



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 26, 2017 – 12:22 AM EDT

PDB ID : 1C8C  
Title : CRYSTAL STRUCTURES OF THE CHROMOSOMAL PROTEINS  
SSO7D/SAC7D BOUND TO DNA CONTAINING T-G MISMATCHED  
BASE PAIRS  
Authors : Su, S.; Gao, Y.-G.; Robinson, H.; Liaw, Y.-C.; Edmondson, S.P.; Shriver,  
J.W.; Wang, A.H.-J.  
Deposited on : unknown  
Resolution : 1.45 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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	:	rb-20030345
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)	:	rb-20030345

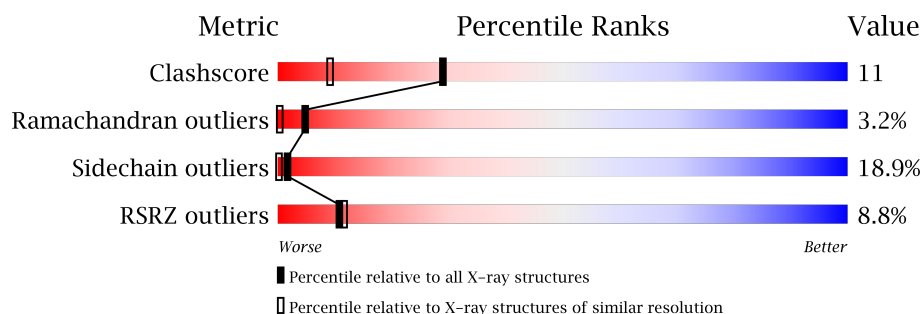
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.45 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))
Clashscore	112137	1573 (1.48-1.44)
Ramachandran outliers	110173	1555 (1.48-1.44)
Sidechain outliers	110143	1555 (1.48-1.44)
RSRZ outliers	101464	1516 (1.48-1.44)

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	B	8	<div> <div>13%</div> <div>75%</div> <div>13%</div> </div>
1	C	8	<div> <div>25%</div> <div>75%</div> </div>
2	A	64	<div> <div>11%</div> <div>67%</div> <div>20%</div> <div>8%</div> <div>5%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 1469 atoms, of which 482 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 DNA chain called 5'-D(\*GP\*TP\*GP\*AP\*TP\*CP\*GP\*C)-3'.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	B	8	Total	C	H	N	O	P	0	0	0
			181	78	19	30	47	7			
1	C	8	Total	C	H	N	O	P	0	0	0
			181	78	19	30	47	7			

- Molecule 2 is a protein called DNA-BINDING PROTEIN 7A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	A	64	Total	C	H	N	O	S	0	0	0
			636	323	130	88	92	3			

- Molecule 3 is water.

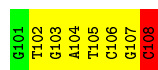
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	36	Total	H	O	0	0
			108	72	36		
3	C	36	Total	H	O	0	0
			108	72	36		
3	A	85	Total	H	O	0	0
			255	170	85		

### 3 Residue-property plots

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.

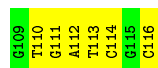
- Molecule 1: 5'-D(\*GP\*TP\*GP\*AP\*TP\*CP\*GP\*C)-3'

Chain B: 



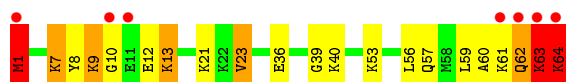
- Molecule 1: 5'-D(\*GP\*TP\*GP\*AP\*TP\*CP\*GP\*C)-3'

Chain C: 



- Molecule 2: DNA-BINDING PROTEIN 7A

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.48 Å 49.80 Å 37.68 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 1.45 34.36 – 1.30	Depositor EDS
% Data completeness (in resolution range)	72.2 (8.00-1.45) 77.7 (34.36-1.30)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.14 (at 1.30 Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, $R_{free}$	0.229 , 0.287 0.203 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	13.3	Xtriage
Anisotropy	0.287	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.22 , 34.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.010 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	1469	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.98% 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	B	0.96	0/181	2.13	8/278 (2.9%)
1	C	0.91	0/181	2.43	15/278 (5.4%)
2	A	20.43	16/512 (3.1%)	4.17	22/674 (3.3%)
All	All	15.65	16/874 (1.8%)	3.45	45/1230 (3.7%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	3
2	A	1	0
All	All	1	3

