



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 27, 2016 – 10:39 AM EDT

PDB ID : 5AWT  
Title : Crystal structure of the SGIP1 mu homology domain in complex with an Eps15 fragment containing two DPF motifs (YDPFGGDPFKG)  
Authors : Shimada, A.; Yamaguchi, A.; Kohda, D.  
Deposited on : 2015-07-08  
Resolution : 2.70 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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  
Mogul : unknown  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027790  
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-20027790

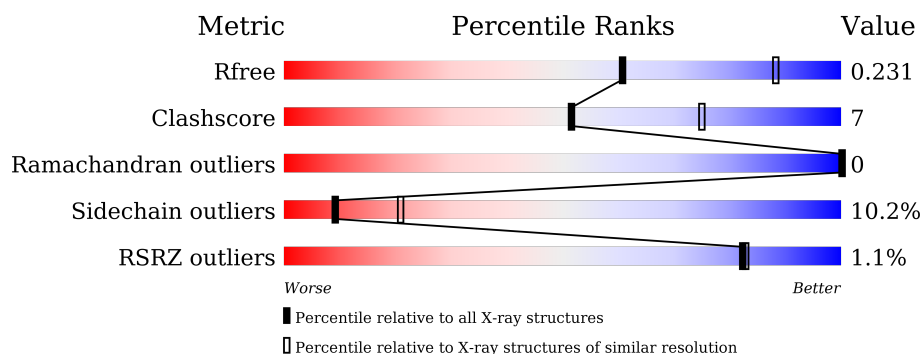
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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))
$R_{free}$	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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	282	
2	B	11	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2179 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 SH3-containing GRB2-like protein 3-interacting protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	266	Total	C	N	O	S	0	0	0
			2057	1314	349	384	10			

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	547	GLY	-	expression tag	UNP Q9BQI5
A	548	PRO	-	expression tag	UNP Q9BQI5
A	549	LEU	-	expression tag	UNP Q9BQI5
A	550	GLY	-	expression tag	UNP Q9BQI5
A	551	SER	-	expression tag	UNP Q9BQI5

- Molecule 2 is a protein called Epidermal growth factor receptor substrate 15.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	10	Total	C	N	O	0	0	0
			81	55	11	15			

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	6	Total	Zn	0	0
			6	6		

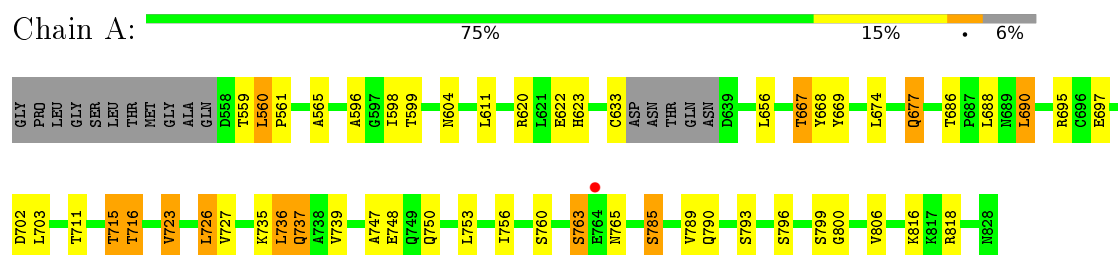
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	35	Total	O	0	0
			35	35		

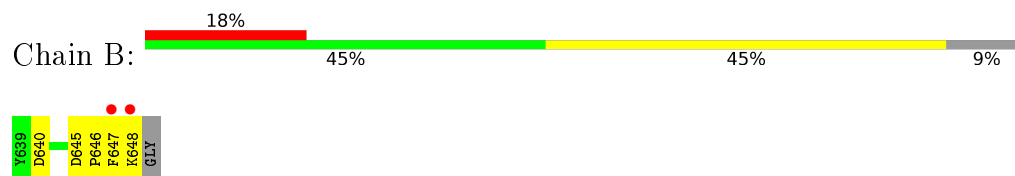
### 3 Residue-property plots [i](#)

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: SH3-containing GRB2-like protein 3-interacting protein 1



- Molecule 2: Epidermal growth factor receptor substrate 15



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.72Å 107.72Å 79.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.27 – 2.70 48.17 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.8 (41.27-2.70) 99.9 (48.17-2.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.72 (at 2.69Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.188 , 0.234 0.186 , 0.231	Depositor DCC
$R_{free}$ test set	660 reflections (4.93%)	DCC
Wilson B-factor (Å <sup>2</sup> )	66.2	Xtriage
Anisotropy	0.499	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 38.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2179	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.36% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN

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	0.57	0/2104	0.71	1/2864 (0.0%)
2	B	0.41	0/85	0.50	0/114
All	All	0.56	0/2189	0.70	1/2978 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	726	LEU	CA-CB-CG	6.59	130.46	115.30

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	2057	0	2062	27	0
2	B	81	0	67	4	0
3	A	6	0	0	0	0
4	A	35	0	0	3	1
All	All	2179	0	2129	29	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 7.

