

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

Nov 7, 2023 – 06:37 PM JST

PDB ID : 8JUG

> Title : Crystal structure of human MMP-7 in complex with inhibitor

Authors : Kamitani, M.; Abe-Sato, K.; Oka, Y.

2023-06-26 Deposited on

1.30 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 types of validation reports are described at http://www.wwpdb.org/validation/2017/FAQs#types.

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 2.36

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

> Refmac 5.8.0158

CCP4 7.0.044 (Gargrove)

Ideal geometry (proteins) Engh & Huber (2001) Ideal geometry (DNA, RNA) Parkinson et al. (1996)

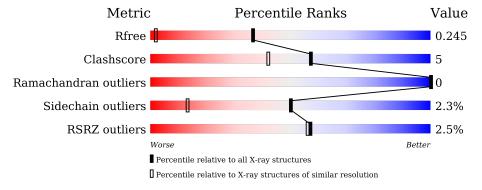
Validation Pipeline (wwPDB-VP) 2.36

# 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 1.30 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	Similar resolution
Metric	$(\# \mathrm{Entries})$	$(\#  ext{Entries},  ext{ resolution range}( ext{Å}))$
$R_{free}$	130704	1058 (1.30-1.30)
Clashscore	141614	1101 (1.30-1.30)
Ramachandran outliers	138981	1058 (1.30-1.30)
Sidechain outliers	138945	1058 (1.30-1.30)
RSRZ outliers	127900	1029 (1.30-1.30)

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 of 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	175	2%	87%	5% • 7%
2	С	7	29%	43%	29%



# 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 1520 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 Matrilysin.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	A	162	Total 1275	C 817	N 222	O 232	S 4	0	1	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	93	MET	-	initiating methionine	UNP P09237
A	94	GLY	-	expression tag	UNP P09237
A	215	GLN	GLU	engineered mutation	UNP P09237

• Molecule 2 is a protein called Peptide Inhibitor.

Mol	Chain	Residues		Atoms			ZeroOcc	AltConf	Trace			
2	С	7	Total 128	C 84	Cl 2	F 6	N 16	O 18	S 2	0	7	1

• Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	2	Total Ca 2 2	0	0

• Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	2	Total Zn 2 2	0	0

• Molecule 5 is water.



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	107	Total O 107 107	0	0
5	С	6	Total O 6 6	0	0



# 3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Matrilysin

Chain A:

87%

5% - 7%

Molecule 2: Peptide Inhibitor

Chain C:

29%

43%

29%



# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4	Depositor
Cell constants	76.06Å 76.06Å 61.89Å	Donositon
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	48.05 - 1.30	Depositor
Resolution (A)	48.01 - 1.30	EDS
% Data completeness	99.8 (48.05-1.30)	Depositor
(in resolution range)	99.8 (48.01-1.30)	EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	1.51  (at  1.31Å)	Xtriage
Refinement program	REFMAC 5.8.0189	Depositor
D.D.	0.221 , $0.238$	Depositor
$R, R_{free}$	0.227 , $0.245$	DCC
$R_{free}$ test set	2108  reflections  (4.91%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	11.1	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.41 , 45.6	EDS
L-test for twinning <sup>2</sup>	$< L > = 0.48, < L^2> = 0.31$	Xtriage
Estimated twinning fraction	0.039 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	1520	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 16.17% 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: EOE, CA, 7SF, NH2, GGL, ZN, TBG, V1C

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	0.71	0/1314	0.87	0/1782	
2	С	0.97	0/6	2.18	0/4	
All	All	0.72	0/1320	0.88	0/1786	

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
2	С	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	С	5[B]	TBG	Peptide

#### 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	1275	0	1224	11	0
2	С	128	0	36	5	0
3	A	2	0	0	0	0
4	A	2	0	0	0	0
5	A	107	0	0	5	0
5	С	6	0	0	2	0
All	All	1520	0	1260	13	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 5.

All (13) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$\begin{array}{c} {\rm Interatomic} \\ {\rm distance} \ ({\rm \AA}) \end{array}$	Clash overlap (Å)
1:A:165:ASP:HB2	5:A:401:HOH:O	1.56	1.02
1:A:166:SER:N	5:A:401:HOH:O	1.82	0.93
1:A:165:ASP:CA	5:A:401:HOH:O	2.31	0.77
1:A:165:ASP:CB	5:A:401:HOH:O	2.21	0.72
1:A:98:PHE:HZ	2:C:3[B]:GLY:O	1.75	0.69
1:A:107:LYS:HD3	5:A:489:HOH:O	1.94	0.67
1:A:98:PHE:CZ	2:C:3[B]:GLY:O	2.54	0.60
1:A:204:LEU:HD23	1:A:205:GLY:N	2.17	0.60
1:A:197:ARG:HD3	1:A:204:LEU:HD22	1.83	0.59
2:C:7[B]:NH2:N	5:C:101:HOH:O	2.42	0.52
1:A:204:LEU:HD23	1:A:204:LEU:C	2.36	0.46
1:A:180:PHE:HA	2:C:4[B]:V1C:O	2.19	0.43
2:C:7[A]:NH2:N	5:C:101:HOH:O	2.52	0.42

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	159/175 (91%)	157 (99%)	2 (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	133/143 (93%)	130 (98%)	3 (2%)	50 13	

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	118	ARG
1	A	122	HIS
1	A	166	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	229	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)

