



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 15, 2023 – 12:13 PM EDT

PDB ID : 1XHU  
Title : HincII bound to cleaved, cognate DNA containing GTCGAC  
Authors : Etzkorn, C.; Horton, N.C.  
Deposited on : 2004-09-20  
Resolution : 2.95 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  
Xtrriage (Phenix) : **NOT EXECUTED**  
EDS : **NOT EXECUTED**  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.35

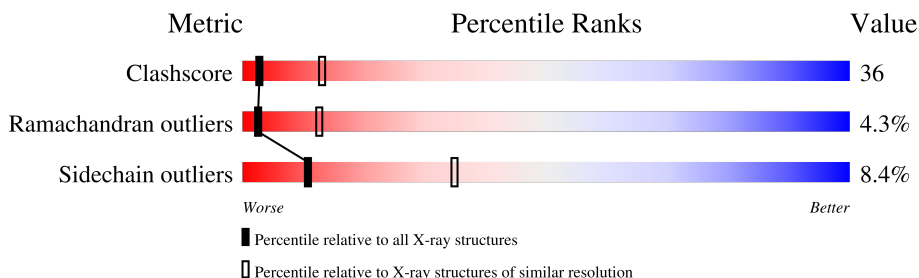
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)




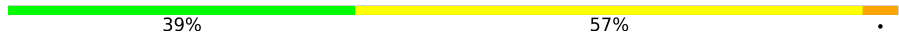
The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria 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  $\leq 5\%$

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	E	7	43% 57%
1	G	7	14% 71% 14%
1	I	7	29% 71%
1	K	7	100%
2	F	6	17% 50% 17% 17%
2	H	6	67% 17% 17%
2	J	6	67% 17% 17%
2	L	6	67% 17% 17%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	A	257	 39% 52% 9%
3	B	257	 45% 48% 7%
3	C	257	 39% 53% 7%
3	D	257	 39% 57% 4%

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 9758 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 5'-D(\*GP\*CP\*CP\*GP\*GP\*TP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	7	Total	C	N	O	P	0	0	0
			140	67	26	41	6			
1	G	7	Total	C	N	O	P	0	0	0
			140	67	26	41	6			
1	I	7	Total	C	N	O	P	0	0	0
			140	67	26	41	6			
1	K	7	Total	C	N	O	P	0	0	0
			140	67	26	41	6			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	6	Total	C	N	O	P	0	0	0
			126	58	26	36	6			
2	H	6	Total	C	N	O	P	0	0	0
			126	58	26	36	6			
2	J	6	Total	C	N	O	P	0	0	0
			126	58	26	36	6			
2	L	6	Total	C	N	O	P	0	0	0
			126	58	26	36	6			

- Molecule 3 is a protein called Type II restriction enzyme HincII.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	257	Total	C	N	O	S	0	0	0
			2079	1348	335	390	6			
3	B	256	Total	C	N	O	S	0	0	0
			2064	1339	334	385	6			
3	C	254	Total	C	N	O	S	0	0	0
			1967	1282	320	359	6			
3	D	256	Total	C	N	O	S	0	0	0
			2025	1315	327	377	6			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	130	THR	ARG	conflict	UNP P17743
A	173	TRP	SER	conflict	UNP P17743
B	130	THR	ARG	conflict	UNP P17743
B	173	TRP	SER	conflict	UNP P17743
C	130	THR	ARG	conflict	UNP P17743
C	173	TRP	SER	conflict	UNP P17743
D	130	THR	ARG	conflict	UNP P17743
D	173	TRP	SER	conflict	UNP P17743

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	13	Total O 13 13	0	0
4	F	11	Total O 11 11	0	0
4	G	14	Total O 14 14	0	0
4	H	15	Total O 15 15	0	0
4	I	15	Total O 15 15	0	0
4	J	2	Total O 2 2	0	0
4	K	9	Total O 9 9	0	0
4	L	9	Total O 9 9	0	0
4	A	146	Total O 146 146	0	0
4	B	192	Total O 192 192	0	0
4	C	73	Total O 73 73	0	0
4	D	60	Total O 60 60	0	0

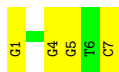
### 3 Residue-property plots

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. 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.

Note EDS was not executed.

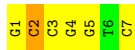
- Molecule 1: 5'-D(\*GP\*CP\*CP\*GP\*GP\*TP\*C)-3'

Chain E: 



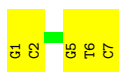
- Molecule 1: 5'-D(\*GP\*CP\*CP\*GP\*GP\*TP\*C)-3'

Chain G: 

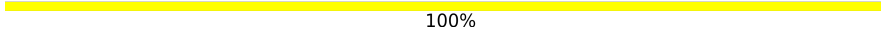


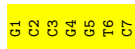
- Molecule 1: 5'-D(\*GP\*CP\*CP\*GP\*GP\*TP\*C)-3'

Chain I: 




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

Chain K: 




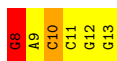
- Molecule 2: 5'-D(P\*GP\*AP\*CP\*CP\*GP\*G)-3'

