



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 11:21 am GMT

PDB ID : 5EN7  
Title : Crystal structure of the Smu1-RED complex (native) of *Caenorhabditis elegans*.  
Authors : Ulrich, A.K.C.; Wahl, M.C.  
Deposited on : 2015-11-09  
Resolution : 2.94 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

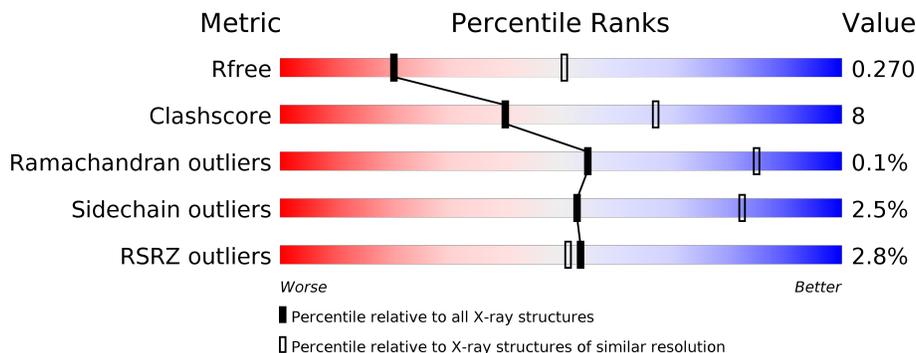
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.94 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	2289 (2.98-2.90)
Clashscore	112137	2543 (2.98-2.90)
Ramachandran outliers	110173	2475 (2.98-2.90)
Sidechain outliers	110143	2477 (2.98-2.90)
RSRZ outliers	101464	2301 (2.98-2.90)

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	184	
1	C	184	
1	E	184	
1	G	184	
2	B	63	
2	D	63	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	63	 <p>2% 22% 8% 70%</p>
2	H	63	 <p>17% 32% 11% 56%</p>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6461 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 SMU-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	177	1418	900	246	270	2	0	0	0
1	C	177	1418	900	246	270	2	0	0	0
1	E	176	1406	891	245	268	2	0	0	0
1	G	177	1417	898	246	271	2	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP G5EEG7
A	-1	ALA	-	expression tag	UNP G5EEG7
A	0	MET	-	expression tag	UNP G5EEG7
A	1	GLY	-	expression tag	UNP G5EEG7
C	-2	GLY	-	expression tag	UNP G5EEG7
C	-1	ALA	-	expression tag	UNP G5EEG7
C	0	MET	-	expression tag	UNP G5EEG7
C	1	GLY	-	expression tag	UNP G5EEG7
E	-2	GLY	-	expression tag	UNP G5EEG7
E	-1	ALA	-	expression tag	UNP G5EEG7
E	0	MET	-	expression tag	UNP G5EEG7
E	1	GLY	-	expression tag	UNP G5EEG7
G	-2	GLY	-	expression tag	UNP G5EEG7
G	-1	ALA	-	expression tag	UNP G5EEG7
G	0	MET	-	expression tag	UNP G5EEG7
G	1	GLY	-	expression tag	UNP G5EEG7

- Molecule 2 is a protein called Suppressor of Mec and Unc defects.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	16	Total	C	N	O	S	0	0	0
			140	87	31	21	1			
2	D	15	Total	C	N	O	S	0	0	0
			131	81	29	20	1			
2	F	19	Total	C	N	O	S	0	0	0
			163	99	36	27	1			
2	H	28	Total	C	N	O	S	0	0	0
			227	140	45	41	1			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	161	GLY	-	expression tag	UNP Q9N4U5
B	162	ALA	-	expression tag	UNP Q9N4U5
D	161	GLY	-	expression tag	UNP Q9N4U5
D	162	ALA	-	expression tag	UNP Q9N4U5
F	161	GLY	-	expression tag	UNP Q9N4U5
F	162	ALA	-	expression tag	UNP Q9N4U5
H	161	GLY	-	expression tag	UNP Q9N4U5
H	162	ALA	-	expression tag	UNP Q9N4U5

