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

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 7, 2016 – 01:33 AM EDT

PDB ID	:	5K23
Title	:	Crystal structure of the complex between human phosphatase PRL-2 in the oxidized state with the Bateman domain of human magnesium transporter CNNM3
Authors	:	Gulerez, I.; Kozlov, G.; Gehring, K.
Deposited on	:	2016-05-18
Resolution	:	2.96 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

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

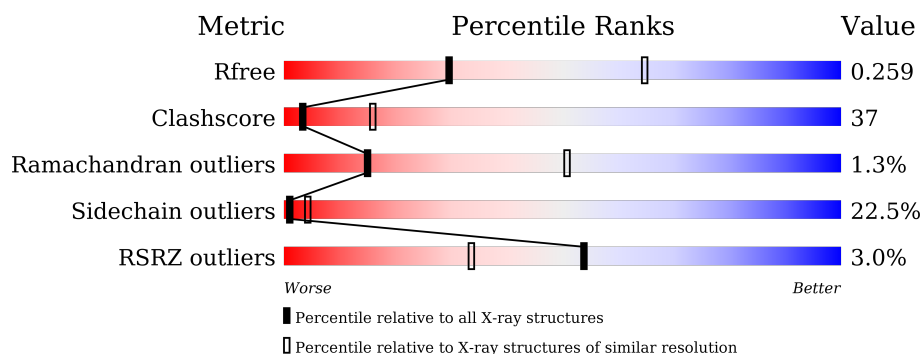|		
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 2.96 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	Percentile Ranks	Value
Rfree		0.259
Clashscore		37
Ramachandran outliers		1.3%
Sidechain outliers		22.5%
RSRZ outliers		3.0%

Worse — *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2184 (3.00-2.92)
Clashscore	102246	2552 (3.00-2.92)
Ramachandran outliers	100387	2468 (3.00-2.92)
Sidechain outliers	100360	2471 (3.00-2.92)
RSRZ outliers	91569	2201 (3.00-2.92)

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. 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	189	3% / 38% 33% 11% • 17%
2	C	155	2% / 54% 32% 8% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2401 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 Protein tyrosine phosphatase type IVA 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	156	Total 1226	C 783	N 215	O 219	S 9	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-21	MET	-	initiating methionine	UNP Q12974
A	-20	SER	-	expression tag	UNP Q12974
A	-19	TYR	-	expression tag	UNP Q12974
A	-18	TYR	-	expression tag	UNP Q12974
A	-17	HIS	-	expression tag	UNP Q12974
A	-16	HIS	-	expression tag	UNP Q12974
A	-15	HIS	-	expression tag	UNP Q12974
A	-14	HIS	-	expression tag	UNP Q12974
A	-13	HIS	-	expression tag	UNP Q12974
A	-12	HIS	-	expression tag	UNP Q12974
A	-11	LEU	-	expression tag	UNP Q12974
A	-10	GLU	-	expression tag	UNP Q12974
A	-9	SER	-	expression tag	UNP Q12974
A	-8	THR	-	expression tag	UNP Q12974
A	-7	SER	-	expression tag	UNP Q12974
A	-6	LEU	-	expression tag	UNP Q12974
A	-5	TYR	-	expression tag	UNP Q12974
A	-4	LYS	-	expression tag	UNP Q12974
A	-3	LYS	-	expression tag	UNP Q12974
A	-2	ALA	-	expression tag	UNP Q12974
A	-1	GLY	-	expression tag	UNP Q12974
A	0	PHE	-	expression tag	UNP Q12974

- Molecule 2 is a protein called Metal transporter CNNM3.

| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	149	1175	751	187	230	7	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
C	298	GLY	-	expression tag	UNP Q8NE01
C	299	PRO	-	expression tag	UNP Q8NE01
C	300	LEU	-	expression tag	UNP Q8NE01
C	301	ASN	-	expression tag	UNP Q8NE01
C	302	MET	-	expression tag	UNP Q8NE01
C	303	ILE	-	expression tag	UNP Q8NE01
C	304	GLN	-	expression tag	UNP Q8NE01
C	305	GLY	-	expression tag	UNP Q8NE01
C	306	VAL	-	expression tag	UNP Q8NE01
C	307	LEU	-	expression tag	UNP Q8NE01
C	308	GLU	-	expression tag	UNP Q8NE01

