W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Sep 10, 2023 – 12:58 PM EDT

|		
PDB ID	:	4K5M
Title	:	Phosphonic Arginine Mimetics as Inhibitors of the M1 Aminopeptidases from Plasmodium falciparum
Authors	:	McGowan, S.
Deposited on	:	2013-04-14
Resolution	:	1.75 Å(reported)

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 1.75 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.
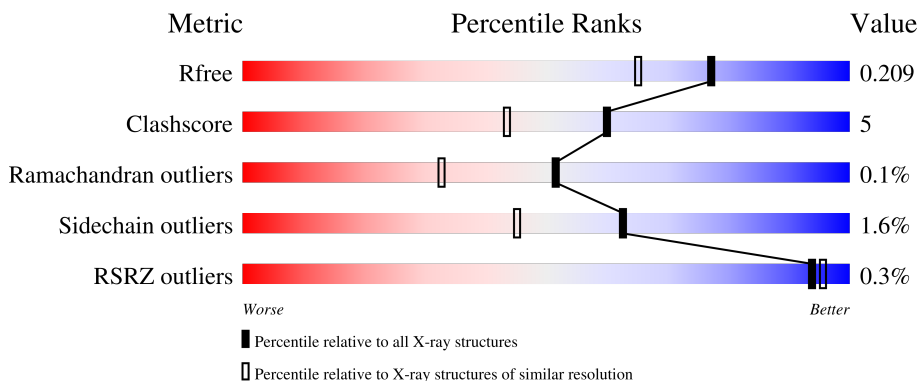
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	898	89% 10% ··

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 8458 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 M1 family aminopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	889	7272	4683	1178	1383	28	7	12	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	194	MET	-	expression tag	UNP O96935
A	213	GLN	ASN	engineered mutation	UNP O96935
A	223	GLN	ASN	engineered mutation	UNP O96935
A	378	PRO	HIS	engineered mutation	UNP O96935
A	501	GLN	ASN	engineered mutation	UNP O96935
A	745	GLN	ASN	engineered mutation	UNP O96935
A	795	GLN	ASN	engineered mutation	UNP O96935
A	1069	GLN	ASN	engineered mutation	UNP O96935
A	1086	HIS	-	expression tag	UNP O96935
A	1087	HIS	-	expression tag	UNP O96935
A	1088	HIS	-	expression tag	UNP O96935
A	1089	HIS	-	expression tag	UNP O96935
A	1090	HIS	-	expression tag	UNP O96935
A	1091	HIS	-	expression tag	UNP O96935

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
2	A	1	1	1	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

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

- Molecule 5 is [(1R)-1-amino-4-carbamimidamidobutyl]phosphonic acid (three-letter code: 1OV) (formula: $C_5H_{15}N_4O_3P$).

1OV
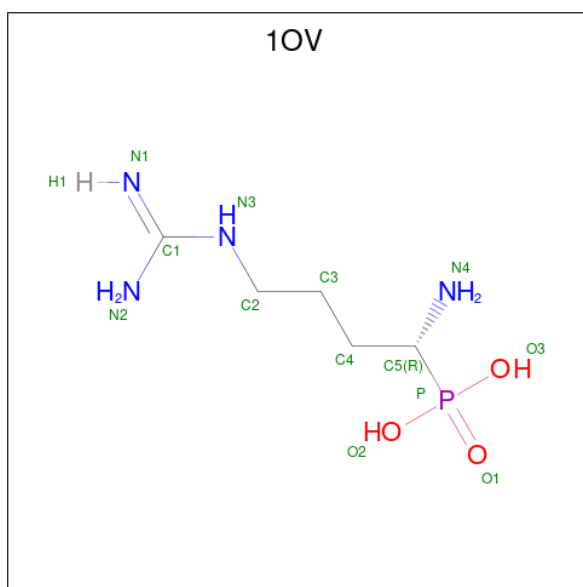
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total 18	C 7	N 7	O 3	P 1	0	1

• Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1147	Total 1147	O 1147	0	0

3 Residue-property plots ⓘ

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: M1 family aminopeptidase
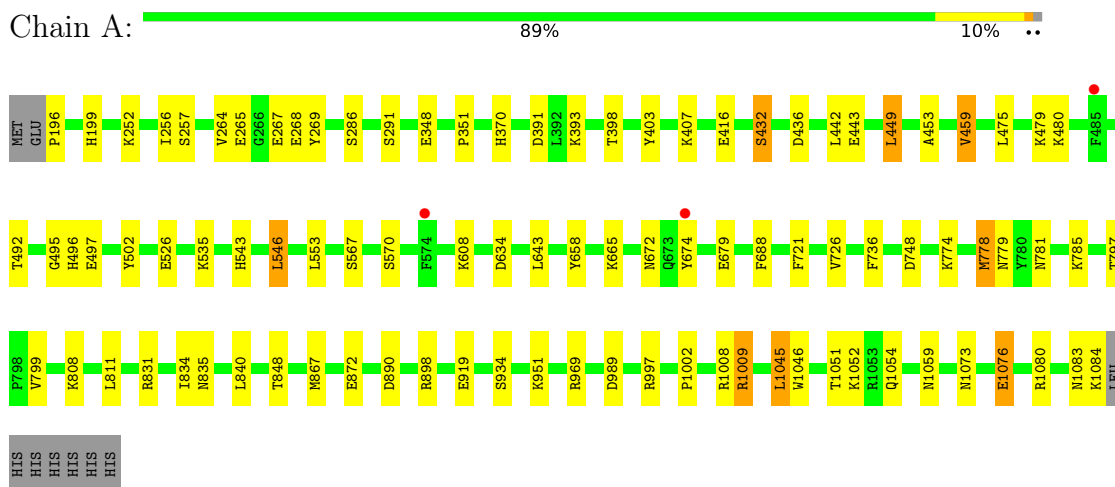
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	76.00Å 109.03Å 118.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	80.26 – 1.75	Depositor
	80.26 – 1.75	EDS
% Data completeness	98.8 (80.26-1.75)	Depositor
(in resolution range)	98.8 (80.26-1.75)	EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.98 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.163 , 0.207	Depositor
	0.163 , 0.209	DCC
R_{free} test set	4952 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	17.4	Xtriage
Anisotropy	0.464	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 48.1	EDS
L-test for twinning[2]	$< |L| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8458	wwPDB-VP
Average B, all atoms (Å^2)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.23% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]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.

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: MG, GOL, ZN, 1OV

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	1.17	6/7471 (0.1%)	0.96	14/10116 (0.1%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1076	GLU	CB-CG	7.77	1.67	1.52
1	A	934	SER	CB-OG	-5.74	1.34	1.42
1	A	658	TYR	CD2-CE2	5.55	1.47	1.39
1	A	736	PHE	CE2-CZ	5.32	1.47	1.37
1	A	799	VAL	CB-CG1	5.16	1.63	1.52

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	778[A]	MET	CG-SD-CE	7.51	112.22	100.20
1	A	778[B]	MET	CG-SD-CE	7.51	112.22	100.20
1	A	546	LEU	CB-CG-CD2	6.98	122.86	111.00
1	A	898	ARG	NE-CZ-NH2	-6.00	117.30	120.30
1	A	748	ASP	CB-CG-OD1	5.94	123.64	118.30

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)	H(added)	Clashes	Symm-Clashes
1	A	7272	0	7141	66	0
2	A	1	0	0	0	0
3	A	2	0	0	0	0
4	A	18	0	24	1	0
5	A	18	0	14	1	0
6	A	1147	0	0	40	1
All	All	8458	0	7179	66	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 5.

