



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 12, 2017 – 08:09 pm GMT

PDB ID : 2HDD  
Title : ENGRAILED HOMEODOMAIN Q50K VARIANT DNA COMPLEX  
Authors : Tucker-Kellogg, L.; Rould, M.A.; Chambers, K.A.; Ades, S.E.; Sauer, R.T.; Pabo, C.O.  
Deposited on : 1998-02-10  
Resolution : 1.90 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) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

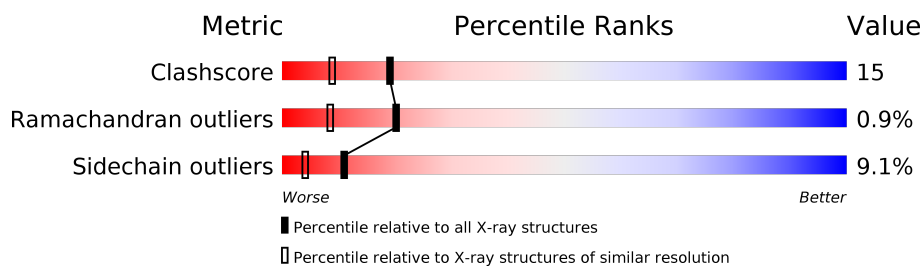
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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))
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)

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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	21	 67% 33%
2	D	21	 67% 24% 10%
3	A	61	 67% 18% 5% 10%
3	B	61	 70% 18% • 8%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 1990 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 DNA chain called DNA (5'-D(\*TP\*TP\*TP\*TP\*GP\*CP\*CP\*AP\*TP\*GP\*TP\*AP\*AP\*TP\*CP\*CP\*CP\*CP\*GP\*GP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	21	Total	C	N	O	P	0	0	0
			423	204	72	127	20			

- Molecule 2 is a DNA chain called DNA (5'-D(\*AP\*TP\*CP\*CP\*GP\*GP\*GP\*GP\*AP\*TP\*TP\*AP\*CP\*AP\*TP\*GP\*GP\*GP\*GP\*AP\*AP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	21	Total	C	N	O	P	0	0	0
			432	206	85	121	20			

- Molecule 3 is a protein called PROTEIN (ENGRAILED HOMEODOMAIN Q50K).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	A	55	Total	C	N	O	0	1	0
			476	297	96	83			
3	B	56	Total	C	N	O	0	0	0
			476	297	98	81			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	50	LYS	GLN	ENGINEERED	UNP P02836
B	50	LYS	GLN	ENGINEERED	UNP P02836

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	39	Total	O	0	0
			39	39		
4	B	45	Total	O	0	0
			45	45		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	51	Total	O	0	0
			51	51		
4	D	48	Total	O	0	0
			48	48		

### 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 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.

Note EDS was not executed.

- Molecule 1: DNA (5'-D(\*TP\*TP\*TP\*TP\*GP\*CP\*CP\*AP\*TP\*GP\*TP\*AP\*AP\*TP\*CP\*CP\*CP\*CP\*GP\*GP\*A)-3')

Chain C: 



- Molecule 2: DNA (5'-D(\*AP\*TP\*CP\*CP\*GP\*GP\*GP\*GP\*AP\*TP\*TP\*AP\*CP\*AP\*TP\*GP\*GP\*CP\*AP\*AP\*A)-3')

Chain D: 



- Molecule 3: PROTEIN (ENGRAILED HOMEODOMAIN Q50K)

Chain A: 



- Molecule 3: PROTEIN (ENGRAILED HOMEODOMAIN Q50K)

Chain B: 



## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	127.72Å 45.34Å 72.47Å 90.00° 119.53° 90.00°	Depositor
Resolution (Å)	6.00 – 1.90	Depositor
% Data completeness (in resolution range)	94.0 (6.00-1.90)	Depositor
$R_{merge}$	0.04	Depositor
$R_{sym}$	0.04	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.205 , 0.251	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	1990	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

## 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	C	0.76	0/472	1.02	0/726
2	D	0.69	0/486	0.96	0/749
3	A	0.59	0/486	0.73	0/642
3	B	0.61	0/482	0.74	0/640
All	All	0.67	0/1926	0.88	0/2757

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	C	0	2
2	D	0	2
All	All	0	4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	16	DC	Sidechain
1	C	17	DC	Sidechain
2	D	22	DA	Sidechain
2	D	37	DG	Sidechain

### 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	C	423	0	240	11	0
2	D	432	0	237	15	0
3	A	476	0	502	12	0
3	B	476	0	493	11	0
4	A	39	0	0	3	0
4	B	45	0	0	0	0
4	C	51	0	0	2	0
4	D	48	0	0	1	0
All	All	1990	0	1472	48	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 15.