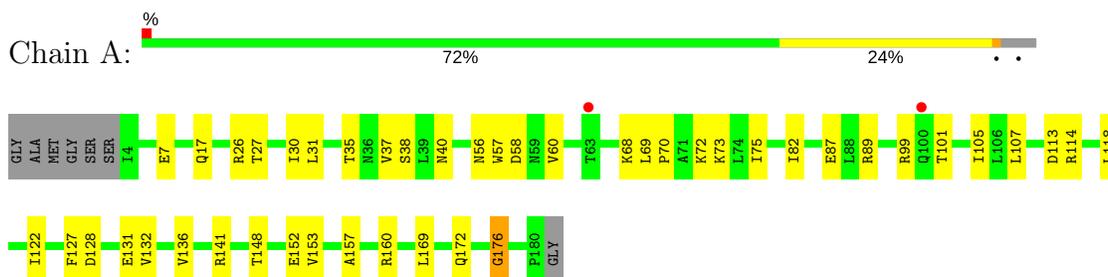
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	20	Total	O	0	0
			20	20		
3	B	4	Total	O	0	0
			4	4		
3	C	25	Total	O	0	0
			25	25		
3	D	3	Total	O	0	0
			3	3		
3	E	46	Total	O	0	0
			46	46		
3	F	5	Total	O	0	0
			5	5		
3	G	36	Total	O	0	0
			36	36		
3	H	2	Total	O	0	0
			2	2		

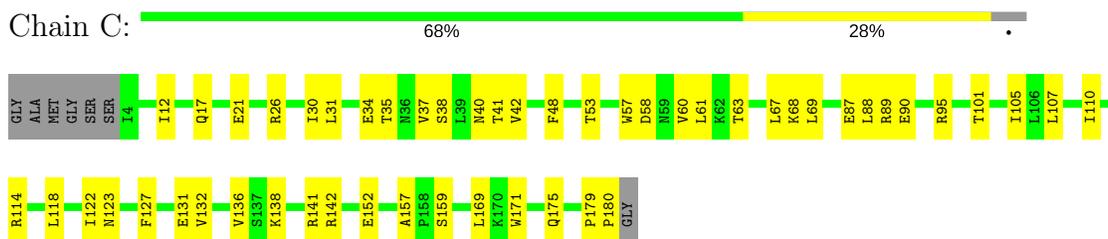
### 3 Residue-property plots [i](#)

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.

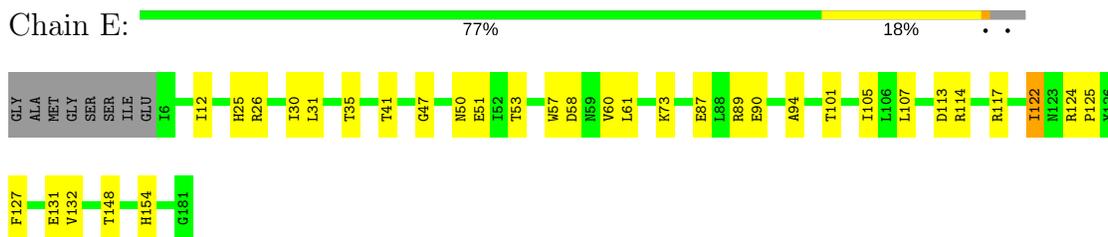
- Molecule 1: SMU-1



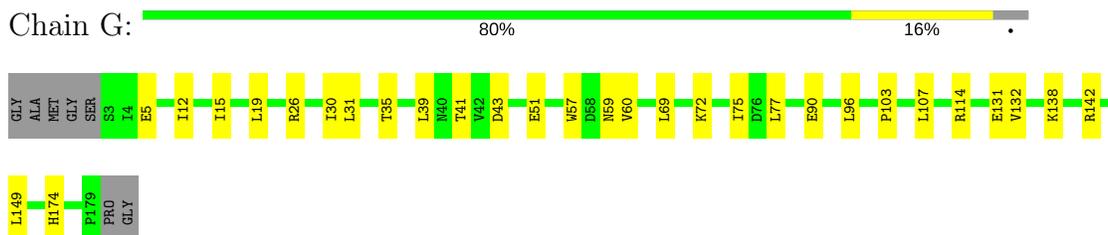
- Molecule 1: SMU-1



- Molecule 1: SMU-1



- Molecule 1: SMU-1





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	122.32Å 41.04Å 126.38Å 90.00° 95.92° 90.00°	Depositor
Resolution (Å)	46.16 – 2.94 46.16 – 2.94	Depositor EDS
% Data completeness (in resolution range)	98.3 (46.16-2.94) 98.3 (46.16-2.94)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.12	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 2.96Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.222 , 0.269 0.224 , 0.270	Depositor DCC
$R_{free}$ test set	1349 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	61.2	Xtrriage
Anisotropy	0.167	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 39.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	0.098 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6461	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.76% 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](#)

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	0.26	0/1439	0.49	0/1951
1	C	0.28	0/1439	0.48	0/1951
1	E	0.27	0/1427	0.47	0/1933
1	G	0.28	0/1437	0.48	0/1947
2	B	0.23	0/142	0.46	0/187
2	D	0.23	0/133	0.43	0/176
2	F	0.26	0/165	0.44	0/218
2	H	0.24	0/229	0.39	0/305
All	All	0.27	0/6411	0.47	0/8668

There are no bond length outliers.

