



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 30, 2023 – 09:30 PM JST

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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.36  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.36

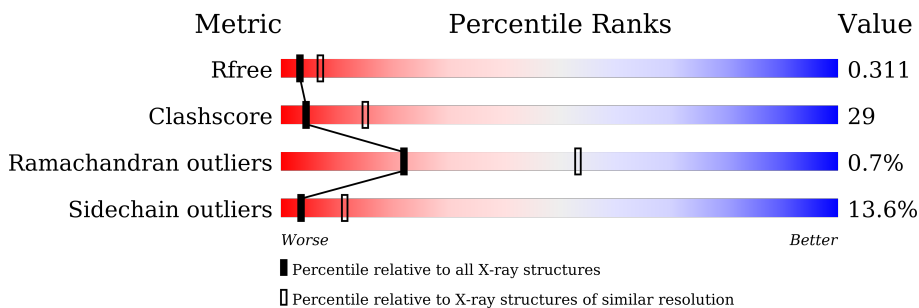
# 1 Overall quality at a glance

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))
$R_{free}$	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	307	46% (green) 39% (yellow) 8% (orange) • 6% (red) 6% (grey)
1	B	307	46% (green) 40% (yellow) 7% (orange) 7% (red) 7% (grey)
2	C	21	43% (green) 52% (yellow) 5% (orange)
3	D	21	48% (green) 33% (yellow) 19% (orange)

## 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
			Total	C	N	O	S			
1	B	286	2368	1514	414	423	17	0	0	0
1	A	288	2384	1524	418	425	17	0	0	0

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\*GP\*TP\*C P\*TP\*AP\*AP\*CP\*CP\*CP\*CP\*TP\*CP\*AP\*T)-3').

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

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

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

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

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

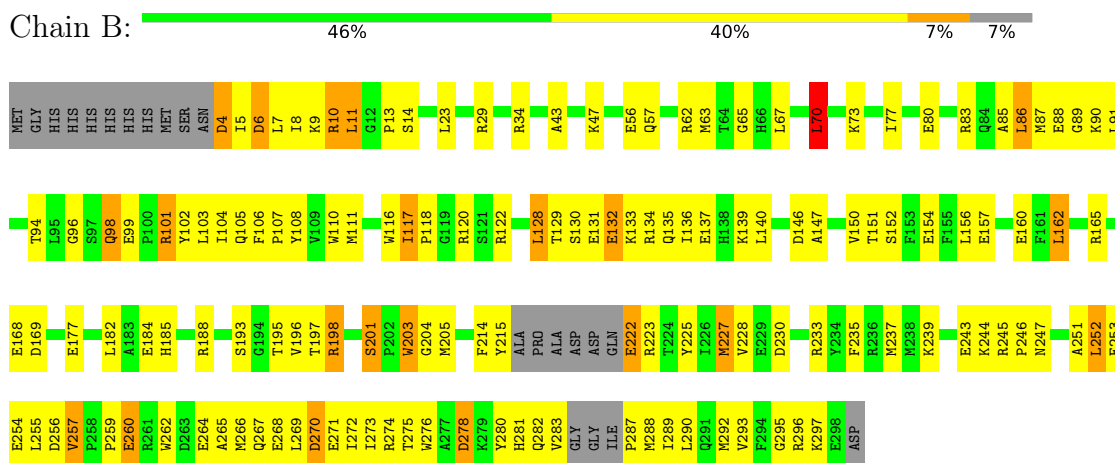
- Molecule 5 is water.

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

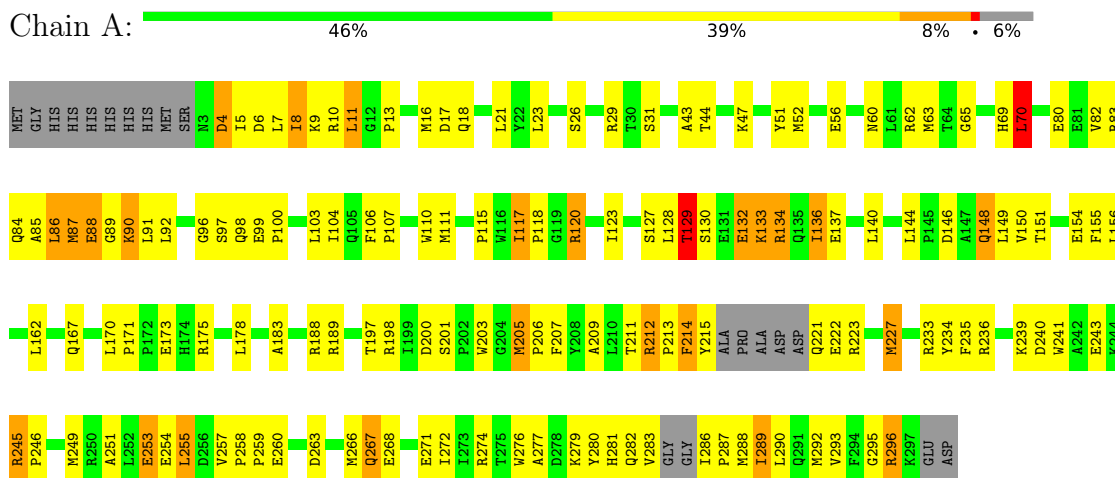
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

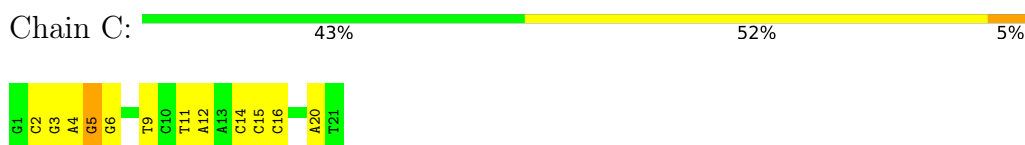
- Molecule 1: Heterocyst differentiation control protein



- Molecule 1: Heterocyst differentiation control protein

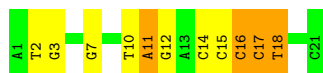


- Molecule 2: DNA (5'-D(P\*GP\*CP\*GP\*AP\*GP\*GP\*GP\*TP\*CP\*TP\*AP\*AP\*CP\*CP\*C P\*CP\*TP\*CP\*AP\*T)-3')



- 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, $\alpha$ , $\beta$ , $\gamma$	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
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.58 (at 2.81Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.1_743)	Depositor
R, $R_{free}$	0.203 , 0.264 0.269 , 0.311	Depositor DCC
$R_{free}$ test set	1297 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	75.4	Xtriage
Anisotropy	0.461	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 71.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	5648	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	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.*

<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

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(°)	Ideal(°)
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%)	14	41
1	B	280/307 (91%)	259 (92%)	20 (7%)	1 (0%)	34	66
All	All	562/614 (92%)	510 (91%)	48 (8%)	4 (1%)	22	53

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%)	3	9
1	B	253/269 (94%)	222 (88%)	31 (12%)	4	15
All	All	508/538 (94%)	439 (86%)	69 (14%)	3	11

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

Mol	Chain	Res	Type
1	A	221	GLN
1	A	235	PHE
1	A	267	GLN
1	B	245	ARG
1	B	235	PHE

Sometimes 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 monosaccharides 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

### 6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.