



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 01:24 PM GMT

PDB ID : 1QA7  
Title : CRYSTAL COMPLEX OF THE 3C PROTEINASE FROM HEPATITIS A VIRUS WITH ITS INHIBITOR AND IMPLICATIONS FOR THE POLYPROTEIN PROCESSING IN HAV  
Authors : Bergmann, E.M.; Cherney, M.M.; Mckendrick, J.; Vederas, J.C.; James, M.N.G.  
Deposited on : 1999-04-15  
Resolution : 1.90 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>

---

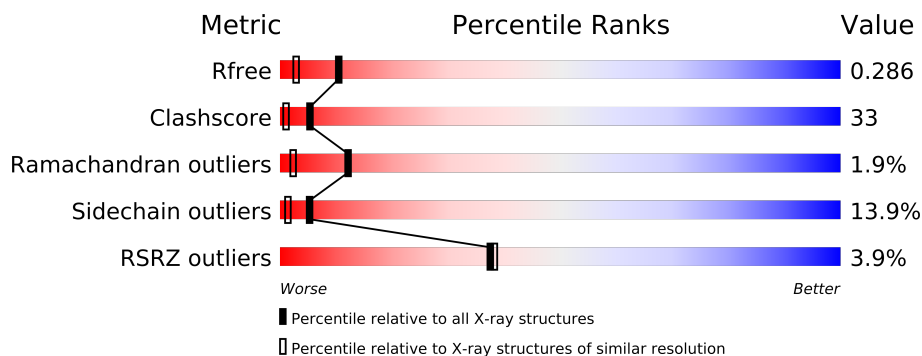
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
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 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))
$R_{free}$	66092	3684 (1.90-1.90)
Clashscore	79885	4465 (1.90-1.90)
Ramachandran outliers	78287	4413 (1.90-1.90)
Sidechain outliers	78261	4414 (1.90-1.90)
RSRZ outliers	66119	3686 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	217	
1	B	217	
1	C	217	
1	D	217	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	IVF	B	301	-	X
3	DMS	D	302	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7228 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.

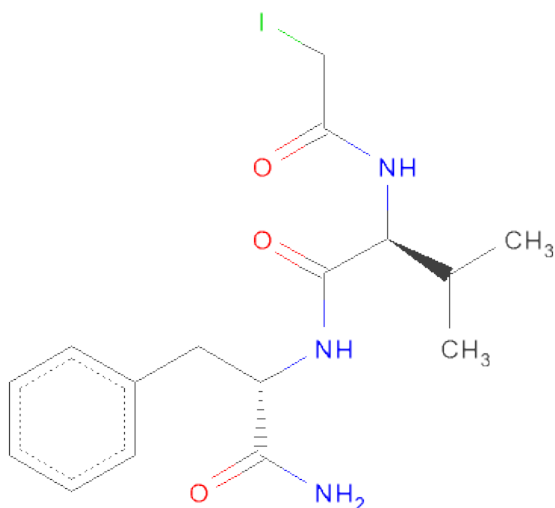
- Molecule 1 is a protein called HAV 3C PROTEINASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	213	Total	C	N	O	S	0	0	0
			1636	1040	281	306	9			
1	B	217	Total	C	N	O	S	0	0	0
			1672	1063	287	313	9			
1	C	213	Total	C	N	O	S	0	0	0
			1636	1040	281	306	9			
1	D	217	Total	C	N	O	S	0	0	0
			1672	1063	287	313	9			

There are 12 discrepancies between the modelled and reference sequences:

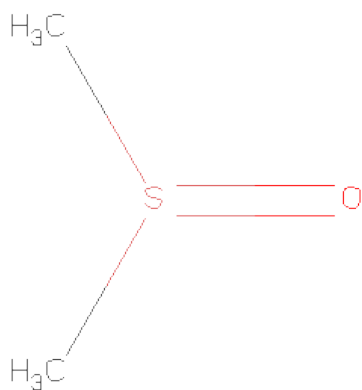
Chain	Residue	Modelled	Actual	Comment	Reference
A	24	SER	CYS	ENGINEERED	UNP P26582
A	82	ALA	PHE	ENGINEERED	UNP P26582
A	101	GLN	GLU	CONFLICT	UNP P26582
B	24	SER	CYS	ENGINEERED	UNP P26582
B	82	ALA	PHE	ENGINEERED	UNP P26582
B	101	GLN	GLU	CONFLICT	UNP P26582
C	24	SER	CYS	ENGINEERED	UNP P26582
C	82	ALA	PHE	ENGINEERED	UNP P26582
C	101	GLN	GLU	CONFLICT	UNP P26582
D	24	SER	CYS	ENGINEERED	UNP P26582
D	82	ALA	PHE	ENGINEERED	UNP P26582
D	101	GLN	GLU	CONFLICT	UNP P26582

- Molecule 2 is N-(IODOACETYL)-L-VALYL-L-PHENYLALANINAMIDE (three-letter code: IVF) (formula: C<sub>16</sub>H<sub>22</sub>IN<sub>3</sub>O<sub>3</sub>).



