

# Full wwPDB X-ray Structure Validation Report (i)

May 16, 2020 – 05:32 am BST

PDB ID : 3WG5

> Title 1510-N membrane-bound stomatin-specific protease K138A mutant in complex

> > with a substrate peptide under heat treatment

Authors : Yokoyama, H.; Fujii, S.; Matsui, I.

Deposited on 2013-07-26

2.40 Å(reported) Resolution

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 A user guide is available at

https://www.wwpdb.org/validation/2017/XrayValidationReportHelp with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity 4.02b-467

> 1.8.5 (274361), CSD as541be (2020) Mogul

Xtriage (Phenix) 1.13 EDS 2.11

Percentile statistics 20191225.v01 (using entries in the PDB archive December 25th 2019)

> Refmac 5.8.0158

7.0.044 (Gargrove) CCP4 Engh & Huber (2001)

Ideal geometry (proteins) Ideal geometry (DNA, RNA) Parkinson et al. (1996)

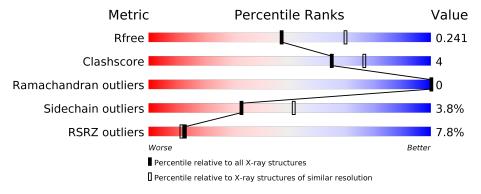
Validation Pipeline (wwPDB-VP) 2.11

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure: X- $RAY\ DIFFRACTION$ 

The reported resolution of this entry is 2.40 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	$egin{aligned}  ext{Whole archive} \ (\# ext{Entries}) \end{aligned}$	$\begin{array}{c} {\rm Similar \; resolution} \\ (\#{\rm Entries, \; resolution \; range(\AA)}) \end{array}$
$R_{free}$	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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 >=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 <=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	230	6%	85%		9%	• 5%
1	В	230	7%	86%		10%	5%
2	С	10	10%	50%	20%	20%	_



# 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3603 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 441aa long hypothetical nfeD protein.

$\mathbf{Mol}$	Chain	Residues	${f Atoms}$			ZeroOcc	AltConf	Trace		
1	Λ	219	Total	С	N	О	S	0	0	0
1	Λ	219	1687	1082	277	323	5	0	U	
1	B	219	Total	С	N	О	S	0	0	0
1	Б	219	1687	1082	277	323	5	0	0	U

There are 20 discrepancies between the modelled and reference sequences:

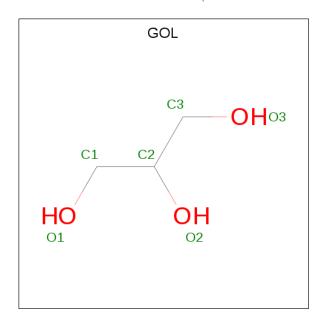
Chain	Residue	Modelled	Actual	Comment	Reference
A	15	MET	-	expression tag	UNP O59179
A	138	ALA	LYS	engineered mutation	UNP O59179
A	237	LEU	-	expression tag	UNP O59179
A	238	GLU	_	expression tag	UNP O59179
A	239	HIS	_	expression tag	UNP O59179
A	240	HIS	_	expression tag	UNP O59179
A	241	HIS	_	expression tag	UNP O59179
A	242	HIS	-	expression tag	UNP O59179
A	243	HIS	_	expression tag	UNP O59179
A	244	HIS	_	expression tag	UNP O59179
В	15	MET	_	expression tag	UNP O59179
В	138	ALA	LYS	engineered mutation	UNP O59179
В	237	LEU	-	expression tag	UNP O59179
В	238	GLU	_	expression tag	UNP O59179
В	239	HIS	_	expression tag	UNP O59179
В	240	HIS	_	expression tag	UNP O59179
В	241	HIS	_	expression tag	UNP O59179
В	242	HIS	-	expression tag	UNP O59179
В	243	HIS	-	expression tag	UNP O59179
В	244	HIS	_	expression tag	UNP O59179

• Molecule 2 is a protein called PH1511 stomatin.



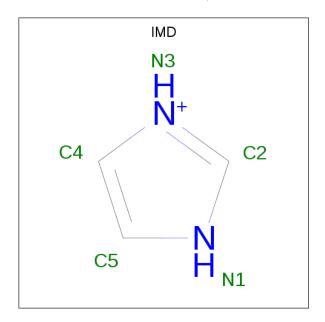
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
9	С	0	Total	С	Ν	О	S	0	0	0
		0	61	42	9	9	1	0	U	U

• Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	В	1	Total C O 6 3 3	0	0

• Molecule 4 is IMIDAZOLE (three-letter code: IMD) (formula:  $C_3H_5N_2$ ).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total C N 5 3 2	0	0
4	A	1	Total C N 5 3 2	0	0
4	В	1	Total C N 5 3 2	0	0

• Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Cl 1 1	0	0

• Molecule 6 is water.

