wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 2, 2024 – 10:12 am BST

PDB ID	:	8PN4
Title	:	transcription factor BARHL2 bound to DNA sequences
Authors	:	Morgunova, E.; Yin, Y.; Popov, A.; Taipale, J.
Deposited on	:	2023-06-29
Resolution	:	2.60 Å(reported)

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

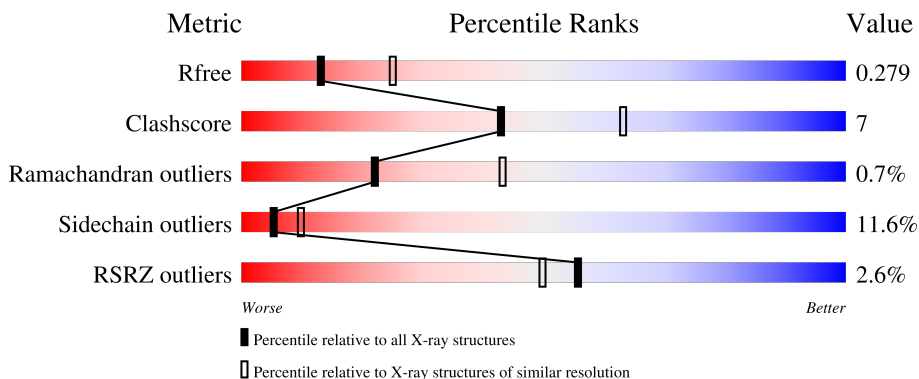|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.60 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	Percentile Ranks	Value
Rfree		0.279
Clashscore		7
Ramachandran outliers		0.7%
Sidechain outliers		11.6%
RSRZ outliers		2.6%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 >=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 <=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	62	85% 11% •
1	D	62	2% / 79% 15% • 5%
1	G	62	74% 21% • •
1	J	62	2% / 79% 13% • 6%
1	M	62	13% / 42% 24% 6% • 26%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	B	12	67% 33%
2	E	12	92% 8%
2	H	12	92% 8%
2	K	12	83% 17%
3	C	12	75% 25%
3	F	12	75% 25%
3	I	12	83% 17%
3	L	12	42% 58%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 4562 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 BarH-like 2 homeobox protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	60	Total 522	C 323	N 103	O 95	S 1	0	1	0
1	D	59	Total 509	C 315	N 102	O 91	S 1	0	0	0
1	G	60	Total 514	C 317	N 103	O 93	S 1	0	0	0
1	J	58	Total 500	C 310	N 100	O 89	S 1	0	0	0
1	M	46	Total 401	C 250	N 78	O 72	S 1	0	0	0

- Molecule 2 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	12	Total 240	C 115	N 43	O 71	P 11	0	0	0
2	E	12	Total 240	C 116	N 43	O 70	P 11	0	0	0
2	H	12	Total 241	C 116	N 43	O 71	P 11	0	0	0
2	K	12	Total 241	C 116	N 43	O 71	P 11	0	0	0

- Molecule 3 is a DNA chain called DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	12	Total 245	C 117	N 48	O 69	P 11	0	0	0
3	F	12	Total 245	C 117	N 48	O 69	P 11	0	0	0
3	I	12	Total 245	C 117	N 48	O 69	P 11	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	L	12	245	117	48	69	11	0	0	0

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	22	Total 22	O 22	0	0
5	B	13	Total 13	O 13	0	0
5	C	14	Total 14	O 14	0	0
5	D	10	Total 10	O 10	0	0
5	E	3	Total 3	O 3	0	0
5	F	3	Total 3	O 3	0	0
5	G	27	Total 27	O 27	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	H	10	Total 10	O 10	0	0
5	I	11	Total 11	O 11	0	0
5	J	18	Total 18	O 18	0	0
5	K	13	Total 13	O 13	0	0
5	L	11	Total 11	O 11	0	0
5	M	15	Total 15	O 15	0	0