All (29) 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:816:LYS:HE2	1:A:818:ARG:HH21	1.36	0.89
1:A:560:LEU:HD11	1:A:598:ILE:HG22	1.61	0.81
1:A:747:ALA:O	1:A:750:GLN:NE2	2.22	0.72
1:A:667:THR:OG1	2:B:645:ASP:OD1	2.10	0.68
1:A:748:GLU:OE2	4:A:1001:HOH:O	2.13	0.66
1:A:785:SER:O	4:A:1002:HOH:O	2.14	0.65
1:A:790:GLN:HG3	4:A:1011:HOH:O	2.03	0.59
1:A:760:SER:H	1:A:763:SER:HB2	1.69	0.58
1:A:716:THR:HG23	1:A:796:SER:HB2	1.86	0.58
1:A:695:ARG:NH1	1:A:702:ASP:OD2	2.37	0.57
1:A:598:ILE:HD13	1:A:656:LEU:HD13	1.88	0.55
1:A:620:ARG:HD2	1:A:799:SER:O	2.06	0.55
1:A:623:HIS:HB2	1:A:677:GLN:HB2	1.91	0.53
1:A:735:LYS:HE3	1:A:737:GLN:HG2	1.91	0.51
1:A:736:LEU:HD12	1:A:739:VAL:HG21	1.94	0.49
1:A:668:TYR:OH	2:B:640:ASP:OD1	2.14	0.49
1:A:598:ILE:HD12	1:A:599:THR:N	2.27	0.49
1:A:816:LYS:HE2	1:A:818:ARG:NH2	2.18	0.47
2:B:646:PRO:HG2	2:B:647:PHE:CD2	2.50	0.47
1:A:596:ALA:HA	1:A:669:TYR:HE2	1.80	0.45
1:A:560:LEU:HD23	1:A:561:PRO:HD2	1.99	0.45
1:A:622:GLU:HB2	1:A:623:HIS:ND1	2.33	0.44
1:A:793:SER:O	1:A:818:ARG:HB2	2.18	0.43
1:A:565:ALA:HA	1:A:816:LYS:O	2.18	0.43
2:B:645:ASP:O	2:B:648:LYS:HE2	2.19	0.42
1:A:715:THR:HG21	1:A:800:GLY:HA2	2.02	0.42
1:A:723:VAL:HG22	1:A:756:ILE:HB	2.02	0.42
1:A:726:LEU:HD23	1:A:753:LEU:HD13	2.02	0.41
1:A:690:LEU:HD21	1:A:789:VAL:HG21	2.02	0.41

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)
4:A:1014:HOH:O	4:A:1029:HOH:O[3_545]	2.17	0.03

## 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	262/282 (93%)	252 (96%)	10 (4%)	0	100	100
2	B	8/11 (73%)	8 (100%)	0	0	100	100
All	All	270/293 (92%)	260 (96%)	10 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	227/239 (95%)	203 (89%)	24 (11%)	8	19
2	B	8/8 (100%)	8 (100%)	0	100	100
All	All	235/247 (95%)	211 (90%)	24 (10%)	9	21

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	559	THR
1	A	560	LEU
1	A	604	ASN
1	A	611	LEU
1	A	633	CYS
1	A	667	THR
1	A	674	LEU
1	A	677	GLN

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	686	THR
1	A	688	LEU
1	A	690	LEU
1	A	697	GLU
1	A	703	LEU
1	A	711	THR
1	A	715	THR
1	A	716	THR
1	A	723	VAL
1	A	727	VAL
1	A	736	LEU
1	A	737	GLN
1	A	763	SER
1	A	765	ASN
1	A	785	SER
1	A	806	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	266/282 (94%)	0.11	1 (0%) 93 94	33, 60, 103, 125	0
2	B	10/11 (90%)	1.12	2 (20%) 1 1	70, 95, 112, 131	0
All	All	276/293 (94%)	0.15	3 (1%) 82 83	33, 61, 104, 131	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	648	LYS	3.3
2	B	647	PHE	3.1
1	A	764	GLU	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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZN	A	905	1/1	0.71	0.12	-	154,154,154,154	0
3	ZN	A	904	1/1	0.92	0.14	-	170,170,170,170	0
3	ZN	A	903	1/1	0.92	0.27	-	97,97,97,97	0
3	ZN	A	902	1/1	0.96	0.17	-	85,85,85,85	0
3	ZN	A	906	1/1	0.78	0.08	-	129,129,129,129	0
3	ZN	A	901	1/1	0.96	0.15	-	95,95,95,95	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.