wwPDB X-ray Structure Validation Summary Report (i)

Oct 5, 2023 – 04:11 AM EDT

PDB ID	:	6VGG
Title	:	Crystal structure of the DNA binding domains of human transcription factor ERG, human Runx2 bound to core binding factor beta (Cbfb), and mithramycin, in complex with 16mer DNA CAGAGGATGTGGCTTC
Authors	:	Hou, C.; Rohr, J.; Tsodikov, O.V.
Deposited on	:	2020-01-08
Resolution	:	4.31 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	FAILED
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 4.31 Å.

There are no overall percentile quality scores available for this entry.

MolProbity and EDS failed to run properly - the sequence quality summary graphics cannot be shown.

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 3515 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 Transcriptional regulator ERG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	98	806	512	143	146	5	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	302	GLY	-	expression tag	UNP P11308
A	303	PRO	-	expression tag	UNP P11308
A	304	HIS	-	expression tag	UNP P11308
A	305	MET	-	expression tag	UNP P11308
A	420	SER	-	expression tag	UNP P11308
A	421	TYR	-	expression tag	UNP P11308
A	422	HIS	-	expression tag	UNP P11308
A	423	ALA	-	expression tag	UNP P11308
A	424	HIS	-	expression tag	UNP P11308
A	425	PRO	-	expression tag	UNP P11308
A	426	GLN	-	expression tag	UNP P11308
A	427	LYS	-	expression tag	UNP P11308
A	428	MET	-	expression tag	UNP P11308
A	429	ASN	-	expression tag	UNP P11308

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	B	16	332	157	62	97	16	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	16	324	154	59	95	16	0	0	0

- Molecule 4 is a protein called Runt-related transcription factor 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	117	901	567	164	166	4	0	0	0

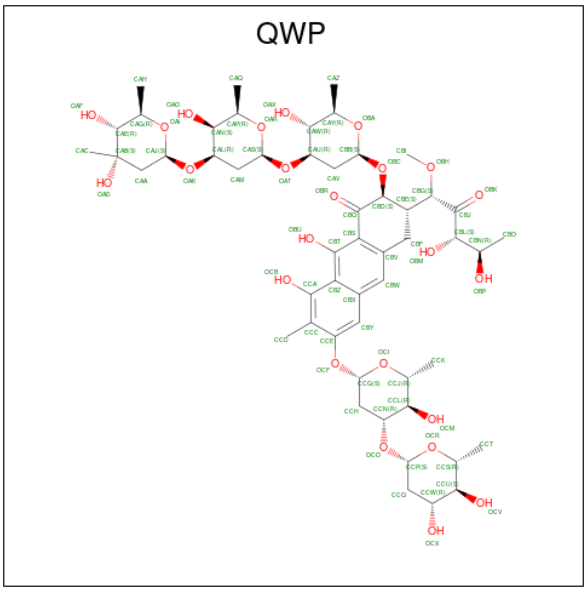
- Molecule 5 is a protein called Core-binding factor subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	G	121	998	622	181	189	6	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-13	MET	-	expression tag	UNP Q13951
G	-12	GLY	-	expression tag	UNP Q13951
G	-11	SER	-	expression tag	UNP Q13951
G	-10	SER	-	expression tag	UNP Q13951
G	-9	HIS	-	expression tag	UNP Q13951
G	-8	HIS	-	expression tag	UNP Q13951
G	-7	HIS	-	expression tag	UNP Q13951
G	-6	HIS	-	expression tag	UNP Q13951
G	-5	HIS	-	expression tag	UNP Q13951
G	-4	HIS	-	expression tag	UNP Q13951
G	-3	SER	-	expression tag	UNP Q13951
G	-2	GLN	-	expression tag	UNP Q13951
G	-1	ASP	-	expression tag	UNP Q13951
G	0	PRO	-	expression tag	UNP Q13951

- Molecule 6 is mithramycin (three-letter code: QWP) (formula: $C_{52}H_{76}O_{24}$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
6	B	1	76	52	24	0	0
			Total	C	O		
6	C	1	75	52	23	0	0

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mg		
7	B	1	1	1	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	B	1	1	1	0	0
			Total	O		
8	C	1	1	1	0	0

MolProbity and EDS failed to run properly - this section is therefore empty.

3 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, α, β, γ	104.68Å 104.68Å 322.84Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	35.00 – 4.31	Depositor
% Data completeness (in resolution range)	98.1 (35.00-4.31)	Depositor
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.28 (at 4.28Å)	Xtriage
Refinement program	REFMAC 5.8.0238	Depositor
R, R_{free}	0.251 , 0.285	Depositor
Wilson B-factor (Å^2)	195.7	Xtriage
Anisotropy	0.592	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	3515	wwPDB-VP
Average B, all atoms (Å^2)	308.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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.

