



# Full wwPDB X-ray Structure Validation Report ⓘ

May 25, 2020 – 07:35 PM EDT

PDB ID : 5R3E  
Title : PanDDA analysis group deposition – Auto-refined data of Endothiapepsin for ground state model 38, DMSO-Free  
Authors : Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G.M.A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M.S.  
Deposited on : 2020-02-13  
Resolution : 1.01 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.10.1
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.10.1

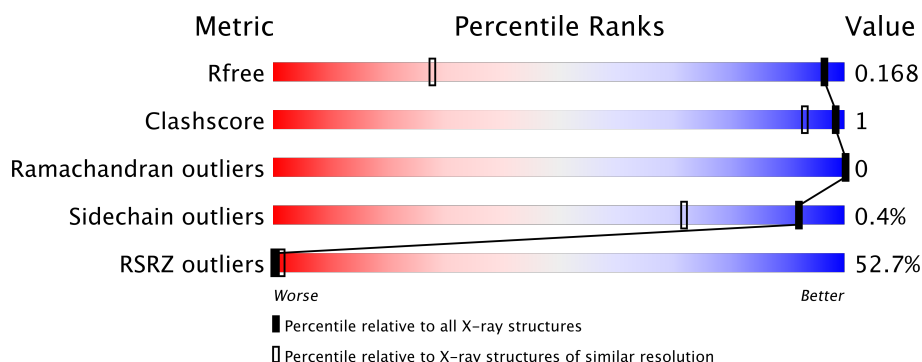
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.01 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))
$R_{free}$	111664	1020 (1.08-0.92)
Clashscore	122126	1084 (1.08-0.92)
Ramachandran outliers	120053	1010 (1.08-0.92)
Sidechain outliers	120020	1012 (1.08-0.92)
RSRZ outliers	108989	1485 (1.10-0.90)

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 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  $\leq 5\%$ . 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	419	<div> <div>42%</div> <div> <div></div> <div>77%</div> <div>•</div> <div>21%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5094 atoms, of which 2341 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 Endothiapepsin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	330	Total	C	H	N	O	S	0	25	0
			4803	1566	2341	367	527	2			

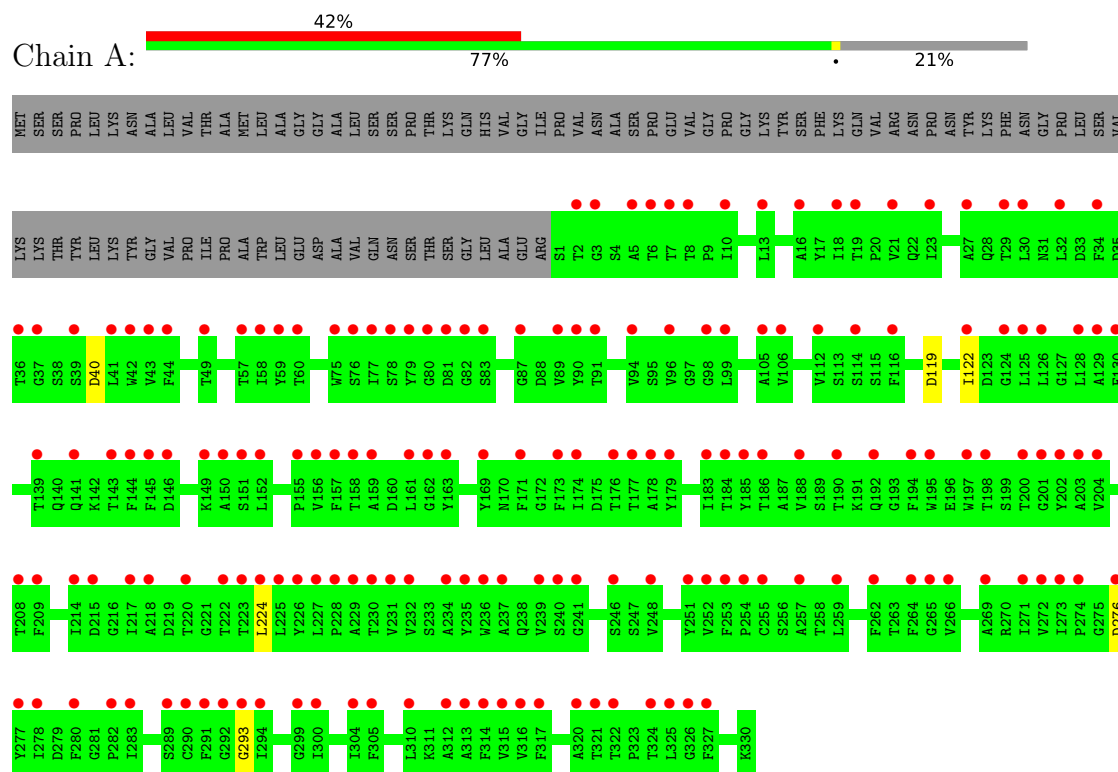
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	291	Total	O	0	0
			291	291		

