



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 6, 2017 – 05:15 AM EST

PDB ID : 4IHT  
Title : Crystal Structure of BenM\_DBD/benA site 1 DNA Complex  
Authors : Alanazi, A.; Momany, C.; Neidle, E.L.  
Deposited on : unknown  
Resolution : 3.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

<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  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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-20030345

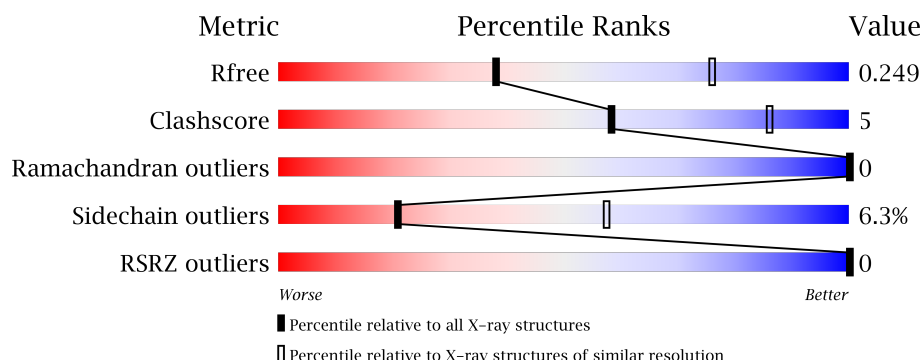
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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  $\geq 3$ , 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	94	
1	B	94	
1	C	94	
1	D	94	
2	E	25	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	25	 72% 28%
3	F	25	 68% 32%
3	H	25	 76% 24%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 9001 atoms, of which 4088 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 HTH-type transcriptional regulator BenM.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	89	Total	C	H	N	O	S	0	0	0
			1453	456	735	127	131	4			
1	B	90	Total	C	H	N	O	S	0	0	0
			1471	462	743	130	132	4			
1	C	86	Total	C	H	N	O	S	0	0	0
			1417	445	719	122	127	4			
1	D	90	Total	C	H	N	O	S	0	0	0
			1471	462	743	130	132	4			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	88	GLY	-	EXPRESSION TAG	UNP O68014
A	89	HIS	-	EXPRESSION TAG	UNP O68014
A	90	HIS	-	EXPRESSION TAG	UNP O68014
A	91	HIS	-	EXPRESSION TAG	UNP O68014
A	92	HIS	-	EXPRESSION TAG	UNP O68014
A	93	HIS	-	EXPRESSION TAG	UNP O68014
A	94	HIS	-	EXPRESSION TAG	UNP O68014
B	88	GLY	-	EXPRESSION TAG	UNP O68014
B	89	HIS	-	EXPRESSION TAG	UNP O68014
B	90	HIS	-	EXPRESSION TAG	UNP O68014
B	91	HIS	-	EXPRESSION TAG	UNP O68014
B	92	HIS	-	EXPRESSION TAG	UNP O68014
B	93	HIS	-	EXPRESSION TAG	UNP O68014
B	94	HIS	-	EXPRESSION TAG	UNP O68014
C	88	GLY	-	EXPRESSION TAG	UNP O68014
C	89	HIS	-	EXPRESSION TAG	UNP O68014
C	90	HIS	-	EXPRESSION TAG	UNP O68014
C	91	HIS	-	EXPRESSION TAG	UNP O68014
C	92	HIS	-	EXPRESSION TAG	UNP O68014
C	93	HIS	-	EXPRESSION TAG	UNP O68014
C	94	HIS	-	EXPRESSION TAG	UNP O68014

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	88	GLY	-	EXPRESSION TAG	UNP O68014
D	89	HIS	-	EXPRESSION TAG	UNP O68014
D	90	HIS	-	EXPRESSION TAG	UNP O68014
D	91	HIS	-	EXPRESSION TAG	UNP O68014
D	92	HIS	-	EXPRESSION TAG	UNP O68014
D	93	HIS	-	EXPRESSION TAG	UNP O68014
D	94	HIS	-	EXPRESSION TAG	UNP O68014

- Molecule 2 is a DNA chain called benA site 1 DNA.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	E	25	Total	C	H	N	O	P	0	0	0
			795	247	287	89	148	24			
2	G	25	Total	C	H	N	O	P	0	0	0
			795	247	287	89	148	24			

- Molecule 3 is a DNA chain called benA site 1 DNA - complement.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	F	25	Total	C	H	N	O	P	0	0	0
			798	248	287	91	148	24			
3	H	25	Total	C	H	N	O	P	0	0	0
			798	248	287	91	148	24			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total	O	0	0
			3	3		



