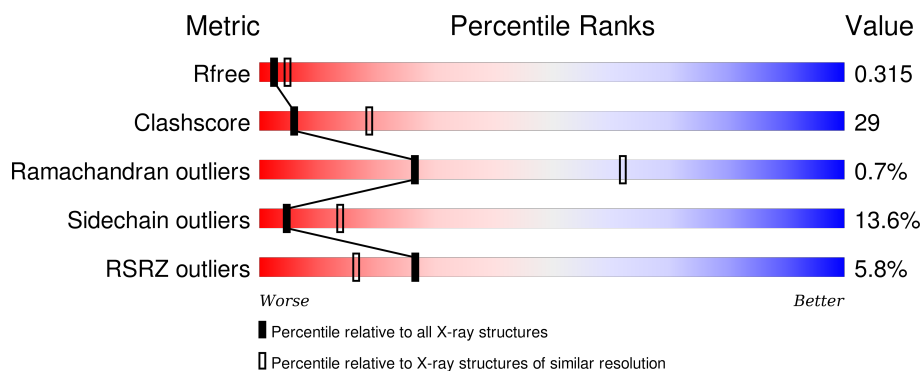
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	307	 6% 46% 39% 8% • 6%
1	B	307	 6% 46% 40% 7% 7%
2	C	21	 43% 52% 5%
3	D	21	 48% 33% 19%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 5648 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 Heterocyst differentiation control protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	286	Total	C	N	O	S	0	0	0
			2368	1514	414	423	17			
1	A	288	Total	C	N	O	S	0	0	0
			2384	1524	418	425	17			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-7	MET	-	expression tag	UNP P27709
B	-6	GLY	-	expression tag	UNP P27709
B	-5	HIS	-	expression tag	UNP P27709
B	-4	HIS	-	expression tag	UNP P27709
B	-3	HIS	-	expression tag	UNP P27709
B	-2	HIS	-	expression tag	UNP P27709
B	-1	HIS	-	expression tag	UNP P27709
B	0	HIS	-	expression tag	UNP P27709
A	-7	MET	-	expression tag	UNP P27709
A	-6	GLY	-	expression tag	UNP P27709
A	-5	HIS	-	expression tag	UNP P27709
A	-4	HIS	-	expression tag	UNP P27709
A	-3	HIS	-	expression tag	UNP P27709
A	-2	HIS	-	expression tag	UNP P27709
A	-1	HIS	-	expression tag	UNP P27709
A	0	HIS	-	expression tag	UNP P27709

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	21	Total	C	N	O	P	0	0	0
			429	203	79	126	21			

- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*TP*GP*AP*GP*GP*GP*GP*TP*TP*AP*GP*AP*CP*CP*CP*CP*TP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	21	Total	C	N	O	P	0	0	0
			432	204	81	126	21			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Ca	0	0
			1	1		
4	D	1	Total	Ca	0	0
			1	1		
4	C	2	Total	Ca	0	0
			2	2		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	13	Total	O	0	0
			13	13		
5	A	13	Total	O	0	0
			13	13		
5	C	3	Total	O	0	0
			3	3		
5	D	2	Total	O	0	0
			2	2		



● Molecule 3: DNA (5'-D(P*AP*TP*GP*AP*GP*GP*GP*GP*TP*TP*AP*GP*AP*CP*CP*CP*CP*TP*CP*GP*C)-3')

Chain D: 48% 33% 19%



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	90.56Å 90.56Å 242.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.95 – 2.80 39.95 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.2 (39.95-2.80) 99.2 (39.95-2.80)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.58 (at 2.81Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.1_743)	Depositor
R, R_{free}	0.203 , 0.264 0.271 , 0.315	Depositor DCC
R_{free} test set	1298 reflections (5.37%)	DCC
Wilson B-factor (Å ²)	75.4	Xtriage
Anisotropy	0.461	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 71.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 25452 reflections	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	5648	wwPDB-VP
Average B, all atoms (Å ²)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.32% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	0.42	0/2443	0.61	2/3298 (0.1%)
1	B	0.38	0/2427	0.58	1/3275 (0.0%)
2	C	0.90	2/480 (0.4%)	1.40	8/738 (1.1%)
3	D	0.88	3/484 (0.6%)	1.49	11/745 (1.5%)
All	All	0.51	5/5834 (0.1%)	0.82	22/8056 (0.3%)