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	1418	0	1462	31	0
1	C	1418	0	1462	34	0
1	E	1406	0	1448	23	0
1	G	1417	0	1460	18	0
2	B	140	0	145	2	0
2	D	131	0	132	5	0
2	F	163	0	164	3	0
2	H	227	0	231	5	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	20	0	0	1	0
3	B	4	0	0	0	0
3	C	25	0	0	1	0
3	D	3	0	0	0	0
3	E	46	0	0	1	0
3	F	5	0	0	0	0
3	G	36	0	0	0	0
3	H	2	0	0	0	0
All	All	6461	0	6504	107	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 8.

The worst 5 of 107 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:C:90:GLU:HG2	2:D:212:VAL:HG11	1.58	0.86
1:A:40:ASN:HD21	1:A:68:LYS:HG2	1.51	0.76
1:E:30:ILE:HG13	1:G:30:ILE:HG13	1.67	0.76
1:G:107:LEU:HD11	1:G:114:ARG:HD3	1.71	0.73
1:C:107:LEU:HD11	1:C:114:ARG:HD3	1.71	0.72

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	A	175/184 (95%)	171 (98%)	3 (2%)	1 (1%)	28	63
1	C	175/184 (95%)	172 (98%)	3 (2%)	0	100	100
1	E	174/184 (95%)	173 (99%)	1 (1%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	G	175/184 (95%)	172 (98%)	3 (2%)	0	100	100
2	B	14/63 (22%)	14 (100%)	0	0	100	100
2	D	13/63 (21%)	13 (100%)	0	0	100	100
2	F	17/63 (27%)	17 (100%)	0	0	100	100
2	H	26/63 (41%)	26 (100%)	0	0	100	100
All	All	769/988 (78%)	758 (99%)	10 (1%)	1 (0%)	55	85

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	176	GLY

### 5.3.2 Protein sidechains [i](#)

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	163/166 (98%)	161 (99%)	2 (1%)	75	92
1	C	163/166 (98%)	160 (98%)	3 (2%)	64	87
1	E	161/166 (97%)	160 (99%)	1 (1%)	89	97
1	G	163/166 (98%)	160 (98%)	3 (2%)	64	87
2	B	16/57 (28%)	14 (88%)	2 (12%)	5	15
2	D	15/57 (26%)	14 (93%)	1 (7%)	19	47
2	F	19/57 (33%)	17 (90%)	2 (10%)	8	24
2	H	25/57 (44%)	21 (84%)	4 (16%)	3	8
All	All	725/892 (81%)	707 (98%)	18 (2%)	53	82

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	122	ILE
2	F	211	MET

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
2	H	202	LEU
1	C	110	ILE
2	D	211	MET

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	40	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 [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	177/184 (96%)	0.12	2 (1%) 80 81	30, 73, 103, 122	0
1	C	177/184 (96%)	0.02	0 100 100	28, 68, 102, 115	0
1	E	176/184 (95%)	-0.03	0 100 100	27, 53, 81, 101	0
1	G	177/184 (96%)	-0.03	0 100 100	23, 54, 79, 112	0
2	B	16/63 (25%)	1.52	5 (31%) 0 0	119, 134, 144, 152	0
2	D	15/63 (23%)	1.31	3 (20%) 1 1	107, 116, 130, 137	0
2	F	19/63 (30%)	0.77	1 (5%) 27 25	63, 86, 114, 114	0
2	H	28/63 (44%)	2.14	11 (39%) 0 0	67, 103, 143, 156	0
All	All	785/988 (79%)	0.17	22 (2%) 53 51	23, 63, 115, 156	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	219	LEU	7.6
2	H	202	LEU	7.0
2	H	200	GLN	6.1
2	H	201	GLU	6.1
2	H	196	SER	6.1

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.