



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 02:48 pm BST

PDB ID : 6PXB  
Title : N-Terminal SH2 domain of the p120RasGAP  
Authors : Jaber Chehayeb, R.; Stiegler, A.L.; Boggon, T.J.  
Deposited on : 2019-07-25  
Resolution : 1.75 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

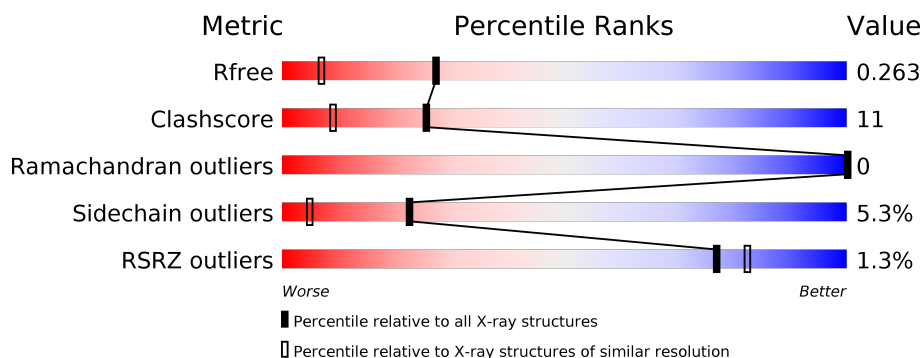
# 1 Overall quality at a glance

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	3764 (1.76-1.72)
Clashscore	141614	3923 (1.76-1.72)
Ramachandran outliers	138981	3878 (1.76-1.72)
Sidechain outliers	138945	3878 (1.76-1.72)
RSRZ outliers	127900	3705 (1.76-1.72)

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 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  $\leq 5\%$ . 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	109	<div> <div></div> <div>79%14%6%</div> </div>
1	B	109	<div> <div></div> <div>78%12%7%</div> </div>
1	C	109	<div> <div></div> <div>78%13%6%</div> </div>
1	D	109	<div> <div></div> <div>3%83%6%7%</div> </div>
1	E	109	<div> <div></div> <div>66%25%6%</div> </div>
1	F	109	<div> <div></div> <div>2%70%22%7%</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5192 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 Ras GTPase-activating protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	F	101	Total	C	N	O	S	0	0	0
			812	515	144	151	2			
1	C	103	Total	C	N	O	S	0	0	0
			829	527	147	153	2			
1	A	103	Total	C	N	O	S	0	0	0
			829	527	147	153	2			
1	D	101	Total	C	N	O	S	0	0	0
			814	519	144	149	2			
1	E	102	Total	C	N	O	S	0	0	0
			822	523	146	151	2			
1	B	101	Total	C	N	O	S	0	0	0
			814	519	144	149	2			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	172	GLY	-	expression tag	UNP P20936
F	173	SER	-	expression tag	UNP P20936
F	236	SER	CYS	engineered mutation	UNP P20936
F	261	SER	CYS	engineered mutation	UNP P20936
C	172	GLY	-	expression tag	UNP P20936
C	173	SER	-	expression tag	UNP P20936
C	236	SER	CYS	engineered mutation	UNP P20936
C	261	SER	CYS	engineered mutation	UNP P20936
A	172	GLY	-	expression tag	UNP P20936
A	173	SER	-	expression tag	UNP P20936
A	236	SER	CYS	engineered mutation	UNP P20936
A	261	SER	CYS	engineered mutation	UNP P20936
D	172	GLY	-	expression tag	UNP P20936
D	173	SER	-	expression tag	UNP P20936
D	236	SER	CYS	engineered mutation	UNP P20936
D	261	SER	CYS	engineered mutation	UNP P20936
E	172	GLY	-	expression tag	UNP P20936

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	173	SER	-	expression tag	UNP P20936
E	236	SER	CYS	engineered mutation	UNP P20936
E	261	SER	CYS	engineered mutation	UNP P20936
B	172	GLY	-	expression tag	UNP P20936
B	173	SER	-	expression tag	UNP P20936
B	236	SER	CYS	engineered mutation	UNP P20936
B	261	SER	CYS	engineered mutation	UNP P20936

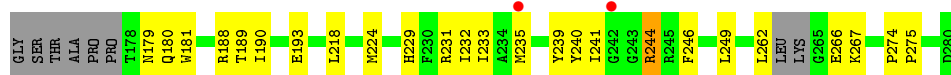
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	F	43	Total O 43 43	0	0
2	C	50	Total O 50 50	0	0
2	A	51	Total O 51 51	0	0
2	D	39	Total O 39 39	0	0
2	E	39	Total O 39 39	0	0
2	B	50	Total O 50 50	0	0

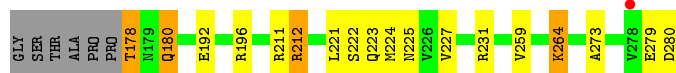
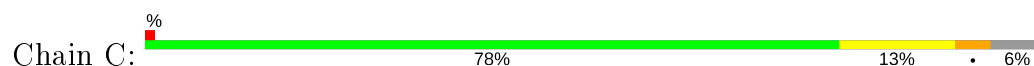
### 3 Residue-property plots [i](#)

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.