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	3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	17	DC	O3'-P	-6.72	1.53	1.61
3	D	2	DT	O3'-P	-6.02	1.53	1.61
2	C	20	DA	O3'-P	-5.71	1.54	1.61
2	C	9	DT	O3'-P	-5.29	1.54	1.61
3	D	18	DT	O3'-P	-5.13	1.54	1.61

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	5	DG	C1'-O4'-C4'	-8.81	101.29	110.10
2	C	15	DC	O4'-C4'-C3'	-8.34	100.99	106.00
2	C	16	DC	O4'-C4'-C3'	-8.21	101.08	106.00
3	D	11	DA	O4'-C1'-N9	-7.89	102.47	108.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	C	5	DG	O4'-C4'-C3'	-7.20	101.62	104.50

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	129	THR	Peptide
1	A	88	GLU	Peptide
1	A	89	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	2384	0	2377	180	0
1	B	2368	0	2359	174	0
2	C	429	0	236	9	0
3	D	432	0	236	9	0
4	B	1	0	0	0	0
4	C	2	0	0	0	0
4	D	1	0	0	0	0
5	A	13	0	0	3	0
5	B	13	0	0	2	0
5	C	3	0	0	1	0
5	D	2	0	0	0	0
All	All	5648	0	5208	309	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 29.

The worst 5 of 309 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:150:VAL:CG1	1:A:154:GLU:HB3	1.59	1.30
1:A:150:VAL:HG13	1:A:154:GLU:CB	1.62	1.28

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:150:VAL:CG1	1:B:154:GLU:HB2	1.65	1.26
1:A:213:PRO:O	1:A:214:PHE:HD1	1.16	1.24
1:A:150:VAL:CG1	1:A:154:GLU:CB	2.14	1.19

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	282/307 (92%)	251 (89%)	28 (10%)	3 (1%)	17	50
1	B	280/307 (91%)	259 (92%)	20 (7%)	1 (0%)	39	74
All	All	562/614 (92%)	510 (91%)	48 (8%)	4 (1%)	26	62

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	129	THR
1	A	90	LYS
1	A	258	PRO
1	B	89	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	255/269 (95%)	217 (85%)	38 (15%)	4	11
1	B	253/269 (94%)	222 (88%)	31 (12%)	6	18
All	All	508/538 (94%)	439 (86%)	69 (14%)	5	14

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	4	ASP
1	A	86	LEU
1	A	260	GLU
1	A	8	ILE
1	A	23	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	98	GLN

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

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, 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	288/307 (93%)	0.60	19 (6%) 22 13	53, 85, 149, 177	0
1	B	286/307 (93%)	0.53	17 (5%) 26 16	52, 84, 132, 149	0
2	C	21/21 (100%)	-0.02	0 100 100	20, 70, 79, 97	0
3	D	21/21 (100%)	-0.17	0 100 100	20, 69, 87, 90	0
All	All	616/656 (93%)	0.52	36 (5%) 26 16	20, 83, 137, 177	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	174	HIS	6.1
1	A	221	GLN	5.6
1	A	135	GLN	5.3
1	B	125	GLY	5.0
1	A	131	GLU	5.0

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](#)

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
4	CA	B	301	1/1	0.83	0.13	-1.19	107,107,107,107	0
4	CA	D	101	1/1	0.97	0.34	-	96,96,96,96	0
4	CA	C	101	1/1	0.85	0.51	-	112,112,112,112	0
4	CA	C	102	1/1	0.89	0.34	-	97,97,97,97	0

6.5 Other polymers [i](#)

There are no such residues in this entry.