



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 04:53 AM GMT

PDB ID : 1AIH  
Title : CATALYTIC DOMAIN OF BACTERIOPHAGE HP1 INTEGRASE  
Authors : Hickman, A.B.; Waninger, S.; ScoCCA, J.J.; Dyda, F.  
Deposited on : 1997-04-17  
Resolution : 2.50 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

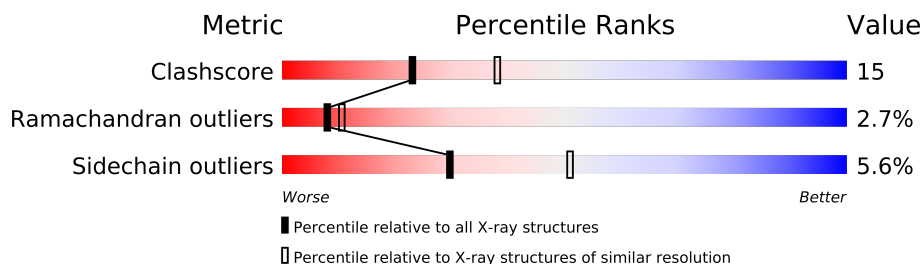
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	21963
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

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	79885	3562 (2.50-2.50)
Ramachandran outliers	78287	3480 (2.50-2.50)
Sidechain outliers	78261	3482 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	A	170	
1	B	170	
1	C	170	
1	D	170	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5545 atoms, of which 0 are hydrogen and 0 are deuterium.

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 HP1 INTEGRASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	170	Total	C	N	O	S	0	0	0
			1349	854	235	252	8			
1	B	168	Total	C	N	O	S	0	0	0
			1330	845	233	244	8			
1	C	168	Total	C	N	O	S	0	0	0
			1330	845	233	244	8			
1	D	170	Total	C	N	O	S	0	0	0
			1339	849	232	250	8			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	241	ASP	LYS	CONFLICT	UNP P21442
B	241	ASP	LYS	CONFLICT	UNP P21442
C	241	ASP	LYS	CONFLICT	UNP P21442
D	241	ASP	LYS	CONFLICT	UNP P21442

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	C	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Mg	0	0
			1	1		
3	A	1	Total	Mg	0	0
			1	1		
3	D	1	Total	Mg	0	0
			1	1		
3	C	2	Total	Mg	0	0
			2	2		

- Molecule 4 is water.

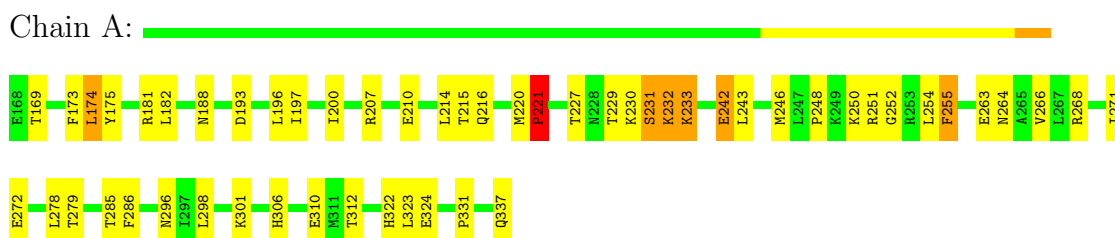
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	54	Total 54	O 54	0	0
4	B	33	Total 33	O 33	0	0
4	C	39	Total 39	O 39	0	0
4	D	26	Total 26	O 26	0	0

### 3 Residue-property plots

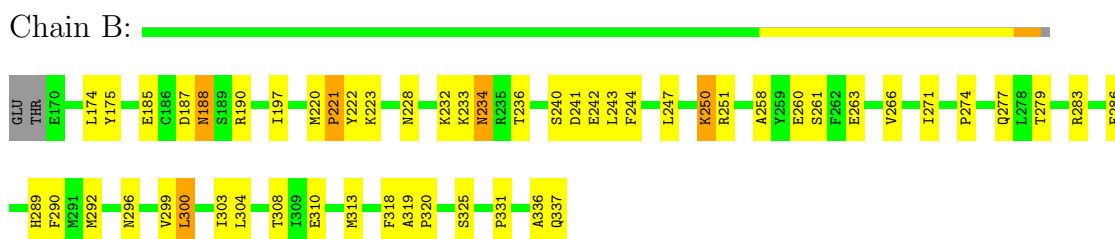
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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.