All (16) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	64	LYS	CD-CE	264.95	8.13	1.51
2	A	64	LYS	CE-NZ	215.38	6.87	1.49
2	A	64	LYS	N-CA	149.39	4.45	1.46
2	A	63	LYS	CB-CG	126.32	4.93	1.52
2	A	64	LYS	CA-CB	122.61	4.23	1.53
2	A	63	LYS	CE-NZ	92.05	3.79	1.49
2	A	63	LYS	CD-CE	86.40	3.67	1.51
2	A	63	LYS	CA-C	85.32	3.74	1.52
2	A	63	LYS	CA-CB	80.19	3.30	1.53
2	A	64	LYS	CA-C	68.23	3.30	1.52
2	A	64	LYS	CB-CG	66.12	3.31	1.52
2	A	63	LYS	CG-CD	61.23	3.60	1.52
2	A	1	MET	CG-SD	29.33	2.57	1.81
2	A	63	LYS	N-CA	-16.89	1.12	1.46
2	A	63	LYS	C-O	9.73	1.41	1.23

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	64	LYS	CG-CD	5.63	1.71	1.52

All (45) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	63	LYS	C-N-CA	-46.06	6.56	121.70
2	A	63	LYS	CA-C-N	-34.38	41.57	117.20
2	A	63	LYS	CA-CB-CG	-32.64	41.60	113.40
2	A	64	LYS	CB-CA-C	-30.16	50.09	110.40
2	A	64	LYS	CG-CD-CE	-27.30	30.00	111.90
2	A	64	LYS	N-CA-CB	-25.30	65.05	110.60
2	A	63	LYS	CA-C-O	-24.83	67.95	120.10
2	A	1	MET	N-CA-CB	24.68	155.02	110.60
2	A	64	LYS	CA-C-O	-23.61	70.51	120.10
2	A	64	LYS	CD-CE-NZ	-21.72	61.75	111.70
2	A	62	GLN	C-N-CA	18.39	167.68	121.70
2	A	63	LYS	CB-CG-CD	-16.69	68.20	111.60
2	A	1	MET	CG-SD-CE	16.43	126.50	100.20
2	A	63	LYS	N-CA-C	-14.06	73.04	111.00
2	A	64	LYS	CA-CB-CG	-12.34	86.26	113.40
2	A	63	LYS	CG-CD-CE	-11.70	76.82	111.90
2	A	64	LYS	N-CA-C	-11.28	80.56	111.00
1	C	110	DT	N3-C4-O4	-9.98	113.91	119.90
1	C	116	DC	C1'-O4'-C4'	-9.54	100.56	110.10
2	A	63	LYS	N-CA-CB	-9.53	93.44	110.60
2	A	1	MET	CB-CA-C	-9.40	91.61	110.40
2	A	63	LYS	CD-CE-NZ	-9.04	90.91	111.70
2	A	63	LYS	O-C-N	-8.24	109.52	122.70
1	C	116	DC	O4'-C4'-C3'	-8.18	101.09	106.00
1	C	114	DC	O4'-C4'-C3'	-8.03	101.18	106.00
1	B	102	DT	N3-C4-O4	7.73	124.54	119.90
1	B	102	DT	C5-C4-O4	-7.53	119.63	124.90
1	C	112	DA	O4'-C1'-N9	-7.19	102.97	108.00
2	A	1	MET	CB-CG-SD	7.03	133.50	112.40
1	C	110	DT	C5-C4-O4	7.03	129.82	124.90
1	C	114	DC	C5-C6-N1	6.55	124.28	121.00
1	B	102	DT	O4'-C4'-C3'	-6.29	101.98	104.50
1	C	114	DC	N3-C4-N4	6.03	122.22	118.00
1	C	110	DT	O4'-C1'-N1	-5.95	103.84	108.00
1	B	105	DT	C1'-O4'-C4'	-5.92	104.18	110.10
1	C	113	DT	C1'-O4'-C4'	-5.78	104.32	110.10
1	C	114	DC	C5-C4-N4	-5.76	116.17	120.20

