



wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 06:27 am BST

PDB ID : 4ZPR
Title : Crystal Structure of the Heterodimeric HIF-1a:ARNT Complex with HRE DNA
Authors : Wu, D.; Potluri, N.; Lu, J.; Kim, Y.; Rastinejad, F.
Deposited on : 2015-05-08
Resolution : 3.90 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	:	2.11
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)	:	2.11

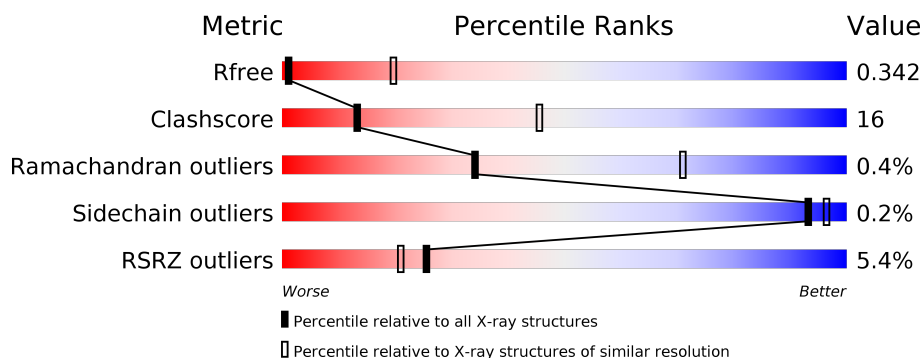
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.90 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	1002 (4.14-3.66)
Clashscore	141614	1004 (4.12-3.68)
Ramachandran outliers	138981	1021 (4.14-3.66)
Sidechain outliers	138945	1014 (4.14-3.66)
RSRZ outliers	127900	1275 (4.20-3.60)

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	384	 2% 40% 22% 39%
2	B	345	 6% 49% 30% 19%
3	C	21	 5% 48% 52%
4	D	21	 5% 71% 29%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5042 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
1	A	235	Total	C	N	O	S	0	0	0
			1915	1207	341	354	13			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	81	MET	-	initiating methionine	UNP P53762

- Molecule 2 is a protein called Hypoxia-inducible factor 1-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	279	Total	C	N	O	S	0	0	0
			2272	1434	397	420	21			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	21	Total	C	N	O	P	0	0	0
			433	205	80	128	20			

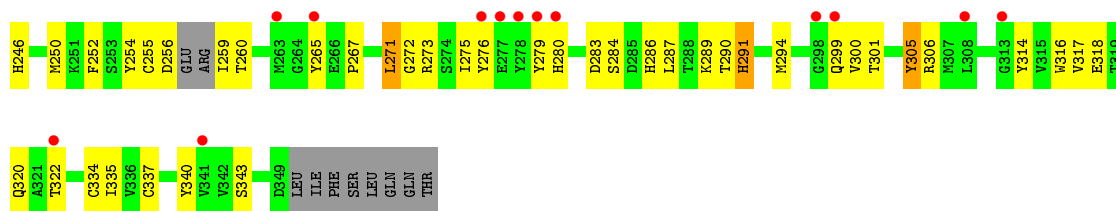
- Molecule 4 is a DNA chain called DNA (5'-D(*CP*AP*CP*GP*AP*CP*CP*CP*GP*CP*AP*CP*GP*TP*AP*CP*GP*CP*AP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	21	Total	C	N	O	P	0	0	0
			422	200	82	120	20			

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 ($\text{RSRZ} > 2$). 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.