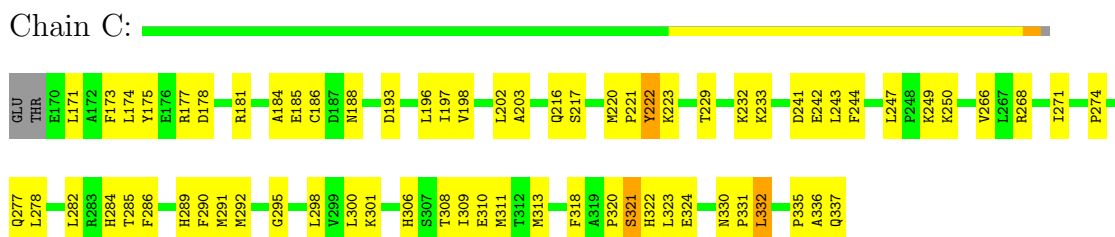
#### • Molecule 1: HP1 INTEGRASE



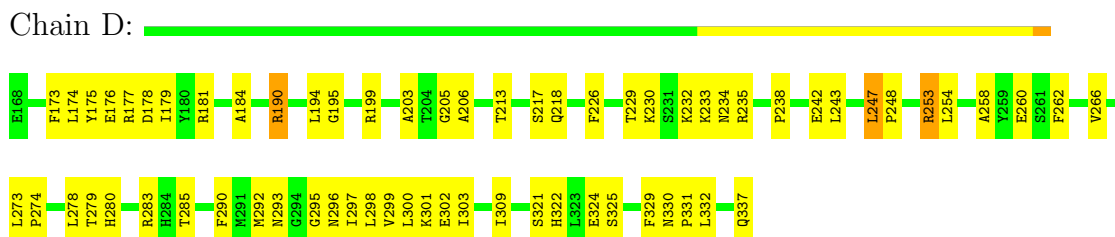
#### • Molecule 1: HP1 INTEGRASE



#### • Molecule 1: HP1 INTEGRASE



#### • Molecule 1: HP1 INTEGRASE



## 4 Data and refinement statistics

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

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.40Å 129.30Å 234.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.50	Depositor
% Data completeness (in resolution range)	97.5 (30.00-2.50)	Depositor
$R_{merge}$	0.09	Depositor
$R_{sym}$	0.09	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.209 , 0.272	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5545	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	48.0	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: MG, SO4

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	0/1378	0.66	0/1868
1	B	0.42	0/1359	0.64	0/1841
1	C	0.41	0/1359	0.61	0/1841
1	D	0.38	0/1368	0.61	0/1856
All	All	0.42	0/5464	0.63	0/7406

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1349	0	1320	34	7
1	B	1330	0	1310	45	0
1	C	1330	0	1310	47	7
1	D	1339	0	1305	53	0
2	A	15	0	0	1	0
2	B	10	0	0	1	0
2	C	5	0	0	0	0
2	D	10	0	0	3	3

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	2	0	0	0	0
3	D	1	0	0	0	0
4	A	54	0	0	2	3
4	B	33	0	0	4	0
4	C	39	0	0	5	0
4	D	26	0	0	11	0
All	All	5545	0	5245	155	10

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 15.

