



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 08:23 am BST

PDB ID : 4C5E  
Title : Crystal structure of the minimal Pho-Sfmbt complex (P21 spacegroup)  
Authors : Alfieri, C.; Glatt, S.; Mueller, C.W.  
Deposited on : 2013-09-11  
Resolution : 1.95 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

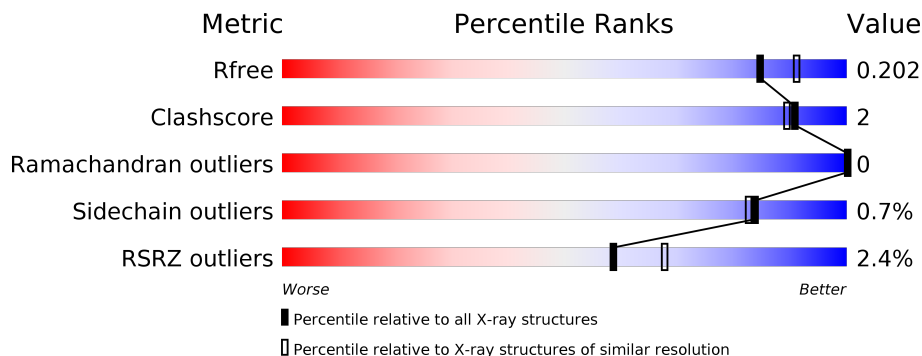
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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.



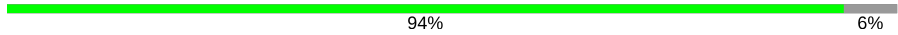

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	451	 2% 93% 5% 2%
1	B	451	 2% 93% 5% 2%
1	C	451	 1% 92% 5% 2%
1	D	451	 5% 92% 5% 2%
2	E	32	 88% 6% 6%
2	F	32	 91% 6% 3%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	G	32	 94% 6%
2	H	32	 91% 6%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 16286 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 POLYCOMB PROTEIN SFMBT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	440	Total 3479	C 2229	N 578	O 646	S 26	1	2	0
1	B	439	Total 3480	C 2229	N 578	O 647	S 26	0	5	0
1	C	438	Total 3456	C 2216	N 573	O 641	S 26	0	2	0
1	D	437	Total 3409	C 2186	N 568	O 630	S 25	0	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	530	MET	-	expression tag	UNP Q9VK33
B	530	MET	-	expression tag	UNP Q9VK33
C	530	MET	-	expression tag	UNP Q9VK33
D	530	MET	-	expression tag	UNP Q9VK33

- Molecule 2 is a protein called POLYCOMB PROTEIN PHO.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	30	Total 238	C 150	N 43	O 42	S 3	0	0	0
2	F	30	Total 232	C 147	N 40	O 42	S 3	0	0	0
2	G	30	Total 231	C 147	N 40	O 41	S 3	1	0	0
2	H	30	Total 233	C 148	N 40	O 42	S 3	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	141	GLY	-	expression tag	UNP Q8ST83
E	142	ALA	-	expression tag	UNP Q8ST83
E	143	MET	-	expression tag	UNP Q8ST83
E	144	ALA	-	expression tag	UNP Q8ST83
F	141	GLY	-	expression tag	UNP Q8ST83
F	142	ALA	-	expression tag	UNP Q8ST83
F	143	MET	-	expression tag	UNP Q8ST83
F	144	ALA	-	expression tag	UNP Q8ST83
G	141	GLY	-	expression tag	UNP Q8ST83
G	142	ALA	-	expression tag	UNP Q8ST83
G	143	MET	-	expression tag	UNP Q8ST83
G	144	ALA	-	expression tag	UNP Q8ST83
H	141	GLY	-	expression tag	UNP Q8ST83
H	142	ALA	-	expression tag	UNP Q8ST83
H	143	MET	-	expression tag	UNP Q8ST83
H	144	ALA	-	expression tag	UNP Q8ST83

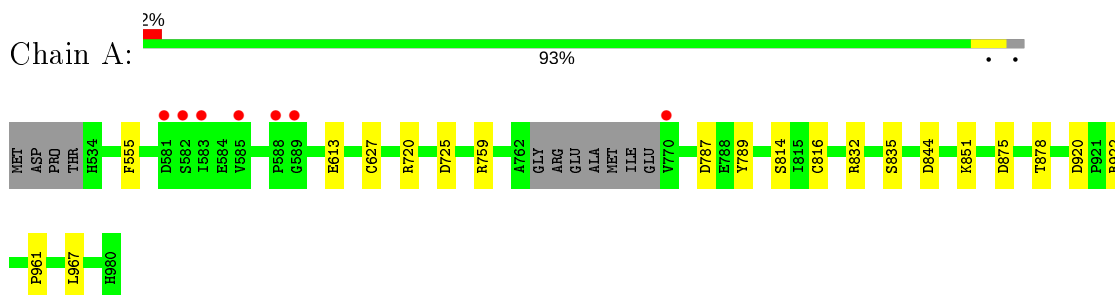
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	427	Total O 427 427	0	0
3	B	409	Total O 409 409	0	0
3	C	384	Total O 384 384	0	0
3	D	256	Total O 256 256	0	0
3	E	21	Total O 21 21	0	0
3	F	10	Total O 10 10	0	0
3	G	7	Total O 7 7	0	0
3	H	14	Total O 14 14	0	0