- Molecule 1: HTH-type transcriptional regulator BenM



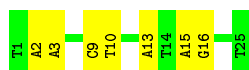
- |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| M1 | Q16 | A22 | L32 | S33 | R34 | Q37 | G44 | I45 | R53 | P54 | V55 | E60 | Y67 | A68 | I69 | K70 | L71 | V75 | V79 | S80 | M81 | T82 | H90 | H15 | H15 | H15 | H15 |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

- |    |  |     |  |     |  |     |     |     |     |     |     |     |  |     |  |     |     |     |     |     |     |     |     |     |
|----|--|-----|--|-----|--|-----|-----|-----|-----|-----|-----|-----|--|-----|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| M1 |  | E15 |  | R34 |  | Y65 | Q66 | Y67 | A68 | I69 | K70 | L71 |  | V75 |  | A86 | SER | GLY | HIS | HIS | HIS | HIS | HIS | HIS |
|----|--|-----|--|-----|--|-----|-----|-----|-----|-----|-----|-----|--|-----|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|

- |    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| M1 | V13 | Q16 | L32 | S33 | R34 | L39 | E40 | G44 | I45 | Q46 | R53 | P54 | V55 | E60 | G61 | H62 | Y65 | Q66 | Y67 | A68 | I69 | D76 | Q77 | M78 | W79 | S80 | K83 | R84 | H90 | H15 | H15 | H15 | H15 |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

- 
- Diagram illustrating a sequence of nodes or components:
- Box 1: T1, A2, A3, A4
  - Box 2: C9, T10
  - Box 3: T25

- 
- WORLD WIDE  
PDB  
PROTEIN DATA BANK




- Molecule 3: benA site 1 DNA - complement

Chain F:  68% 32%



- Molecule 3: benA site 1 DNA - complement

Chain H:  76% 24%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	58.96 Å   300.30 Å   46.00 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	31.66 – 3.00 31.66 – 3.00	Depositor EDS
% Data completeness (in resolution range)	92.5 (31.66-3.00) 92.8 (31.66-3.00)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.49 (at 3.00 Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.184   ,   0.245 0.186   ,   0.249	Depositor DCC
$R_{free}$ test set	820 reflections (5.40%)	DCC
Wilson B-factor (Å <sup>2</sup> )	80.7	Xtriage
Anisotropy	0.258	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 47.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	9001	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	95.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.67% 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 ⓘ

### 5.1 Standard geometry ⓘ

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	0.45	0/731	0.57	0/984
1	B	0.47	0/742	0.61	0/999
1	C	0.43	0/710	0.59	0/956
1	D	0.43	0/742	0.58	0/999
2	E	0.26	0/569	0.51	0/876
2	G	0.28	0/569	0.54	0/876
3	F	0.28	0/573	0.57	0/883
3	H	0.28	0/573	0.61	0/883
All	All	0.38	0/5209	0.58	0/7456

There are no bond length outliers.

There are no bond angle outliers.

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	718	735	732	6	0
1	B	728	743	739	8	0
1	C	698	719	717	3	0
1	D	728	743	739	12	0
2	E	508	287	287	6	0
2	G	508	287	287	5	0
3	F	511	287	287	6	0
3	H	511	287	287	8	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	3	0	0	0	0
All	All	4913	4088	4075	46	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 5.

All (46) 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)
3:H:8:DA:H2''	3:H:9:DC:O5'	1.86	0.75
1:A:86:ALA:HB2	1:B:60:GLU:HG2	1.75	0.67
2:E:1:DT:H2'	2:E:2:DA:C8	2.31	0.65
1:B:44:GLY:O	1:B:45:ILE:HD12	1.98	0.63
1:D:34:ARG:NH2	3:H:17:DG:O6	2.32	0.62
1:B:34:ARG:NH2	3:F:17:DG:O6	2.33	0.61
2:G:13:DA:C2	3:H:14:DG:N2	2.73	0.57
2:G:9:DC:H2''	2:G:10:DT:OP1	2.06	0.54
1:D:76:ASP:O	1:D:80:SER:OG	2.22	0.52
3:H:8:DA:H2''	3:H:9:DC:C5'	2.40	0.52
3:H:24:DT:H2'	3:H:25:DA:C8	2.44	0.51
2:E:1:DT:H4'	2:E:2:DA:OP1	2.10	0.51
1:D:62:HIS:NE2	1:D:66:GLN:OE1	2.43	0.51
1:D:13:VAL:HG11	1:D:65:TYR:CD1	2.46	0.51
2:E:3:DA:H2'	2:E:4:DA:C8	2.47	0.50
3:F:12:DA:H2''	3:F:13:DT:H71	1.93	0.50
1:A:75:VAL:O	1:A:79:VAL:HG23	2.11	0.50
2:G:9:DC:H2'	2:G:10:DT:C6	2.46	0.50
3:F:24:DT:H1'	3:F:25:DA:C8	2.48	0.49
2:E:1:DT:H2'	2:E:2:DA:N7	2.29	0.47
1:B:16:GLN:O	1:B:55:VAL:HG12	2.15	0.46
3:H:24:DT:C2'	3:H:25:DA:C8	2.99	0.46
2:E:9:DC:H2'	2:E:10:DT:C6	2.50	0.46
1:D:44:GLY:C	1:D:45:ILE:HD12	2.37	0.46
1:D:53:ARG:HB2	3:H:25:DA:H4'	1.97	0.46
1:A:71:LEU:O	1:A:75:VAL:HG23	2.16	0.45
1:A:45:ILE:HD13	1:B:82:THR:HG22	1.99	0.45
1:D:16:GLN:O	1:D:55:VAL:HG12	2.17	0.45
1:A:23:ASP:C	1:A:23:ASP:OD2	2.55	0.44
1:D:39:LEU:O	1:D:40:GLU:C	2.56	0.44
1:B:22:ALA:HA	1:B:32:LEU:HD12	2.00	0.43
3:H:8:DA:H2'	3:H:9:DC:C6	2.54	0.43