- Chain A:
-
- Sequence logo for Chain A. The y-axis represents information content in bits (0.00 to 1.46). The x-axis shows positions 1 to 100. The color scale at the top indicates conservation levels: 2% (red), 40% (green), 22% (yellow), and 39% (grey). Red dots above the sequence indicate specific residues of interest.
- | Position | Amino Acid | Information Content (bits) |
|----------|------------|----------------------------|
| 1 | I468 | 0.10 |
| 2 | V464 | 0.10 |
| 3 | D877 | 0.10 |
| 4 | Q387 | 0.10 |
| 5 | P388 | 0.10 |
| 6 | Q389 | 0.10 |
| 7 | E390 | 0.10 |
| 8 | L391 | 0.10 |
| 9 | L392 | 0.10 |
| 10 | N395 | 0.10 |
| 11 | I396 | 0.10 |
| 12 | F399 | 0.10 |
| 13 | C400 | 0.10 |
| 14 | Q406 | 0.10 |
| 15 | L407 | 0.10 |
| 16 | L408 | 0.10 |
| 17 | R409 | 0.10 |
| 18 | D410 | 0.10 |
| 19 | S411 | 0.10 |
| 20 | F412 | 0.10 |
| 21 | V415 | 0.10 |
| 22 | V416 | 0.10 |
| 23 | K417 | 0.10 |
| 24 | L418 | 0.10 |
| 25 | Q421 | 0.10 |
| 26 | V422 | 0.10 |
| 27 | L423 | 0.10 |
| 28 | S424 | 0.10 |
| 29 | V425 | 0.10 |
| 30 | M426 | 0.10 |
| 31 | F427 | 0.10 |
| 32 | ARG | 0.10 |
| 33 | PHE | 0.10 |
| 34 | ARG | 0.10 |
| 35 | SER | 0.10 |
| 36 | LYS | 0.10 |
| 37 | THR | 0.10 |
| 38 | ARG | 0.10 |
| 39 | GLU | 0.10 |
| 40 | TRP | 0.10 |
| 41 | LEU | 0.10 |
| 42 | W438 | 0.10 |
| 43 | M439 | 0.10 |
| 44 | R440 | 0.10 |
| 45 | F444 | 0.10 |
| 46 | T445 | 0.10 |
| 47 | F446 | 0.10 |
| 48 | Q447 | 0.10 |
| 49 | N448 | 0.10 |
| 50 | P449 | 0.10 |
| 51 | D452 | 0.10 |
| 52 | E453 | 0.10 |
| 53 | I454 | 0.10 |
| 54 | ALA | 0.10 |
| 55 | TRP | 0.10 |
| 56 | PRO | 0.10 |
| 57 | ALA | 0.10 |
| 58 | GLY | 0.10 |
| 59 | VAL | 0.10 |
| 60 | SER | 0.10 |
| 61 | LEU | 0.10 |
| 62 | R260 | 0.10 |
| 63 | F263 | 0.10 |
| 64 | L264 | 0.10 |
| 65 | R266 | 0.10 |
| 66 | C265 | 0.10 |
| 67 | M267 | 0.10 |
| 68 | R268 | 0.10 |
| 69 | C269 | 0.10 |
| 70 | GLY | 0.10 |
| 71 | THR | 0.10 |
| 72 | SER | 0.10 |
| 73 | LYS | 0.10 |
| 74 | VAL | 0.10 |
| 75 | ASP | 0.10 |
| 76 | CYS | 0.10 |
| 77 | PRO | 0.10 |
| 78 | VAL | 0.10 |
| 79 | SER | 0.10 |
| 80 | ASN | 0.10 |
| 81 | ARG | 0.10 |
| 82 | CYS | 0.10 |
| 83 | ASP | 0.10 |
| 84 | MET | 0.10 |
| 85 | ASN | 0.10 |
| 86 | ARG | 0.10 |
| 87 | GLY | 0.10 |
| 88 | LEU | 0.10 |
| 89 | GLY | 0.10 |
| 90 | SER | 0.10 |
| 91 | VAL | 0.10 |
| 92 | LYS | 0.10 |
| 93 | GLY | 0.10 |
| 94 | THR | 0.10 |
| 95 | ASP | 0.10 |
| 96 | LEU | 0.10 |
| 97 | GLY | 0.10 |
| 98 | PRO | 0.10 |
| 99 | PHE | 0.10 |
| 100 | VAL | 0.10 |

