



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 9, 2017 – 10:57 am GMT

PDB ID : 5TSR  
Title : Crystal structure of PRL-3 phosphatase in complex with the Bateman domain of CNNM3 magnesium transporter  
Authors : Kozlov, G.; Zhang, H.; Gehring, K.  
Deposited on : 2016-10-31  
Resolution : 3.19 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](mailto: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
Xtriage (Phenix)	:	1.9-1692
EDS	:	recalc28906
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	recalc28906

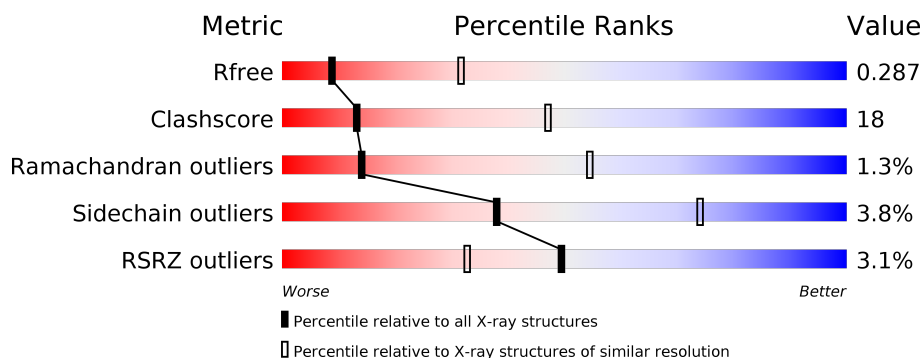
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.19 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))
$R_{free}$	100719	1123 (3.20-3.16)
Clashscore	112137	1255 (3.20-3.16)
Ramachandran outliers	110173	1233 (3.20-3.16)
Sidechain outliers	110143	1232 (3.20-3.16)
RSRZ outliers	101464	1128 (3.20-3.16)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	172	<div> <div>4%</div> <div> <div></div> <div>51%</div> <div>36%</div> <div>• • 10%</div> </div> </div>
1	C	172	<div> <div>2%</div> <div> <div></div> <div>57%</div> <div>28%</div> <div>• • 10%</div> </div> </div>
2	B	155	<div> <div>3%</div> <div> <div></div> <div>68%</div> <div>25%</div> <div>• •</div> </div> </div>
2	D	155	<div> <div>2%</div> <div> <div></div> <div>66%</div> <div>24%</div> <div>• • 5%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4767 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 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	155	Total	C	N	O	S	0	0	0
			1216	779	213	219	5			
1	C	154	Total	C	N	O	S	0	0	0
			1201	770	208	218	5			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP O75365
A	-1	SER	-	expression tag	UNP O75365
A	0	HIS	-	expression tag	UNP O75365
A	104	ALA	CYS	engineered mutation	UNP O75365
C	-2	GLY	-	expression tag	UNP O75365
C	-1	SER	-	expression tag	UNP O75365
C	0	HIS	-	expression tag	UNP O75365
C	104	ALA	CYS	engineered mutation	UNP O75365

- Molecule 2 is a protein called Metal transporter CNNM3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	149	Total	C	N	O	S	0	0	0
			1180	754	192	227	7			
2	D	147	Total	C	N	O	S	0	0	0
			1167	747	190	223	7			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	298	GLY	-	expression tag	UNP Q8NE01
B	299	PRO	-	expression tag	UNP Q8NE01
B	300	LEU	-	expression tag	UNP Q8NE01
B	301	ASN	-	expression tag	UNP Q8NE01