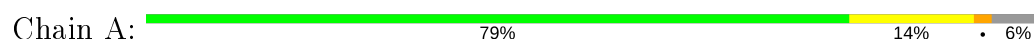
- Molecule 1: Ras GTPase-activating protein 1



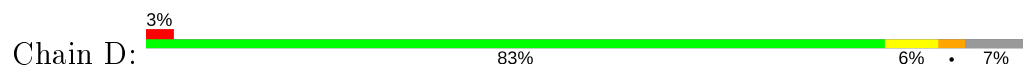
- Molecule 1: Ras GTPase-activating protein 1



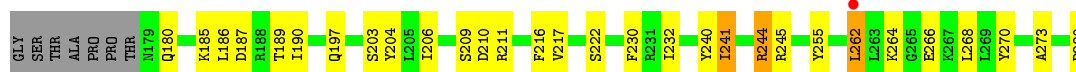
- Molecule 1: Ras GTPase-activating protein 1



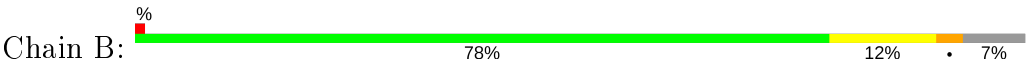
- Molecule 1: Ras GTPase-activating protein 1



- Molecule 1: Ras GTPase-activating protein 1



- Molecule 1: Ras GTPase-activating protein 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.02Å 64.02Å 119.45Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	39.82 – 1.75 39.82 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.3 (39.82-1.75) 99.3 (39.82-1.75)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.78 (at 1.75Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, $R_{free}$	0.220 , 0.263 0.220 , 0.263	Depositor DCC
$R_{free}$ test set	2810 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.9	Xtriage
Anisotropy	0.567	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 34.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.044 for -h,-k,l 0.043 for h,-h-k,-l 0.480 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	5192	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 70.04 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.3391e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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](#)

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	0.87	1/849 (0.1%)	0.77	0/1145
1	B	0.74	1/834 (0.1%)	0.69	0/1124
1	C	0.65	0/849	0.76	0/1145
1	D	0.64	2/834 (0.2%)	0.72	0/1124
1	E	0.69	0/842	0.72	0/1135
1	F	0.68	0/831	0.75	0/1120
All	All	0.72	4/5039 (0.1%)	0.73	0/6793

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	266	GLU	CD-OE2	-6.12	1.19	1.25
1	A	208	GLU	CD-OE2	-5.17	1.20	1.25
1	D	266	GLU	CD-OE1	-5.13	1.20	1.25
1	B	266	GLU	CD-OE2	-5.11	1.20	1.25

There are no bond angle outliers.

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	829	0	816	16	0
1	B	814	0	803	11	0
1	C	829	0	816	19	1

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	814	0	803	17	0
1	E	822	0	809	28	0
1	F	812	0	791	27	1
2	A	51	0	0	4	0
2	B	50	0	0	3	0
2	C	50	0	0	3	0
2	D	39	0	0	3	0
2	E	39	0	0	6	0
2	F	43	0	0	2	0
All	All	5192	0	4838	103	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 11.

The worst 5 of 103 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:E:222:SER:OG	1:E:266:GLU:OE2	1.65	1.14
1:F:188:ARG:NH2	1:D:280:ASP:O	1.80	1.14
1:C:224:MET:CE	1:D:280:ASP:HB3	1.81	1.10
1:F:188:ARG:HH22	1:D:280:ASP:C	1.68	0.96
1:E:280:ASP:OD2	2:E:301:HOH:O	1.84	0.95

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:F:244:ARG:NH2	1:C:178:THR:O[2_754]	2.18	0.02

## 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	101/109 (93%)	99 (98%)	2 (2%)	0	100	100
1	B	99/109 (91%)	98 (99%)	1 (1%)	0	100	100
1	C	101/109 (93%)	100 (99%)	1 (1%)	0	100	100
1	D	99/109 (91%)	99 (100%)	0	0	100	100
1	E	100/109 (92%)	98 (98%)	2 (2%)	0	100	100
1	F	97/109 (89%)	95 (98%)	2 (2%)	0	100	100
All	All	597/654 (91%)	589 (99%)	8 (1%)	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	90/94 (96%)	86 (96%)	4 (4%)	28	8
1	B	88/94 (94%)	81 (92%)	7 (8%)	12	1
1	C	90/94 (96%)	83 (92%)	7 (8%)	12	2
1	D	88/94 (94%)	86 (98%)	2 (2%)	50	27
1	E	89/94 (95%)	83 (93%)	6 (7%)	16	3
1	F	88/94 (94%)	86 (98%)	2 (2%)	50	27
All	All	533/564 (94%)	505 (95%)	28 (5%)	22	5

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	267	LYS
1	E	206	ILE
1	B	263	LEU
1	D	263	LEU
1	D	267	LYS

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

### 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	103/109 (94%)	-0.40	0	100   100	29, 42, 81, 90	0
1	B	101/109 (92%)	-0.39	1 (0%)	82   87	27, 42, 80, 100	0
1	C	103/109 (94%)	-0.36	1 (0%)	82   87	30, 46, 82, 105	0
1	D	101/109 (92%)	-0.29	3 (2%)	50   56	30, 49, 79, 101	0
1	E	102/109 (93%)	-0.26	1 (0%)	82   87	33, 55, 88, 97	0
1	F	101/109 (92%)	-0.25	2 (1%)	65   71	29, 48, 74, 88	0
All	All	611/654 (93%)	-0.33	8 (1%)	77   82	27, 47, 84, 105	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	278	VAL	3.2
1	B	264	LYS	3.0
1	F	242	GLY	2.4
1	E	262	LEU	2.3
1	D	263	LEU	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.