



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:37 PM GMT

PDB ID : 4FLN  
Title : Crystal structure of plant protease Deg2  
Authors : Gong, W.; Liu, L.; Sun, R.; Gao, F.  
Deposited on : 2012-06-15  
Resolution : 2.80 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

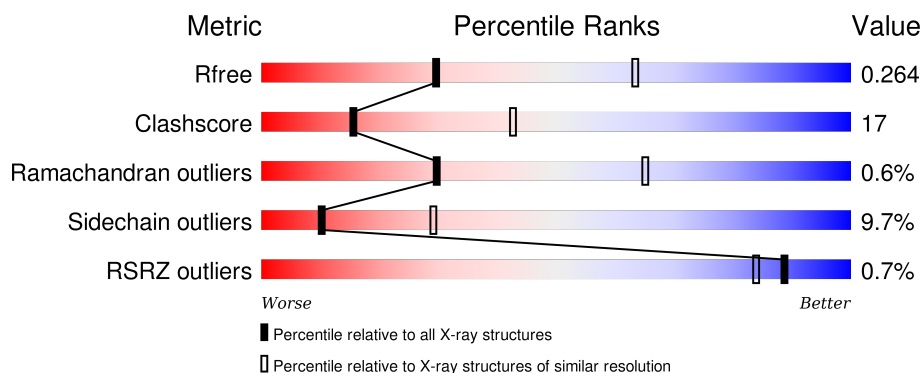
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	539	<div> <div></div> <div>55% 27% • 14%</div> </div>
1	B	539	<div> <div></div> <div>58% 26% • 14%</div> </div>
1	C	539	<div> <div></div> <div>53% 30% • 14%</div> </div>
2	D	4	<div> <div></div> <div>100%</div> </div>
2	F	4	<div> <div></div> <div>100%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	E	20	 90%10%

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10921 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 Do-like 2, chloroplastic.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	463	Total	C	N	O	S	8	4	0
			3553	2262	613	666	12			
1	B	464	Total	C	N	O	S	0	0	0
			3517	2240	604	661	12			
1	C	463	Total	C	N	O	S	0	2	0
			3506	2236	603	655	12			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	69	ASN	-	EXPRESSION TAG	UNP O82261
A	70	ALA	-	EXPRESSION TAG	UNP O82261
B	69	ASN	-	EXPRESSION TAG	UNP O82261
B	70	ALA	-	EXPRESSION TAG	UNP O82261
C	69	ASN	-	EXPRESSION TAG	UNP O82261
C	70	ALA	-	EXPRESSION TAG	UNP O82261

- Molecule 2 is a protein called Unknown peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	4	Total	C	N	O	0	0	0
			29	20	5	4			
2	F	4	Total	C	N	O	0	0	0
			29	20	5	4			

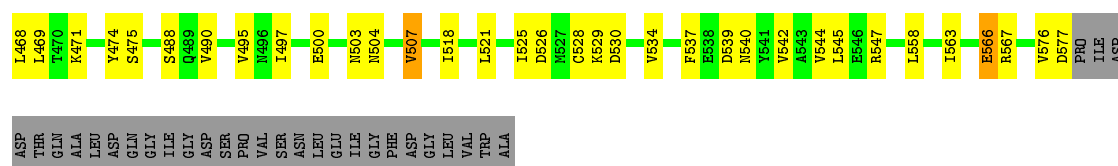
- Molecule 3 is a protein called Unknown peptide.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	E	20	Total	C	N	O	0	0	0
			109	66	21	22			

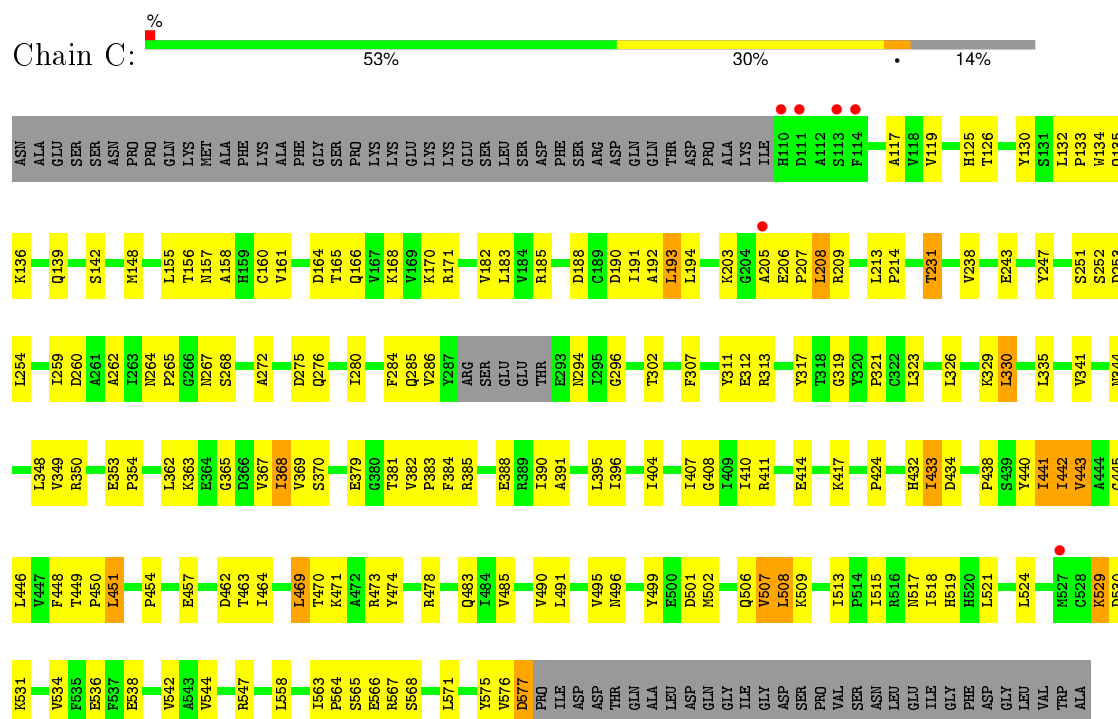
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	75	Total 75	O 75	0	0
4	B	61	Total 61	O 61	0	0
4	C	42	Total 42	O 42	0	0