4 Model quality (i)

4.1 Standard geometry (i)

MolProbity failed to run properly - this section is therefore empty.

4.2 Too-close contacts (i)

MolProbity failed to run properly - this section is therefore empty.

4.3 Torsion angles (i)

4.3.1 Protein backbone (i)

MolProbity failed to run properly - this section is therefore empty.

4.3.2 Protein sidechains (i)

MolProbity failed to run properly - this section is therefore empty.

4.3.3 RNA (i)

MolProbity failed to run properly - this section is therefore empty.

4.4 Non-standard residues in protein, DNA, RNA chains (i)

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

4.5 Carbohydrates (i)

There are no monosaccharides in this entry.

4.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
6	QWP	C	101	7	80,82,83	1.32	4 (5%)	110,125,127	1.12	8 (7%)
6	QWP	B	101	7	81,83,83	1.29	4 (4%)	112,127,127	1.04	5 (4%)

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
6	QWP	C	101	7	-	12/38/134/137	0/8/8/8
6	QWP	B	101	7	-	3/38/137/137	0/8/8/8

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	B	101	QWP	CBF-CBV	-5.78	1.40	1.51
6	C	101	QWP	CBF-CBV	-5.68	1.40	1.51
6	B	101	QWP	CCD-CCC	-5.49	1.40	1.51
6	C	101	QWP	CCD-CCC	-5.49	1.40	1.51
6	C	101	QWP	CBD-CBQ	-3.68	1.40	1.50

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	B	101	QWP	OCF-CCE-CCC	3.52	117.56	114.19
6	B	101	QWP	CAB-CAA-CAJ	-3.37	108.35	114.82
6	C	101	QWP	OCF-CCE-CCC	3.04	117.10	114.19
6	C	101	QWP	CAB-CAA-CAJ	-2.95	109.17	114.82
6	C	101	QWP	CAH-CAG-CAE	-2.79	107.81	112.57

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	B	101	QWP	OBK-CBJ-CBL-OBM
6	C	101	QWP	CBE-CBG-OBH-CBI
6	C	101	QWP	CBJ-CBG-OBH-CBI

Continued on next page...

Continued from previous page...

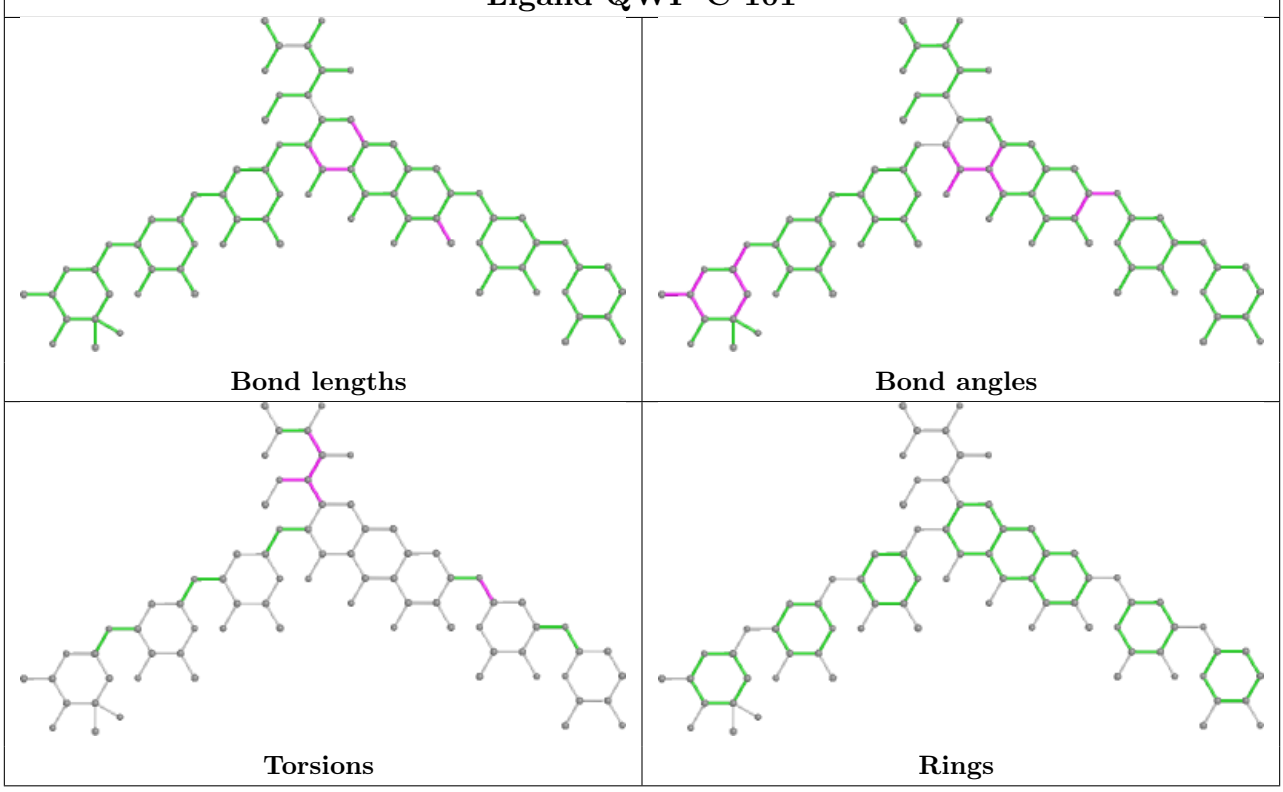
Mol	Chain	Res	Type	Atoms
6	C	101	QWP	CBG-CBJ-CBL-OBM
6	C	101	QWP	CBG-CBJ-CBL-CBN

