



wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 12:33 am BST

PDB ID : 6OAJ
Title : HUaE34K 19bp SYM DNA
Authors : Remesh, S.G.; Hammel, M.
Deposited on : 2019-03-16
Resolution : 4.09 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

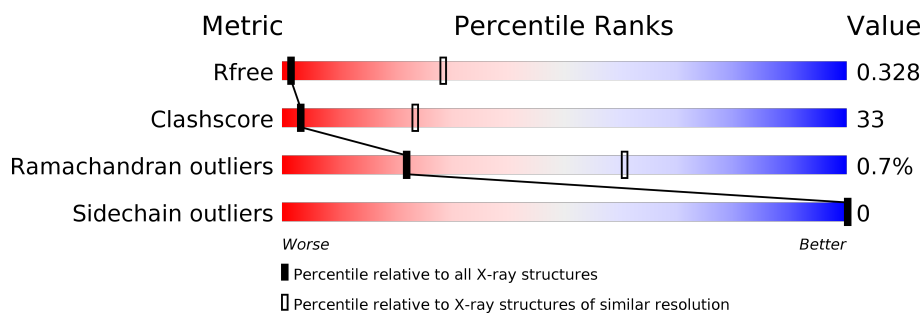
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.09 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	1193 (4.50-3.70)
Clashscore	141614	1003 (4.44-3.76)
Ramachandran outliers	138981	1005 (4.48-3.72)
Sidechain outliers	138945	1199 (4.50-3.70)

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 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	90	<div> <div>50%</div> <div>29%</div> <div>21%</div> </div>
1	B	90	<div> <div>49%</div> <div>37%</div> <div>• 12%</div> </div>
1	C	90	<div> <div>61%</div> <div>22%</div> <div>• 16%</div> </div>
1	D	90	<div> <div>63%</div> <div>14%</div> <div>22%</div> </div>
2	K	19	<div> <div>5%</div> <div>79%</div> <div>16%</div> </div>
3	L	20	<div> <div>10%</div> <div>65%</div> <div>25%</div> </div>

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2925 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 DNA-binding protein HU-alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	71	Total	C	N	O	S	0	0	0
			522	332	90	99	1			
1	B	79	Total	C	N	O	S	0	0	0
			545	345	92	107	1			
1	C	76	Total	C	N	O	S	0	0	0
			556	351	96	108	1			
1	D	70	Total	C	N	O	S	0	0	0
			504	322	83	98	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	34	LYS	GLU	engineered mutation	UNP P0ACF0
B	34	LYS	GLU	engineered mutation	UNP P0ACF0
C	34	LYS	GLU	engineered mutation	UNP P0ACF0
D	34	LYS	GLU	engineered mutation	UNP P0ACF0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	K	19	Total	C	N	O	P	0	0	0
			389	184	71	115	19			

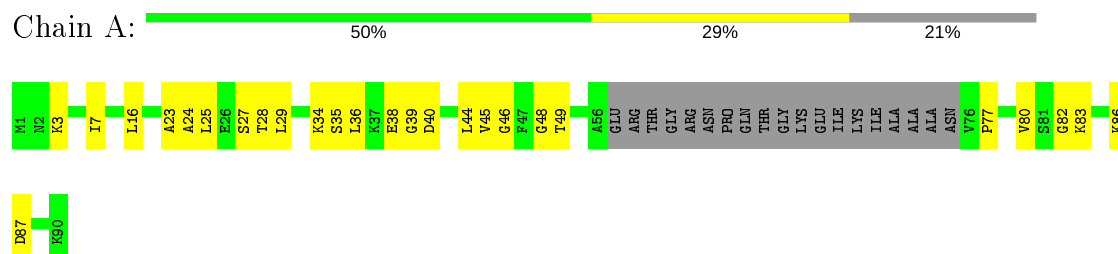
- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*CP*GP*CP*GP*TP*GP*CP*CP*AP*AP*TP*TP*GP*AP*AP*CP*CP*GP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	L	20	Total	C	N	O	P	0	0	0
			409	193	77	119	20			

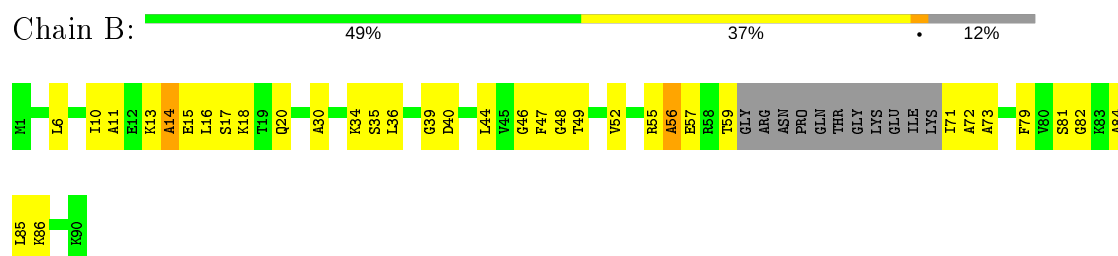
3 Residue-property plots [i](#)

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. 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.

