



wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 11:35 pm BST

PDB ID : 5FIQ
Title : Exonuclease domain-containing 1 (Exd1) in the native conformation
Authors : Yang, Z.; Chen, K.M.; Pandey, R.R.; Homolka, D.; Reuter, M.; Rodino Janeiro, B.K.; Sachidanandam, R.; Fauvarque, M.O.; McCarthy, A.A.; Pillai, R.S.
Deposited on : 2015-10-01
Resolution : 2.40 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

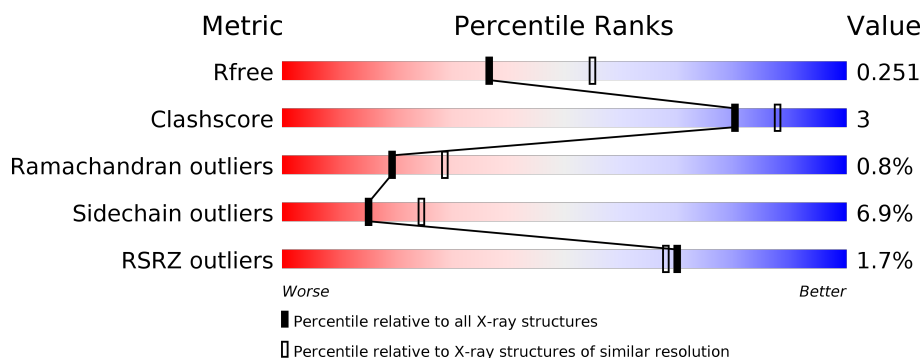
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	243	<div> <div>2%</div> <div> <div></div> <div>85%</div> <div>13%</div> <div>.</div> </div> </div>
1	C	243	<div> <div>%</div> <div> <div></div> <div>85%</div> <div>12%</div> <div>..</div> </div> </div>
1	E	243	<div> <div>%</div> <div> <div></div> <div>84%</div> <div>12%</div> <div>..</div> </div> </div>
1	G	243	<div> <div>%</div> <div> <div></div> <div>84%</div> <div>14%</div> <div>..</div> </div> </div>
1	I	243	<div> <div>3%</div> <div> <div></div> <div>85%</div> <div>12%</div> <div>..</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9738 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 EXD1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	242	Total	C	N	O	S	0	0	0
			1950	1242	332	366	10			
1	C	240	Total	C	N	O	S	0	0	0
			1933	1230	329	364	10			
1	E	240	Total	C	N	O	S	0	0	0
			1933	1230	329	364	10			
1	G	240	Total	C	N	O	S	0	0	0
			1936	1233	329	364	10			
1	I	240	Total	C	N	O	S	0	0	0
			1933	1230	329	364	10			

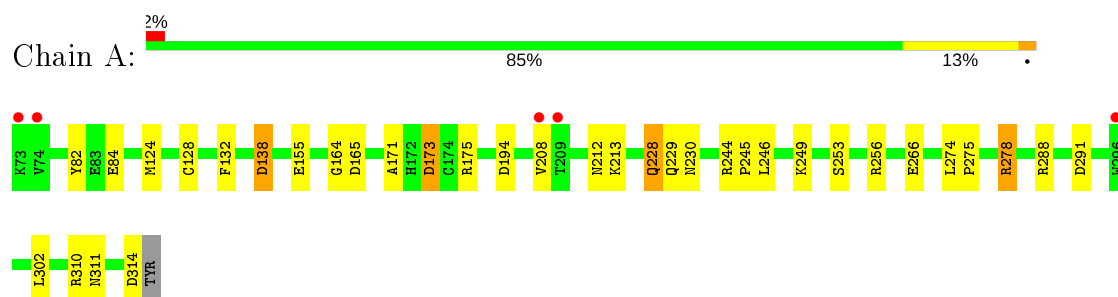
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total	O	0	0
			21	21		
2	C	13	Total	O	0	0
			13	13		
2	E	12	Total	O	0	0
			12	12		
2	G	6	Total	O	0	0
			6	6		
2	I	1	Total	O	0	0
			1	1		