The worst 5 of 155 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:D:194:LEU:HD11	4:D:2120:HOH:O	1.80	0.81
1:A:227:THR:HA	1:A:233:LYS:O	1.82	0.80
1:A:323:LEU:HD12	1:B:299:VAL:HG13	1.67	0.74
1:D:262:PHE:HA	4:D:2120:HOH:O	1.87	0.73
1:A:251:ARG:HD2	2:A:109:SO4:O2	1.89	0.72

The worst 5 of 10 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	Distance(Å)	Clash(Å)
2:D:111:SO4:O2	4:A:2008:HOH:O[2_465]	1.23	0.97
1:A:188:ASN:O	1:C:188:ASN:OD1[3_555]	1.57	0.63
1:A:188:ASN:OD1	1:C:188:ASN:ND2[3_555]	1.64	0.56
1:A:181:ARG:NH2	1:C:181:ARG:NH2[3_555]	1.84	0.36
1:A:272:GLU:OE1	1:C:177:ARG:NH2[3_555]	1.85	0.35

## 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	168/170 (99%)	155 (92%)	9 (5%)	4 (2%)	9	13
1	B	166/170 (98%)	151 (91%)	11 (7%)	4 (2%)	9	13
1	C	166/170 (98%)	150 (90%)	12 (7%)	4 (2%)	9	13
1	D	168/170 (99%)	147 (88%)	15 (9%)	6 (4%)	5	6
All	All	668/680 (98%)	603 (90%)	47 (7%)	18 (3%)	8	10

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	231	SER
1	C	233	LYS
1	D	321	SER
1	A	232	LYS
1	B	188	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	146/150 (97%)	134 (92%)	12 (8%)	17	29
1	B	143/150 (95%)	138 (96%)	5 (4%)	48	74
1	C	143/150 (95%)	135 (94%)	8 (6%)	30	51
1	D	144/150 (96%)	137 (95%)	7 (5%)	35	59
All	All	576/600 (96%)	544 (94%)	32 (6%)	30	51

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	261	SER
1	C	217	SER
1	D	253	ARG
1	B	300	LEU
1	C	221	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	188	ASN
1	C	188	ASN
1	D	188	ASN
1	D	234	ASN
1	D	280	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains 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 13 ligands modelled in this entry, 5 are monoatomic - leaving 8 for Mogul analysis.

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	$\# Z  > 2$
2	SO4	A	105	-	4,4,4	1.19	0	6,6,6	0.12	0
2	SO4	A	109	-	4,4,4	1.05	0	6,6,6	0.09	0
2	SO4	A	112	-	4,4,4	1.08	0	6,6,6	0.12	0
2	SO4	B	106	-	4,4,4	1.30	0	6,6,6	0.07	0
2	SO4	B	110	-	4,4,4	1.12	0	6,6,6	0.11	0
2	SO4	C	107	-	4,4,4	1.24	0	6,6,6	0.06	0
2	SO4	D	108	-	4,4,4	1.24	0	6,6,6	0.09	0
2	SO4	D	111	-	4,4,4	1.06	0	6,6,6	0.17	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	SO4	A	105	-	-	0/0/0/0	0/0/0/0
2	SO4	A	109	-	-	0/0/0/0	0/0/0/0
2	SO4	A	112	-	-	0/0/0/0	0/0/0/0
2	SO4	B	106	-	-	0/0/0/0	0/0/0/0
2	SO4	B	110	-	-	0/0/0/0	0/0/0/0
2	SO4	C	107	-	-	0/0/0/0	0/0/0/0
2	SO4	D	108	-	-	0/0/0/0	0/0/0/0
2	SO4	D	111	-	-	0/0/0/0	0/0/0/0

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.

## 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 will therefore be empty.

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

EDS was not executed - this section will therefore be empty.

### 6.3 Carbohydrates ⓘ

EDS was not executed - this section will therefore be empty.

### 6.4 Ligands ⓘ

EDS was not executed - this section will therefore be empty.

### 6.5 Other polymers ⓘ

EDS was not executed - this section will therefore be empty.