



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 21, 2023 – 01:11 PM EDT

PDB ID : 8EJR
Title : Kelch domain of human KEAP1 bound to Nrf2 linear peptide, Ac-GDPETGE-NH₂
Authors : Muellers, S.N.; Allen, K.N.
Deposited on : 2022-09-18
Resolution : 2.08 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
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (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

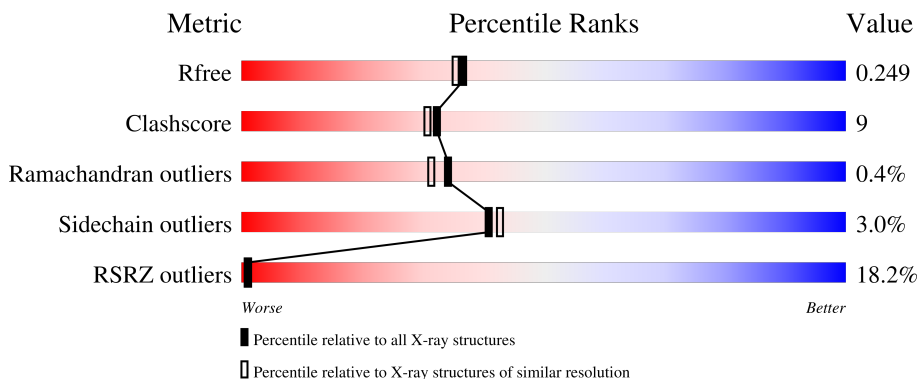
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.08 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	6189 (2.10-2.06)
Clashscore	141614	6738 (2.10-2.06)
Ramachandran outliers	138981	6663 (2.10-2.06)
Sidechain outliers	138945	6664 (2.10-2.06)
RSRZ outliers	127900	6057 (2.10-2.06)

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	336	 7% 74% 9% • 16%
1	B	336	 24% 63% 18% •• 16%
2	C	9	 89% 11%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 4674 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 Kelch-like ECH-associated protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	283	2170	1351	392	412	15	0	0	0
1	B	283	2170	1351	392	412	15	0	0	0

There are 58 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	289	MET	-	expression tag	UNP Q14145
A	290	GLY	-	expression tag	UNP Q14145
A	291	SER	-	expression tag	UNP Q14145
A	292	SER	-	expression tag	UNP Q14145
A	293	HIS	-	expression tag	UNP Q14145
A	294	HIS	-	expression tag	UNP Q14145
A	295	HIS	-	expression tag	UNP Q14145
A	296	HIS	-	expression tag	UNP Q14145
A	297	HIS	-	expression tag	UNP Q14145
A	298	HIS	-	expression tag	UNP Q14145
A	299	SER	-	expression tag	UNP Q14145
A	300	SER	-	expression tag	UNP Q14145
A	301	GLY	-	expression tag	UNP Q14145
A	302	GLY	-	expression tag	UNP Q14145
A	303	GLU	-	expression tag	UNP Q14145
A	304	ASN	-	expression tag	UNP Q14145
A	305	LEU	-	expression tag	UNP Q14145
A	306	TYR	-	expression tag	UNP Q14145
A	307	PHE	-	expression tag	UNP Q14145
A	308	GLN	-	expression tag	UNP Q14145
A	309	GLY	-	expression tag	UNP Q14145
A	310	HIS	-	expression tag	UNP Q14145
A	311	MET	-	expression tag	UNP Q14145
A	319	SER	CYS	conflict	UNP Q14145
A	540	ALA	GLU	engineered mutation	UNP Q14145

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	542	ALA	GLU	engineered mutation	UNP Q14145
A	613	SER	CYS	engineered mutation	UNP Q14145
A	622	SER	CYS	conflict	UNP Q14145
A	624	SER	CYS	conflict	UNP Q14145
B	289	MET	-	expression tag	UNP Q14145
B	290	GLY	-	expression tag	UNP Q14145
B	291	SER	-	expression tag	UNP Q14145
B	292	SER	-	expression tag	UNP Q14145
B	293	HIS	-	expression tag	UNP Q14145
B	294	HIS	-	expression tag	UNP Q14145
B	295	HIS	-	expression tag	UNP Q14145
B	296	HIS	-	expression tag	UNP Q14145
B	297	HIS	-	expression tag	UNP Q14145
B	298	HIS	-	expression tag	UNP Q14145
B	299	SER	-	expression tag	UNP Q14145
B	300	SER	-	expression tag	UNP Q14145
B	301	GLY	-	expression tag	UNP Q14145
B	302	GLY	-	expression tag	UNP Q14145
B	303	GLU	-	expression tag	UNP Q14145
B	304	ASN	-	expression tag	UNP Q14145
B	305	LEU	-	expression tag	UNP Q14145
B	306	TYR	-	expression tag	UNP Q14145
B	307	PHE	-	expression tag	UNP Q14145
B	308	GLN	-	expression tag	UNP Q14145
B	309	GLY	-	expression tag	UNP Q14145
B	310	HIS	-	expression tag	UNP Q14145
B	311	MET	-	expression tag	UNP Q14145
B	319	SER	CYS	conflict	UNP Q14145
B	540	ALA	GLU	engineered mutation	UNP Q14145
B	542	ALA	GLU	engineered mutation	UNP Q14145
B	613	SER	CYS	engineered mutation	UNP Q14145
B	622	SER	CYS	conflict	UNP Q14145
B	624	SER	CYS	conflict	UNP Q14145

- Molecule 2 is a protein called Linear peptide from Nuclear factor erythroid 2-related factor 2.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace	
			Total	C	N				O
2	C	9	52	29	8	15	0	0	1

