wwPDB X-ray Structure Validation Summary Report (i)

Sep 24, 2023 – 09:28 PM EDT

PDB ID	:	5V0L
Title	:	Crystal structure of the AHR-ARNT heterodimer in complex with the DRE
Authors	:	Seok, S.-H.; Lee, W.; Jiang, L.; Bradfield, C.A.; Xing, Y.
Deposited on	:	2017-02-28
Resolution	:	4.00 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 (i)) were used in the production of this report:

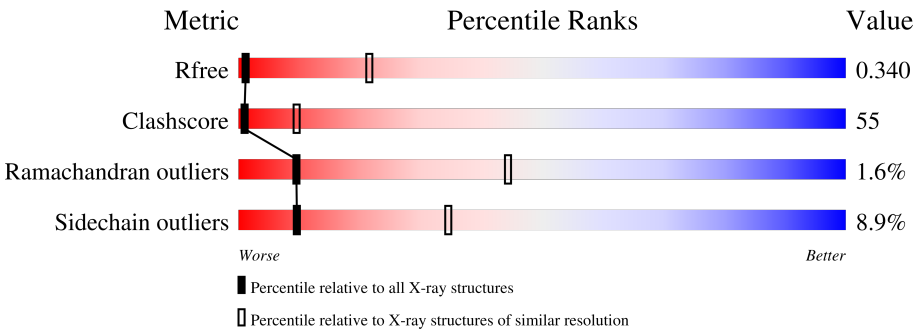
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 4.00 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	1087 (4.30-3.70)
Clashscore	141614	1148 (4.30-3.70)
Ramachandran outliers	138981	1108 (4.30-3.70)
Sidechain outliers	138945	1099 (4.30-3.70)

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 of 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	279	14% 43% 5% 37%
2	B	241	21% 43% 5% 31%
3	C	17	100%
4	D	14	7% 93%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 3418 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 Aryl hydrocarbon receptor nuclear translocator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	176	1391	865	252	263	11	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	68	GLY	-	expression tag	UNP P27540
A	69	SER	-	expression tag	UNP P27540

- Molecule 2 is a protein called Aryl hydrocarbon receptor.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	166	1347	866	232	245	4	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	27	GLY	-	expression tag	UNP P30561
B	28	SER	-	expression tag	UNP P30561

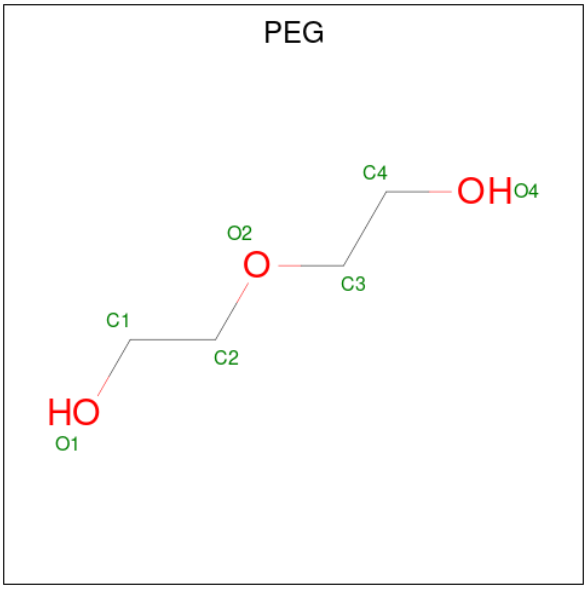
- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*GP*AP*TP*TP*GP*CP*GP*TP*GP*AP*GP*AP*AP*CP*TP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	C	17	356	168	69	102	17	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(P*AP*GP*TP*TP*CP*TP*CP*AP*CP*GP*CP*AP*AP*T)-3').

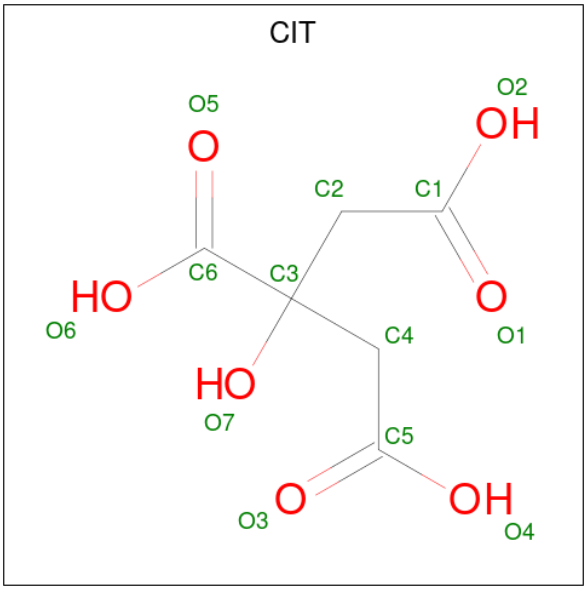
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	D	14	284	136	50	84	14	0	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	A	1	7	4	3	0	0
5	B	1	7	4	3	0	0

- Molecule 6 is CITRIC ACID (three-letter code: CIT) (formula: $C_6H_8O_7$).



