



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 04:42 am BST

PDB ID : 4PK7  
Title : crystal structure of human Stromal Antigen 2 (SA2) in complex with Sister Chromatid Cohesion protein 1 (Scc1) with bound MES, native proteins  
Authors : Hara, K.; Chen, Z.; Tomchick, D.R.; Yu, H.  
Deposited on : 2014-05-13  
Resolution : 2.95 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
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

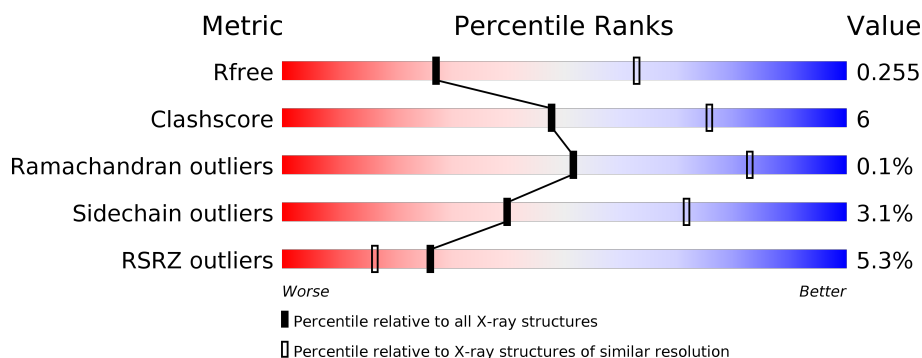
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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	994	<div> <div>5%</div> <div> <div></div> <div>76%</div> <div>14%</div> <div>• 9%</div> </div> </div>
2	B	148	<div> <div>41%</div> <div>9%</div> <div>•</div> <div>49%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 16049 atoms, of which 8063 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 Cohesin subunit SA-2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	903	Total	C	H	N	O	S	0	0	0
			14756	4702	7393	1226	1381	54			

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	67	GLY	-	expression tag	UNP Q8N3U4
A	68	ALA	-	expression tag	UNP Q8N3U4
A	69	MET	-	expression tag	UNP Q8N3U4
A	70	ASP	-	expression tag	UNP Q8N3U4
A	71	PRO	-	expression tag	UNP Q8N3U4
A	72	GLU	-	expression tag	UNP Q8N3U4
A	73	PHE	-	expression tag	UNP Q8N3U4
A	74	GLY	-	expression tag	UNP Q8N3U4
A	75	ARG	-	expression tag	UNP Q8N3U4
A	76	PRO	-	expression tag	UNP Q8N3U4
A	77	GLY	-	expression tag	UNP Q8N3U4
A	78	ARG	-	expression tag	UNP Q8N3U4
A	79	PRO	-	expression tag	UNP Q8N3U4

- Molecule 2 is a protein called Double-strand-break repair protein rad21 homolog.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	76	Total	C	H	N	O	S	0	0	0
			1268	396	657	103	109	3			

There are 8 discrepancies between the modelled and reference sequences:

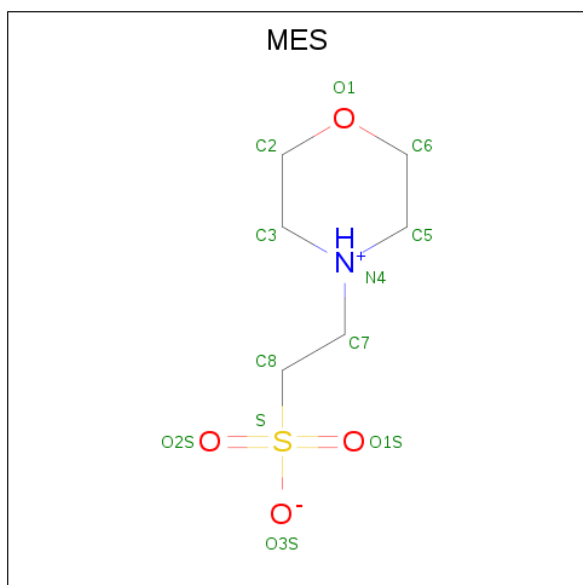
Chain	Residue	Modelled	Actual	Comment	Reference
B	273	GLY	-	expression tag	UNP O60216
B	274	PRO	-	expression tag	UNP O60216
B	275	LEU	-	expression tag	UNP O60216