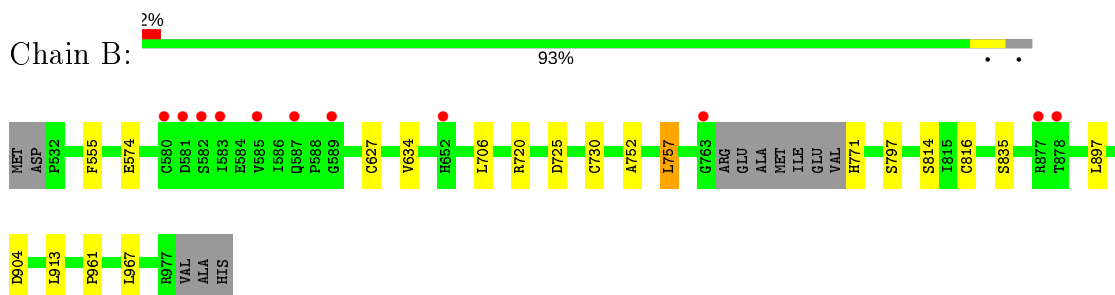
### 3 Residue-property plots [i](#)

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.

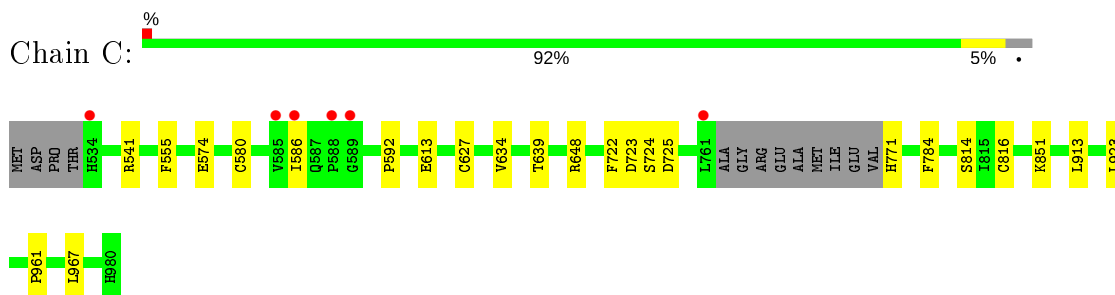
- Molecule 1: POLYCOMB PROTEIN SFMBT



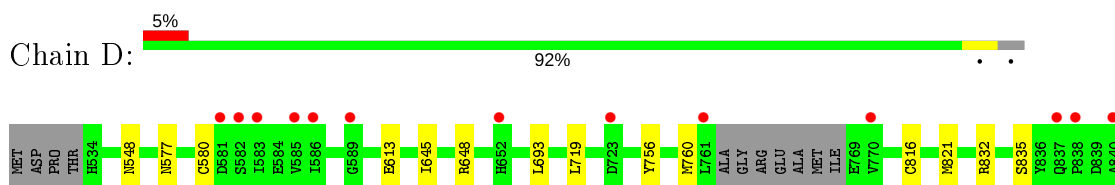
- Molecule 1: POLYCOMB PROTEIN SFMBT

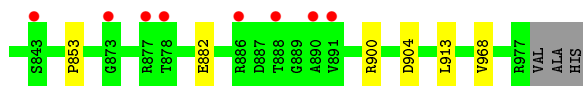


- Molecule 1: POLYCOMB PROTEIN SFMBT

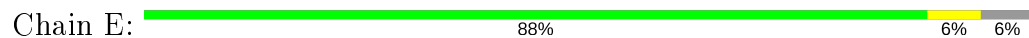


- Molecule 1: POLYCOMB PROTEIN SFMBT

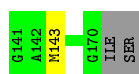
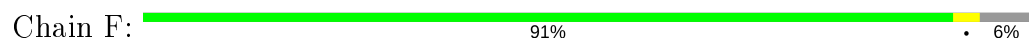




