



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2020 – 09:34 PM BST

PDB ID : 6EDB
Title : Crystal structure of SRY.hcGAS-21bp dsDNA complex
Authors : Xie, W.; Lama, L.; Adura, C.; Glickman, J.F.; Tuschl, T.; Patel, D.J.
Deposited on : 2018-08-09
Resolution : 3.21 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

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	:	2.13
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)	:	2.13

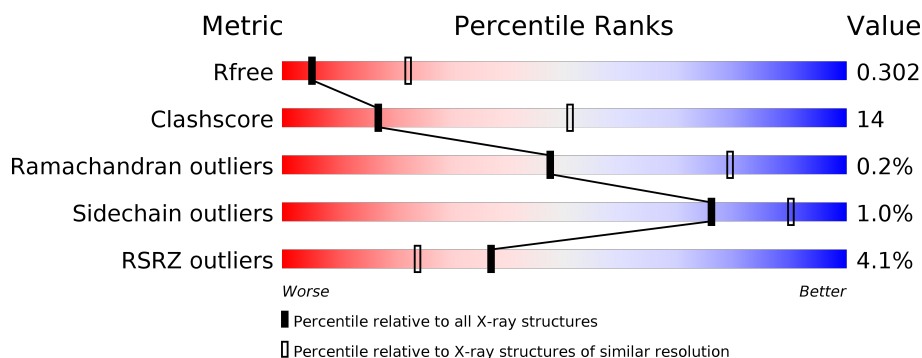
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.21 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 ≥ 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	455	<div> <div>3%</div> <div>55%</div> <div>23%</div> <div>•</div> <div>22%</div> </div>
1	B	455	<div> <div>%</div> <div>12%</div> <div>•</div> <div>85%</div> </div>
2	C	22	<div> <div>18%</div> <div>77%</div> <div>5%</div> </div>
3	D	22	<div> <div>55%</div> <div>41%</div> <div>5%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4379 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 Sex-determining region Y protein,Cyclic GMP-AMP synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	357	Total	C	N	O	S	0	0	0
			2944	1882	510	537	15			
1	B	67	Total	C	N	O	S	0	0	0
			579	367	110	97	5			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	68	SER	-	expression tag	UNP Q05066
A	144	LYS	-	linker	UNP Q05066
A	145	GLY	-	linker	UNP Q05066
A	146	GLU	-	linker	UNP Q05066
A	147	THR	-	linker	UNP Q05066
A	148	LYS	-	linker	UNP Q05066
A	149	LYS	-	linker	UNP Q05066
A	150	LYS	-	linker	UNP Q05066
A	151	PHE	-	linker	UNP Q05066
A	152	LYS	-	linker	UNP Q05066
A	153	ASP	-	linker	UNP Q05066
A	154	PRO	-	linker	UNP Q05066
A	155	ASN	-	linker	UNP Q05066
A	156	ALA	-	linker	UNP Q05066
B	0	SER	-	expression tag	UNP Q05066
B	76	LYS	-	linker	UNP Q05066
B	77	GLY	-	linker	UNP Q05066
B	78	GLU	-	linker	UNP Q05066
B	79	THR	-	linker	UNP Q05066
B	80	LYS	-	linker	UNP Q05066
B	81	LYS	-	linker	UNP Q05066
B	82	LYS	-	linker	UNP Q05066
B	83	PHE	-	linker	UNP Q05066
B	84	LYS	-	linker	UNP Q05066
B	85	ASP	-	linker	UNP Q05066

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	86	PRO	-	linker	UNP Q05066
B	87	ASN	-	linker	UNP Q05066
B	88	ALA	-	linker	UNP Q05066

- Molecule 2 is a DNA chain called DNA (5'-D(*CP*GP*AP*TP*CP*CP*GP*GP*GP*AP*TP*CP*TP*AP*AP*AP*CP*AP*AP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	21	Total	C	N	O	P	0	0	0
			429	205	83	121	20			

- Molecule 3 is a DNA chain called DNA (5'-D(*GP*CP*AP*TP*TP*GP*TP*TP*TP*AP*GP*AP*TP*CP*CP*CP*GP*GP*AP*TP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	21	Total	C	N	O	P	0	0	0
			426	205	74	127	20			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Zn	0	0
			1	1		

ASP
CYS
LEU
LYS
LEU
MET
LYS
TYR
LEU
GLU
GLN
LYS
LEU
LYS
GLU
ASN
ARG
PHE
LYS
ASP
LYS
HIS
LEU
ASP
LYS
PHE
SER
SER
ASN
LEU
TYR
HIS
VAL
LYS
SER
THR
LYS
ALA
PHE
PHE
HIS
VAL
THR
LYS
GLN
ILE
GLN
ASN
PRO
GLN
ASP
SER
ASN
GLN
TRP
PHE
ASP
ARG
LYS
ASP
LEU
GLY
LEU
CYS
PHE
ASN
CYS