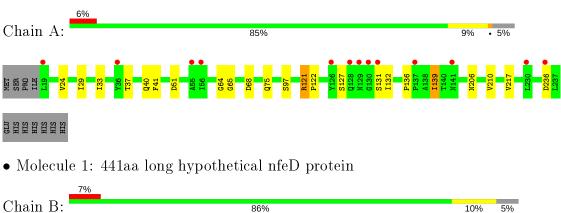
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	91	Total O 91 91	0	0
6	В	49	Total O 49 49	0	0

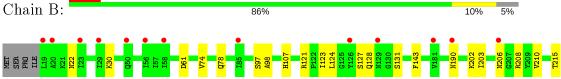


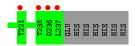
# 3 Residue-property plots (i)

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: 441aa long hypothetical nfeD protein







• Molecule 2: PH1511 stomatin







# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants	111.45Å 111.45Å 91.74Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 - 2.40	Depositor
Resolution (A)	19.93 - 2.40	EDS
% Data completeness	99.6 (20.00-2.40)	Depositor
(in resolution range)	99.9 (19.93-2.40)	EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	$7.05 \; (at \; 2.41 \text{Å})$	Xtriage
Refinement program	REFMAC 5.8.0044	Depositor
D D.	0.202 , 0.241	Depositor
$R, R_{free}$	0.202 , $0.241$	DCC
$R_{free}$ test set	2346 reflections (10.16%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	52.5	Xtriage
Anisotropy	0.091	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.36, 40.9	EDS
L-test for twinning <sup>2</sup>	$ < L > = 0.51, < L^2> = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3603	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 3.86% of the height of the origin peak. No significant pseudotranslation is detected.

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



<sup>&</sup>lt;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, IMD, CL

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		
IVIOI	Chain	RMSZ	# Z >5	RMSZ	# Z  > 5	
1	A	0.45	0/1717	0.56	0/2334	
1	В	0.44	0/1717	0.55	0/2334	
2	С	0.59	0/61	0.95	0/83	
All	All	0.45	0/3495	0.57	0/4751	

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)	$\mathbf{H}(\mathbf{added})$	Clashes	Symm-Clashes
1	Α	1687	0	1723	16	0
1	В	1687	0	1723	14	0
2	С	61	0	72	9	0
3	A	6	0	8	0	0
3	В	6	0	8	0	0
4	A	10	0	10	2	0
4	В	5	0	5	1	0
5	A	1	0	0	0	0
6	A	91	0	0	0	0
6	В	49	0	0	0	0
All	All	3603	0	3549	29	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 4.

All (29) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic	Clash
		${\rm distance} \; ({\rm \AA})$	overlap (Å)
1:B:121:ARG:O	2:C:241:PRO:HD2	1.74	0.86
1:B:123:ILE:HG12	2:C:241:PRO:HD3	1.59	0.84
1:A:24:VAL:HG23	1:A:217:VAL:HG21	1.61	0.82
1:A:139:ILE:HD11	2:C:237:VAL:HG21	1.74	0.68
1:A:68:ASP:HB3	1:B:124:LEU:O	2.01	0.61
1:A:24:VAL:HG23	1:A:217:VAL:CG2	2.32	0.58
1:B:97:SER:OG	1:B:98:ALA:N	2.37	0.57
1:B:127:SER:HB3	1:B:131:SER:HB2	1.85	0.57
1:B:74:VAL:O	1:B:78:GLN:HG2	2.06	0.55
1:B:202:LYS:HE2	1:B:202:LYS:HA	1.88	0.54
1:A:139:ILE:CD1	2:C:237:VAL:HG11	2.39	0.53
1:A:122:PRO:HA	2:C:235:VAL:HG23	1.92	0.51
1:B:121:ARG:O	2:C:241:PRO:CD	2.54	0.50
1:B:30:LYS:HG2	1:B:61:ASP:HB3	1.94	0.49
1:B:97:SER:OG	2:C:240:LEU:HB2	2.14	0.47
1:A:29:ILE:HD13	1:A:41:PHE:CE1	2.50	0.47
1:B:143:PHE:CE1	2:C:238:LEU:HD22	2.51	0.46
1:A:210:VAL:HG11	4:A:302:IMD:H4	1.98	0.45
1:A:121:ARG:O	2:C:234:ASN:HA	2.17	0.45
1:A:139:ILE:O	1:A:139:ILE:HG13	2.16	0.45
1:B:210:VAL:HG11	4:B:302:IMD:H4	1.99	0.44
1:B:22:ASN:ND2	1:B:215:THR:O	2.50	0.44
1:A:127:SER:OG	1:A:131:SER:HB2	2.18	0.43
1:A:64:GLY:HA2	1:A:97:SER:HB3	1.99	0.43
1:A:210:VAL:HG21	4:A:302:IMD:H5	2.01	0.42
1:B:107:HIS:CD2	1:B:203:ILE:HG13	2.54	0.42
1:A:136:PRO:O	1:A:139:ILE:HG23	2.20	0.41
1:A:33:ILE:HB	1:A:65:GLY:HA3	2.02	0.41
1:A:37:THR:HA	1:A:40:GLN:HE21	1.85	0.40

