



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 06:07 am BST

PDB ID : 5XJQ  
Title : Crystal Structure of the Gemin2-binding domain of SMN, Gemin2 in Complex with SmD1(1-82)/D2/F/E/G from Human  
Authors : Yi, H.; Zhang, R.  
Deposited on : 2017-05-04  
Resolution : 3.28 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

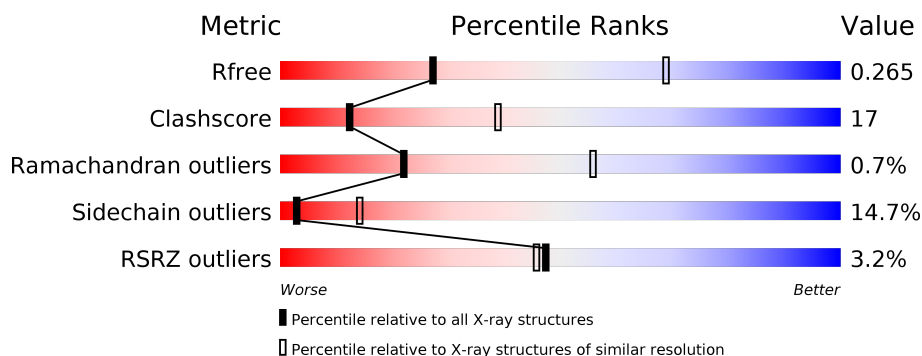
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.28 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	1177 (3.32-3.24)
Clashscore	141614	1044 (3.30-3.26)
Ramachandran outliers	138981	1026 (3.30-3.26)
Sidechain outliers	138945	1025 (3.30-3.26)
RSRZ outliers	127900	1141 (3.32-3.24)

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	2	280	<div> <div>50%</div> <div>18%</div> <div>•</div> <div>28%</div> </div>
2	A	82	<div> <div>73%</div> <div>17%</div> <div>7%</div> <div>•</div> </div>
3	B	118	<div> <div>50%</div> <div>23%</div> <div>8%</div> <div>19%</div> </div>
4	E	92	<div> <div>47%</div> <div>30%</div> <div>• •</div> <div>18%</div> </div>
5	F	86	<div> <div>69%</div> <div>14%</div> <div>•</div> <div>14%</div> </div>
6	G	76	<div> <div>25%</div> <div>43%</div> <div>28%</div> <div>5%</div> <div>24%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	M	37	<div><div></div><div></div><div></div><div></div></div> <div>30%14%•54%</div>

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 4793 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 Gem-associated protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	2	201	Total	C	N	O	S	0	0	0
			1614	1022	286	298	8			

- Molecule 2 is a protein called Small nuclear ribonucleoprotein Sm D1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	80	Total	C	N	O	S	0	0	0
			633	404	111	115	3			

- Molecule 3 is a protein called Small nuclear ribonucleoprotein Sm D2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	95	Total	C	N	O	S	0	0	0
			770	483	141	141	5			

- Molecule 4 is a protein called Small nuclear ribonucleoprotein E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	75	Total	C	N	O	S	0	0	0
			623	395	111	113	4			

- Molecule 5 is a protein called Small nuclear ribonucleoprotein F.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	74	Total	C	N	O	S	0	0	0
			576	372	95	104	5			

- Molecule 6 is a protein called Small nuclear ribonucleoprotein G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	58	Total	C	N	O	S	0	0	0
			447	285	80	76	6			

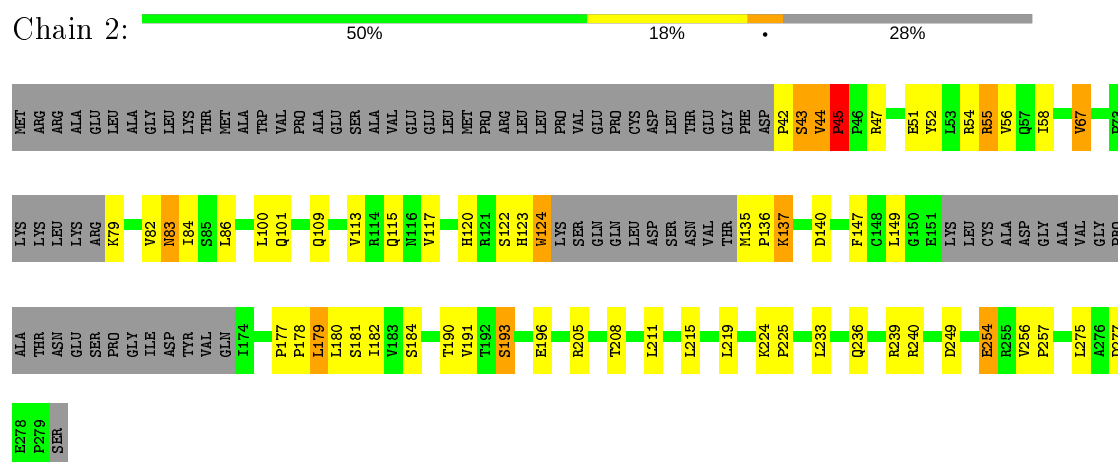
- Molecule 7 is a protein called Survival motor neuron protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
7	M	17	Total	C	N	O	0	0	0
			130	84	20	26			

