



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:16 AM GMT

PDB ID : 2GEF  
Title : Crystal structure of a Novel viral protease with a serine/lysine catalytic dyad mechanism  
Authors : Paetzel, M.; Feldman, A.R.; Lee, J.; Delmas, B.  
Deposited on : 2006-03-20  
Resolution : 2.20 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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

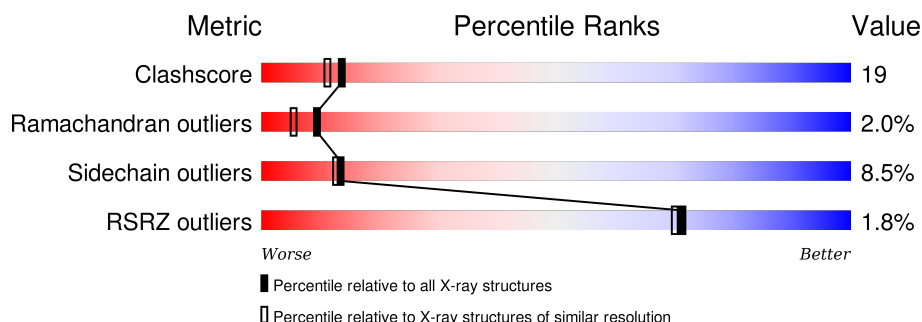
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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.

Mol	Chain	Length	Quality of chain
1	A	217	<div> <div>2%</div> <div>54% 30% 7% 9%</div> </div>
1	B	217	<div> <div>2%</div> <div>53% 32% 6% 6%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3285 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 protein called Protease VP4.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	198	Total	C	N	O	S	Se	0	0	0
			1488	935	258	289	2	4			
1	B	203	Total	C	N	O	S	Se	0	0	0
			1519	954	263	296	2	4			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	557	MET	-	INITIATING METHIONINE	UNP Q8AZM0
A	615	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
A	685	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
A	743	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
A	767	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
B	557	MET	-	INITIATING METHIONINE	UNP Q8AZM0
B	615	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
B	685	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
B	743	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0
B	767	MSE	MET	MODIFIED RESIDUE	UNP Q8AZM0

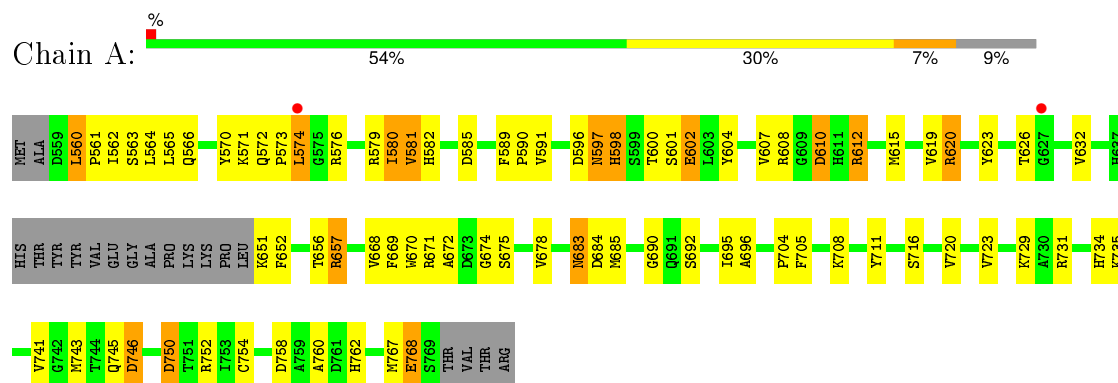
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	162	Total	O	0	0
			162	162		
2	B	116	Total	O	0	0
			116	116		