*Continued on next page...*

*Continued from previous page...*

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

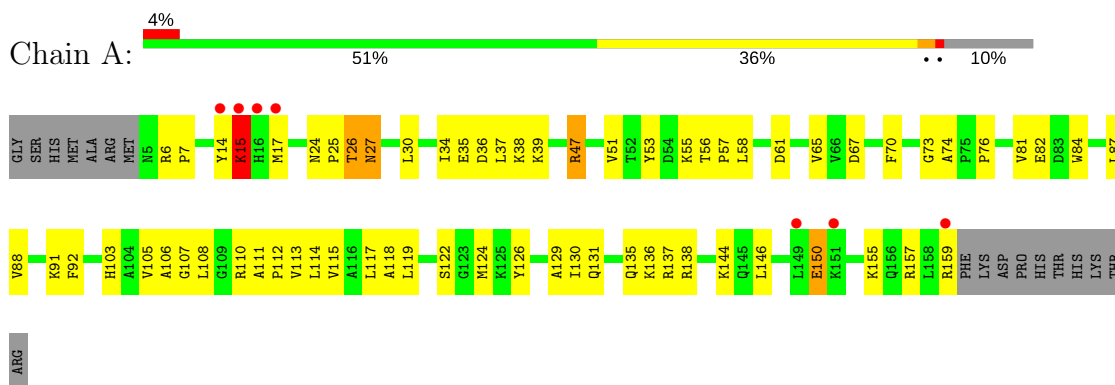
- Molecule 3 is water.

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

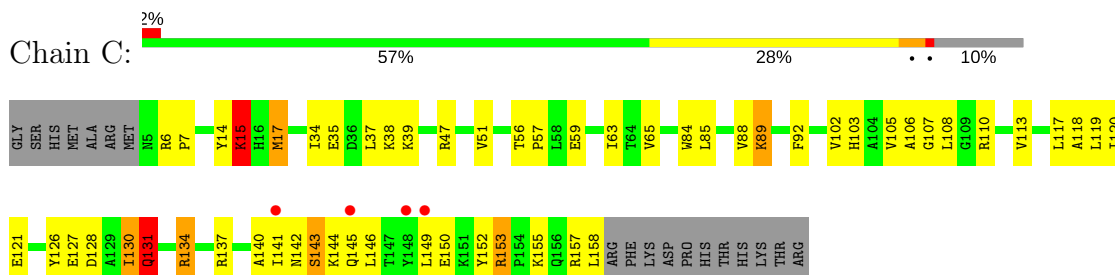
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

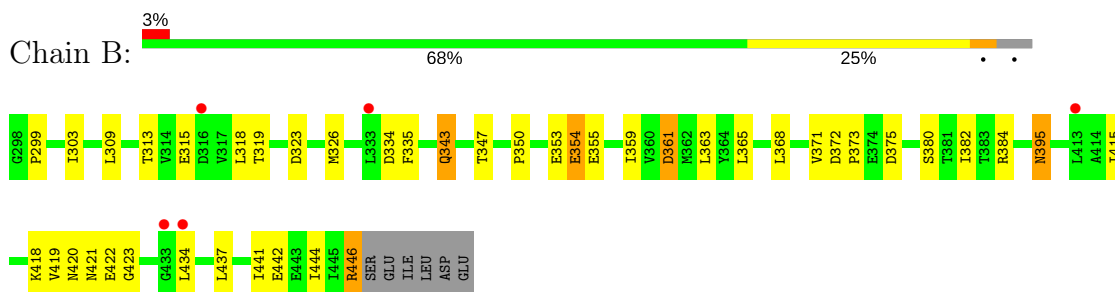
#### • Molecule 1: Protein tyrosine phosphatase type IVA 3



#### • Molecule 1: Protein tyrosine phosphatase type IVA 3

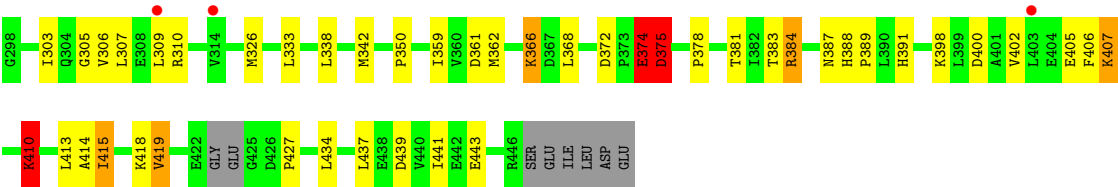


#### • Molecule 2: Metal transporter CNNM3



#### • Molecule 2: Metal transporter CNNM3





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	155.34Å 125.15Å 52.04Å 90.00° 102.21° 90.00°	Depositor
Resolution (Å)	27.74 – 3.19 48.28 – 3.19	Depositor EDS
% Data completeness (in resolution range)	95.9 (27.74-3.19) 96.0 (48.28-3.19)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.83 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.4_1496)	Depositor
R, $R_{free}$	0.239 , 0.287 0.239 , 0.287	Depositor DCC
$R_{free}$ test set	805 reflections (5.16%)	DCC
Wilson B-factor (Å <sup>2</sup> )	111.9	Xtriage
Anisotropy	0.071	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 78.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	4767	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	113.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.44% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<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.

