



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2023 – 12:06 PM EDT

PDB ID : 5VHE
Title : DHX36 in complex with the c-Myc G-quadruplex
Authors : Chen, M.; Ferre-D'Amare, A.
Deposited on : 2017-04-13
Resolution : 3.79 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Xtrriage (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

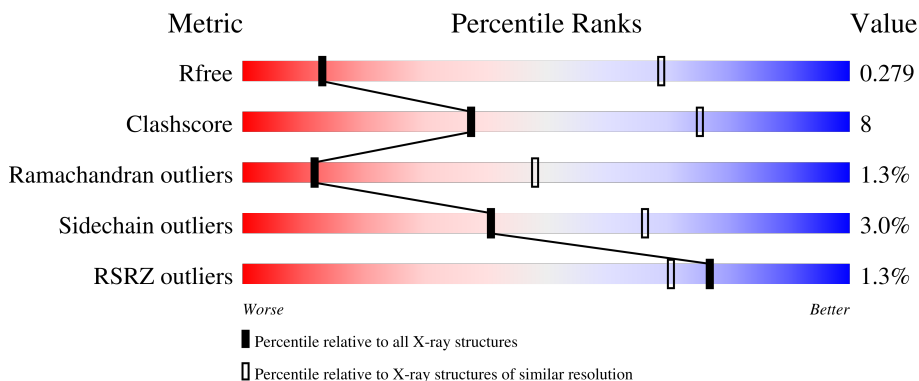
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.79 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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 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 $\leq 5\%$. 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	933	 % 74% 18% • 7%
2	B	24	 4% 67% 33%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7202 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 DEAH (Asp-Glu-Ala-His) box polypeptide 36.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	871	6697	4261	1147	1254	35	0	0	0

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	55	GLY	-	expression tag	UNP Q05B79
A	?	-	PRO	deletion	UNP Q05B79
A	?	-	ALA	deletion	UNP Q05B79
A	?	-	GLU	deletion	UNP Q05B79
A	?	-	ASN	deletion	UNP Q05B79
A	?	-	LYS	deletion	UNP Q05B79
A	?	-	PRO	deletion	UNP Q05B79
A	?	-	ASN	deletion	UNP Q05B79
A	?	-	SER	deletion	UNP Q05B79
A	?	-	VAL	deletion	UNP Q05B79
A	?	-	LYS	deletion	UNP Q05B79
A	?	-	ASN	deletion	UNP Q05B79
A	?	-	VAL	deletion	UNP Q05B79
A	?	-	GLU	deletion	UNP Q05B79
A	?	-	HIS	deletion	UNP Q05B79
A	?	-	GLN	deletion	UNP Q05B79
A	?	-	GLU	deletion	UNP Q05B79
A	?	-	LYS	deletion	UNP Q05B79
A	?	-	LYS	deletion	UNP Q05B79
A	?	-	MET	deletion	UNP Q05B79
A	?	-	ILE	deletion	UNP Q05B79
A	?	-	ASN	deletion	UNP Q05B79
A	?	-	GLN	deletion	UNP Q05B79
A	?	-	GLU	deletion	UNP Q05B79
A	?	-	LYS	deletion	UNP Q05B79
A	?	-	ARG	deletion	UNP Q05B79
A	?	-	PRO	deletion	UNP Q05B79

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	PHE	deletion	UNP Q05B79
A	?	-	ARG	deletion	UNP Q05B79
A	?	-	ILE	deletion	UNP Q05B79
A	?	-	ARG	deletion	UNP Q05B79
A	?	-	ASP	deletion	UNP Q05B79
A	?	-	LYS	deletion	UNP Q05B79
A	435	TYR	GLU	conflict	UNP Q05B79
A	436	TYR	GLU	conflict	UNP Q05B79
A	437	TYR	LYS	conflict	UNP Q05B79
A	752	ALA	LYS	conflict	UNP Q05B79
A	753	ALA	ASP	conflict	UNP Q05B79
A	755	ALA	LYS	conflict	UNP Q05B79
A	1011	PRO	-	expression tag	UNP Q05B79
A	1012	HIS	-	expression tag	UNP Q05B79
A	1013	HIS	-	expression tag	UNP Q05B79
A	1014	HIS	-	expression tag	UNP Q05B79
A	1015	HIS	-	expression tag	UNP Q05B79
A	1016	HIS	-	expression tag	UNP Q05B79
A	1017	HIS	-	expression tag	UNP Q05B79
A	1018	HIS	-	expression tag	UNP Q05B79
A	1019	HIS	-	expression tag	UNP Q05B79