All (48) 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:D:37:DG:H2''	2:D:38:DG:C5'	1.92	0.98
2:D:37:DG:H2''	2:D:38:DG:H5''	1.56	0.86
2:D:37:DG:C2'	2:D:38:DG:H5''	2.08	0.84
2:D:25:DC:H2''	2:D:26:DG:C8	2.12	0.84
2:D:37:DG:H2''	2:D:38:DG:H5'	1.66	0.76
3:A:29:ARG:O	3:A:33:GLN:HG3	1.93	0.68
3:A:24:ARG:HD2	4:A:102:HOH:O	1.96	0.64
3:A:15:ARG:NH2	3:A:30:ARG:HH21	1.95	0.64
2:D:25:DC:C2'	2:D:26:DG:C8	2.82	0.62
1:C:1:DT:H2'	1:C:2:DT:H72	1.82	0.62
1:C:1:DT:H5'	1:C:1:DT:H6	1.65	0.61
3:B:15:ARG:HG2	3:B:38:LEU:HD21	1.80	0.61
1:C:1:DT:H2'	1:C:2:DT:C7	2.30	0.61
3:A:19:GLU:HG3	3:A:34:LEU:HD21	1.83	0.61
1:C:6:DC:H2''	1:C:7:DC:C6	2.37	0.59
4:C:39:HOH:O	3:B:46:LYS:HE2	2.04	0.57
1:C:6:DC:H2''	1:C:7:DC:C5	2.43	0.54
3:A:30:ARG:NH1	3:A:33:GLN:OE1	2.41	0.54
3:B:19:GLU:HG2	3:B:34:LEU:HD11	1.90	0.54
2:D:37:DG:H1'	2:D:38:DG:H5''	1.91	0.52
4:D:46:HOH:O	3:B:3:ARG:HD2	2.09	0.52
3:A:15:ARG:NH2	3:A:30:ARG:NH2	2.57	0.52
2:D:37:DG:C1'	2:D:38:DG:H5''	2.41	0.51

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:6:DC:H5	4:C:62:HOH:O	1.92	0.50
1:C:1:DT:C6	1:C:1:DT:H5'	2.46	0.50
2:D:22:DA:H2'	2:D:23:DT:H72	1.92	0.50
1:C:1:DT:H2'	1:C:2:DT:C5	2.48	0.49
3:A:52:LYS:O	3:A:56:ILE:HG12	2.12	0.48
2:D:25:DC:H2''	2:D:26:DG:H8	1.74	0.47
3:B:15:ARG:HH22	3:B:30:ARG:HH21	1.62	0.47
1:C:1:DT:H2'	1:C:2:DT:C6	2.50	0.46
3:A:15:ARG:HH21	3:A:30:ARG:HH21	1.64	0.46
3:A:58:LYS:HD2	4:A:111:HOH:O	2.15	0.46
2:D:37:DG:C2'	2:D:38:DG:C5'	2.71	0.46
3:B:34:LEU:HA	3:B:34:LEU:HD23	1.76	0.45
3:B:15:ARG:HD2	3:B:15:ARG:HA	1.77	0.44
3:A:55:LYS:O	3:A:58:LYS:HG2	2.17	0.44
1:C:6:DC:H1'	1:C:7:DC:C6	2.54	0.43
3:B:52:LYS:O	3:B:56:ILE:HG12	2.19	0.43
2:D:22:DA:H2'	2:D:23:DT:C7	2.49	0.43
3:B:40:LEU:HA	3:B:40:LEU:HD23	1.84	0.42
3:A:15:ARG:O	3:A:19:GLU:HG2	2.20	0.42
1:C:8:DA:C2	2:D:37:DG:C2	3.08	0.42
2:D:22:DA:H2''	2:D:23:DT:O5'	2.19	0.42
3:B:15:ARG:NH2	3:B:19:GLU:OE1	2.51	0.41
2:D:24:DC:H2''	2:D:25:DC:O5'	2.21	0.41
3:A:24:ARG:NE	4:A:79:HOH:O	2.45	0.41
3:B:31:ARG:NH1	3:B:46:LYS:HD3	2.36	0.41

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	
3	A	54/61 (88%)	52 (96%)	2 (4%)	0	100	100

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
3	B	54/61 (88%)	52 (96%)	1 (2%)	1 (2%)	9 2
All	All	108/122 (88%)	104 (96%)	3 (3%)	1 (1%)	20 8

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	B	3	ARG

### 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
3	A	51/55 (93%)	45 (88%)	6 (12%)	6 2
3	B	49/55 (89%)	46 (94%)	3 (6%)	22 11
All	All	100/110 (91%)	91 (91%)	9 (9%)	11 4

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	6	THR
3	A	11	GLU
3	A	30	ARG
3	A	55	LYS
3	A	58	LYS
3	A	59	SER
3	B	11	GLU
3	B	15	ARG
3	B	33	GLN

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

Mol	Chain	Res	Type
3	A	44	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	B	33	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](#)

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

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.