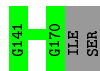
- Molecule 2: POLYCOMB PROTEIN PHO



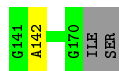
- Molecule 2: POLYCOMB PROTEIN PHO



- Molecule 2: POLYCOMB PROTEIN PHO



- Molecule 2: POLYCOMB PROTEIN PHO



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	87.51Å 147.88Å 97.88Å 90.00° 112.26° 90.00°	Depositor
Resolution (Å)	48.60 – 1.95 49.29 – 1.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.60-1.95) 99.8 (49.29-1.95)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.66 (at 1.95Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.179 , 0.198 0.184 , 0.202	Depositor DCC
$R_{free}$ test set	8327 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	28.8	Xtrriage
Anisotropy	0.256	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 52.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	16286	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 30.08 % 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 1.3892e-03. 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	0.23	0/3582	0.43	0/4880
1	B	0.23	0/3583	0.43	0/4883
1	C	0.22	0/3556	0.42	0/4843
1	D	0.22	0/3509	0.42	0/4783
2	E	0.25	0/243	0.40	0/324
2	F	0.26	0/237	0.40	0/317
2	G	0.22	0/236	0.38	0/316
2	H	0.24	0/238	0.40	0/318
All	All	0.23	0/15184	0.42	0/20664

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	3479	0	3272	13	0
1	B	3480	0	3267	12	0
1	C	3456	0	3245	13	0
1	D	3409	0	3182	10	0
2	E	238	0	233	1	0
2	F	232	0	222	1	0
2	G	231	0	219	0	0
2	H	233	0	224	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	427	0	0	2	0
3	B	409	0	0	0	0
3	C	384	0	0	2	0
3	D	256	0	0	2	0
3	E	21	0	0	0	0
3	F	10	0	0	1	0
3	G	7	0	0	0	0
3	H	14	0	0	0	0
All	All	16286	0	13864	49	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 2.

The worst 5 of 49 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:722:PHE:CG	1:C:722:PHE:CE2	2.41	0.99
1:B:814:SER:OG	1:B:816[B]:CYS:SG	2.51	0.69
1:B:720:ARG:NH1	1:B:725:ASP:O	2.27	0.67
1:A:832:ARG:NH1	1:A:844:ASP:OD2	2.26	0.67
1:D:853:PRO:O	1:D:900:ARG:NH2	2.27	0.67

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	438/451 (97%)	426 (97%)	12 (3%)	0	100 100
1	B	440/451 (98%)	431 (98%)	9 (2%)	0	100 100
1	C	436/451 (97%)	426 (98%)	10 (2%)	0	100 100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	435/451 (96%)	424 (98%)	11 (2%)	0	100	100
2	E	28/32 (88%)	27 (96%)	1 (4%)	0	100	100
2	F	28/32 (88%)	26 (93%)	2 (7%)	0	100	100
2	G	28/32 (88%)	27 (96%)	1 (4%)	0	100	100
2	H	28/32 (88%)	27 (96%)	1 (4%)	0	100	100
All	All	1861/1932 (96%)	1814 (98%)	47 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	372/392 (95%)	372 (100%)	0	100	100
1	B	372/392 (95%)	367 (99%)	5 (1%)	69	65
1	C	368/392 (94%)	365 (99%)	3 (1%)	81	80
1	D	358/392 (91%)	355 (99%)	3 (1%)	81	80
2	E	24/26 (92%)	24 (100%)	0	100	100
2	F	23/26 (88%)	23 (100%)	0	100	100
2	G	22/26 (85%)	22 (100%)	0	100	100
2	H	23/26 (88%)	23 (100%)	0	100	100
All	All	1562/1672 (93%)	1551 (99%)	11 (1%)	84	82

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	913	LEU
1	C	771	HIS
1	D	882	GLU
1	B	771	HIS
1	C	923	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	566	ASN
1	D	692	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	440/451 (97%)	-0.20	7 (1%) 72 79	13, 25, 50, 67	1 (0%)
1	B	439/451 (97%)	-0.28	11 (2%) 57 66	15, 27, 56, 73	0
1	C	438/451 (97%)	-0.22	6 (1%) 75 82	18, 30, 52, 75	0
1	D	437/451 (96%)	-0.03	21 (4%) 30 40	18, 39, 65, 77	0
2	E	30/32 (93%)	-0.35	0 100 100	14, 25, 46, 53	0
2	F	30/32 (93%)	0.05	0 100 100	18, 33, 61, 77	0
2	G	30/32 (93%)	-0.07	0 100 100	24, 36, 55, 62	1 (3%)
2	H	30/32 (93%)	-0.02	0 100 100	17, 29, 52, 58	0
All	All	1874/1932 (96%)	-0.18	45 (2%) 59 68	13, 30, 58, 77	2 (0%)

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	583	ILE	4.8
1	D	589	GLY	4.5
1	B	585	VAL	4.4
1	D	582	SER	3.9
1	B	582	SER	3.7

### 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.