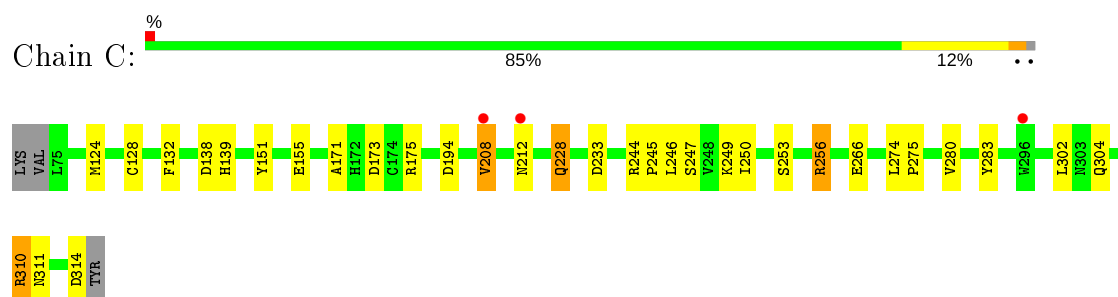
3 Residue-property plots [i](#)

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 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.

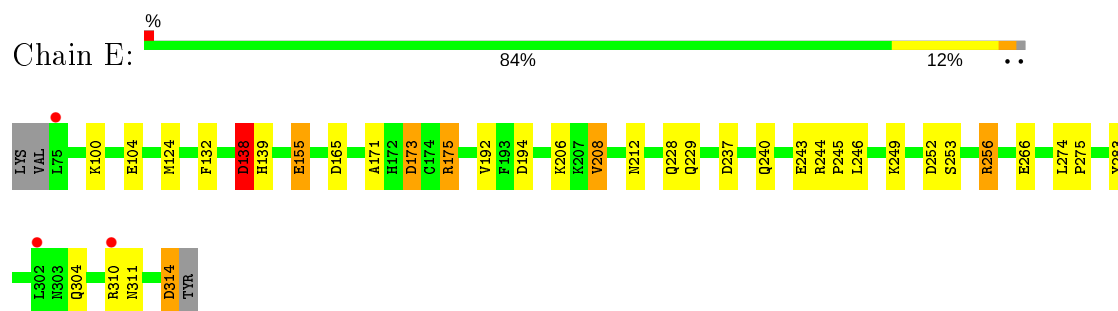
• Molecule 1: EXD1



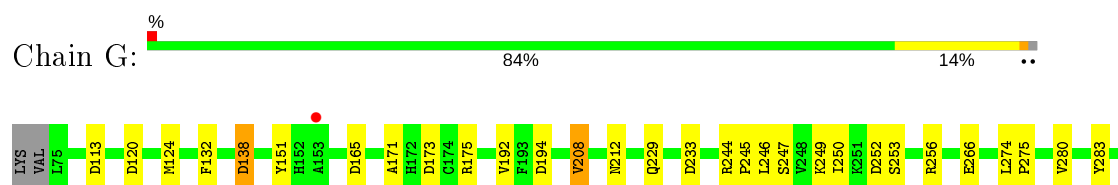
• Molecule 1: EXD1

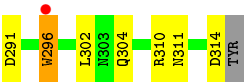


• Molecule 1: EXD1

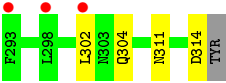
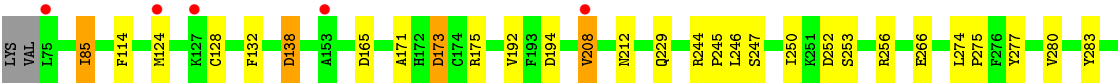
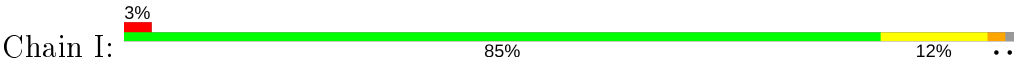