- Molecule 2 is a DNA chain called DNA (5'-D(*AP*GP*GP*GP*TP*GP*GP*GP*TP*AP*GP*GP*GP*TP*GP*GP*GP*TP*TP*TP*TP*TP*T)-3').

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
			Total	C	N	O	P			
2	B	24	503	240	90	150	23	0	0	0

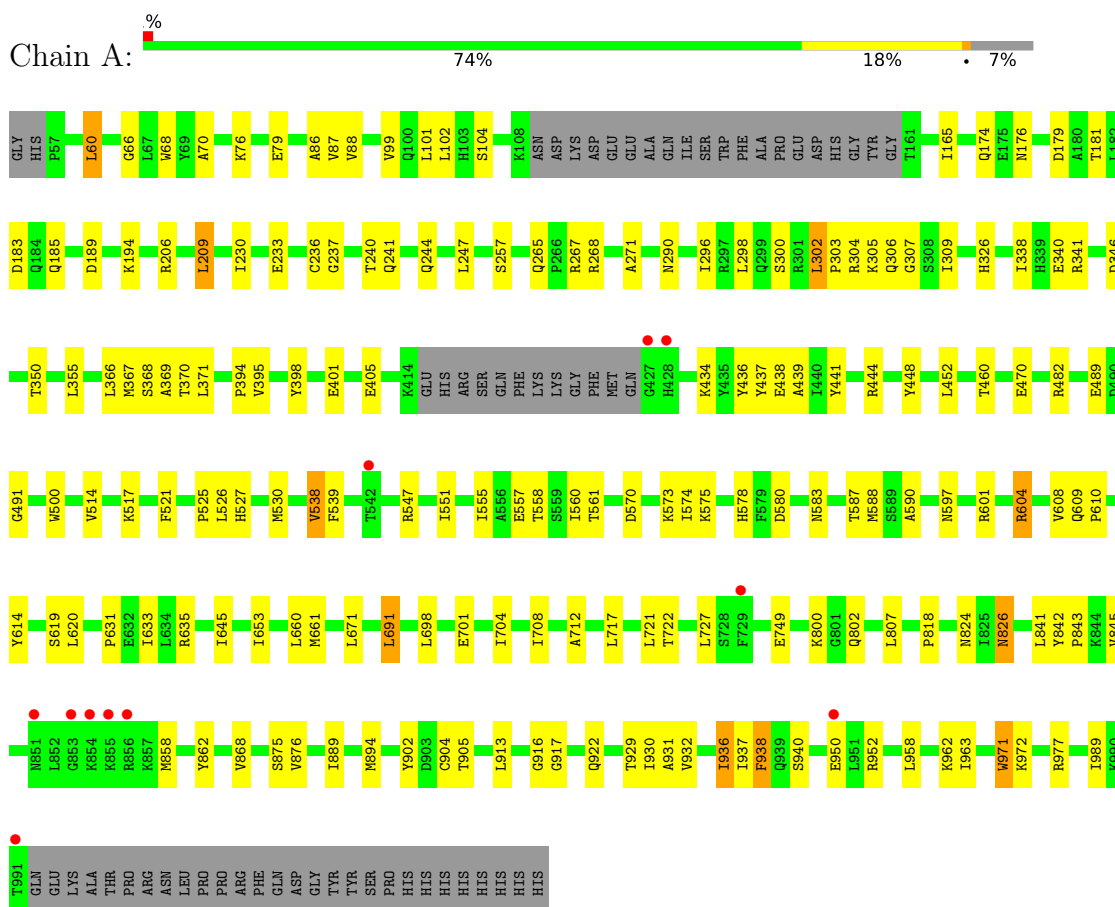
- Molecule 3 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	K		
3	B	2	2	2	0	0