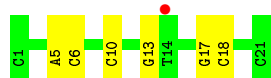
- Chain B:
-
- | Amino Acid | Percentage |
|------------|------------|
| MET | 6% |
| SER | 49% |
| S15 | 49% |
| R13 | 49% |
| K21 | 49% |
| S28 | 49% |
| R29 | 49% |
| R30 | 49% |
| S31 | 49% |
| V36 | 49% |
| F37 | 49% |
| L40 | 49% |
| P45 | 49% |
| L46 | 49% |
| P47 | 49% |
| H46 | 49% |
| N49 | 49% |
| V50 | 49% |
| S51 | 49% |
| S52 | 49% |
| D55 | 49% |
| K56 | 49% |
| A57 | 49% |
| S58 | 49% |
| V59 | 49% |
| M60 | 49% |
| R61 | 49% |
| L62 | 49% |
| S65 | 49% |
| R68 | 49% |
| V69 | 49% |
| K71 | 49% |
| L72 | 49% |
| L73 | 49% |
| ASP | 30% |
| ALA | 30% |
| GLY | 30% |
| LEU | 30% |
| SER | 30% |
| ASP | 30% |
| GLU | 30% |
| ASP | 30% |
| MET | 30% |
| LYS | 30% |
| ALA | 30% |
| C1N | 30% |



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



- Molecule 4: DNA (5'-D(*CP*AP*CP*GP*AP*CP*CP*CP*GP*CP*AP*CP*GP*TP*AP*CP*GP*CP*AP*GP*C)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α , β , γ	66.73Å 66.73Å 243.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.84 – 3.90 46.33 – 3.90	Depositor EDS
% Data completeness (in resolution range)	99.3 (29.84-3.90) 99.2 (46.33-3.90)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.82 (at 3.88Å)	Xtriage
Refinement program	PHENIX, CORELS	Depositor
R, R_{free}	0.265 , 0.341 0.273 , 0.342	Depositor DCC
R_{free} test set	963 reflections (10.05%)	wwPDB-VP
Wilson B-factor (Å ²)	205.0	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 146.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.058 for h,-k,-l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	5042	wwPDB-VP
Average B, all atoms (Å ²)	194.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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](#)

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	0/1945	0.53	1/2616 (0.0%)
2	B	0.30	0/2318	0.54	1/3120 (0.0%)
3	C	0.62	0/485	0.93	0/749
4	D	0.67	0/473	0.80	0/726
All	All	0.39	0/5221	0.62	2/7211 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	305	TYR	CB-CA-C	-7.51	95.37	110.40
1	A	125	LYS	N-CA-C	7.05	130.03	111.00

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	1915	0	1895	65	0
2	B	2272	0	2230	91	0
3	C	433	0	238	9	0
4	D	422	0	234	4	0
All	All	5042	0	4597	154	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 16.

The worst 5 of 154 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:214:HIS:CD2	1:A:215:PRO:HD2	1.57	1.38
1:A:105:MET:O	1:A:109:ILE:HG12	1.35	1.23
2:B:252:PHE:HB2	2:B:271:LEU:O	1.42	1.19
1:A:214:HIS:HD2	1:A:215:PRO:HD2	1.15	0.93
2:B:119:TYR:CE1	2:B:226:PRO:HG3	2.06	0.90

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	217/384 (56%)	183 (84%)	34 (16%)	0	100	100
2	B	265/345 (77%)	213 (80%)	50 (19%)	2 (1%)	19	57
All	All	482/729 (66%)	396 (82%)	84 (17%)	2 (0%)	34	71

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	120	MET
2	B	271	LEU

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	217/347 (62%)	217 (100%)	0	100	100
2	B	258/316 (82%)	257 (100%)	1 (0%)	91	94
All	All	475/663 (72%)	474 (100%)	1 (0%)	93	96

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	291	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	344	GLN
2	B	229	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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 ⓘ

There are no ligands in this entry.

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 [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	235/384 (61%)	0.23	8 (3%) 45 35	176, 191, 214, 222	0
2	B	279/345 (80%)	0.39	20 (7%) 15 11	175, 192, 210, 231	0
3	C	21/21 (100%)	0.02	1 (4%) 30 25	178, 197, 219, 225	0
4	D	21/21 (100%)	0.04	1 (4%) 30 25	186, 201, 212, 221	0
All	All	556/771 (72%)	0.30	30 (5%) 25 21	175, 192, 213, 231	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	299	GLN	5.2
2	B	308	LEU	5.1
1	A	94	HIS	4.4
2	B	169	LEU	4.3
2	B	101	MET	4.2

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 carbohydrates in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.