Chain F: 

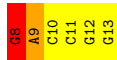


- Molecule 2: 5'-D(P\*GP\*AP\*CP\*CP\*GP\*G)-3'

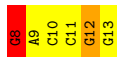
Chain H: 



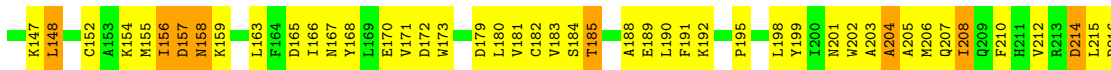
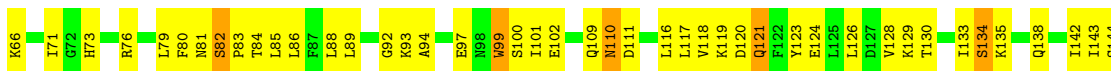
- Molecule 2: 5'-D(P\*GP\*AP\*CP\*CP\*GP\*G)-3'



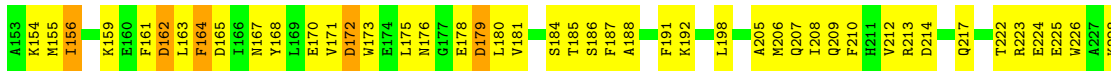
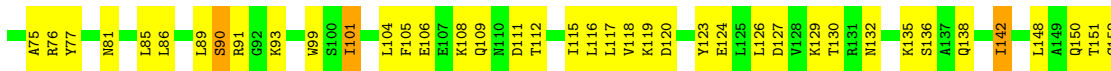
- Molecule 2: 5'-D(P\*GP\*AP\*CP\*CP\*GP\*G)-3'



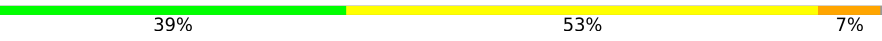
- Molecule 3: Type II restriction enzyme HincII

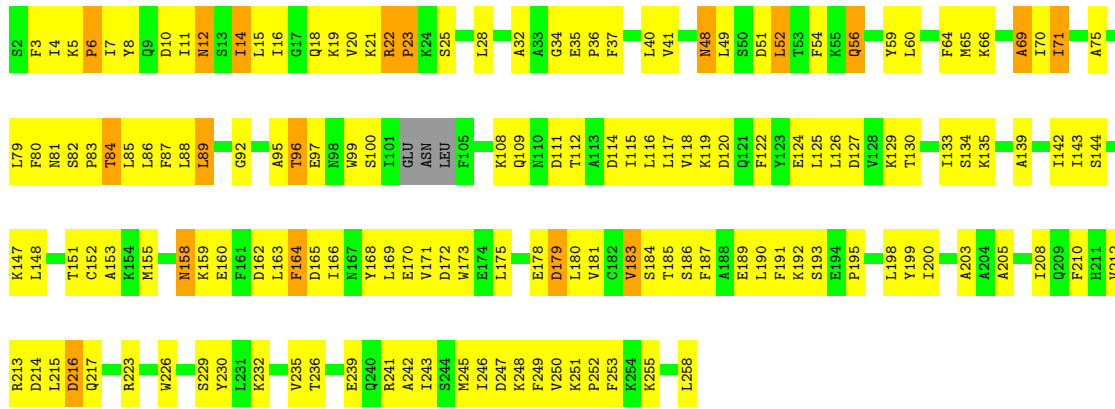


- Molecule 3: Type II restriction enzyme HincII



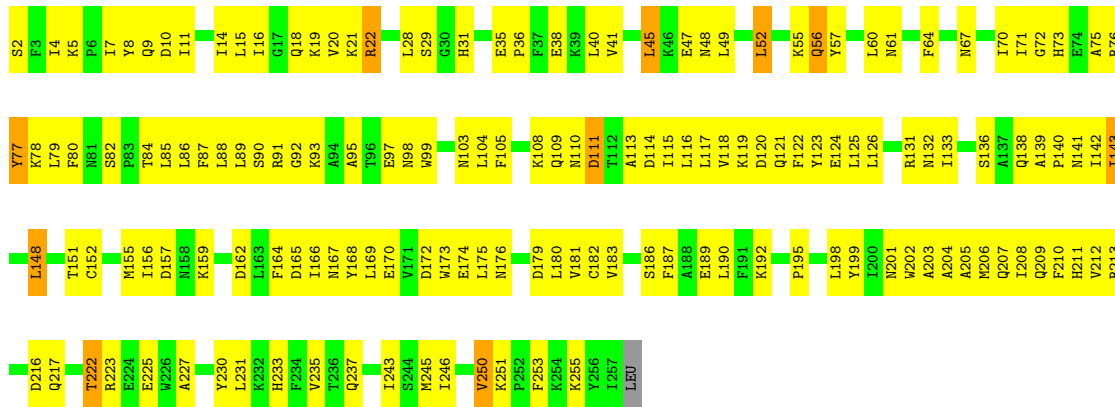
- Molecule 3: Type II restriction enzyme HincII

