WORLDWIDE PDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 02:12 am GMT

|		
PDB ID	:	5GXI
Title	:	Strucutre of the Gemin5 WD40 domain in complex with AAUUUUUGAG
Authors	:	Xu, C.; He, H.; Li, Y.; Dong, A.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.M.; Min, J.; Structural Genomics Consortium (SGC)
Deposited on	:	2016-09-18
Resolution	:	1.85 Å(reported)

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 1.85 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.
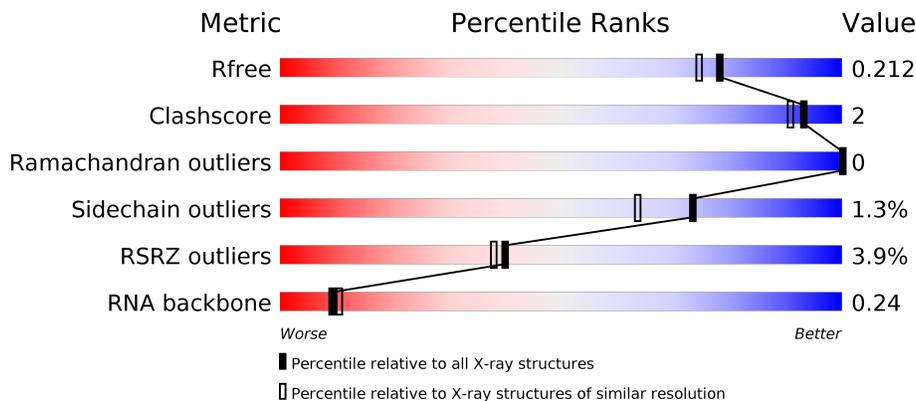
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	100719	1923 (1.86-1.86)
Clashscore	112137	2083 (1.86-1.86)
Ramachandran outliers	110173	2060 (1.86-1.86)
Sidechain outliers	110143	2060 (1.86-1.86)
RSRZ outliers	101464	1932 (1.86-1.86)
RNA backbone	2435	1018 (2.60-1.12)

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. 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 <=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	757	
2	B	10	
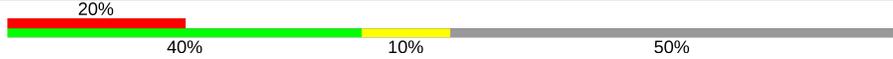
The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	UNX	A	809	-	-	-	X
3	UNX	A	810	-	-	-	X
3	UNX	A	811	-	-	-	X
3	UNX	A	812	-	-	-	X
3	UNX	A	813	-	-	-	X
3	UNX	A	814	-	-	-	X
3	UNX	A	816	-	-	-	X
3	UNX	A	818	-	-	-	X

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 5735 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 Gem-associated protein 5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	667	Total 5247	C 3360	N 890	O 967	S 30	0	19	0

There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-17	MET	-	expression tag	UNP Q8TEQ6
A	-16	HIS	-	expression tag	UNP Q8TEQ6
A	-15	HIS	-	expression tag	UNP Q8TEQ6
A	-14	HIS	-	expression tag	UNP Q8TEQ6
A	-13	HIS	-	expression tag	UNP Q8TEQ6
A	-12	HIS	-	expression tag	UNP Q8TEQ6
A	-11	HIS	-	expression tag	UNP Q8TEQ6
A	-10	SER	-	expression tag	UNP Q8TEQ6
A	-9	SER	-	expression tag	UNP Q8TEQ6
A	-8	GLY	-	expression tag	UNP Q8TEQ6
A	-7	ARG	-	expression tag	UNP Q8TEQ6
A	-6	GLU	-	expression tag	UNP Q8TEQ6
A	-5	ASN	-	expression tag	UNP Q8TEQ6
A	-4	LEU	-	expression tag	UNP Q8TEQ6
A	-3	TYR	-	expression tag	UNP Q8TEQ6
A	-2	PHE	-	expression tag	UNP Q8TEQ6
A	-1	GLN	-	expression tag	UNP Q8TEQ6
A	0	GLY	-	expression tag	UNP Q8TEQ6
A	682	GLN	ARG	natural variant	UNP Q8TEQ6

