



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 05:44 am GMT

PDB ID : 1HWT  
Title : STRUCTURE OF A HAP1/DNA COMPLEX REVEALS DRAMATICALLY  
ASYMMETRIC DNA BINDING BY A HOMODIMERIC PROTEIN  
Authors : King, D.A.; Zhang, L.; Guarente, L.; Marmorstein, R.  
Deposited on : 1998-09-17  
Resolution : 2.50 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

<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

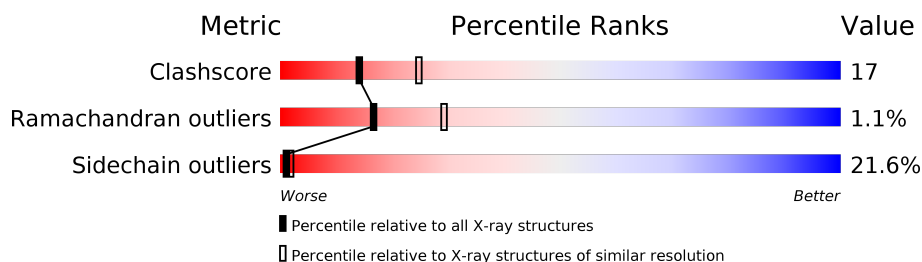
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)


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	A	20	
1	E	20	
2	B	20	
2	F	20	
3	C	81	
3	D	81	
3	G	81	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	H	81	 A horizontal bar chart showing the quality of chain H. The bar is divided into four segments: green (53%), yellow (28%), orange (9%), and grey (10%).

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 4062 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(\*GP\*CP\*GP\*CP\*TP\*AP\*TP\*TP\*AP\*TP\*CP\*GP\*CP\*TP\*AP\*TP\*TP\*AP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	20	Total	C	N	O	P	0	0	0
			404	195	69	121	19			
1	E	20	Total	C	N	O	P	0	0	0
			404	195	69	121	19			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	20	Total	C	N	O	P	0	0	0
			410	196	80	115	19			
2	F	20	Total	C	N	O	P	0	0	0
			410	196	80	115	19			

- Molecule 3 is a protein called PROTEIN (HEME ACTIVATOR PROTEIN).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	70	Total	C	N	O	S	0	0	0
			577	358	110	102	7			
3	D	74	Total	C	N	O	S	0	0	0
			614	378	122	107	7			
3	G	70	Total	C	N	O	S	0	0	0
			577	358	110	102	7			
3	H	73	Total	C	N	O	S	0	0	0
			599	369	117	106	7			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	H	2	Total	Zn	0	0
			2	2		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	G	2	Total 2	Zn 2	0	0
4	D	3	Total 3	Zn 3	0	0
4	C	2	Total 2	Zn 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	4	Total 4	O 4	0	0
5	B	4	Total 4	O 4	0	0
5	C	3	Total 3	O 3	0	0
5	D	15	Total 15	O 15	0	0
5	E	4	Total 4	O 4	0	0
5	F	9	Total 9	O 9	0	0
5	G	10	Total 10	O 10	0	0
5	H	9	Total 9	O 9	0	0

### 3 Residue-property plots

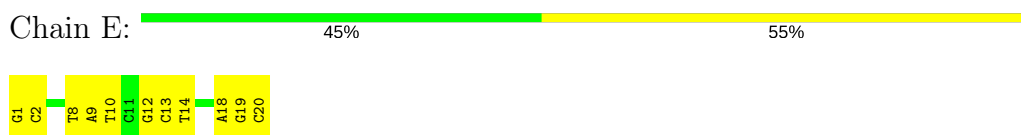
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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(\*GP\*CP\*GP\*CP\*TP\*AP\*TP\*TP\*AP\*TP\*CP\*GP\*CP\*TP\*AP\*TP\*TP\*AP\*GP\*C)-3')



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



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



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



- Molecule 3: PROTEIN (HEME ACTIVATOR PROTEIN)