3 Residue-property plots ⓘ

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 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.
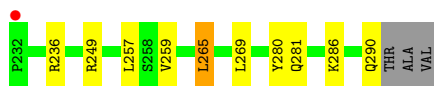
• Molecule 1: BarH-like 2 homeobox protein

Chain A:

• Molecule 1: BarH-like 2 homeobox protein

Chain D:

• Molecule 1: BarH-like 2 homeobox protein

Chain G:

• Molecule 1: BarH-like 2 homeobox protein

Chain J:

• Molecule 1: BarH-like 2 homeobox protein

Chain M:
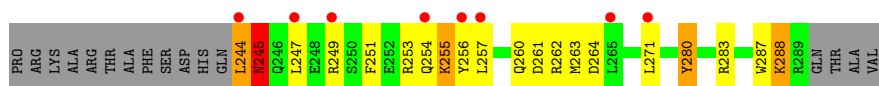
• Molecule 2: DNA

Chain B:

67%   33%

C1 G2 C3 T4 A5 T11 T12

• Molecule 2: DNA

Chain E:

92%   8%

C1 T4 T12
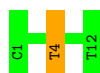
• Molecule 2: DNA

Chain H:

92%   8%

C1 T4 T12
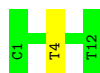
• Molecule 2: DNA

Chain K:

83%   17%

C1 G10 T11 T12

• Molecule 3: DNA

Chain C:

75%   25%

A13 A14 C15 C16 G17 G24

• Molecule 3: DNA

Chain F:

75%   25%

A13 A14 C15 C16 G17 G24

• Molecule 3: DNA

Chain I:

83%   17%

A13 C16 G17 G24

• Molecule 3: DNA

Chain L:

42%   58%

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	39.43Å 134.56Å 64.49Å 90.00° 93.18° 90.00°	Depositor
Resolution (Å)	39.40 – 2.60 39.37 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.5 (39.40-2.60) 99.5 (39.37-2.60)	Depositor EDS
R_{merge}	0.22	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.28 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0415	Depositor
R, R_{free}	0.243 , 0.274 0.246 , 0.279	Depositor DCC
R_{free} test set	1034 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	51.8	Xtriage
Anisotropy	0.476	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 27.2	EDS
L-test for twinning[2]	$< \mid L \mid > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	4562	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.66% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< \mid L \mid >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: ACT

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	0.77	3/534 (0.6%)	0.72	0/716
1	D	0.29	0/518	0.64	0/694
1	G	0.33	0/522	0.63	0/700
1	J	0.29	0/509	0.60	0/682
1	M	0.30	0/407	0.68	0/546
2	B	0.57	0/266	1.21	1/406 (0.2%)
2	E	0.56	0/268	1.10	1/412 (0.2%)
2	H	0.55	0/269	1.26	1/413 (0.2%)
2	K	0.60	0/269	1.18	0/413
3	C	0.59	0/275	1.24	0/423
3	F	0.56	0/275	1.17	0/423
3	I	0.66	0/275	1.22	0/423
3	L	0.66	0/275	1.24	0/423
All	All	0.52	3/4662 (0.1%)	0.97	3/6674 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	239	PHE	CG-CD2	11.48	1.55	1.38
1	A	239	PHE	CE1-CZ	7.88	1.52	1.37
1	A	239	PHE	CG-CD1	5.76	1.47	1.38

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	H	4	DT	O4'-C1'-N1	7.82	113.47	108.00
2	B	4	DT	O4'-C1'-N1	7.14	113.00	108.00
2	E	4	DT	O4'-C1'-N1	6.00	112.20	108.00

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	522	0	522	3	0
1	D	509	0	509	5	0
1	G	514	0	513	7	0
1	J	500	0	501	7	0
1	M	401	0	404	20	0
2	B	240	0	133	4	0
2	E	240	0	134	2	0
2	H	241	0	137	0	0
2	K	241	0	137	2	0
3	C	245	0	136	4	0
3	F	245	0	136	2	0
3	I	245	0	136	1	0
3	L	245	0	136	8	0
4	B	4	0	3	1	0
5	A	22	0	0	1	0
5	B	13	0	0	1	0
5	C	14	0	0	1	0
5	D	10	0	0	0	0
5	E	3	0	0	0	0
5	F	3	0	0	0	0
5	G	27	0	0	2	0
5	H	10	0	0	0	0
5	I	11	0	0	0	0
5	J	18	0	0	2	0
5	K	13	0	0	1	0
5	L	11	0	0	0	0
5	M	15	0	0	4	0
All	All	4562	0	3537	54	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 7.