There are no ring outliers.

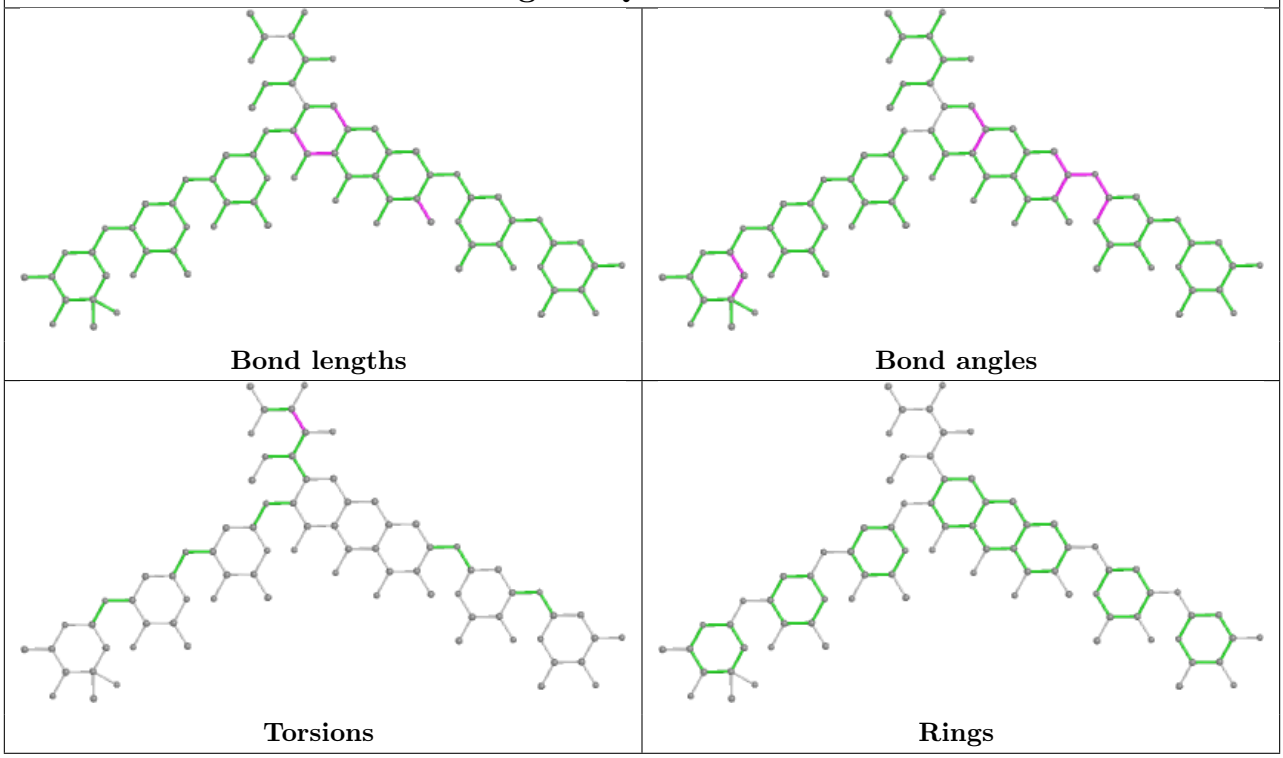
No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand QWP C 101



Ligand QWP B 101



4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data (i)

5.1 Protein, DNA and RNA chains (i)

EDS failed to run properly - this section is therefore empty.

5.2 Non-standard residues in protein, DNA, RNA chains (i)

EDS failed to run properly - this section is therefore empty.

5.3 Carbohydrates (i)

EDS failed to run properly - this section is therefore empty.

5.4 Ligands (i)

EDS failed to run properly - this section is therefore empty.

5.5 Other polymers (i)

EDS failed to run properly - this section is therefore empty.