



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 11:56 am GMT

PDB ID : 4EMO  
Title : Crystal structure of the PH domain of SHARPIN  
Authors : Stieglitz, B.; Haire, L.F.; Dikic, I.; Rittinger, K.  
Deposited on : 2012-04-12  
Resolution : 2.00 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	:	trunk28620
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)	:	recalc28949

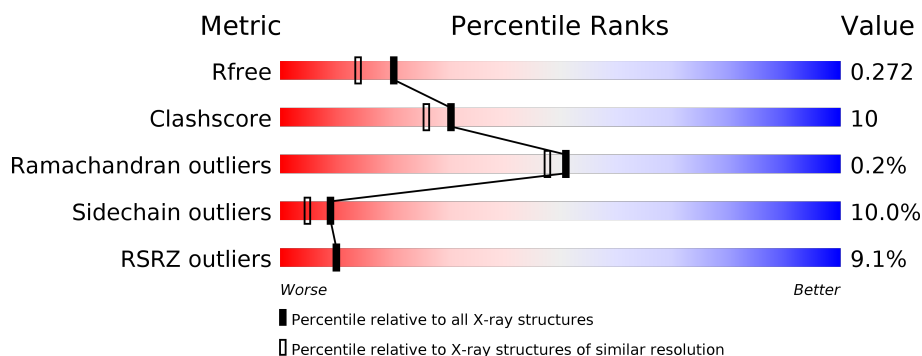
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	6609 (2.00-2.00)
Clashscore	112137	7775 (2.00-2.00)
Ramachandran outliers	110173	7679 (2.00-2.00)
Sidechain outliers	110143	7678 (2.00-2.00)
RSRZ outliers	101464	6696 (2.00-2.00)

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	129	<div> <div>5%</div> <div> <div>69%</div> <div>10%</div> <div>•</div> <div>19%</div> </div> </div>
1	B	129	<div> <div>9%</div> <div> <div>59%</div> <div>15%</div> <div>•</div> <div>22%</div> </div> </div>
1	C	129	<div> <div>7%</div> <div> <div>67%</div> <div>19%</div> <div>5%</div> <div>9%</div> </div> </div>
1	D	129	<div> <div>9%</div> <div> <div>56%</div> <div>20%</div> <div>•</div> <div>•</div> <div>19%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3272 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 Sharpin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	104	Total	C	N	O	Se	0	0	0
			782	494	147	139	2			
1	B	100	Total	C	N	O	Se	0	0	0
			764	482	144	136	2			
1	C	117	Total	C	N	O	Se	0	0	0
			855	538	160	154	3			
1	D	105	Total	C	N	O	Se	0	0	0
			785	496	148	139	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP Q9H0F6
A	0	SER	-	EXPRESSION TAG	UNP Q9H0F6
A	22	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
A	101	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
B	-1	GLY	-	EXPRESSION TAG	UNP Q9H0F6
B	0	SER	-	EXPRESSION TAG	UNP Q9H0F6
B	22	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
B	101	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
C	-1	GLY	-	EXPRESSION TAG	UNP Q9H0F6
C	0	SER	-	EXPRESSION TAG	UNP Q9H0F6
C	22	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
C	101	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
D	-1	GLY	-	EXPRESSION TAG	UNP Q9H0F6
D	0	SER	-	EXPRESSION TAG	UNP Q9H0F6
D	22	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6
D	101	MSE	LEU	ENGINEERED MUTATION	UNP Q9H0F6

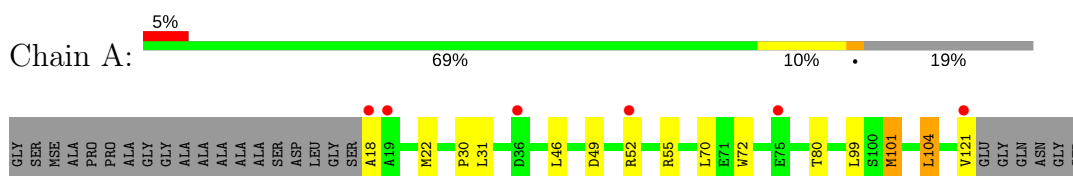
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	27	Total 27	O 27	0	0
2	B	18	Total 18	O 18	0	0
2	C	30	Total 30	O 30	0	0
2	D	11	Total 11	O 11	0	0

