

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

#### Sep 20, 2023 – 08:53 PM EDT

PDB ID	:	5IMT
Title	:	Toxin receptor complex
Authors	:	Morton, C.J.; Lawrence, S.L.; Feil, S.C.; Parker, M.W.
Deposited on	:	2016-03-06
Resolution	:	2.70  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 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:

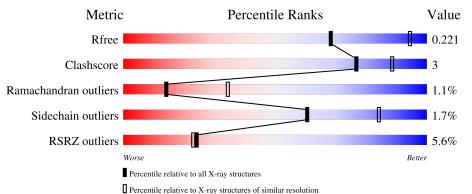
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

# 1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.70 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	$\begin{array}{c} \textbf{Whole archive} \\ \textbf{(\#Entries)} \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	А	534	81%	14%						
2	D	77	91%	9%						



# 2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 8516 atoms, of which 4159 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 Intermedilysin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
1	А	460	Total 7158	C 2261	Н 3571	N 631	O 690	${S \atop 5}$	0	2	0

Chain	Residue	Modelled	Actual	Comment	Reference
А	-1	GLY	-	expression tag	UNP Q9LCB8
А	0	GLY	-	expression tag	UNP Q9LCB8
А	1	SER	-	expression tag	UNP Q9LCB8
А	2	HIS	-	expression tag	UNP Q9LCB8
А	3	HIS	-	expression tag	UNP Q9LCB8
А	4	HIS	-	expression tag	UNP Q9LCB8
А	5	HIS	-	expression tag	UNP Q9LCB8
А	6	HIS	-	expression tag	UNP Q9LCB8
А	7	HIS	-	expression tag	UNP Q9LCB8
А	8	GLY	-	expression tag	UNP Q9LCB8
А	9	MET	-	expression tag	UNP Q9LCB8
А	10	ALA	-	expression tag	UNP Q9LCB8
А	11	SER	-	expression tag	UNP Q9LCB8
А	12	MET	-	expression tag	UNP Q9LCB8
А	13	THR	-	expression tag	UNP Q9LCB8
А	14	GLY	-	expression tag	UNP Q9LCB8
А	15	GLY	-	expression tag	UNP Q9LCB8
А	16	GLN	-	expression tag	UNP Q9LCB8
А	17	GLN	-	expression tag	UNP Q9LCB8
А	18	MET	-	expression tag	UNP Q9LCB8
А	19	GLY	-	expression tag	UNP Q9LCB8
А	20	ARG	-	expression tag	UNP Q9LCB8
А	21	ASP	-	expression tag	UNP Q9LCB8
А	22	LEU	-	expression tag	UNP Q9LCB8
А	23	TYR	-	expression tag	UNP Q9LCB8
А	24	ASP	-	expression tag	UNP Q9LCB8
А	25	ASP	-	expression tag	UNP Q9LCB8

There are 39 discrepancies between the modelled and reference sequences:

Continued on next page...



	0 1	vious page			
Chain	Residue	Modelled	Actual	Comment	Reference
А	26	ASP	-	expression tag	UNP Q9LCB8
A	27	ASP	-	expression tag	UNP Q9LCB8
А	28	LYS	-	expression tag	UNP Q9LCB8
A	29	ASP	-	expression tag	UNP Q9LCB8
А	30	ARG	-	expression tag	UNP Q9LCB8
А	31	TRP	-	expression tag	UNP Q9LCB8
А	32	GLY	-	expression tag	UNP Q9LCB8
A	33	SER	-	expression tag	UNP Q9LCB8
А	48	PRO	THR	conflict	UNP Q9LCB8
А	346	CYS	THR	conflict	UNP Q9LCB8
А	361	CYS	ILE	conflict	UNP Q9LCB8
А	445	GLU	ASP	conflict	UNP Q9LCB8

Continued from previous page...