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:75:VAL:O	1:B:79:VAL:HG23	2.19	0.43
1:C:71:LEU:O	1:C:75:VAL:HG23	2.18	0.43
2:E:1:DT:C2'	2:E:2:DA:C8	3.01	0.43
1:D:79:VAL:HG12	1:D:83:LYS:HD2	2.01	0.42
1:C:67:TYR:HB3	1:D:78:MET:HG3	2.01	0.41
1:D:46:GLN:HE21	1:D:46:GLN:HA	1.86	0.41
3:F:2:DT:H1'	3:F:3:DA:C8	2.56	0.41
2:G:2:DA:H2''	2:G:3:DA:C8	2.56	0.41
1:A:53:ARG:HD3	3:F:4:DA:H4'	2.03	0.40
3:F:3:DA:H2''	3:F:4:DA:C8	2.56	0.40
1:C:65:TYR:O	1:C:69:ILE:HG23	2.20	0.40
1:D:60:GLU:O	1:D:61:GLY:C	2.60	0.40
2:G:15:DA:C6	2:G:16:DG:C6	3.10	0.40
1:B:33:SER:O	1:B:37:GLN:HG3	2.21	0.40

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	87/94 (93%)	82 (94%)	5 (6%)	0	100	100
1	B	88/94 (94%)	81 (92%)	7 (8%)	0	100	100
1	C	84/94 (89%)	81 (96%)	3 (4%)	0	100	100
1	D	88/94 (94%)	77 (88%)	11 (12%)	0	100	100
All	All	347/376 (92%)	321 (92%)	26 (8%)	0	100	100

There are no Ramachandran outliers to report.

### 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	79/84 (94%)	76 (96%)	3 (4%)	38	75
1	B	80/84 (95%)	74 (92%)	6 (8%)	16	49
1	C	77/84 (92%)	72 (94%)	5 (6%)	20	56
1	D	80/84 (95%)	74 (92%)	6 (8%)	16	49
All	All	316/336 (94%)	296 (94%)	20 (6%)	21	57

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1	MET
1	A	34	ARG
1	A	78	MET
1	B	32	LEU
1	B	53	ARG
1	B	67	TYR
1	B	69	ILE
1	B	71	LEU
1	B	80	SER
1	C	1	MET
1	C	15	GLU
1	C	34	ARG
1	C	67	TYR
1	C	69	ILE
1	D	32	LEU
1	D	46	GLN
1	D	53	ARG
1	D	67	TYR
1	D	69	ILE
1	D	84	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	62	HIS
1	A	74	ASN
1	B	5	HIS
1	B	35	GLN
1	B	74	ASN
1	C	16	GLN
1	C	62	HIS
1	C	66	GLN
1	D	74	ASN

### 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 ⓘ

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	89/94 (94%)	-0.25	0 100 100	57, 76, 114, 142	0
1	B	90/94 (95%)	-0.27	0 100 100	56, 83, 130, 168	0
1	C	86/94 (91%)	-0.16	0 100 100	55, 79, 140, 163	0
1	D	90/94 (95%)	-0.17	0 100 100	62, 100, 146, 162	0
2	E	25/25 (100%)	-0.29	0 100 100	73, 101, 142, 155	0
2	G	25/25 (100%)	-0.26	0 100 100	63, 102, 139, 155	0
3	F	25/25 (100%)	-0.32	0 100 100	76, 89, 125, 134	0
3	H	25/25 (100%)	-0.14	0 100 100	68, 88, 151, 160	0
All	All	455/476 (95%)	-0.22	0 100 100	55, 87, 142, 168	0

There are no RSRZ outliers to report.

### 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 carbohydrates in this entry.

### 6.4 Ligands ⓘ

There are no ligands in this entry.

### 6.5 Other polymers ⓘ

There are no such residues in this entry.