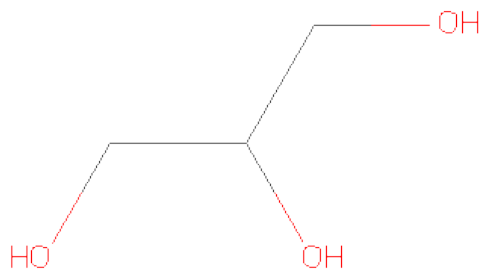
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			22	16	3	3		
2	B	1	Total	C	N	O	0	0
			22	16	3	3		
2	C	1	Total	C	N	O	0	0
			22	16	3	3		
2	D	1	Total	C	N	O	0	0
			22	16	3	3		

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C<sub>2</sub>H<sub>6</sub>OS).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	D	1	Total	C	O	S	0	0
			4	2	1	1		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	D	1	Total	C	O	0	0
			6	3	3		

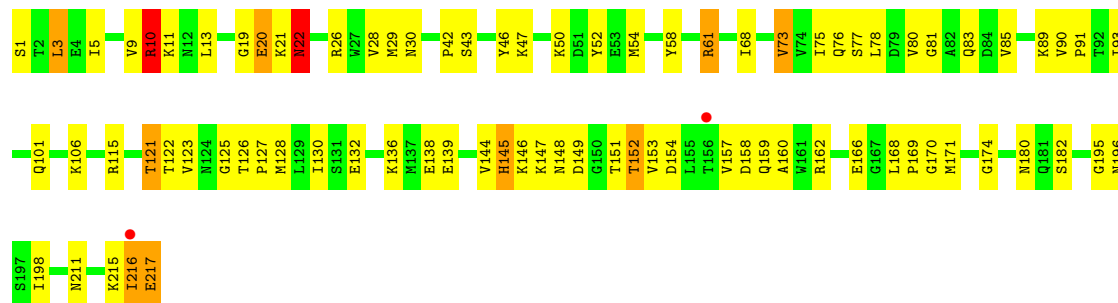
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	131	Total	O	0	0
			131	131		
5	B	129	Total	O	0	0
			129	129		
5	C	121	Total	O	0	0
			121	121		
5	D	133	Total	O	0	0
			133	133		



● Molecule 1: HAV 3C PROTEINASE

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.56Å 78.36Å 105.29Å 90.00° 97.50° 90.00°	Depositor
Resolution (Å)	15.00 – 1.90 14.96 – 1.89	Depositor EDS
% Data completeness (in resolution range)	92.3 (15.00-1.90) 91.6 (14.96-1.89)	Depositor EDS
$R_{merge}$	0.03	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.42 (at 1.90Å)	Xtriage
Refinement program	TNT V. 5-E	Depositor
R, $R_{free}$	0.204 , 0.298 0.199 , 0.286	Depositor DCC
$R_{free}$ test set	3902 reflections (7.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	26.4	Xtriage
Anisotropy	0.295	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.47 , 121.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 59384 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7228	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.73% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality i

### 5.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, IVF, DMS

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	0.75	0/1665	0.92	1/2252 (0.0%)
1	B	3.39	1/1701 (0.1%)	0.96	4/2297 (0.2%)
1	C	0.70	0/1665	0.95	1/2252 (0.0%)
1	D	4.90	2/1701 (0.1%)	0.96	3/2297 (0.1%)
All	All	3.04	3/6732 (0.0%)	0.95	9/9098 (0.1%)

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	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	217	GLU	C-OXT	199.38	5.02	1.23
1	B	217	GLU	C-OXT	136.55	3.82	1.23
1	D	46	TYR	CD1-CE1	-5.26	1.31	1.39

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	10	ARG	NE-CZ-NH2	-8.28	116.16	120.30
1	B	150	GLY	N-CA-C	-7.59	94.11	113.10
1	C	113	LEU	CA-CB-CG	7.11	131.65	115.30
1	B	148	ASN	N-CA-C	6.40	128.28	111.00
1	D	10	ARG	NE-CZ-NH1	5.98	123.29	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	132	GLU	Mainchain

## 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	1636	0	1661	112	0
1	B	1672	0	1704	89	0
1	C	1636	0	1661	157	0
1	D	1672	0	1704	90	0
2	A	22	0	20	7	0
2	B	22	0	20	0	0
2	C	22	0	20	3	0
2	D	22	0	20	4	0
3	D	4	0	6	0	0
4	D	6	0	8	0	0
5	A	131	0	0	7	0
5	B	129	0	0	18	0
5	C	121	0	0	16	0
5	D	133	0	0	17	0
All	All	7228	0	6824	442	0

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 33.