• Molecule 1: EXD1





● Molecule 1: EXD1



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	122.54 Å 81.58 Å 154.75 Å 90.00° 111.10° 90.00°	Depositor
Resolution (Å)	50.01 – 2.40 48.12 – 2.40	Depositor EDS
% Data completeness (in resolution range)	96.1 (50.01-2.40) 96.1 (48.12-2.40)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.27 (at 2.39 Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.221 , 0.247 0.224 , 0.251	Depositor DCC
R_{free} test set	2596 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å ²)	41.8	Xtriage
Anisotropy	1.160	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 26.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.013 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	9738	wwPDB-VP
Average B, all atoms (Å ²)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.21% 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	1.13	3/1984 (0.2%)	1.08	6/2669 (0.2%)
1	C	1.02	0/1967	1.02	3/2647 (0.1%)
1	E	1.04	2/1967 (0.1%)	1.03	8/2647 (0.3%)
1	G	0.94	1/1970 (0.1%)	1.04	7/2651 (0.3%)
1	I	0.98	3/1967 (0.2%)	1.03	6/2647 (0.2%)
All	All	1.02	9/9855 (0.1%)	1.04	30/13261 (0.2%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	82	TYR	CZ-OH	7.58	1.50	1.37
1	A	84	GLU	CG-CD	6.90	1.62	1.51
1	I	277	TYR	CG-CD2	6.71	1.47	1.39
1	E	155	GLU	CD-OE2	6.35	1.32	1.25
1	A	138	ASP	CB-CG	-6.05	1.39	1.51

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	124	MET	CG-SD-CE	7.27	111.83	100.20
1	I	85	ILE	CA-CB-CG1	7.05	124.39	111.00
1	G	165	ASP	CB-CG-OD1	-6.98	112.02	118.30
1	A	288	ARG	NE-CZ-NH2	-6.43	117.08	120.30
1	A	84	GLU	OE1-CD-OE2	-6.38	115.65	123.30

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	1950	0	1966	8	0
1	C	1933	0	1942	16	0
1	E	1933	0	1942	16	0
1	G	1936	0	1951	17	0
1	I	1933	0	1942	14	0
2	A	21	0	0	0	0
2	C	13	0	0	0	0
2	E	12	0	0	1	0
2	G	6	0	0	0	0
2	I	1	0	0	0	0
All	All	9738	0	9743	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 3.

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:G:296:TRP:O	1:G:296:TRP:HE3	1.57	0.88
1:G:296:TRP:O	1:G:296:TRP:CE3	2.41	0.72
1:C:208:VAL:HG22	1:E:208:VAL:HG22	1.79	0.65
1:A:274:LEU:HD13	1:A:278:ARG:NH1	2.13	0.63
1:G:175:ARG:NH2	1:I:304:GLN:O	2.31	0.62

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	240/243 (99%)	226 (94%)	12 (5%)	2 (1%)	19 29

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	238/243 (98%)	225 (94%)	12 (5%)	1 (0%)	34	48
1	E	238/243 (98%)	225 (94%)	11 (5%)	2 (1%)	19	29
1	G	238/243 (98%)	224 (94%)	12 (5%)	2 (1%)	19	29
1	I	238/243 (98%)	223 (94%)	13 (6%)	2 (1%)	19	29
All	All	1192/1215 (98%)	1123 (94%)	60 (5%)	9 (1%)	19	29

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	I	173	ASP
1	I	229	GLN
1	A	173	ASP
1	A	229	GLN
1	C	173	ASP

5.3.2 Protein sidechains ⓘ

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	220/224 (98%)	202 (92%)	18 (8%)	11	17
1	C	218/224 (97%)	204 (94%)	14 (6%)	17	28
1	E	218/224 (97%)	203 (93%)	15 (7%)	15	25
1	G	219/224 (98%)	205 (94%)	14 (6%)	17	28
1	I	218/224 (97%)	204 (94%)	14 (6%)	17	28
All	All	1093/1120 (98%)	1018 (93%)	75 (7%)	15	25

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	138	ASP
1	E	253	SER
1	I	253	SER
1	E	155	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	228	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	97	GLN
1	E	183	HIS
1	G	183	HIS
1	C	139	HIS
1	G	110	ASN

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	242/243 (99%)	-0.13	5 (2%) 63 61	33, 47, 83, 107	0
1	C	240/243 (98%)	-0.12	3 (1%) 77 75	36, 53, 91, 115	0
1	E	240/243 (98%)	-0.08	3 (1%) 77 75	40, 57, 90, 115	0
1	G	240/243 (98%)	-0.10	2 (0%) 86 84	40, 61, 99, 120	0
1	I	240/243 (98%)	0.11	8 (3%) 46 45	41, 66, 103, 121	0
All	All	1202/1215 (98%)	-0.06	21 (1%) 70 68	33, 56, 96, 121	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	296	TRP	6.1
1	A	208	VAL	3.6
1	I	75	LEU	3.5
1	A	209	THR	3.5
1	I	302	LEU	3.4

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.