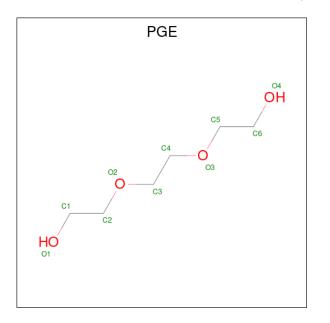
• Molecule 2 is a protein called CD59 glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
2	D	77	Total 1192	C 383	Н 574	N 105	O 120	S 10	0	0	0

There is a discrepancy between the modelled and reference sequences:

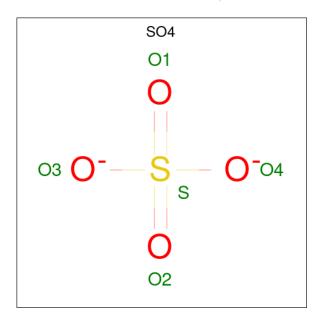
Chain	Residue	Modelled	Actual	Comment	Reference
D	22	ALA	ASP	$\operatorname{conflict}$	UNP P13987

• Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula:  $C_6H_{14}O_4$ ).





Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	А	1	Total 24	С 6	H 14	0 4	0	0



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
4	А	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0
4	D	1	$\begin{array}{ccc} \text{Total} & \text{O} & \text{S} \\ 5 & 4 & 1 \end{array}$	0	0

• Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	5	Total Zn 5 5	0	0
5	D	2	Total Zn 2 2	0	0

• Molecule 6 is COPPER (II) ION (three-letter code: CU) (formula: Cu).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	А	2	Total 2	Cu 2	0	0



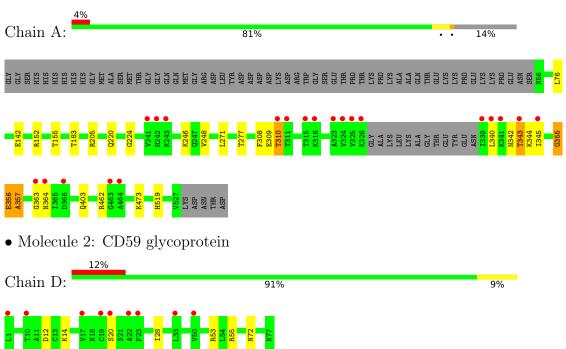
• Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	А	103	Total O 103 103	0	0
7	D	15	Total O 15 15	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: Intermedilysin



# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants	93.73Å 166.56Å 118.25Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	48.21 - 2.70	Depositor
nesolution (A)	48.21 - 2.70	EDS
% Data completeness	98.5 (48.21-2.70)	Depositor
(in resolution range)	93.3 (48.21 - 2.70)	EDS
R <sub>merge</sub>	(Not available)	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.51 (at 2.69 \text{\AA})$	Xtriage
Refinement program	PHENIX	Depositor
D D	0.208 , $0.257$	Depositor
$R, R_{free}$	0.213 , $0.221$	DCC
$R_{free}$ test set	1260 reflections $(4.98\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	30.6	Xtriage
Anisotropy	0.880	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.37, $30.8$	EDS
L-test for twinning <sup>2</sup>	$< L >=0.46, < L^2>=0.29$	Xtriage
Estimated twinning fraction	0.026 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l	Xtriage
Estimated twinning fraction	0.042 for $1/2$ *h+ $1/2$ *k, $3/2$ *h- $1/2$ *k,-l	Attrage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	8516	wwPDB-VP
Average B, all atoms $(Å^2)$	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 4.07% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<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: CU, ZN, SO4, PGE

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
	Unain	RMSZ	# Z  > 5	RMSZ	# Z  > 5
1	А	0.25	0/3666	0.47	0/4975
2	D	0.25	0/630	0.43	0/853
All	All	0.25	0/4296	0.46	0/5828

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	А	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
1	А	462	ARG	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.



OINT
------

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	А	3587	3571	3566	19	0
2	D	618	574	574	5	0
3	А	10	14	14	0	0
4	А	10	0	0	0	0
4	D	5	0	0	1	0
5	А	5	0	0	0	0
5	D	2	0	0	0	0
6	А	2	0	0	0	0
7	А	103	0	0	3	1
7	D	15	0	0	4	1
All	All	4357	4159	4154	25	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 3.