The worst 5 of 442 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:138:GLU:HG3	1:C:162:ARG:HD2	1.36	1.06
1:D:76:GLN:HE22	1:D:211:ASN:HB2	1.14	1.04
1:C:128:MET:HE1	1:C:168:LEU:HG	1.43	0.99
1:A:167:GLY:HA3	1:A:171:MET:HE1	1.49	0.92
1:B:144:VAL:HG21	1:B:196:ASN:ND2	1.86	0.91

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	
1	A	211/217 (97%)	193 (92%)	14 (7%)	4 (2%)	12	2
1	B	215/217 (99%)	195 (91%)	16 (7%)	4 (2%)	12	2
1	C	211/217 (97%)	195 (92%)	10 (5%)	6 (3%)	8	1
1	D	215/217 (99%)	196 (91%)	17 (8%)	2 (1%)	25	10
All	All	852/868 (98%)	779 (91%)	57 (7%)	16 (2%)	12	2

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	82	ALA
1	B	53	GLU
1	B	83	GLN
1	C	146	LYS
1	C	149	ASP

### 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	178/182 (98%)	148 (83%)	30 (17%)	3	1
1	B	182/182 (100%)	163 (90%)	19 (10%)	10	3
1	C	178/182 (98%)	145 (82%)	33 (18%)	2	0
1	D	182/182 (100%)	164 (90%)	18 (10%)	11	4
All	All	720/728 (99%)	620 (86%)	100 (14%)	5	1

5 of 100 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	149	ASP
1	C	61	ARG
1	D	132	GLU
1	B	154	ASP
1	C	8	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	206	GLN
1	B	210	GLN
1	D	72	ASN
1	B	148	ASN
1	D	76	GLN

### 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 ⓘ

6 ligands are modelled in this entry.

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	IVF	A	301	1	21,22,23	1.36	1 (4%)	27,29,30	0.91	0
2	IVF	B	301	1	21,22,23	1.31	1 (4%)	27,29,30	0.59	0
2	IVF	C	301	1	21,22,23	1.48	1 (4%)	27,29,30	0.91	1 (3%)
2	IVF	D	301	1	21,22,23	1.45	1 (4%)	27,29,30	0.92	3 (11%)
3	DMS	D	302	-	3,3,3	0.13	0	3,3,3	0.45	0
4	GOL	D	303	-	5,5,5	0.55	0	5,5,5	1.16	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	IVF	A	301	1	-	0/24/24/26	0/1/1/1
2	IVF	B	301	1	-	0/24/24/26	0/1/1/1
2	IVF	C	301	1	-	0/24/24/26	0/1/1/1
2	IVF	D	301	1	-	0/24/24/26	0/1/1/1
3	DMS	D	302	-	-	0/0/0/0	0/0/0/0
4	GOL	D	303	-	-	0/4/4/4	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	301	IVF	C-NXT	6.32	1.46	1.33
2	D	301	IVF	C-NXT	5.85	1.45	1.33
2	A	301	IVF	C-NXT	5.64	1.44	1.33
2	B	301	IVF	C-NXT	5.42	1.44	1.33

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	301	IVF	C2-CA2-N2	3.58	120.26	110.53
2	D	301	IVF	C-CA-N	-2.80	104.67	111.28
2	D	301	IVF	O1-C1-N2	-2.22	118.71	122.26
2	D	301	IVF	C2-CA2-N2	2.11	116.27	110.53

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 ⓘ

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	213/217 (98%)	-0.03	11 (5%) 26 26	13, 27, 53, 71	0
1	B	217/217 (100%)	-0.06	7 (3%) 45 46	14, 26, 50, 63	0
1	C	213/217 (98%)	0.24	14 (6%) 18 17	16, 31, 55, 67	0
1	D	217/217 (100%)	-0.17	2 (0%) 81 83	11, 23, 47, 55	0
All	All	860/868 (99%)	-0.01	34 (3%) 37 37	11, 27, 52, 71	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	149	ASP	6.2
1	A	152	THR	6.0
1	D	216	ILE	5.5
1	C	147	LYS	5.3
1	C	148	ASN	5.3

### 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 ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	DMS	D	302	4/4	0.17	3.88	22,25,28,32	0
2	IVF	B	301	22/23	0.12	2.84	19,26,30,32	0
4	GOL	D	303	6/6	0.12	1.79	33,36,38,44	0
2	IVF	A	301	22/23	0.20	0.96	29,41,49,55	0
2	IVF	D	301	22/23	0.11	0.04	18,28,38,43	0
2	IVF	C	301	22/23	0.12	-0.43	21,33,40,43	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.