- Molecule 1: Protease Do-like 2, chloroplastic



- Molecule 2: Unknown peptide



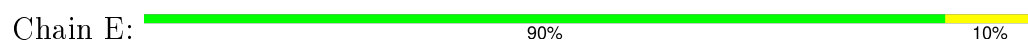
There are no outlier residues recorded for this chain.

- Molecule 2: Unknown peptide



There are no outlier residues recorded for this chain.

- Molecule 3: Unknown peptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	125.40 Å 188.16 Å 166.13 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.74 – 2.80 39.63 – 2.79	Depositor EDS
% Data completeness (in resolution range)	98.9 (29.74-2.80) 95.6 (39.63-2.79)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.43 (at 2.81 Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.1_743)	Depositor
R, $R_{free}$	0.203 , 0.274 0.193 , 0.264	Depositor DCC
$R_{free}$ test set	2336 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	62.2	Xtriage
Anisotropy	0.151	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 57.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 48432 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	10921	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% 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 [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	0.38	0/3634	0.65	1/4941 (0.0%)
1	B	0.35	0/3585	0.58	0/4878
1	C	0.31	0/3581	0.55	0/4874
2	D	0.44	0/15	0.44	0/20
2	F	0.33	0/15	0.34	0/20
3	E	0.33	0/22	0.34	0/28
All	All	0.35	0/10852	0.59	1/14761 (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	A	0	2
1	C	0	3
All	All	0	5

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	458	GLU	N-CA-C	-5.15	97.10	111.00

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	250	GLY	Peptide
1	A	575	TYR	Peptide
1	C	251	SER	Peptide
1	C	501	ASP	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	C	507	VAL	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	3553	0	3516	135	1
1	B	3517	0	3457	99	0
1	C	3506	0	3427	136	0
2	D	29	0	15	0	0
2	F	29	0	15	0	0
3	E	109	0	34	2	0
4	A	75	0	0	3	0
4	B	61	0	0	3	0
4	C	42	0	0	4	0
All	All	10921	0	10464	364	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 17.

The worst 5 of 364 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:C:341:VAL:HG12	1:C:410:ILE:HD12	1.51	0.91
1:A:526:ASP:OD1	1:A:567:ARG:NH2	2.04	0.89
1:A:461:GLU:HA	1:A:464:ILE:HD11	1.56	0.86
1:A:443:VAL:HG21	1:A:525:ILE:HD13	1.58	0.86
1:C:238:VAL:H	1:C:260:ASP:HB2	1.45	0.82

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)
1:A:139:GLN:NE2	1:A:490:VAL:O[3_555]	2.15	0.05

## 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	463/539 (86%)	423 (91%)	36 (8%)	4 (1%)	21	55
1	B	460/539 (85%)	430 (94%)	29 (6%)	1 (0%)	52	84
1	C	461/539 (86%)	427 (93%)	31 (7%)	3 (1%)	26	62
2	D	1/4 (25%)	1 (100%)	0	0	100	100
2	F	1/4 (25%)	1 (100%)	0	0	100	100
3	E	2/20 (10%)	2 (100%)	0	0	100	100
All	All	1388/1645 (84%)	1284 (92%)	96 (7%)	8 (1%)	30	65

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	433	ILE
1	B	433	ILE
1	C	433	ILE
1	A	442	ILE
1	A	576	VAL

### 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	381/463 (82%)	340 (89%)	41 (11%)	8	23
1	B	373/463 (81%)	334 (90%)	39 (10%)	8	24
1	C	367/463 (79%)	336 (92%)	31 (8%)	14	37

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	D	1/1 (100%)	1 (100%)	0	100	100
2	F	1/1 (100%)	1 (100%)	0	100	100
3	E	2/2 (100%)	2 (100%)	0	100	100
All	All	1125/1393 (81%)	1014 (90%)	111 (10%)	10	28

5 of 111 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	279	CYS
1	B	387	SER
1	C	469	LEU
1	B	295	ILE
1	B	368	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	432	HIS
1	C	267	ASN
1	C	519	HIS

### 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	463/539 (85%)	-0.47	3 (0%) 90 86	30, 53, 93, 140	0
1	B	464/539 (86%)	-0.39	1 (0%) 95 94	34, 60, 104, 137	0
1	C	463/539 (85%)	-0.13	6 (1%) 79 71	38, 78, 121, 144	0
2	D	1/4 (25%)	-0.65	0 100 100	72, 72, 72, 72	0
2	F	1/4 (25%)	-1.04	0 100 100	73, 73, 73, 73	0
3	E	2/20 (10%)	-0.67	0 100 100	56, 56, 56, 75	0
All	All	1394/1645 (84%)	-0.33	10 (0%) 89 84	30, 64, 111, 144	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	110	HIS	4.0
1	A	460	CYS	3.6
1	C	114	PHE	3.3
1	A	463	THR	2.8
1	C	113	SER	2.6

### 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.