VAL
THR
PHE
LEU
GLN
CYS
LEU
ARG
THR
GLU
LYS
LEU
GLU
ASN
TYR
PHE
ILE
PRO
GLU
PHE
ASN
LEU
PHE
SER
SER
ASN
LEU
ILE
ASP
LYS
ARG
SER
LYS
GLU
PHE
PHE
LEU
THR
LYS
GLN
ILE
GLU
TYR
GLU
ARG
ASN
ASN
GLU
PHE
PRO
VAL
PHE
ASP
GLU
PHE

● Molecule 2: DNA (5'-D(*CP*GP*AP*TP*CP*CP*GP*GP*GP*AP*TP*CP*TP*AP*AP*AP*CP*AP*AP*TP*G)-3')



G1
G2
A3
T4
C5
C6
C7
G8
G9
A10
T11
C12
T13
A14
A15
A16
G17
A18
A19
T20
G21
DC

● Molecule 3: DNA (5'-D(*GP*CP*AP*TP*TP*GP*TP*TP*TP*AP*GP*AP*TP*CP*CP*CP*GP*GP*AP*TP*C)-3')



DC
G2
C3
T6
T10
A11
G12
C17
G18
G19
G22

4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, α , β , γ	90.31Å 115.83Å 159.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.01 – 3.21 30.01 – 3.21	Depositor EDS
% Data completeness (in resolution range)	90.3 (30.01-3.21) 82.5 (30.01-3.21)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.83 (at 3.24Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.255 , 0.302 0.261 , 0.302	Depositor DCC
R_{free} test set	1264 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å ²)	120.7	Xtriage
Anisotropy	0.498	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 28.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4379	wwPDB-VP
Average B, all atoms (Å ²)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.12% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: ZN

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	0.31	0/3003	0.54	1/4027 (0.0%)
1	B	0.24	0/594	0.35	0/794
2	C	0.45	0/482	0.84	0/742
3	D	0.52	0/476	0.96	0/733
All	All	0.35	0/4555	0.62	1/6296 (0.0%)

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	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	330	TRP	C-N-CD	-7.61	103.85	120.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	330	TRP	Mainchain

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	A	2944	0	2990	91	0
1	B	579	0	575	11	0
2	C	429	0	237	15	0
3	D	426	0	240	7	0
4	A	1	0	0	0	0
All	All	4379	0	4042	117	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 14.

The worst 5 of 117 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:444:HIS:O	1:A:447:THR:HG22	1.40	1.18
1:A:430:LEU:HD21	1:A:504:LEU:HD13	1.40	1.03
1:A:241:GLU:OE1	1:A:246:ARG:HD2	1.68	0.92
1:A:416:LEU:O	1:A:420:LEU:HD13	1.72	0.88
1:A:220:ILE:HD11	1:A:418:GLU:OE2	1.79	0.82

There are no symmetry-related clashes.

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
1	A	355/455 (78%)	324 (91%)	30 (8%)	1 (0%)	41 74

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	65/455 (14%)	64 (98%)	1 (2%)	0	100	100
All	All	420/910 (46%)	388 (92%)	31 (7%)	1 (0%)	47	79

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	303	GLY

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	A	331/417 (79%)	327 (99%)	4 (1%)	71	88
1	B	60/417 (14%)	60 (100%)	0	100	100
All	All	391/834 (47%)	387 (99%)	4 (1%)	76	90

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	245	THR
1	A	370	PHE
1	A	410	LEU
1	A	483	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	27	ASN

5.3.3 RNA ⓘ

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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95th 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 ²)	Q<0.9
1	A	357/455 (78%)	0.08	13 (3%) 42 27	63, 100, 137, 148	0
1	B	67/455 (14%)	0.43	6 (8%) 9 5	90, 116, 135, 157	0
2	C	21/22 (95%)	-0.51	0 100 100	87, 101, 123, 127	0
3	D	21/22 (95%)	-0.17	0 100 100	87, 102, 121, 141	0
All	All	466/954 (48%)	0.09	19 (4%) 37 24	63, 103, 137, 157	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	430	LEU	4.1
1	A	368	ASN	3.8
1	B	71	ASN	3.7
1	A	201	SER	3.3
1	B	20	ARG	3.3

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 monosaccharides 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, 95th 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(\AA^2)	Q<0.9
4	ZN	A	601	1/1	0.85	0.08	30,30,30,30	0

6.5 Other polymers [i](#)

There are no such residues in this entry.