### 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 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.

#### • Molecule 1: Endothiapepsin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.24Å 72.99Å 52.54Å 90.00° 109.23° 90.00°	Depositor
Resolution (Å)	42.72 – 1.01 49.61 – 1.01	Depositor EDS
% Data completeness (in resolution range)	95.2 (42.72-1.01) 95.3 (49.61-1.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.98 (at 1.01Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, $R_{free}$	0.139 , 0.153 0.152 , 0.168	Depositor DCC
$R_{free}$ test set	2099 reflections (1.30%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	12.8	Xtriage
Anisotropy	0.268	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.45 , 55.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	5094	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.53% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 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	A	0.34	0/2594	0.60	0/3553

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	2462	2341	2300	3	0
2	A	291	0	0	0	0
All	All	2753	2341	2300	3	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 1.

All (3) 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:A:276[B]:ASP:N	1:A:276[B]:ASP:OD1	2.37	0.55
1:A:119:ASP:OD2	1:A:122:ILE:HD12	2.18	0.42
1:A:224:LEU:HD22	1:A:293:GLY:HA2	2.01	0.41

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	
1	A	353/419 (84%)	349 (99%)	4 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	
1	A	273/336 (81%)	272 (100%)	1 (0%)	92	72

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	40	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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

### 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	330/419 (78%)	2.09	174 (52%) <b>0</b> <b>1</b>	11, 15, 22, 32	0