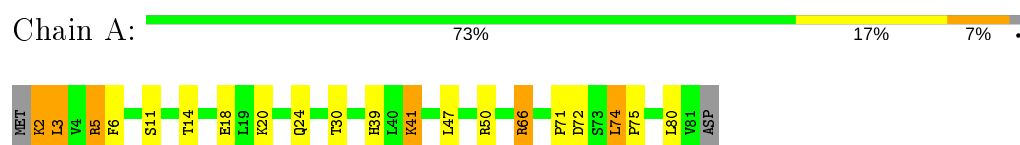
### 3 Residue-property plots [i](#)

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.

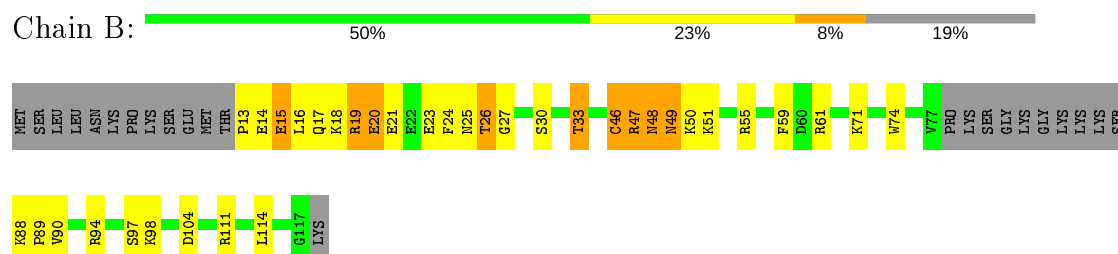
- Molecule 1: Gem-associated protein 2



- Molecule 2: Small nuclear ribonucleoprotein Sm D1

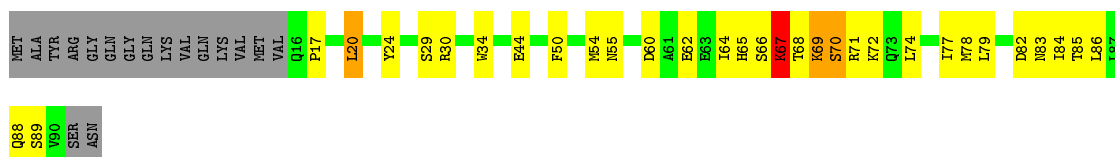


- Molecule 3: Small nuclear ribonucleoprotein Sm D2



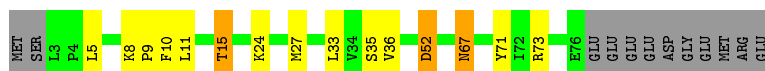
- Molecule 4: Small nuclear ribonucleoprotein E





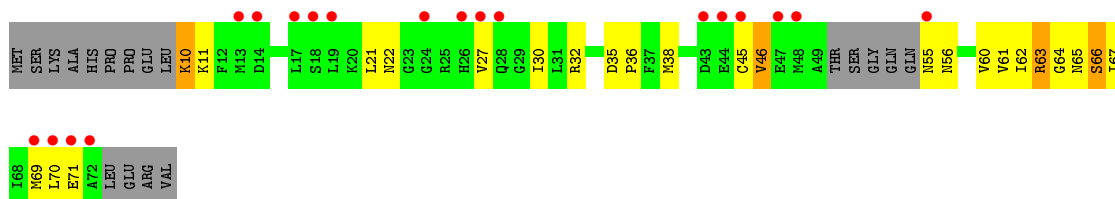
- Molecule 5: Small nuclear ribonucleoprotein F

Chain F: 69% 14% 14%



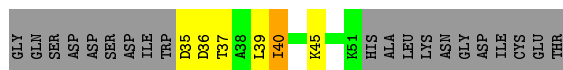
- Molecule 6: Small nuclear ribonucleoprotein G