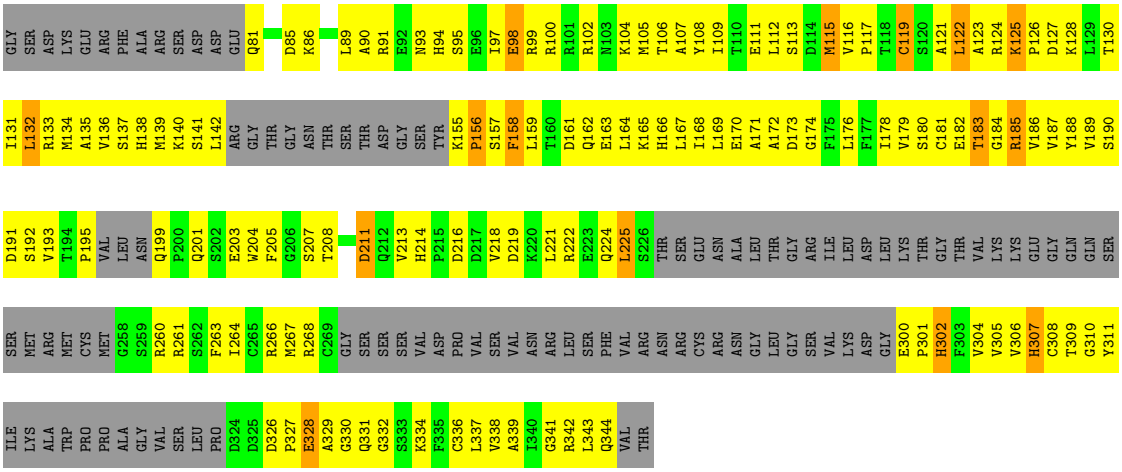
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
6	B	1	13	6	7	0	0
			Total	C	O		
6	C	1	13	6	7	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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.

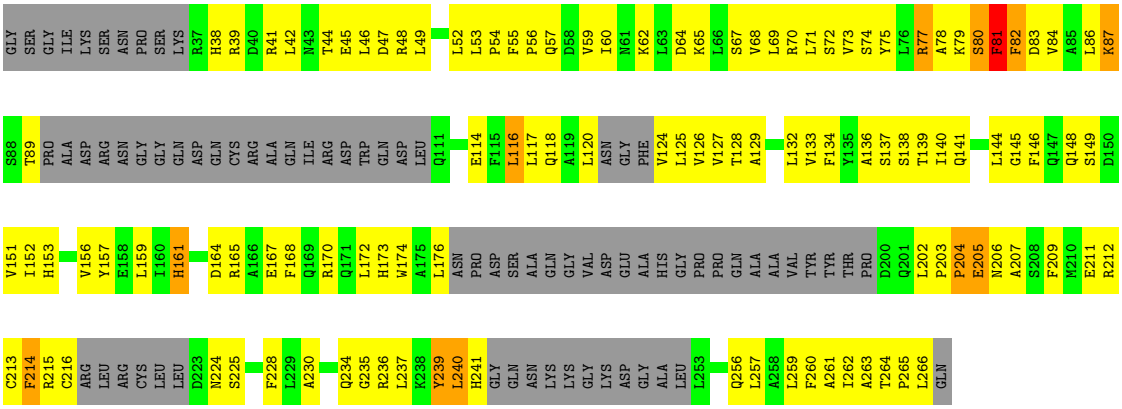
• Molecule 1: Aryl hydrocarbon receptor nuclear translocator

Chain A: 14% 43% 5% 37%



• Molecule 2: Aryl hydrocarbon receptor

Chain B: 21% 43% 5% 31%



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

Chain C: 100%

G4 G5 A6 T7 T8 G9 C10 G11 T12 G13 A14 G15 A16 A17 C18 T19 G20

• Molecule 4: DNA (5'-D(P*AP*GP*TP*TP*CP*TP*CP*AP*CP*GP*CP*AP*AP*T)-3')