All (174) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	80	GLY	6.0
1	A	300[A]	ILE	5.9
1	A	150[A]	ALA	5.6
1	A	299[A]	GLY	4.7
1	A	78	SER	4.6
1	A	77	ILE	4.4
1	A	246[A]	SER	4.1
1	A	321	THR	4.0
1	A	259	LEU	3.9
1	A	241	GLY	3.9
1	A	79	TYR	3.8
1	A	231	VAL	3.6
1	A	177	THR	3.5
1	A	169	TYR	3.5
1	A	315	VAL	3.5
1	A	139	THR	3.4
1	A	236	TRP	3.4
1	A	145	PHE	3.3
1	A	314	PHE	3.3
1	A	276[A]	ASP	3.3
1	A	114	SER	3.2
1	A	222	THR	3.2
1	A	81	ASP	3.2
1	A	58	ILE	3.2
1	A	37	GLY	3.2
1	A	82	GLY	3.1
1	A	34	PHE	3.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	312	ALA	3.1
1	A	44	PHE	3.1
1	A	313	ALA	3.1
1	A	156	VAL	3.1
1	A	248	VAL	3.1
1	A	149[A]	LYS	3.0
1	A	257	ALA	3.0
1	A	43	VAL	3.0
1	A	41	LEU	3.0
1	A	174[A]	ILE	3.0
1	A	194	PHE	3.0
1	A	59	TYR	3.0
1	A	197	TRP	3.0
1	A	294	ILE	3.0
1	A	290	CYS	2.9
1	A	264	PHE	2.9
1	A	32	LEU	2.9
1	A	146	ASP	2.9
1	A	8	THR	2.9
1	A	188	VAL	2.8
1	A	225	LEU	2.8
1	A	18	ILE	2.8
1	A	269	ALA	2.8
1	A	163	TYR	2.8
1	A	253	PHE	2.8
1	A	151	SER	2.8
1	A	122	ILE	2.8
1	A	90	TYR	2.8
1	A	21	VAL	2.7
1	A	96	VAL	2.7
1	A	126	LEU	2.7
1	A	161	LEU	2.7
1	A	144	PHE	2.7
1	A	173	PHE	2.7
1	A	325[A]	LEU	2.7
1	A	255	CYS	2.7
1	A	23	ILE	2.7
1	A	217	ILE	2.7
1	A	305	PHE	2.7
1	A	240[A]	SER	2.6
1	A	7	THR	2.6
1	A	190	THR	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	42	TRP	2.6
1	A	226	TYR	2.6
1	A	130	PHE	2.6
1	A	293	GLY	2.6
1	A	57	THR	2.6
1	A	224	LEU	2.6
1	A	234	ALA	2.6
1	A	195	TRP	2.6
1	A	280	PHE	2.6
1	A	129	ALA	2.6
1	A	327	PHE	2.6
1	A	252	VAL	2.5
1	A	266	VAL	2.5
1	A	277[A]	TYR	2.5
1	A	83	SER	2.5
1	A	317	PHE	2.5
1	A	10	ILE	2.5
1	A	271	ILE	2.5
1	A	304	ILE	2.5
1	A	36	THR	2.5
1	A	176	THR	2.5
1	A	178	ALA	2.5
1	A	215[A]	ASP	2.5
1	A	75	TRP	2.5
1	A	60	THR	2.5
1	A	204	VAL	2.4
1	A	87	GLY	2.4
1	A	273	ILE	2.4
1	A	198	THR	2.4
1	A	157	PHE	2.4
1	A	159	ALA	2.4
1	A	232	VAL	2.4
1	A	239	VAL	2.4
1	A	322	THR	2.4
1	A	278	ILE	2.4
1	A	283	ILE	2.4
1	A	99	LEU	2.4
1	A	125	LEU	2.4
1	A	152	LEU	2.4
1	A	235	TYR	2.4
1	A	98	GLY	2.4
1	A	116	PHE	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	262	PHE	2.4
1	A	192	GLN	2.4
1	A	316	VAL	2.4
1	A	185	TYR	2.4
1	A	209	PHE	2.4
1	A	6	THR	2.3
1	A	89	VAL	2.3
1	A	105	ALA	2.3
1	A	237	ALA	2.3
1	A	214	ILE	2.3
1	A	310	LEU	2.3
1	A	202	TYR	2.3
1	A	16	ALA	2.3
1	A	186	THR	2.3
1	A	27	ALA	2.3
1	A	320	ALA	2.3
1	A	76	SER	2.3
1	A	272	VAL	2.3
1	A	19	THR	2.3
1	A	155	PRO	2.3
1	A	39	SER	2.2
1	A	218	ALA	2.2
1	A	201	GLY	2.2
1	A	2	THR	2.2
1	A	158	THR	2.2
1	A	208	THR	2.2
1	A	30	LEU	2.2
1	A	227	LEU	2.2
1	A	162	GLY	2.2
1	A	203	ALA	2.2
1	A	91	THR	2.2
1	A	143	THR	2.2
1	A	94	VAL	2.2
1	A	106	VAL	2.2
1	A	251	TYR	2.2
1	A	13	LEU	2.2
1	A	128	LEU	2.2
1	A	274	PRO	2.2
1	A	184	THR	2.2
1	A	220	THR	2.2
1	A	265	GLY	2.1
1	A	112	VAL	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	179	TYR	2.1
1	A	141	GLN	2.1
1	A	183[A]	ILE	2.1
1	A	228	PRO	2.1
1	A	282	PRO	2.1
1	A	229	ALA	2.1
1	A	49	THR	2.1
1	A	171	PHE	2.1
1	A	200	THR	2.1
1	A	3	GLY	2.1
1	A	124	GLY	2.1
1	A	292	GLY	2.1
1	A	5	ALA	2.0
1	A	223	THR	2.0
1	A	324	THR	2.0
1	A	289[A]	SER	2.0
1	A	291	PHE	2.0
1	A	326[A]	GLY	2.0
1	A	29	THR	2.0
1	A	230	THR	2.0
1	A	254	PRO	2.0

## 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.