- Molecule 2 is a RNA chain called RNA (5'-R(*A*AP*UP*UP*UP*UP*UP*GP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	5	Total 82	C 37	N 11	O 30	P 4	0	0	0

WORLDWIDE
wwPDB
PROTEIN DATA BANK

• Molecule 3 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	22	Total X 22 22		0	0

• Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	376	Total O 378 378		0	2
4	B	5	Total O 6 6		0	2

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: Gem-associated protein 5

Chain A:
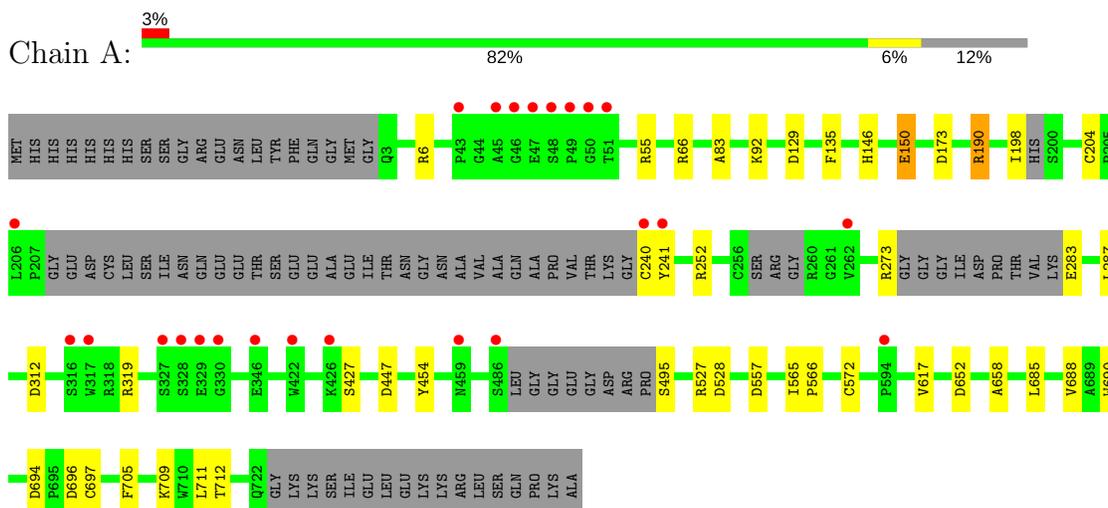
• Molecule 2: RNA (5'-R(*A*AP*UP*UP*UP*UP*UP*GP*AP*G)-3')

Chain B:
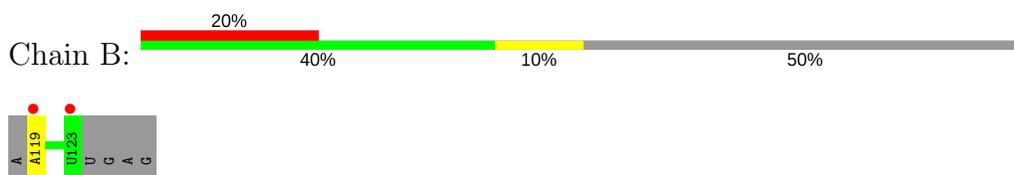
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	57.85Å 124.52Å 60.60Å 90.00° 116.77° 90.00°	Depositor
Resolution (Å)	28.90 – 1.85 28.16 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.9 (28.90-1.85) 100.0 (28.16-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.01 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.164 , 0.207 0.175 , 0.212	Depositor DCC
R_{free} test set	3294 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	20.7	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 46.6	EDS
L-test for twinning [2]	$<|L|> = 0.48, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	0.033 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	5735	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.29% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<|L|>$, $<L^2>$ 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 ⓘ

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

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	1.01	2/5461 (0.0%)	0.98	17/7451 (0.2%)
2	B	0.66	0/90	0.96	0/139
All	All	1.00	2/5551 (0.0%)	0.98	17/7590 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	283	GLU	CD-OE1	5.27	1.31	1.25
1	A	150	GLU	CD-OE1	-5.19	1.20	1.25

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	66	ARG	NE-CZ-NH2	-7.11	116.74	120.30
1	A	652	ASP	CB-CG-OD1	6.89	124.50	118.30
1	A	129	ASP	CB-CG-OD2	-6.59	112.37	118.30
1	A	528	ASP	CB-CG-OD1	6.59	124.23	118.30
1	A	319	ARG	NE-CZ-NH1	6.38	123.49	120.30

There are no chirality outliers.