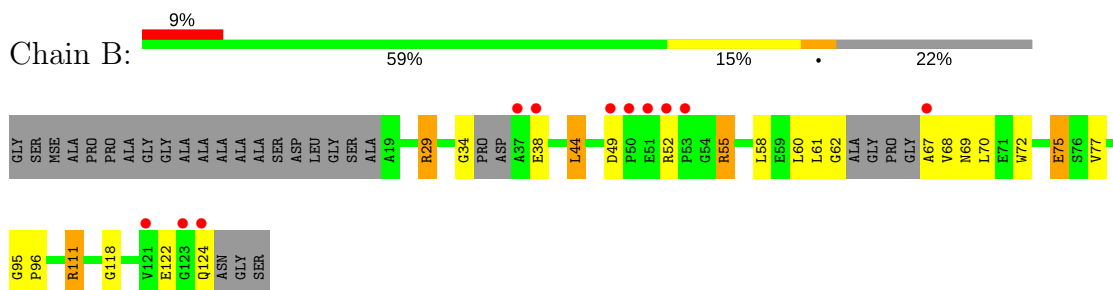
### 3 Residue-property plots [i](#)

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.

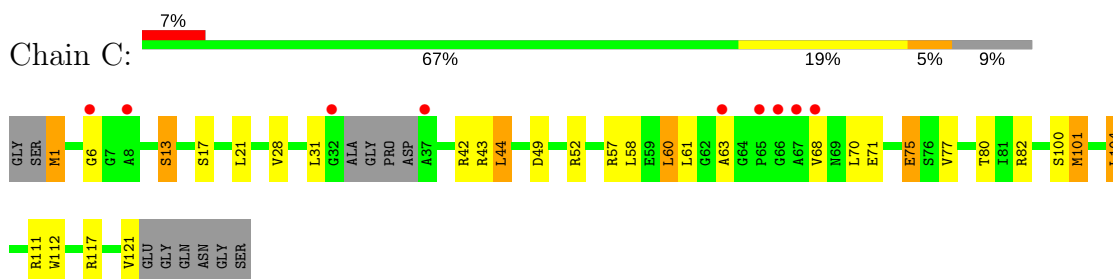
#### • Molecule 1: Sharpin



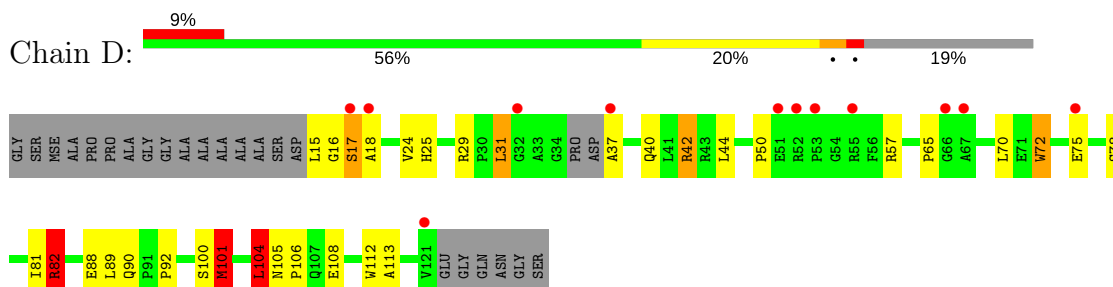
#### • Molecule 1: Sharpin



#### • Molecule 1: Sharpin



#### • Molecule 1: Sharpin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.55Å 61.55Å 222.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.00 28.43 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.8 (30.00-2.00) 99.0 (28.43-2.00)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.09 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.209 , 0.269 0.210 , 0.272	Depositor DCC
$R_{free}$ test set	1507 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	30.4	Xtriage
Anisotropy	0.115	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 54.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3272	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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	0.90	1/801 (0.1%)	1.04	2/1092 (0.2%)
1	B	1.02	2/779 (0.3%)	1.31	5/1056 (0.5%)
1	C	0.91	0/874	1.09	8/1190 (0.7%)
1	D	0.90	1/802 (0.1%)	1.20	5/1090 (0.5%)
All	All	0.93	4/3256 (0.1%)	1.16	20/4428 (0.5%)

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

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	72	TRP	CD2-CE2	5.81	1.48	1.41
1	A	72	TRP	CD2-CE2	5.53	1.48	1.41
1	D	72	TRP	CD2-CE2	5.30	1.47	1.41
1	B	111	ARG	CD-NE	-5.15	1.37	1.46

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	111	ARG	NE-CZ-NH2	-20.66	109.97	120.30
1	B	111	ARG	NE-CZ-NH1	16.26	128.43	120.30
1	D	82	ARG	NE-CZ-NH1	15.15	127.87	120.30
1	D	82	ARG	NE-CZ-NH2	-12.27	114.17	120.30
1	C	101	MSE	CA-CB-CG	-11.16	94.33	113.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	18	ALA	Peptide