The worst 5 of 25 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)
2:D:14:LYS:NZ	7:D:202:HOH:O	2.20	0.65
1:A:473:LYS:NZ	7:A:705:HOH:O	2.38	0.56
1:A:152[B]:ARG:NH1	1:A:271:LEU:O	2.39	0.56
1:A:152[B]:ARG:NH2	7:A:707:HOH:O	2.40	0.55
1:A:309:GLU:N	1:A:343:THR:HG23	2.23	0.54

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)
7:A:734:HOH:O	7:D:202:HOH:O[5_545]	1.99	0.21

#### 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	Percentile	$\mathbf{es}$
1	А	458/534~(86%)	425~(93%)	29~(6%)	4 (1%)	17 40	
2	D	75/77~(97%)	72 (96%)	1 (1%)	2(3%)	5 12	
All	All	533/611 (87%)	497~(93%)	30 (6%)	6 (1%)	14 34	

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	343	THR
1	А	357	ALA
2	D	12	ASP
2	D	20	SER
1	А	356	GLU

#### 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	А	396/456~(87%)	389~(98%)	7~(2%)	59 83
2	D	71/71~(100%)	70~(99%)	1 (1%)	67 86
All	All	467/527~(89%)	459~(98%)	8 (2%)	60 84

5 of 8 residues with a non-rotameric side chain are listed below:

Mol	Chain	Res	Type
2	D	55	ARG
1	А	519	HIS
1	А	344	LYS
1	А	310	THR
1	А	403	GLN

Sometimes 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 monosaccharides in this entry.

### 5.6 Ligand geometry (i)

Of 13 ligands modelled in this entry, 9 are monoatomic - leaving 4 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	Type	Chain	Dec	Res Link	B	Bond lengths			Bond angles		
	Type	Ullalli	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z >2	
4	SO4	А	603	-	4,4,4	0.14	0	$6,\!6,\!6$	0.06	0	
3	PGE	А	601	-	9,9,9	0.29	0	8,8,8	0.51	0	
4	SO4	А	602	-	4,4,4	0.14	0	$6,\!6,\!6$	0.06	0	
4	SO4	D	101	-	4,4,4	0.14	0	$6,\!6,\!6$	0.05	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
3	PGE	А	601	-	-	6/7/7/7	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.



Mol	Chain	Res	Type	Atoms
3	А	601	PGE	O3-C5-C6-O4
3	А	601	PGE	O1-C1-C2-O2
3	А	601	PGE	O2-C3-C4-O3
3	А	601	PGE	C1-C2-O2-C3
3	А	601	PGE	C4-C3-O2-C2

5 of 6 torsion outliers are listed below:

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	D	101	SO4	1	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	$\langle RSRZ \rangle$	#RSRZ>2		$\mathbf{OWAB}(\mathbf{\mathring{A}}^2)$	Q<0.9
1	А	460/534~(86%)	0.24	21 (4%) 32 3	31	17, 33, 67, 78	0
2	D	77/77~(100%)	0.79	9 (11%) 4 3	3	28, 46, 63, 69	0
All	All	537/611~(87%)	0.32	30 (5%) 24 2	23	17,  35,  67,  78	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	340	LEU	4.4
1	А	364	ASN	4.1
2	D	1	LEU	3.9
1	А	324	VAL	3.8
2	D	22	ALA	3.8

#### 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 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	Res	Atoms	RSCC	RSR	$B-factors(A^2)$	Q<0.9
5	ZN	А	608	1/1	0.75	0.38	120,120,120,120	0
4	SO4	А	602	5/5	0.87	0.13	31,32,55,82	0
3	PGE	А	601	10/10	0.88	0.27	$28,\!56,\!69,\!78$	0
6	CU	А	610	1/1	0.88	0.21	113,113,113,113	0
5	ZN	D	103	1/1	0.90	0.15	$115,\!115,\!115,\!115$	0
4	SO4	А	603	5/5	0.91	0.19	$58,\!66,\!79,\!90$	0
4	SO4	D	101	5/5	0.93	0.13	42,55,65,99	0
6	CU	А	609	1/1	0.97	0.07	$52,\!52,\!52,\!52$	0
5	ZN	А	604	1/1	0.98	0.22	60,60,60,60	0
5	ZN	А	606	1/1	0.98	0.20	$50,\!50,\!50,\!50$	0
5	ZN	А	607	1/1	0.99	0.18	$53,\!53,\!53,\!53$	0
5	ZN	А	605	1/1	0.99	0.22	38,38,38,38	0
5	ZN	D	102	1/1	0.99	0.13	52,52,52,52	0

## 6.5 Other polymers (i)

There are no such residues in this entry.