- Molecule 3: PROTEIN (HEME ACTIVATOR PROTEIN)

Chain D: 53% 31% 6% • 9%



- Molecule 3: PROTEIN (HEME ACTIVATOR PROTEIN)

Chain G: 53% 26% 7% 14%



- Molecule 3: PROTEIN (HEME ACTIVATOR PROTEIN)

Chain H: 53% 28% 9% 10%



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.60Å 85.70Å 94.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 2.50	Depositor
% Data completeness (in resolution range)	82.2 (8.00-2.50)	Depositor
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
Refinement program	X-PLOR 3.851	Depositor
R, $R_{free}$	0.246 , 0.291	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4062	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP



## 5 Model quality

### 5.1 Standard geometry

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

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	0.53	0/451	0.73	0/694
1	E	0.57	0/451	0.78	0/694
2	B	0.58	0/461	0.78	0/710
2	F	0.57	0/461	0.80	0/710
3	C	0.49	0/584	0.57	0/778
3	D	0.48	0/621	0.62	0/825
3	G	0.46	0/584	0.58	0/778
3	H	0.44	0/606	0.61	0/807
All	All	0.51	0/4219	0.68	0/5996

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	2
2	B	0	3
2	F	0	2
All	All	0	7

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	10	DT	Sidechain
1	A	3	DG	Sidechain

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
2	B	1	DG	Sidechain
2	B	16	DA	Sidechain
2	B	9	DC	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	A	404	0	229	14	0
1	E	404	0	229	17	0
2	B	410	0	226	12	0
2	F	410	0	226	18	0
3	C	577	0	604	26	0
3	D	614	0	642	26	0
3	G	577	0	604	28	0
3	H	599	0	618	20	0
4	C	2	0	0	0	0
4	D	3	0	0	0	0
4	G	2	0	0	0	0
4	H	2	0	0	0	0
5	A	4	0	0	0	0
5	B	4	0	0	0	0
5	C	3	0	0	1	0
5	D	15	0	0	6	0
5	E	4	0	0	0	0
5	F	9	0	0	1	0
5	G	10	0	0	2	0
5	H	9	0	0	0	0
All	All	4062	0	3378	129	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 17.

The worst 5 of 129 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:F:7:DA:H1'	2:F:8:DG:H5'	1.39	1.01
3:D:126:LYS:HE2	5:D:151:HOH:O	1.61	1.01
3:G:113:GLU:HG2	5:G:142:HOH:O	1.67	0.94
3:C:69:LYS:HE2	3:D:103:GLU:OE2	1.75	0.86
3:G:82:GLN:HE21	3:G:86:LYS:HD2	1.39	0.86

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	
3	C	68/81 (84%)	66 (97%)	2 (3%)	0	100	100
3	D	72/81 (89%)	67 (93%)	4 (6%)	1 (1%)	13	23
3	G	68/81 (84%)	65 (96%)	2 (3%)	1 (2%)	12	21
3	H	71/81 (88%)	66 (93%)	4 (6%)	1 (1%)	13	23
All	All	279/324 (86%)	264 (95%)	12 (4%)	3 (1%)	17	29

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	G	126	LYS
3	H	127	THR
3	D	126	LYS

### 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	
3	C	66/77 (86%)	53 (80%)	13 (20%)	1	2
3	D	69/77 (90%)	53 (77%)	16 (23%)	1	1
3	G	66/77 (86%)	54 (82%)	12 (18%)	2	3
3	H	67/77 (87%)	50 (75%)	17 (25%)	0	1
All	All	268/308 (87%)	210 (78%)	58 (22%)	1	2

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	115	LYS
3	G	83	GLN
3	H	115	LYS
3	D	126	LYS
3	G	68	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
3	C	82	GLN
3	C	112	ASN
3	G	82	GLN
3	G	112	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 carbohydrates in this entry.

## 5.6 Ligand geometry

Of 9 ligands modelled in this entry, 9 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

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.