Chain D: 7% 93%

A4 G5 T6 T7 C8 T9 C10 A11 C12 G13 C14 A15 A16 T17

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	78.20Å 64.36Å 157.72Å 90.00° 100.14° 90.00°	Depositor
Resolution (Å)	39.69 – 4.00 39.68 – 4.00	Depositor EDS
% Data completeness (in resolution range)	99.5 (39.69-4.00) 99.5 (39.68-4.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.10 (at 4.00Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.285 , 0.323 0.315 , 0.340	Depositor DCC
R_{free} test set	332 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	194.2	Xtriage
Anisotropy	0.490	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 235.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	3418	wwPDB-VP
Average B, all atoms (Å^2)	221.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.66% 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 (i)

5.1 Standard geometry (i)

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

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	0.32	0/1412	0.61	5/1896 (0.3%)
2	B	0.27	0/1372	0.51	2/1852 (0.1%)
3	C	0.45	0/400	0.76	0/617
4	D	0.44	0/317	0.87	0/486
All	All	0.33	0/3501	0.63	7/4851 (0.1%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	191	ASP	O-C-N	-8.66	108.85	122.70
2	B	81	PHE	CB-CA-C	7.69	125.78	110.40
1	A	191	ASP	CA-C-N	6.23	130.90	117.20
1	A	125	LYS	C-N-CD	5.52	139.98	128.40
1	A	191	ASP	C-N-CA	5.51	135.49	121.70

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	1391	0	1370	209	2
2	B	1347	0	1337	148	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	356	0	192	19	0
4	D	284	0	159	17	1
5	A	7	0	10	0	0
5	B	7	0	10	0	0
6	B	13	0	5	0	0
6	C	13	0	5	0	0
All	All	3418	0	3088	359	2

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 55.

The worst 5 of 359 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:180:SER:CB	1:A:183:THR:CG2	2.05	1.34
1:A:86:LYS:HA	1:A:90:ALA:CB	1.58	1.34
1:A:180:SER:HB3	1:A:183:THR:CG2	1.62	1.30
1:A:86:LYS:O	1:A:90:ALA:HB3	1.32	1.26
1:A:331:GLN:NE2	1:A:332:GLY:O	1.68	1.25

All (2) 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:328:GLU:OE1	2:B:239:TYR:OH[3_545]	1.76	0.44
1:A:81:GLN:N	4:D:17:DT:OP2[4_445]	2.06	0.14

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	164/279 (59%)	131 (80%)	29 (18%)	4 (2%)	6 36
2	B	154/241 (64%)	138 (90%)	15 (10%)	1 (1%)	25 63
All	All	318/520 (61%)	269 (85%)	44 (14%)	5 (2%)	9 44

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	123	ALA
1	A	156	PRO
1	A	327	PRO
2	B	204	PRO
1	A	330	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
1	A	156/245 (64%)	142 (91%)	14 (9%)	9 34
2	B	148/206 (72%)	135 (91%)	13 (9%)	10 35
All	All	304/451 (67%)	277 (91%)	27 (9%)	9 34

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	77	ARG
2	B	86	LEU
2	B	214	PHE
2	B	81	PHE
2	B	87	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	166	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	307	HIS
1	A	344	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

4 ligands are 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	$\#\|Z\| > 2$
6	CIT	B	301	-	12,12,12	1.02	0	17,17,17	1.55	1 (5%)
5	PEG	B	302	-	6,6,6	0.49	0	5,5,5	0.26	0
5	PEG	A	401	-	6,6,6	0.48	0	5,5,5	0.28	0
6	CIT	C	101	-	12,12,12	1.02	0	17,17,17	1.56	2 (11%)

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	CIT	B	301	-	-	6/16/16/16	-
5	PEG	B	302	-	-	0/4/4/4	-
5	PEG	A	401	-	-	0/4/4/4	-
6	CIT	C	101	-	-	7/16/16/16	-

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	B	301	CIT	O6-C6-C3	4.16	120.27	113.05
6	C	101	CIT	O6-C6-C3	4.11	120.19	113.05
6	C	101	CIT	O2-C1-C2	2.04	120.91	114.35

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	B	301	CIT	C2-C3-C6-O5
6	B	301	CIT	C2-C3-C6-O6
6	B	301	CIT	O7-C3-C6-O5
6	B	301	CIT	O7-C3-C6-O6
6	C	101	CIT	C2-C3-C6-O5

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)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.