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	108	DC	C6-N1-C2	5.64	122.56	120.30
1	C	116	DC	O4'-C1'-N1	-5.62	104.06	108.00
1	C	114	DC	OP1-P-OP2	5.44	127.76	119.60
1	C	111	DG	C2-N3-C4	5.36	114.58	111.90
1	B	104	DA	P-O3'-C3'	5.32	126.09	119.70
1	B	107	DG	C1'-O4'-C4'	-5.25	104.85	110.10
1	C	114	DC	C2-N3-C4	5.12	122.46	119.90
1	B	104	DA	C8-N9-C4	-5.07	103.77	105.80

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	A	1	MET	CA

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	103	DG	Sidechain
1	B	106	DC	Sidechain
1	B	108	DC	Sidechain

## 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	B	162	19	92	1	0
1	C	162	19	92	0	0
2	A	506	130	543	16	3
3	A	85	170	0	4	5
3	B	36	72	0	1	1
3	C	36	72	0	0	1
All	All	987	482	727	17	5

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.

All (17) 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)
2:A:1:MET:SD	2:A:1:MET:CG	2.57	0.93
2:A:13:LYS:HB3	3:A:1060:HOH:O	2.01	0.61
2:A:63:LYS:NZ	2:A:63:LYS:HG3	2.19	0.57
2:A:36:GLU:HB3	3:A:1052:HOH:O	2.10	0.50
2:A:60:ALA:HA	3:A:1065:HOH:O	2.11	0.50
2:A:8:TYR:CD2	2:A:9:LYS:HG3	2.47	0.49
2:A:56:LEU:O	2:A:59:LEU:HB3	2.13	0.49
2:A:63:LYS:HG2	2:A:64:LYS:N	2.30	0.46
2:A:57:GLN:O	2:A:61:LYS:HG3	2.16	0.46
2:A:23:VAL:HG22	2:A:59:LEU:HD13	1.98	0.45
2:A:8:TYR:O	2:A:9:LYS:HD3	2.15	0.45
2:A:7:LYS:HB2	2:A:7:LYS:HE2	1.77	0.43
1:B:108:DC:H2"	3:B:1100:HOH:O	2.19	0.42
2:A:8:TYR:HD2	2:A:9:LYS:HD2	1.83	0.42
2:A:62:GLN:HG3	2:A:64:LYS:HD2	2.01	0.41
2:A:61:LYS:HG3	3:A:1096:HOH:O	2.20	0.41
2:A:61:LYS:HE2	2:A:61:LYS:HB3	1.66	0.41

All (5) 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)
2:A:21:LYS:HZ3	3:A:1115:HOH:H1[3_545]	0.93	0.67
2:A:39:GLY:H	3:A:1074:HOH:H2[3_545]	0.96	0.64
3:B:1020:HOH:H2	3:A:1112:HOH:H1[4_456]	1.29	0.31
3:C:1104:HOH:H1	3:A:1036:HOH:H2[2_565]	1.30	0.30
2:A:39:GLY:N	3:A:1074:HOH:H2[3_545]	1.56	0.04

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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
2	A	62/64 (97%)	57 (92%)	3 (5%)	2 (3%)	5 0

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	A	63	LYS
2	A	10	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
2	A	53/53 (100%)	43 (81%)	10 (19%)	<b>2</b> <b>0</b>

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	A	1	MET
2	A	7	LYS
2	A	9	LYS
2	A	12	GLU
2	A	13	LYS
2	A	23	VAL
2	A	40	LYS
2	A	53	LYS
2	A	63	LYS
2	A	64	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 ⓘ

### 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	B	8/8 (100%)	-0.48	0	100 100	14, 17, 20, 28	0
1	C	8/8 (100%)	-0.51	0	100 100	13, 18, 25, 26	0
2	A	64/64 (100%)	0.44	7 (10%)	6 7	9, 20, 64, 86	0
All	All	80/80 (100%)	0.25	7 (8%)	11 11	9, 19, 64, 86	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	1	MET	10.3
2	A	64	LYS	6.4
2	A	63	LYS	3.8
2	A	61	LYS	2.9
2	A	11	GLU	2.7
2	A	10	GLY	2.7
2	A	62	GLN	2.5

### 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 [i](#)

There are no such residues in this entry.