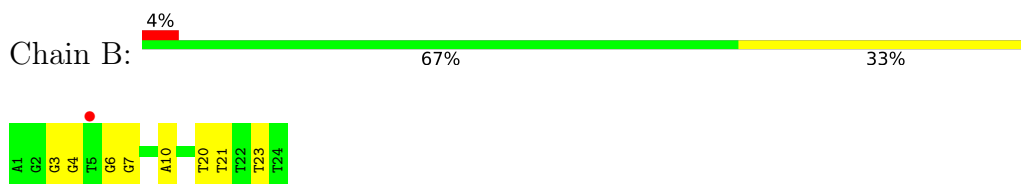
3 Residue-property plots

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 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: DEAH (Asp-Glu-Ala-His) box polypeptide 36



- Molecule 2: DNA (5'-D(*AP*GP*GP*GP*TP*GP*GP*GP*TP*AP*GP*GP*GP*TP*GP*GP*GP*TP*TP*TP*TP*TP*TP*TP*T)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	72.53Å 79.28Å 212.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.64 – 3.79 39.64 – 3.79	Depositor EDS
% Data completeness (in resolution range)	99.3 (39.64-3.79) 99.3 (39.64-3.79)	Depositor EDS
R_{merge}	0.27	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.63 (at 3.76Å)	Xtrriage
Refinement program	PHENIX (1.10.1_2155)	Depositor
R, R_{free}	0.238 , 0.280 0.238 , 0.279	Depositor DCC
R_{free} test set	1261 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å ²)	115.0	Xtrriage
Anisotropy	0.389	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 72.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	7202	wwPDB-VP
Average B, all atoms (Å ²)	120.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.50% 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

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

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	0.23	0/6828	0.40	0/9277
2	B	0.55	0/564	0.97	0/874
All	All	0.27	0/7392	0.48	0/10151

There are no bond length outliers.

There are no bond angle outliers.

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	6697	0	6518	99	1
2	B	503	0	276	11	0
3	B	2	0	0	0	0
All	All	7202	0	6794	106	1

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 8.

The worst 5 of 106 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:561:THR:HA	1:A:604:ARG:HH12	1.15	1.09
1:A:341:ARG:NH1	1:A:346:ASP:OD2	2.14	0.79
1:A:401:GLU:OE2	1:A:448:TYR:OH	2.02	0.78
1:A:268:ARG:HH11	1:A:296:ILE:HD13	1.50	0.77
2:B:6:DG:H1	2:B:10:DA:H62	1.32	0.76

All (1) 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:482:ARG:NH1	1:A:749:GLU:OE2[3_644]	2.02	0.18

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	865/933 (93%)	789 (91%)	65 (8%)	11 (1%)	12 48

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	88	VAL
1	A	538	VAL
1	A	842	TYR
1	A	87	VAL
1	A	395	VAL

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	703/831 (85%)	682 (97%)	21 (3%)	41 66

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	858	MET
1	A	938	PHE
1	A	971	TRP
1	A	940	SER
1	A	936	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	939	GLN
1	A	703	HIS
1	A	306	GLN
1	A	295	GLN
1	A	582	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](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - 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 [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	871/933 (93%)	-0.15	11 (1%) 77 70	58, 109, 193, 326	0
2	B	24/24 (100%)	0.51	1 (4%) 36 30	86, 166, 188, 197	0
All	All	895/957 (93%)	-0.14	12 (1%) 77 70	58, 110, 193, 326	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	854	LYS	4.6
1	A	855	LYS	4.5
1	A	853	GLY	4.0
1	A	856	ARG	3.0
1	A	851	ASN	2.9

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 monosaccharides in this entry.

6.4 Ligands [i](#)

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. 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.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	K	B	102	1/1	0.91	0.10	115,115,115,115	0
3	K	B	101	1/1	0.93	0.04	118,118,118,118	0

6.5 Other polymers [i](#)

There are no such residues in this entry.