8 non-standard protein/DNA/RNA residues 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	Tuno	Chain	Res	Link	Во	ond leng	ths	В	ond ang	les
MIOI	Type	Chain	nes	LIIIK	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	TBG	С	5[A]	2	6,7,8	1.85	1 (16%)	7,10,12	0.91	0
2	GGL	С	2[A]	4,2	7,8,9	0.69	0	7,9,11	1.45	1 (14%)
2	V1C	С	4[A]	2	18,21,22	0.98	1 (5%)	21,28,30	1.44	3 (14%)
2	GGL	С	2[B]	4,2	7,8,9	0.57	0	7,9,11	1.48	1 (14%)
2	TBG	С	5[B]	2	6,7,8	1.64	1 (16%)	7,10,12	0.88	0
2	V1C	С	4[B]	2	18,21,22	0.92	1 (5%)	21,28,30	1.53	2 (9%)
2	EOE	С	6[A]	2	7,8,9	0.64	0	5,9,11	0.56	0
2	EOE	С	6[B]	2	7,8,9	0.73	0	5,9,11	0.53	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	TBG	С	5[A]	2	-	0/6/8/10	-
2	GGL	С	2[A]	4,2	-	1/7/8/9	-
2	V1C	С	4[A]	2	-	2/13/14/16	0/2/2/2
2	GGL	С	2[B]	4,2	-	2/7/8/9	-
2	TBG	С	5[B]	2	-	0/6/8/10	-
2	V1C	С	4[B]	2	-	4/13/14/16	0/2/2/2
2	EOE	С	6[A]	2	-	1/3/10/11	0/1/1/1
2	EOE	С	6[B]	2	-	1/3/10/11	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	Observed(A)	$\operatorname{Ideal}( ext{\AA})$
2	С	5[A]	TBG	CB-CA	-4.39	1.51	1.56
2	С	5[B]	TBG	CB-CA	-3.89	1.52	1.56
2	С	4[A]	V1C	N6-N5	3.28	1.40	1.35
2	С	4[B]	V1C	N6-N5	3.17	1.40	1.35

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$Observed(^o)$	$\mathrm{Ideal}(^{o})$
2	С	4[B]	V1C	CB-CA-C	-4.65	102.76	111.47
2	С	4[A]	V1C	C7-C9-CZ	-3.50	123.15	127.74

Continued on next page...



Continued from previous page...

Mol	Chain	Res	Type	Atoms	$\mathbf{Z}$	$\operatorname{Observed}(^{o})$	$\operatorname{Ideal}({}^{o})$
2	С	4[A]	V1C	CE1-CZ-C9	-3.49	115.31	121.36
2	С	4[B]	V1C	CE1-CZ-C9	-3.13	115.94	121.36
2	С	2[B]	GGL	OXT-C-CA	2.68	122.53	113.38
2	С	2[A]	GGL	OXT-C-CA	2.52	121.97	113.38
2	С	4[A]	V1C	CB-CA-C	-2.06	107.60	111.47

There are no chirality outliers.

All (11) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	С	2[A]	GGL	CA-CB-CG-CD
2	С	2[B]	GGL	CA-CB-CG-CD
2	С	4[A]	V1C	C2-C3-C4-N5
2	С	4[A]	V1C	C1-C3-C4-N5
2	С	4[B]	V1C	O-C-CA-CB
2	С	4[B]	V1C	C2-C3-C4-N5
2	С	4[B]	V1C	C1-C3-C4-N5
2	С	4[B]	V1C	C8-C9-CZ-CE1
2	С	6[B]	EOE	O-C-CA-CB
2	С	6[A]	EOE	O-C-CA-CB
2	С	2[B]	GGL	OXT-C-CA-N

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	С	4[B]	V1C	1	0

### 5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

### 5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.



There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	$\langle { m RSRZ} \rangle$	# RSRZ > 2	$OWAB(A^2)$	Q < 0.9
1	A	162/175 (92%)	0.32	4 (2%) 57 56	8, 15, 24, 43	0
2	C	1/7 (14%)	1.12	0 100 100	10, 10, 10, 10	0
All	All	163/182 (89%)	0.32	4 (2%) 57 56	8, 14, 24, 43	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	243	ASN	5.1
1	A	118	ARG	2.8
1	A	244	PHE	2.4
1	A	235	THR	2.3

### 6.2 Non-standard residues in protein, DNA, RNA chains (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	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
2	EOE	С	6[A]	8/9	0.77	0.17	14,14,19,21	8
2	EOE	С	6[B]	8/9	0.77	0.17	18,19,23,25	8
2	V1C	С	4[A]	20/21	0.88	0.13	7,13,24,25	20
2	V1C	С	4[B]	20/21	0.88	0.13	15,23,36,37	20
2	TBG	С	5[A]	8/9	0.91	0.11	12,13,14,16	8
2	TBG	С	5[B]	8/9	0.91	0.11	18,18,20,21	8
2	GGL	С	2[A]	9/10	0.94	0.10	6,9,11,12	9
2	GGL	С	2[B]	9/10	0.94	0.10	9,11,13,13	9



### 6.3 Carbohydrates (i)

There are no monosaccharides 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	$\operatorname{Res}$	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q<0.9
3	CA	A	302	1/1	0.99	0.05	10,10,10,10	0
3	CA	A	301	1/1	1.00	0.06	10,10,10,10	0
4	ZN	A	303	1/1	1.00	0.04	9,9,9,9	0
4	ZN	A	304	1/1	1.00	0.05	10,10,10,10	0

#### 6.5 Other polymers (i)

There are no such residues in this entry.

