wwPDB X-ray Structure Validation Summary Report (i)

Mar 9, 2026 – 05:20 AM UTC

PDB ID	:	7DW5 / pdb_00007dw5
Title	:	Crystal structure of DUX4 HD1-HD2 domain complexed with ERG sites
Authors	:	Zhang, H.; Cheng, N.; Li, Z.; Zhang, W.; Dong, X.; Huang, J.; Meng, G.
Deposited on	:	2021-01-15
Resolution	:	2.83 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:

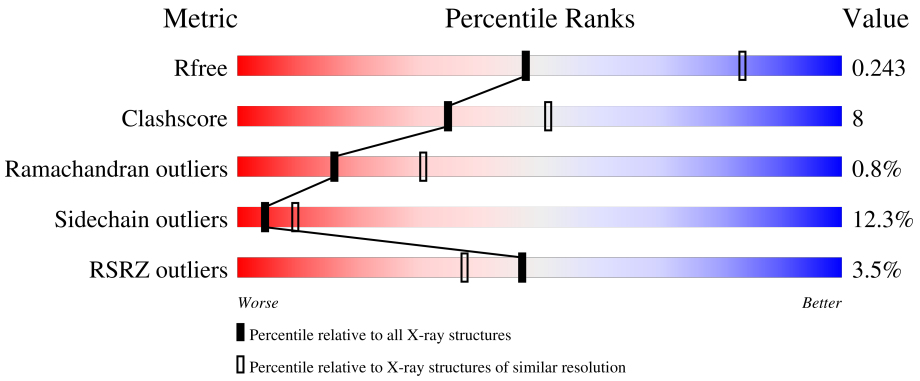
MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

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 2.83 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}	180053	1520 (2.86-2.82)
Clashscore	190562	1559 (2.86-2.82)
Ramachandran outliers	187476	1517 (2.86-2.82)
Sidechain outliers	187428	1518 (2.86-2.82)
RSRZ outliers	180081	1521 (2.86-2.82)

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% 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	150	3% 67% 14% 6% 13%
1	B	150	3% 70% 17% • 13%
2	C	21	10% 62% 38%
2	E	21	71% 29%
3	D	21	5% 43% 57%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	21	43% 57%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7080 atoms, of which 3156 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 Double homeobox protein 4-like protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	131	Total	C	H	N	O	S	0	0	0
			2211	679	1111	237	183	1			
1	B	131	Total	C	H	N	O	S	0	2	0
			2186	675	1093	232	185	1			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	E	21	Total	C	H	N	O	P	0	0	0
			673	207	238	81	126	21			
2	C	21	Total	C	H	N	O	P	0	0	0
			673	207	238	81	126	21			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	F	21	Total	C	H	N	O	P	0	0	0
			664	204	238	75	126	21			
3	D	21	Total	C	H	N	O	P	0	0	0
			664	204	238	75	126	21			

- Molecule 4 is BROMIDE ION (CCD ID: BR) (formula: Br).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	1	Total	Br	0	0
			1	1		

- Molecule 5 is water.

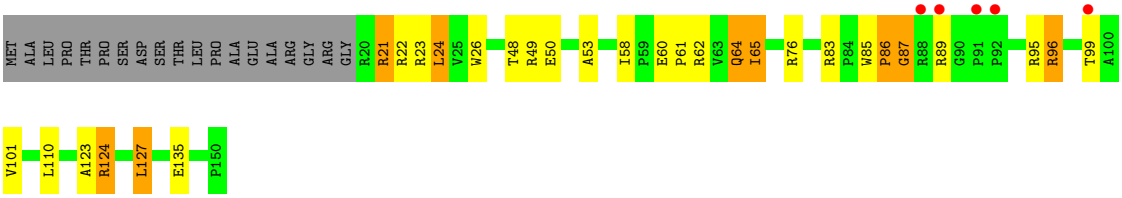
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	3	Total O 3 3	0	0
5	B	2	Total O 2 2	0	0
5	E	2	Total O 2 2	0	0
5	F	1	Total O 1 1	0	0

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 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: Double homeobox protein 4-like protein 2

Chain A:

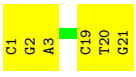


• Molecule 1: Double homeobox protein 4-like protein 2

Chain B:

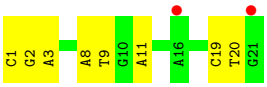
• Molecule 2: DNA (5'-D(P*CP*GP*AP*CP*TP*TP*GP*AP*TP*GP*AP*GP*AP*TP*TP*AP*GP*AP*CP*TP*G)-3')

Chain E:



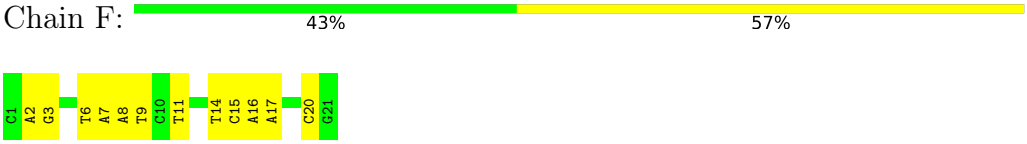
• Molecule 2: DNA (5'-D(P*CP*GP*AP*CP*TP*TP*GP*AP*TP*GP*AP*GP*AP*TP*TP*AP*GP*AP*CP*TP*G)-3')