There are no planarity outliers.

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	5247	0	5042	17	0
2	B	82	0	39	1	0
3	A	22	0	0	0	0
4	A	378	0	0	0	0
4	B	6	0	0	1	0
All	All	5735	0	5081	18	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 18 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:697[A]:CYS:SG	1:A:709:LYS:NZ	2.49	0.85
1:A:697[B]:CYS:SG	1:A:709:LYS:NZ	2.74	0.60
2:B:119:A:N6	4:B:201:HOH:O	2.38	0.55
1:A:694:ASP:OD2	1:A:697[B]:CYS:SG	2.50	0.55
1:A:565:ILE:HB	1:A:566:PRO:HA	1.91	0.53

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	673/757 (89%)	650 (97%)	23 (3%)	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	577/657 (88%)	569 (99%)	8 (1%)	71	61

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	454	TYR
1	A	705	PHE
1	A	690[A]	TRP
1	A	287	LEU
1	A	495	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3   RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	3/10 (30%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

5.4   Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5   Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6   Ligand geometry ⓘ

Of 22 ligands modelled in this entry, 22 are unknown - 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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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	667/757 (88%)	-0.23	24 (3%)	43	41	12, 21, 46, 77	0
2	B	5/10 (50%)	1.42	2 (40%)	0	0	32, 36, 46, 49	5 (100%)
All	All	672/767 (87%)	-0.21	26 (3%)	40	38	12, 21, 46, 77	5 (0%)

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	317	TRP	8.4
1	A	49	PRO	4.9
1	A	594	PRO	4.4
1	A	48	SER	4.2
1	A	327	SER	3.9

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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
3	UNX	A	812	1/1	0.93	0.40	16.78	25,25,25,25	0
3	UNX	A	813	1/1	0.84	0.41	14.61	28,28,28,28	0
3	UNX	A	818	1/1	0.92	0.31	7.91	43,43,43,43	0
3	UNX	A	809	1/1	0.94	0.24	7.51	22,22,22,22	0
3	UNX	A	811	1/1	0.98	0.26	7.48	15,15,15,15	0
3	UNX	A	814	1/1	0.85	0.23	6.64	23,23,23,23	0
3	UNX	A	810	1/1	0.90	0.21	6.32	28,28,28,28	0
3	UNX	A	816	1/1	0.84	0.14	2.04	36,36,36,36	0
3	UNX	A	808	1/1	0.97	0.18	1.99	25,25,25,25	0
3	UNX	A	804	1/1	0.98	0.10	1.18	21,21,21,21	0
3	UNX	A	801	1/1	0.99	0.08	0.12	12,12,12,12	0
3	UNX	A	802	1/1	0.81	0.29	-	20,20,20,20	0
3	UNX	A	822	1/1	0.98	0.08	-	18,18,18,18	0
3	UNX	A	821	1/1	0.95	0.12	-	28,28,28,28	0
3	UNX	A	815	1/1	0.94	0.13	-	34,34,34,34	0
3	UNX	A	817	1/1	0.81	0.16	-	40,40,40,40	0
3	UNX	A	806	1/1	0.67	0.29	-	21,21,21,21	0
3	UNX	A	820	1/1	0.96	0.14	-	24,24,24,24	0
3	UNX	A	803	1/1	0.77	0.25	-	21,21,21,21	0
3	UNX	A	819	1/1	0.99	0.19	-	25,25,25,25	0
3	UNX	A	807	1/1	0.98	0.11	-	9,9,9,9	0
3	UNX	A	805	1/1	0.86	0.11	-	24,24,24,24	0

6.5 Other polymers ⓘ

There are no such residues in this entry.