The worst 5 of 54 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)
2:B:11:DT:C3'	2:B:11:DT:C5'	2.58	0.81
1:M:255:LYS:HG3	1:M:283:ARG:HG2	1.62	0.81
1:M:255:LYS:HD2	5:M:301:HOH:O	1.80	0.80
1:M:244:LEU:HB2	1:M:247:LEU:HD12	1.66	0.78
1:G:240:SER:H	1:G:243:GLN:HE21	1.33	0.73

There are no symmetry-related clashes.

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	59/62 (95%)	59 (100%)	0	0	100	100
1	D	57/62 (92%)	57 (100%)	0	0	100	100
1	G	58/62 (94%)	58 (100%)	0	0	100	100
1	J	56/62 (90%)	56 (100%)	0	0	100	100
1	M	44/62 (71%)	36 (82%)	6 (14%)	2 (4%)	2	3
All	All	274/310 (88%)	266 (97%)	6 (2%)	2 (1%)	22	43

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	M	288	LYS
1	M	245	ASN

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	56/56 (100%)	52 (93%)	4 (7%)	14	29
1	D	54/56 (96%)	49 (91%)	5 (9%)	9	17
1	G	54/56 (96%)	48 (89%)	6 (11%)	6	11
1	J	53/56 (95%)	48 (91%)	5 (9%)	8	17
1	M	43/56 (77%)	33 (77%)	10 (23%)	1	1
All	All	260/280 (93%)	230 (88%)	30 (12%)	5	10

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	280	TYR
1	M	260	GLN
1	J	260	GLN
1	M	280	TYR
1	M	255	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	M	260	GLN
1	M	275	GLN
1	G	246	GLN
1	G	290	GLN
1	M	245	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	$\#|Z| > 2$
4	ACT	B	1101	-	3,3,3	0.97	0	3,3,3	0.78	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
4	B	1101	ACT	1	0

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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(Å²)	Q<0.9
1	A	60/62 (96%)	0.07	0	100	100	31, 45, 73, 91	0
1	D	59/62 (95%)	0.17	1 (1%)	70	66	34, 54, 80, 105	0
1	G	60/62 (96%)	0.05	0	100	100	34, 50, 75, 84	0
1	J	58/62 (93%)	0.08	1 (1%)	70	66	41, 61, 82, 95	0
1	M	46/62 (74%)	1.10	8 (17%)	1	0	58, 88, 115, 121	0
2	B	12/12 (100%)	0.06	0	100	100	33, 57, 70, 72	0
2	E	12/12 (100%)	-0.26	0	100	100	49, 56, 97, 108	0
2	H	12/12 (100%)	-0.20	0	100	100	38, 55, 69, 69	0
2	K	12/12 (100%)	0.01	0	100	100	41, 57, 65, 70	0
3	C	12/12 (100%)	-0.19	0	100	100	39, 48, 57, 67	0
3	F	12/12 (100%)	-0.07	0	100	100	40, 58, 93, 101	0
3	I	12/12 (100%)	0.05	0	100	100	42, 56, 67, 78	0
3	L	12/12 (100%)	-0.11	0	100	100	45, 57, 62, 65	0
All	All	379/406 (93%)	0.17	10 (2%)	56	50	31, 55, 96, 121	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	M	256	TYR	3.3
1	M	254	GLN	3.3
1	M	244	LEU	3.2
1	M	271	LEU	3.1
1	M	247	LEU	2.9

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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. 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	B-factors($Å^2$)	Q<0.9
4	ACT	B	1101	4/4	0.89	0.22	40,43,44,45	0

6.5 Other polymers ⓘ

There are no such residues in this entry.