Chain C:  39% 53% 7%



• Molecule 3: Type II restriction enzyme HincII

Chain D:  39% 57% 4%





## 4 Data and refinement statistics

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

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.20Å 177.20Å 256.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.95	Depositor
% Data completeness (in resolution range)	85.4 (50.00-2.95)	Depositor
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
Refinement program	CNS	Depositor
R, $R_{free}$	0.187 , 0.277	Depositor
Estimated twinning fraction	No twinning to report.	Xtrriage
Total number of atoms	9758	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	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	E	0.76	0/156	0.73	0/239
1	G	1.01	0/156	0.93	0/239
1	I	0.73	0/156	0.86	0/239
1	K	0.66	0/156	0.81	0/239
2	F	1.17	1/141 (0.7%)	1.09	1/214 (0.5%)
2	H	1.04	2/141 (1.4%)	1.12	1/214 (0.5%)
2	J	0.95	1/141 (0.7%)	0.82	0/214
2	L	0.90	1/141 (0.7%)	0.90	0/214
3	A	0.52	0/2127	0.70	0/2877
3	B	0.53	0/2112	0.70	0/2857
3	C	0.38	0/2012	0.54	0/2736
3	D	0.43	0/2073	0.58	0/2815
All	All	0.55	5/9512 (0.1%)	0.68	2/13097 (0.0%)

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	G	0	1
2	F	0	2
2	H	0	4
2	J	0	2
2	L	0	2
3	A	0	1
3	B	0	1
All	All	0	13

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	J	8	DG	OP3-P	-7.20	1.52	1.61

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	8	DG	OP3-P	-7.15	1.52	1.61
2	L	8	DG	OP3-P	-7.04	1.52	1.61
2	H	8	DG	OP3-P	-6.49	1.53	1.61
2	H	8	DG	P-O5'	5.35	1.65	1.59

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	8	DG	OP1-P-OP2	-5.53	111.31	119.60
2	F	9	DA	O4'-C1'-N9	5.05	111.54	108.00

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	8	DG	Sidechain
2	F	9	DA	Sidechain
1	G	2	DC	Sidechain
2	H	8	DG	Sidechain
2	H	9	DA	Sidechain

## 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	E	140	0	80	8	0
1	G	140	0	80	13	0
1	I	140	0	80	8	0
1	K	140	0	80	16	0
2	F	126	0	67	10	0
2	H	126	0	67	4	0
2	J	126	0	67	11	0
2	L	126	0	67	10	0
3	A	2079	0	2031	138	0
3	B	2064	0	2019	144	0
3	C	1967	0	1859	170	0
3	D	2025	0	1930	151	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	146	0	0	10	0
4	B	192	0	0	20	1
4	C	73	0	0	10	0
4	D	60	0	0	5	0
4	E	13	0	0	2	0
4	F	11	0	0	0	0
4	G	14	0	0	4	0
4	H	15	0	0	0	0
4	I	15	0	0	0	0
4	J	2	0	0	0	0
4	K	9	0	0	1	0
4	L	9	0	0	7	0
All	All	9758	0	8427	626	1

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 36.

The worst 5 of 626 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:B:101:ILE:HD13	3:B:101:ILE:H	1.01	1.13
3:B:162:ASP:HB3	4:B:342:HOH:O	1.60	0.99
3:B:101:ILE:H	3:B:101:ILE:CD1	1.77	0.96
3:B:101:ILE:HD13	3:B:101:ILE:N	1.81	0.94
3:B:28:LEU:H	3:B:28:LEU:HD23	1.30	0.94

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:397:HOH:O	4:B:397:HOH:O[8_556]	1.21	0.99

## 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	A	255/257 (99%)	208 (82%)	31 (12%)	16 (6%)	1	6
3	B	254/257 (99%)	208 (82%)	40 (16%)	6 (2%)	6	26
3	C	250/257 (97%)	183 (73%)	50 (20%)	17 (7%)	1	5
3	D	254/257 (99%)	206 (81%)	43 (17%)	5 (2%)	7	30
All	All	1013/1028 (98%)	805 (80%)	164 (16%)	44 (4%)	2	12

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	33	ALA
3	A	56	GLN
3	A	157	ASP
3	C	69	ALA
3	C	158	ASN

### 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	A	222/229 (97%)	200 (90%)	22 (10%)	8	27
3	B	220/229 (96%)	200 (91%)	20 (9%)	9	31
3	C	196/229 (86%)	183 (93%)	13 (7%)	16	46
3	D	208/229 (91%)	192 (92%)	16 (8%)	13	39
All	All	846/916 (92%)	775 (92%)	71 (8%)	11	35

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	22	ARG
3	D	45	LEU
3	D	148	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	B	28	LEU
3	B	10	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	233	HIS
3	D	110	ASN
3	D	237	GLN
3	D	176	ASN
3	B	98	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 monosaccharides 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

EDS was not executed - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates [i](#)

EDS was not executed - this section is therefore empty.

### 6.4 Ligands [i](#)

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers [i](#)

EDS was not executed - this section is therefore empty.