The worst 5 of 66 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:459:VAL:N	6:A:2337:HOH:O	1.83	1.04
1:A:665:LYS:CD	6:A:1914:HOH:O	2.05	1.02
1:A:834:ILE:HB	6:A:1667:HOH:O	1.60	1.00
1:A:781[A]:ASN:ND2	6:A:2297:HOH:O	1.95	0.98
1:A:1080:ARG:NH1	6:A:1365:HOH:O	2.01	0.93

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)
6:A:2128:HOH:O	6:A:2289:HOH:O[2_554]	2.13	0.07

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	899/898 (100%)	883 (98%)	15 (2%)	1 (0%)	51	33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	459	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	805/830 (97%)	791 (98%)	14 (2%)	60 42

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	497	GLU
1	A	546	LEU
1	A	1052	LYS
1	A	840	LEU
1	A	1045	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	212	ASN
1	A	653	HIS
1	A	882	ASN
1	A	915	ASN
1	A	1083	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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 8 ligands modelled in this entry, 3 are 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
4	GOL	A	1105	-	5,5,5	1.04	0	5,5,5	1.07	0
4	GOL	A	1107	-	5,5,5	0.70	0	5,5,5	1.89	1 (20%)
5	1OV	A	1106[A]	-	11,12,12	2.12	4 (36%)	10,16,16	2.43	5 (50%)
5	1OV	A	1106[B]	-	11,12,12	2.14	4 (36%)	10,16,16	2.42	5 (50%)
4	GOL	A	1104	-	5,5,5	0.66	0	5,5,5	1.08	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	GOL	A	1105	-	-	1/4/4/4	-
4	GOL	A	1107	-	-	3/4/4/4	-
5	1OV	A	1106[A]	-	-	4/12/13/13	-
5	1OV	A	1106[B]	-	-	5/12/13/13	-
4	GOL	A	1104	-	-	4/4/4/4	-

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1106[A]	1OV	P-O1	3.90	1.56	1.49
5	A	1106[B]	1OV	P-O1	3.90	1.56	1.49
5	A	1106[A]	1OV	P-O2	3.73	1.60	1.54
5	A	1106[B]	1OV	P-O2	3.73	1.60	1.54

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1106[A]	1OV	P-C5	-2.80	1.82	1.84

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	A	1106[A]	1OV	O2-P-O1	-4.60	101.89	113.45
5	A	1106[B]	1OV	O2-P-O1	-4.60	101.89	113.45
5	A	1106[A]	1OV	O2-P-O3	3.73	117.68	107.64
5	A	1106[B]	1OV	O2-P-O3	3.73	117.68	107.64
4	A	1107	GOL	O2-C2-C1	3.26	123.47	109.12

There are no chirality outliers.

5 of 17 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1104	GOL	O1-C1-C2-C3
4	A	1107	GOL	O1-C1-C2-C3
4	A	1107	GOL	C1-C2-C3-O3
5	A	1106[A]	1OV	N1-C1-N3-C2
5	A	1106[A]	1OV	N2-C1-N3-C2

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1107	GOL	1	0
5	A	1106[B]	1OV	1	0

5.7   Other polymers ⓘ

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 ⓘ

6.1   Protein, DNA and RNA chains ⓘ

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	889/898 (98%)	-0.27	3 (0%)	94	95	7, 17, 31, 50	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	674	TYR	2.5
1	A	485	PHE	2.4
1	A	574	PHE	2.0

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3   Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4   Ligands ⓘ

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, 95th 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(Å^2)	Q<0.9
4	GOL	A	1104	6/6	0.91	0.16	25,26,35,36	0
4	GOL	A	1107	6/6	0.91	0.22	25,32,35,37	0
5	1OV	A	1106[A]	13/13	0.93	0.15	11,16,25,27	5
5	1OV	A	1106[B]	13/13	0.93	0.15	9,15,18,21	5

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	GOL	A	1105	6/6	0.94	0.12	14,16,20,20	0
3	MG	A	1103	1/1	0.96	0.16	41,41,41,41	0
3	MG	A	1102	1/1	0.98	0.20	33,33,33,33	0
2	ZN	A	1101	1/1	1.00	0.07	12,12,12,12	0

6.5 Other polymers ⓘ

There are no such residues in this entry.