



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 8, 2024 – 01:07 AM EDT

PDB ID : 4KPG
Title : Crystal structure of MycP1 from the ESX-1 type VII secretion system
Authors : Solomonson, M.; Wasney, G.A.; Watanabe, N.; Gruninger, R.J.; Prehna, G.; Strynadka, N.C.J.
Deposited on : 2013-05-13
Resolution : 2.15 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

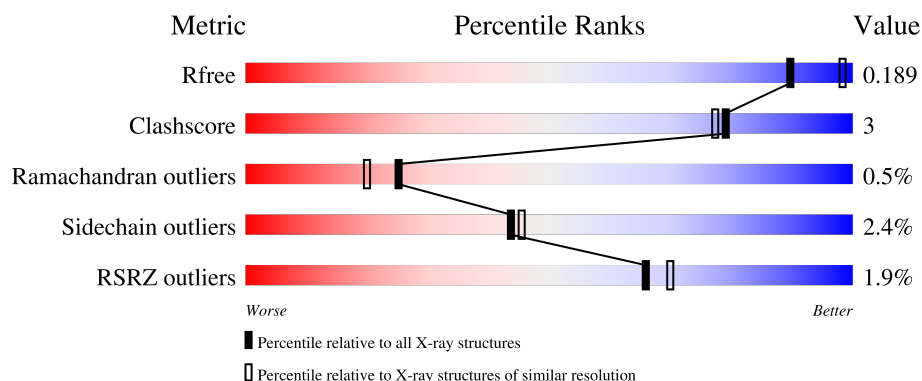
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


The reported resolution of this entry is 2.15 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}	164625	3336 (2.16-2.12)
Clashscore	180529	3585 (2.16-2.12)
Ramachandran outliers	177936	3554 (2.16-2.12)
Sidechain outliers	177891	3553 (2.16-2.12)
RSRZ outliers	164620	3337 (2.16-2.12)

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 $\leq 5\%$. 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	404	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2943 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 Membrane-anchored mycosin mycp1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	383	Total	C	N	O	S	Se	0	0	0
			2764	1720	488	547	4	5			

There are 20 discrepancies between the modelled and reference sequences:

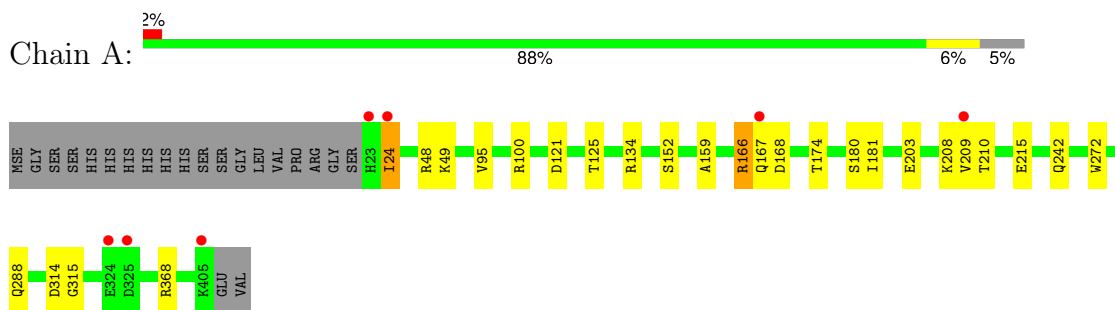
Chain	Residue	Modelled	Actual	Comment	Reference
A	4	MSE	-	expression tag	UNP A0QNL1
A	5	GLY	-	expression tag	UNP A0QNL1
A	6	SER	-	expression tag	UNP A0QNL1
A	7	SER	-	expression tag	UNP A0QNL1
A	8	HIS	-	expression tag	UNP A0QNL1
A	9	HIS	-	expression tag	UNP A0QNL1
A	10	HIS	-	expression tag	UNP A0QNL1
A	11	HIS	-	expression tag	UNP A0QNL1
A	12	HIS	-	expression tag	UNP A0QNL1
A	13	HIS	-	expression tag	UNP A0QNL1
A	14	SER	-	expression tag	UNP A0QNL1
A	15	SER	-	expression tag	UNP A0QNL1
A	16	GLY	-	expression tag	UNP A0QNL1
A	17	LEU	-	expression tag	UNP A0QNL1
A	18	VAL	-	expression tag	UNP A0QNL1
A	19	PRO	-	expression tag	UNP A0QNL1
A	20	ARG	-	expression tag	UNP A0QNL1
A	21	GLY	-	expression tag	UNP A0QNL1
A	22	SER	-	expression tag	UNP A0QNL1
A	23	HIS	-	expression tag	UNP A0QNL1

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	179	Total	O	0	0
			179	179		

i

- Molecule 1: Membrane-anchored mycosin mycp1



4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	81.00Å 94.08Å 119.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.04 – 2.15 47.04 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.04-2.15) 99.5 (47.04-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.53 (at 2.14Å)	Xtriage
Refinement program	PHENIX 1.7.2 _869	Depositor
R, R_{free}	0.166 , 0.197 0.162 , 0.189	Depositor DCC
R_{free} test set	1282 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å ²)	30.9	Xtriage
Anisotropy	0.461	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 44.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	2943	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.43% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0/2827	0.53	0/3877

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	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	A	166	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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2764	0	2666	14	0
2	A	179	0	0	4	0
All	All	2943	0	2666	14	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 3.

The worst 5 of 14 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:166:ARG:HA	1:A:167:GLN:HG2	1.51	0.91
1:A:314:ASP:OD1	2:A:667:HOH:O	1.94	0.85
1:A:166:ARG:HA	1:A:167:GLN:CG	2.24	0.67
1:A:24:ILE:HD11	1:A:209:VAL:HG22	1.75	0.67
1:A:95:VAL:HB	1:A:125:THR:HA	1.80	0.61

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	381/404 (94%)	371 (97%)	8 (2%)	2 (0%)	25	20

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	242	GLN
1	A	168	ASP

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	290/307 (94%)	283 (98%)	7 (2%)	44	45

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	208	LYS
1	A	210	THR
1	A	368	ARG
1	A	272	TRP
1	A	134	ARG

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	43	GLN

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 oligosaccharides 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 ⓘ

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, 95th 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 ²)	Q<0.9
1	A	378/404 (93%)	-0.34	7 (1%) 66 70	17, 33, 60, 108	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	209	VAL	3.3
1	A	24	ILE	3.2
1	A	405	LYS	2.8
1	A	325	ASP	2.7
1	A	167	GLN	2.5

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.