## 5 Model quality

### 5.1 Standard geometry

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	0.46	1/1244 (0.1%)	0.81	2/1690 (0.1%)
1	C	0.43	0/1229	0.77	2/1672 (0.1%)
2	B	0.36	0/1202	0.75	1/1633 (0.1%)
2	D	0.42	0/1188	0.78	3/1613 (0.2%)
All	All	0.42	1/4863 (0.0%)	0.78	8/6608 (0.1%)

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	2
1	C	0	2
2	D	0	2
All	All	0	6

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	150	GLU	CB-CG	-5.24	1.42	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	375	ASP	CB-CG-OD1	7.14	124.73	118.30
1	C	131	GLN	CA-CB-CG	7.02	128.84	113.40
2	D	307	LEU	CA-CB-CG	6.89	131.16	115.30
1	A	26	THR	N-CA-C	-6.25	94.11	111.00
1	A	15	LYS	CA-CB-CG	6.01	126.63	113.40

There are no chirality outliers.



5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	15	LYS	Peptide
1	A	27	ASN	Peptide
1	C	130	ILE	Peptide
1	C	15	LYS	Peptide
2	D	374	GLU	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	1216	0	1240	57	0
1	C	1201	0	1216	49	0
2	B	1180	0	1166	34	0
2	D	1167	0	1156	41	0
3	A	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
All	All	4767	0	4778	174	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 18.

The worst 5 of 174 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:130:ILE:HD11	1:A:146:LEU:HD21	1.32	1.10
1:C:134:ARG:HH12	1:C:141:ILE:H	1.17	0.90
1:A:124:MET:HE2	1:A:129:ALA:HA	1.57	0.87
1:A:15:LYS:HG2	1:A:17:MET:HG2	1.58	0.85
1:C:121:GLU:OE2	1:C:152:TYR:OH	1.98	0.81

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	153/172 (89%)	142 (93%)	11 (7%)	0	100	100
1	C	152/172 (88%)	140 (92%)	9 (6%)	3 (2%)	9	43
2	B	147/155 (95%)	135 (92%)	10 (7%)	2 (1%)	13	52
2	D	143/155 (92%)	133 (93%)	7 (5%)	3 (2%)	8	41
All	All	595/654 (91%)	550 (92%)	37 (6%)	8 (1%)	14	54

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	131	GLN
1	C	143	SER
2	D	374	GLU
2	D	410	LYS
2	B	375	ASP

### 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	131/147 (89%)	128 (98%)	3 (2%)	56	84
1	C	129/147 (88%)	123 (95%)	6 (5%)	30	69
2	B	134/142 (94%)	130 (97%)	4 (3%)	46	79
2	D	133/142 (94%)	126 (95%)	7 (5%)	26	64
All	All	527/578 (91%)	507 (96%)	20 (4%)	38	74

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	134	ARG
1	C	153	ARG
2	D	406	PHE
1	C	89	LYS
1	C	102	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	135	GLN
1	C	135	GLN
2	D	391	HIS

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	155/172 (90%)	0.36	7 (4%) 34 21	79, 112, 153, 171	0
1	C	154/172 (89%)	0.22	4 (2%) 56 41	71, 111, 145, 155	0
2	B	149/155 (96%)	0.31	5 (3%) 46 30	84, 109, 134, 146	0
2	D	147/155 (94%)	0.29	3 (2%) 65 50	81, 108, 133, 156	0
All	All	605/654 (92%)	0.29	19 (3%) 49 33	71, 110, 144, 171	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	16	HIS	5.2
1	C	145	GLN	3.4
1	A	14	TYR	2.9
2	D	403	LEU	2.8
1	C	141	ILE	2.8

### 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 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.