## 5.2 Too-close contacts

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	782	0	784	6	0
1	B	764	0	765	18	0
1	C	855	0	859	16	0
1	D	785	0	791	30	0
2	A	27	0	0	1	0
2	B	18	0	0	1	0
2	C	30	0	0	2	0
2	D	11	0	0	1	0
All	All	3272	0	3199	67	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 10.

The worst 5 of 67 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:D:104:LEU:H	1:D:104:LEU:HD13	1.44	0.80
1:D:18:ALA:HB1	1:D:57:ARG:HH21	1.50	0.77
1:D:18:ALA:HB2	1:D:50:PRO:HD3	1.69	0.74
1:B:44:LEU:HD22	1:B:60:LEU:CD1	2.18	0.74
1:D:31:LEU:HD21	1:D:100:SER:HB2	1.70	0.74

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	102/129 (79%)	102 (100%)	0	0	100	100
1	B	94/129 (73%)	93 (99%)	1 (1%)	0	100	100
1	C	113/129 (88%)	112 (99%)	1 (1%)	0	100	100
1	D	101/129 (78%)	96 (95%)	4 (4%)	1 (1%)	18	10
All	All	410/516 (80%)	403 (98%)	6 (2%)	1 (0%)	51	48

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	65	PRO

### 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	79/88 (90%)	74 (94%)	5 (6%)	21	15
1	B	78/88 (89%)	69 (88%)	9 (12%)	6	3
1	C	84/88 (96%)	73 (87%)	11 (13%)	5	2
1	D	79/88 (90%)	72 (91%)	7 (9%)	11	7
All	All	320/352 (91%)	288 (90%)	32 (10%)	9	5

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1	MSE
1	C	31	LEU
1	D	82	ARG
1	C	17	SER
1	C	44	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	90	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 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	102/129 (79%)	0.15	6 (5%) 23 23	19, 32, 54, 65	0
1	B	98/129 (75%)	0.40	11 (11%) 6 6	16, 29, 60, 84	0
1	C	114/129 (88%)	0.21	9 (7%) 13 13	18, 31, 53, 71	0
1	D	103/129 (79%)	0.51	12 (11%) 5 5	21, 35, 63, 78	0
All	All	417/516 (80%)	0.31	38 (9%) 10 10	16, 32, 59, 84	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	37	ALA	6.2
1	C	66	GLY	5.7
1	B	51	GLU	5.6
1	B	124	GLN	4.5
1	C	68	VAL	4.4

### 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.