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	Perce	ntiles
1	A	217/230 (94%)	213 (98%)	4 (2%)	0	100	100
1	В	217/230 (94%)	209 (96%)	8 (4%)	0	100	100
2	С	6/10 (60%)	6 (100%)	0	0	100	100
All	All	440/470 (94%)	428 (97%)	12 (3%)	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	181/192~(94%)	174 (96%)	7 (4%)	32 50
1	В	181/192 (94%)	177 (98%)	4 (2%)	52 71
2	С	8/10 (80%)	5 (62%)	3 (38%)	0 0
All	All	370/394 (94%)	356 (96%)	14 (4%)	33 51

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	51	ASP
1	A	75	GLN
1	A	121	ARG
1	A	132	ILE
1	A	139	ILE

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type
1	A	206	ASN
1	A	236	ASP
1	В	128	GLN
1	В	190	ASN
1	В	206	ASN
1	В	208	ARG
2	С	235	VAL
2	С	238	LEU
2	С	239	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	40	GLN
1	A	78	GLN
1	A	141	ASN
1	В	22	ASN
1	В	32	GLN
1	В	40	GLN
1	В	52	ASN
1	В	79	GLN
1	В	152	GLN
1	В	206	ASN

#### 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)

Of 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 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	Iol Type Chain E	Chain	Res	Link	Bond lengths			Bond angles		
10101	Type		nes	LIIIK	Counts	RMSZ	# Z >2	Counts	RMSZ	# Z  > 2
4	IMD	В	302	-	3,5,5	0.40	0	4,5,5	0.62	0
4	IMD	A	302	-	3,5,5	0.39	0	4,5,5	0.77	0
3	GOL	A	301	_	5,5,5	0.42	0	5,5,5	0.46	0
3	GOL	В	301	-	5,5,5	0.41	0	5,5,5	0.52	0
4	IMD	A	303	_	3,5,5	0.40	0	4,5,5	0.63	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
4	IMD	В	302	_	-	_	0/1/1/1
4	IMD	A	302	_	-	-	0/1/1/1
3	GOL	A	301	_	-	4/4/4/4	-
3	GOL	В	301	-	-	2/4/4/4	-
4	IMD	A	303	-	-	-	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	301	GOL	C1-C2-C3-O3
3	A	301	GOL	O1-C1-C2-O2
3	A	301	GOL	O1-C1-C2-C3
3	В	301	GOL	O1-C1-C2-C3
3	A	301	GOL	O2-C2-C3-O3
3	В	301	GOL	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 3 short contacts:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	В	302	IMD	1	0
4	A	302	IMD	2	0

# 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,  $95^{th}$  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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(\AA^2)$	Q < 0.9
1	A	$219/230 \ (95\%)$	0.23	13 (5%) 22 21	35, 51, 81, 105	0
1	В	$219/230 \ (95\%)$	0.40	17 (7%) 13 11	38, 56, 89, 120	0
2	С	8/10 (80%)	3.52	5 (62%) 0 0	59, 77, 118, 130	0
All	All	446/470 (94%)	0.37	35 (7%) 13 11	35, 53, 85, 130	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	С	240	LEU	11.4
1	В	19	LEU	4.9
2	С	241	PRO	4.8
1	A	19	LEU	4.6
1	В	236	ASP	4.4
1	В	190	ASN	4.4
1	В	129	ASN	4.3
1	A	131	SER	4.0
2	С	239	MET	3.7
1	В	20	ALA	3.5
1	В	235	THR	3.4
1	A	130	GLY	3.4
1	A	129	ASN	3.4
1	В	221	TYR	3.4
1	В	237	LEU	3.1
1	A	126	TYR	3.0
2	С	234	ASN	3.0
1	A	137	PRO	2.9
1	В	50	GLN	2.8
2	С	238	LEU	2.7
1	В	56	ILE	2.5
1	A	230	LEU	2.5
1	A	236	ASP	2.5

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	RSRZ	
1	A	36	TYR	2.5	
1	A	55	ALA	2.5	
1	В	85	ILE	2.4	
1	В	23	ILE	2.3	
1	В	29	ILE	2.3	
1	A	56	ILE	2.2	
1	В	181	VAL	2.1	
1	A	141	ASN	2.1	
1	В	58	ILE	2.1	
1	В	126	TYR	2.1	
1	В	206	ASN	2.0	
1	A	128	GLN	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)

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. The B-factors column lists the minimum, median,  $95^{th}$  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	RSCC	RSR	${f B\text{-factors}}({f \AA}^2)$	Q < 0.9
3	GOL	A	301	6/6	0.80	0.40	83,87,88,88	0
4	IMD	В	302	5/5	0.81	0.26	94,95,98,98	0
4	IMD	A	303	5/5	0.88	0.26	83,85,89,89	0
3	GOL	В	301	6/6	0.95	0.16	50,52,52,54	0
4	IMD	A	302	5/5	0.98	0.14	44,44,45,46	0
5	CL	A	304	1/1	0.98	0.15	41,41,41,41	1

### 6.5 Other polymers (i)

There are no such residues in this entry.