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	276	GLY	-	expression tag	UNP O60216
B	277	SER	-	expression tag	UNP O60216
B	278	GLY	-	expression tag	UNP O60216
B	279	ARG	-	expression tag	UNP O60216
B	280	PRO	-	expression tag	UNP O60216

- Molecule 3 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	H	N	O	S	0	0
			25	6	13	1	4	1		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.45Å 107.28Å 180.10Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.68 – 2.95 47.68 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.68-2.95) 99.9 (47.68-2.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.22 (at 2.96Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.188 , 0.251 0.194 , 0.255	Depositor DCC
$R_{free}$ test set	1642 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	67.9	Xtriage
Anisotropy	0.248	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 47.0	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	0.92	EDS
Total number of atoms	16049	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.76% 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](#)

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

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	0.45	0/7486	0.61	1/10079 (0.0%)
2	B	0.46	0/622	0.68	0/841
All	All	0.45	0/8108	0.62	1/10920 (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	579	SER	CB-CA-C	-5.42	99.81	110.10

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	7363	7393	7371	90	1
2	B	611	657	657	15	0
3	A	12	13	13	1	0
All	All	7986	8063	8041	97	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 6.

The worst 5 of 97 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:756:GLU:N	1:A:756:GLU:OE2	2.29	0.65
1:A:548:LEU:HD23	1:A:553:LYS:HG3	1.77	0.65
1:A:525:MET:O	1:A:528:THR:OG1	2.15	0.63
2:B:349:THR:OG1	2:B:350:THR:N	2.31	0.63
1:A:981:ALA:HB1	1:A:1000:LEU:HD11	1.81	0.62

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:A:721:GLU:OE2	1:A:1028:PHE:H[3_545]	1.54	0.06

## 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	877/994 (88%)	841 (96%)	35 (4%)	1 (0%)	51	83
2	B	74/148 (50%)	70 (95%)	4 (5%)	0	100	100
All	All	951/1142 (83%)	911 (96%)	39 (4%)	1 (0%)	51	83

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	579	SER

### 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	819/891 (92%)	795 (97%)	24 (3%)	42	73
2	B	71/134 (53%)	67 (94%)	4 (6%)	21	53
All	All	890/1025 (87%)	862 (97%)	28 (3%)	40	71

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

Mol	Chain	Res	Type
1	A	699	ASN
1	A	737	THR
2	B	348	VAL
1	A	708	LEU
1	A	711	CYS

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	744	GLN
1	A	764	GLN

### 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](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	MES	A	1101	-	12,12,12	1.77	1 (8%)	14,16,16	2.21	6 (42%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MES	A	1101	-	-	2/6/14/14	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1101	MES	C8-S	-5.44	1.69	1.77

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1101	MES	O1S-S-C8	4.37	112.18	106.92
3	A	1101	MES	C6-C5-N4	-3.80	104.35	110.10
3	A	1101	MES	O3S-S-C8	3.09	110.76	105.77
3	A	1101	MES	O2S-S-C8	2.63	110.08	106.92
3	A	1101	MES	C5-N4-C3	2.38	114.19	108.83

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1101	MES	C8-C7-N4-C5
3	A	1101	MES	C8-C7-N4-C3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1101	MES	1	0

## 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	903/994 (90%)	0.63	52 (5%) 23 14	25, 59, 105, 130	0
2	B	76/148 (51%)	0.51	0 100 100	30, 49, 85, 103	0
All	All	979/1142 (85%)	0.62	52 (5%) 26 16	25, 59, 104, 130	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1010	LEU	4.6
1	A	807	ARG	4.4
1	A	973	LEU	4.2
1	A	940	ARG	3.9
1	A	936	TYR	3.9

### 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. 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	B-factors( $\text{\AA}^2$ )	Q<0.9
3	MES	A	1101	12/12	0.90	0.28	57,77,118,127	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.