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 errors 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: Protein tyrosine phosphatase type IVA 2
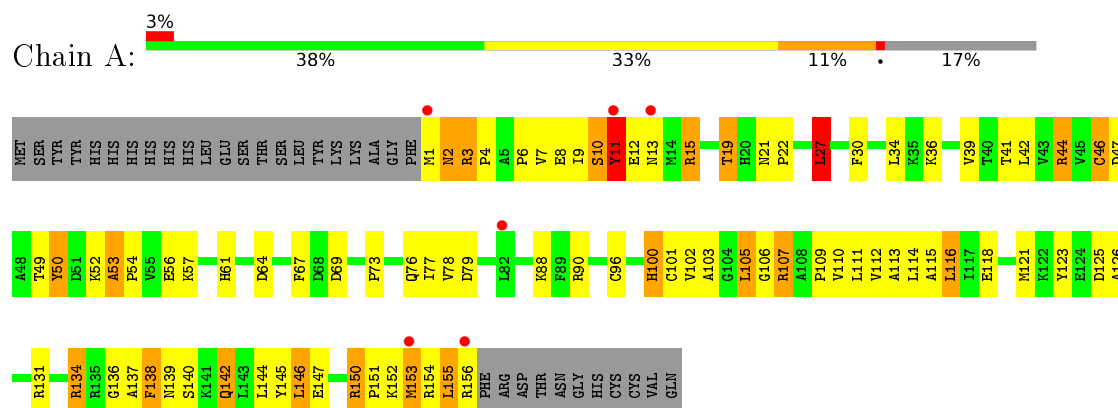
- Molecule 2: Metal transporter CNNM3
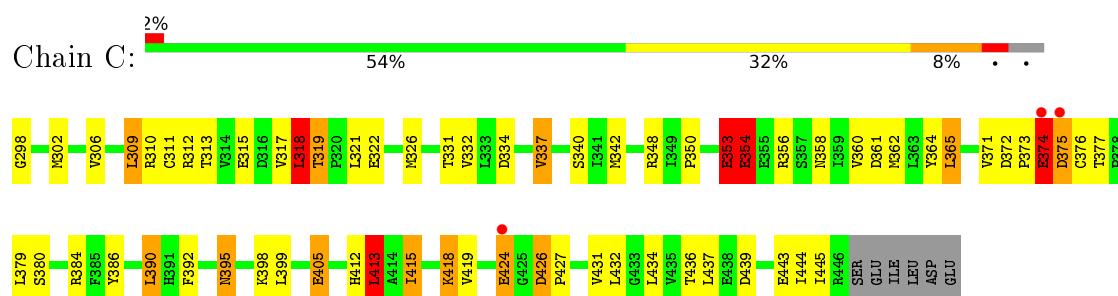
4   Data and refinement statistics (i)

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α, β, γ	51.16Å 124.49Å 159.05Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.02 – 2.96 49.02 – 2.96	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.02-2.96) 99.5 (49.02-2.96)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$<I/\sigma(I)>$ [1]	3.78 (at 2.96Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.205 , 0.264 0.220 , 0.259	Depositor DCC
R_{free} test set	528 reflections (5.06%)	DCC
Wilson B-factor (Å^2)	79.8	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 65.0	EDS
L-test for twinning[2]	$<|L|> = 0.49$, $<L^2> = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2401	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.54% 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)

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	0.75	1/1255 (0.1%)	0.93	6/1708 (0.4%)
2	C	0.73	0/1197	0.96	7/1628 (0.4%)
All	All	0.74	1/2452 (0.0%)	0.95	13/3336 (0.4%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	54	PRO	N-CD	5.05	1.54	1.47

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	46	CYS	N-CA-C	-7.21	91.53	111.00
1	A	50	TYR	N-CA-C	-6.84	92.53	111.00
2	C	413	LEU	CA-CB-CG	6.33	129.85	115.30
2	C	353	GLU	CB-CA-C	6.25	122.91	110.40
2	C	353	GLU	N-CA-C	-6.22	94.19	111.00

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	1226	0	1205	124	1
2	C	1175	0	1148	63	1
All	All	2401	0	2353	176	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 37.

The worst 5 of 176 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:90:ARG:NH2	1:A:156:ARG:NH1	1.61	1.47
1:A:155:LEU:O	1:A:156:ARG:O	1.60	1.18
1:A:7:VAL:HG21	1:A:134:ARG:NH1	1.60	1.17
1:A:121:MET:HE2	1:A:126:ALA:HA	1.24	1.16
1:A:131:ARG:HE	1:A:137:ALA:CB	1.64	1.10

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)
1:A:156:ARG:NH2	2:C:372:ASP:CG[8_455]	2.12	0.08

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	154/189 (82%)	138 (90%)	14 (9%)	2 (1%)	15	51
2	C	147/155 (95%)	141 (96%)	4 (3%)	2 (1%)	14	49
All	All	301/344 (88%)	279 (93%)	18 (6%)	4 (1%)	15	51

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	11	TYR
2	C	374	GLU
1	A	27	LEU
2	C	445	ILE

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	129/166 (78%)	101 (78%)	28 (22%)	1	5
2	C	133/142 (94%)	102 (77%)	31 (23%)	1	4
All	All	262/308 (85%)	203 (78%)	59 (22%)	1	4

5 of 59 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	153	MET
2	C	337	VAL
2	C	419	VAL
1	A	155	LEU
2	C	310	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	61	HIS
2	C	412	HIS
2	C	343	GLN
1	A	20	HIS
2	C	395	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)

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 (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, 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(Å²)	Q<0.9
1	A	156/189 (82%)	0.43	6 (3%)	44	26	43, 75, 111, 127	0
2	C	149/155 (96%)	0.36	3 (2%)	68	48	25, 62, 93, 121	0
All	All	305/344 (88%)	0.40	9 (2%)	54	33	25, 67, 107, 127	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1	MET	3.6
1	A	153	MET	3.3
2	C	375	ASP	2.8
1	A	156	ARG	2.4
2	C	424	GLU	2.3

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.