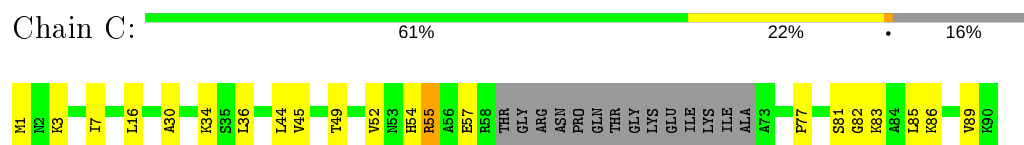
• Molecule 1: DNA-binding protein HU-alpha



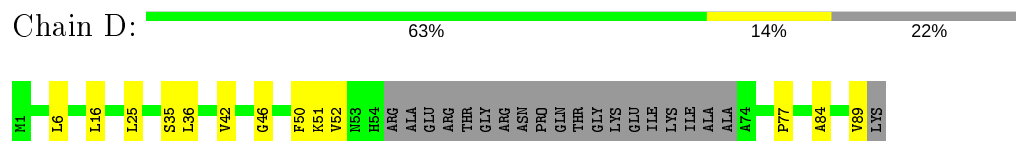
• Molecule 1: DNA-binding protein HU-alpha



• Molecule 1: DNA-binding protein HU-alpha



• Molecule 1: DNA-binding protein HU-alpha

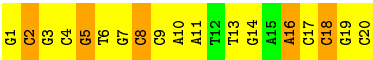
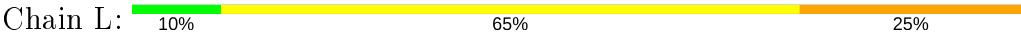


• Molecule 2: DNA (5'-D(P*CP*GP*GP*TP*TP*CP*AP*AP*TP*TP*GP*GP*CP*AP*CP*GP*CP*GP*C)-3')





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



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	64.60Å 86.35Å 91.69Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.49 – 4.09 45.84 – 4.09	Depositor EDS
% Data completeness (in resolution range)	99.2 (40.49-4.09) 89.1 (45.84-4.09)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.66 (at 4.14Å)	Xtriage
Refinement program	PHENIX (1.13 _2998: ???)	Depositor
R, R_{free}	0.312 , 0.329 0.314 , 0.328	Depositor DCC
R_{free} test set	438 reflections (10.11%)	wwPDB-VP
Wilson B-factor (Å ²)	180.6	Xtriage
Anisotropy	0.290	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 65.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	2925	wwPDB-VP
Average B, all atoms (Å ²)	90.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.47% 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

5.1 Standard geometry

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	0.29	0/525	0.51	0/704
1	B	0.30	0/548	0.54	1/741 (0.1%)
1	C	0.39	0/559	0.59	0/749
1	D	0.33	0/506	0.60	0/680
2	K	1.41	1/435 (0.2%)	1.34	5/669 (0.7%)
3	L	1.48	5/458 (1.1%)	1.28	3/704 (0.4%)
All	All	0.83	6/3031 (0.2%)	0.88	9/4247 (0.2%)

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	C	0	1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	K	10	DT	C1'-N1	9.50	1.61	1.49
3	L	18	DC	C1'-N1	6.33	1.57	1.49
3	L	18	DC	C3'-O3'	-6.32	1.35	1.44
3	L	16	DA	C3'-O3'	5.51	1.51	1.44
3	L	2	DC	C3'-O3'	-5.49	1.36	1.44

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	K	13	DC	O4'-C1'-N1	8.21	113.75	108.00
2	K	14	DA	O4'-C4'-C3'	-7.46	101.52	104.50
3	L	5	DG	O5'-P-OP2	6.93	119.02	110.70
3	L	8	DC	O5'-P-OP2	-6.61	99.75	105.70
3	L	5	DG	O4'-C4'-C3'	-6.48	101.91	104.50

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	55	ARG	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	522	0	555	44	0
1	B	545	0	541	64	0
1	C	556	0	581	24	0
1	D	504	0	530	17	0
2	K	389	0	214	46	0
3	L	409	0	224	40	0
All	All	2925	0	2645	178	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 33.

The worst 5 of 178 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:B:57:GLU:CA	1:B:72:ALA:O	1.67	1.41
1:B:59:THR:HA	1:B:71:ILE:N	1.06	1.36
1:B:59:THR:CA	1:B:71:ILE:N	1.88	1.33
1:B:30:ALA:HB1	1:B:34:LYS:NZ	1.52	1.22
1:B:30:ALA:O	1:B:34:LYS:HG2	1.05	1.19

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	67/90 (74%)	62 (92%)	5 (8%)	0	100	100
1	B	75/90 (83%)	71 (95%)	3 (4%)	1 (1%)	12	47
1	C	72/90 (80%)	68 (94%)	3 (4%)	1 (1%)	11	45
1	D	66/90 (73%)	65 (98%)	1 (2%)	0	100	100
All	All	280/360 (78%)	266 (95%)	12 (4%)	2 (1%)	22	60

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	56	ALA
1	C	57	GLU

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	54/69 (78%)	54 (100%)	0	100	100
1	B	50/69 (72%)	50 (100%)	0	100	100
1	C	56/69 (81%)	56 (100%)	0	100	100
1	D	51/69 (74%)	51 (100%)	0	100	100
All	All	211/276 (76%)	211 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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	43	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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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

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

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

6.3 Carbohydrates ⓘ

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

6.4 Ligands ⓘ

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

6.5 Other polymers ⓘ

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