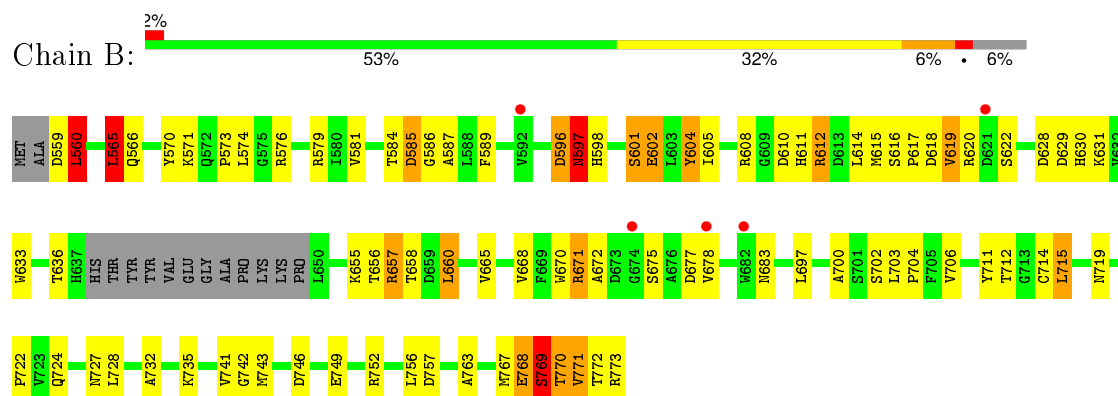
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

#### • Molecule 1: Protease VP4



#### • Molecule 1: Protease VP4



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	144.07Å 144.07Å 144.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.41 – 2.20 36.02 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.41-2.20) 99.9 (36.02-2.20)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	14.15 (at 2.20Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.222 , 0.267 0.235 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	33.9	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 77.9	EDS
Estimated twinning fraction	0.329 for -l,-k,-h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.22$	Xtriage
Outliers	0 of 25350 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	3285	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.87% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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	1.71	19/1516 (1.3%)	1.48	18/2060 (0.9%)
1	B	1.43	10/1547 (0.6%)	1.42	21/2104 (1.0%)
All	All	1.58	29/3063 (0.9%)	1.45	39/4164 (0.9%)

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	B	0	2

The worst 5 of 29 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	570	TYR	CZ-OH	8.27	1.51	1.37
1	B	597	ASN	N-CA	7.15	1.60	1.46
1	A	602	GLU	CB-CG	-7.09	1.38	1.52
1	A	591	VAL	CB-CG1	6.89	1.67	1.52
1	B	770	THR	CB-OG1	6.38	1.56	1.43

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	620	ARG	NE-CZ-NH2	10.24	125.42	120.30
1	A	671	ARG	NE-CZ-NH2	-8.80	115.90	120.30
1	B	752	ARG	NE-CZ-NH2	-8.59	116.00	120.30
1	A	574	LEU	CA-CB-CG	8.29	134.36	115.30
1	A	731	ARG	NE-CZ-NH2	-7.99	116.31	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	636	THR	Peptide
1	B	769	SER	Peptide

## 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	A	1488	0	1434	50	0
1	B	1519	0	1461	70	0
2	A	162	0	0	8	0
2	B	116	0	0	9	0
All	All	3285	0	2895	110	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 19.

The worst 5 of 110 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)
1:B:573:PRO:HG2	1:B:602:GLU:HG3	1.31	1.12
1:B:722:PRO:HB3	1:B:743:MSE:HE2	1.31	1.10
1:A:752:ARG:NH2	2:A:216:HOH:O	1.89	1.05
1:A:565:LEU:HD22	1:B:565:LEU:CD2	1.86	1.05
1:A:565:LEU:HD22	1:B:565:LEU:HD22	1.38	1.05

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	194/217 (89%)	182 (94%)	9 (5%)	3 (2%)	13	9
1	B	199/217 (92%)	181 (91%)	13 (6%)	5 (2%)	7	3
All	All	393/434 (91%)	363 (92%)	22 (6%)	8 (2%)	9	5

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	768	GLU
1	B	768	GLU
1	B	769	SER
1	B	771	VAL
1	A	597	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	158/173 (91%)	145 (92%)	13 (8%)	14	13
1	B	161/173 (93%)	147 (91%)	14 (9%)	13	12
All	All	319/346 (92%)	292 (92%)	27 (8%)	13	13

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	768	GLU
1	B	576	ARG
1	B	660	LEU
1	B	560	LEU
1	A	576	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	745	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	566	GLN
1	B	582	HIS
1	A	719	ASN
1	B	572	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 [i](#)

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

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	194/217 (89%)	-0.25	2 (1%) 84 83	6, 12, 24, 35	0
1	B	199/217 (91%)	0.28	5 (2%) 61 60	7, 24, 42, 48	0
All	All	393/434 (90%)	0.02	7 (1%) 71 70	6, 19, 39, 48	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	574	LEU	3.4
1	B	682	TRP	3.0
1	B	678	VAL	2.9
1	A	627	GLY	2.6
1	B	621	ASP	2.5

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

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

There are no such residues in this entry.