Chain G: 25% 43% 28% 5% 24%



- Molecule 7: Survival motor neuron protein

Chain M: 30% 14% 54%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.32Å 115.76Å 128.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	60.00 – 3.28 47.54 – 3.28	Depositor EDS
% Data completeness (in resolution range)	79.2 (60.00-3.28) 79.2 (47.54-3.28)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.74 (at 3.25Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.198 , 0.262 0.205 , 0.265	Depositor DCC
$R_{free}$ test set	775 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	73.3	Xtriage
Anisotropy	0.101	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 57.3	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	4793	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.37% 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	2	0.63	0/1651	0.82	3/2243 (0.1%)
2	A	0.62	0/641	0.83	0/867
3	B	0.65	0/779	0.83	0/1045
4	E	0.53	0/631	0.77	0/847
5	F	0.59	0/588	0.72	0/795
6	G	0.45	0/450	0.65	0/597
7	M	0.81	1/131 (0.8%)	0.85	0/175
All	All	0.61	1/4871 (0.0%)	0.79	3/6569 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	M	40	ILE	C-N	6.34	1.48	1.34

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	2	44	VAL	C-N-CD	7.04	143.18	128.40
1	2	45	PRO	C-N-CD	6.22	141.46	128.40
1	2	45	PRO	CA-N-CD	-5.62	103.62	111.50

There are no chirality outliers.

There are no planarity outliers.

### 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	2	1614	0	1598	34	0
2	A	633	0	677	18	0
3	B	770	0	798	50	0
4	E	623	0	639	45	0
5	F	576	0	581	19	0
6	G	447	0	472	22	0
7	M	130	0	131	6	0
All	All	4793	0	4896	164	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 17.

The worst 5 of 164 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)
4:E:66:SER:O	4:E:67:LYS:HG2	1.32	1.21
3:B:15:GLU:O	3:B:18:LYS:HG2	1.47	1.10
3:B:13:PRO:HD2	3:B:16:LEU:CB	1.87	1.04
4:E:44:GLU:HB2	4:E:64:ILE:HD11	1.30	1.04
3:B:13:PRO:CD	3:B:16:LEU:HB2	1.90	1.02

There are no symmetry-related clashes.

## 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	2	193/280 (69%)	180 (93%)	12 (6%)	1 (0%)	29	62
2	A	78/82 (95%)	75 (96%)	3 (4%)	0	100	100
3	B	91/118 (77%)	83 (91%)	8 (9%)	0	100	100
4	E	73/92 (79%)	71 (97%)	0	2 (3%)	5	26
5	F	72/86 (84%)	65 (90%)	6 (8%)	1 (1%)	11	40

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
6	G	54/76 (71%)	48 (89%)	6 (11%)	0	100	100
7	M	15/37 (40%)	13 (87%)	2 (13%)	0	100	100
All	All	576/771 (75%)	535 (93%)	37 (6%)	4 (1%)	22	56

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	E	67	LYS
4	E	17	PRO
5	F	52	ASP
1	2	45	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	2	181/248 (73%)	155 (86%)	26 (14%)	3	15
2	A	75/77 (97%)	66 (88%)	9 (12%)	5	21
3	B	89/110 (81%)	72 (81%)	17 (19%)	1	6
4	E	70/84 (83%)	62 (89%)	8 (11%)	5	23
5	F	62/74 (84%)	56 (90%)	6 (10%)	8	29
6	G	49/66 (74%)	38 (78%)	11 (22%)	1	3
7	M	13/30 (43%)	11 (85%)	2 (15%)	2	12
All	All	539/689 (78%)	460 (85%)	79 (15%)	3	14

5 of 79 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	B	14	GLU
3	B	49	ASN
6	G	66	SER
3	B	15	GLU
3	B	33	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	B	25	ASN
4	E	83	ASN
3	B	49	ASN
2	A	39	HIS
3	B	48	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	2	201/280 (71%)	-0.16	0	100	100	33, 75, 117, 139	0
2	A	80/82 (97%)	-0.51	0	100	100	33, 60, 104, 119	0
3	B	95/118 (80%)	-0.39	0	100	100	34, 59, 135, 146	0
4	E	75/92 (81%)	0.02	0	100	100	54, 97, 143, 169	0
5	F	74/86 (86%)	0.05	0	100	100	42, 73, 107, 146	0
6	G	58/76 (76%)	1.54	19 (32%)	0	0	123, 162, 197, 211	0
7	M	17/37 (45%)	0.04	0	100	100	58, 77, 107, 141	0
All	All	600/771 (77%)	-0.03	19 (3%)	47	46	33, 77, 163, 211	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
6	G	18	SER	5.8
6	G	14	ASP	4.0
6	G	26	HIS	3.9
6	G	48	MET	3.8
6	G	69	MET	3.6

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.