Chain C:



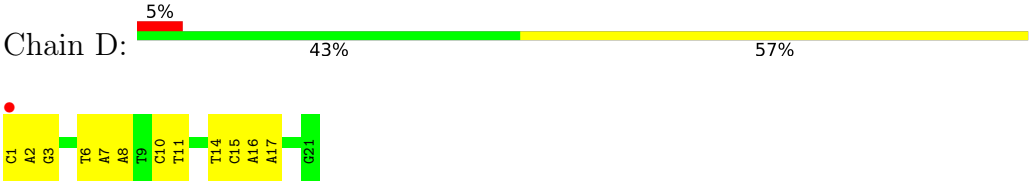
• Molecule 3: DNA (5'-D(P*CP*GP*AP*CP*TP*TP*GP*AP*TP*GP*AP*GP*AP*TP*TP*AP*GP*AP*CP*TP*G)-3')

Chain F:



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

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	42.97Å 64.45Å 69.43Å 95.82° 92.55° 103.31°	Depositor
Resolution (Å)	41.70 – 2.83 41.70 – 2.83	Depositor EDS
% Data completeness (in resolution range)	71.2 (41.70-2.83) 71.4 (41.70-2.83)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.08 (at 2.81Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.223 , 0.242 0.223 , 0.243	Depositor DCC
R_{free} test set	560 reflections (3.25%)	wwPDB-VP
Wilson B-factor (Å^2)	43.6	Xtriage
Anisotropy	0.441	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 17.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	7080	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.42% 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

5.1 Standard geometry

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

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	0.22	0/1127	0.48	0/1517
1	B	0.19	0/1128	0.46	0/1522
2	C	0.27	0/488	0.56	0/752
2	E	0.29	0/488	0.56	0/752
3	D	0.26	0/476	0.56	0/731
3	F	0.31	0/476	0.61	0/731
All	All	0.24	0/4183	0.52	0/6005

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	1100	1111	1111	14	0
1	B	1093	1093	1080	11	0
2	C	435	238	238	11	0
2	E	435	238	238	6	0
3	D	426	238	238	11	0
3	F	426	238	238	8	0
4	E	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	3	0	0	0	0
5	B	2	0	0	0	0
5	E	2	0	0	0	0
5	F	1	0	0	0	0
All	All	3924	3156	3143	52	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 8.

The worst 5 of 52 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:23:ARG:NH2	3:F:6:DT:O2	2.08	0.86
1:B:68[A]:GLN:OE1	1:B:71:ARG:NH1	2.18	0.75
1:A:124:ARG:NH1	1:A:135:GLU:OE2	2.25	0.69
1:B:64:GLN:OE1	1:B:68[B]:GLN:NE2	2.37	0.56
2:E:19:DC:H2"	2:E:20:DT:H5"	1.88	0.54

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	129/150 (86%)	118 (92%)	9 (7%)	2 (2%)	7 16
1	B	131/150 (87%)	118 (90%)	13 (10%)	0	100 100
All	All	260/300 (87%)	236 (91%)	22 (8%)	2 (1%)	16 31

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	86	PRO
1	A	87	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	111/125 (89%)	96 (86%)	15 (14%)	4 7
1	B	110/125 (88%)	98 (89%)	12 (11%)	6 12
All	All	221/250 (88%)	194 (88%)	27 (12%)	4 10

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

Mol	Chain	Res	Type
1	A	127	LEU
1	B	48	THR
1	B	101	VAL
1	B	39	GLU
1	B	50	GLU

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

Mol	Chain	Res	Type
1	A	149	HIS
1	B	149	HIS

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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 (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

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, 95^{th} 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	131/150 (87%)	0.13	5 (3%) 44 35	16, 42, 91, 150	0
1	B	131/150 (87%)	0.07	4 (3%) 51 42	15, 41, 82, 106	1 (0%)
2	C	21/21 (100%)	0.50	2 (9%) 14 11	30, 62, 106, 129	0
2	E	21/21 (100%)	-0.06	0 100 100	22, 36, 84, 92	0
3	D	21/21 (100%)	0.58	1 (4%) 35 29	41, 58, 94, 131	0
3	F	21/21 (100%)	-0.12	0 100 100	20, 36, 81, 93	0
All	All	346/384 (90%)	0.13	12 (3%) 47 37	15, 43, 92, 150	1 (0%)

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	92	PRO	5.5
3	D	1	DC	3.6
1	A	99	THR	3.5
1	A	91	PRO	3.1
1	B	90	GLY	2.6

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

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

6.3 Carbohydrates

There are no oligosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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(Å^2)	Q<0.9
4	BR	E	101	1/1	0.95	0.13	72,72,72,72	0

6.5 Other polymers

There are no such residues in this entry.