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	75	ACE	-	acetylation	UNP Q16236
C	76	GLY	LEU	engineered mutation	UNP Q16236
C	78	PRO	GLU	engineered mutation	UNP Q16236
C	83	NH2	-	amidation	UNP Q16236

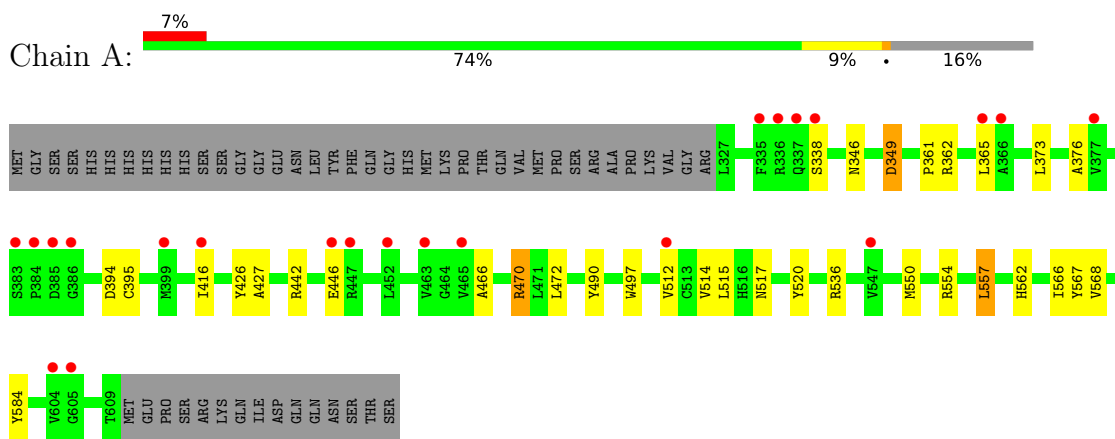
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	179	Total O 179 179	0	0
3	B	96	Total O 96 96	0	0
3	C	7	Total O 7 7	0	0

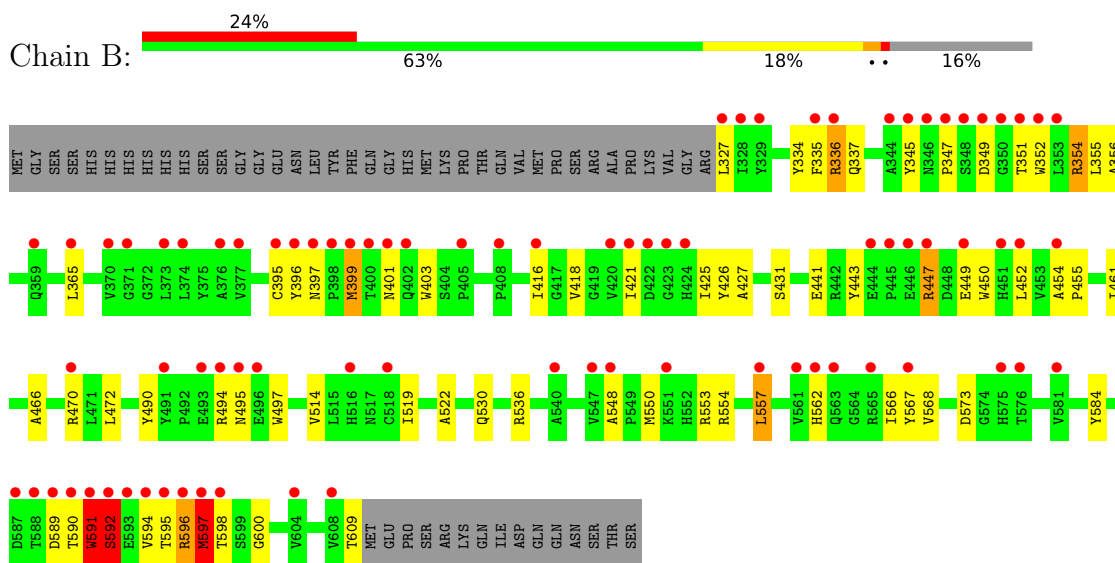
3 Residue-property plots [i](#)

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.

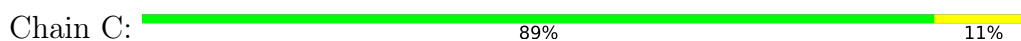
- Molecule 1: Kelch-like ECH-associated protein 1



- Molecule 1: Kelch-like ECH-associated protein 1



- Molecule 2: Linear peptide from Nuclear factor erythroid 2-related factor 2





4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	162.81Å 68.91Å 77.59Å 90.00° 117.66° 90.00°	Depositor
Resolution (Å)	29.41 – 2.08 29.41 – 2.08	Depositor EDS
% Data completeness (in resolution range)	99.0 (29.41-2.08) 99.0 (29.41-2.08)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.91 (at 2.08Å)	Xtrriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, R_{free}	0.222 , 0.249 0.222 , 0.249	Depositor DCC
R_{free} test set	1896 reflections (4.18%)	wwPDB-VP
Wilson B-factor (Å ²)	30.3	Xtrriage
Anisotropy	0.567	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 54.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.010 for -h-2*1,-k,l	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	4674	wwPDB-VP
Average B, all atoms (Å ²)	39.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.39% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NH2, ACE

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	0.37	0/2223	0.58	1/3029 (0.0%)
1	B	0.61	4/2223 (0.2%)	0.78	6/3029 (0.2%)
2	C	0.26	0/49	0.55	0/66
All	All	0.50	4/4495 (0.1%)	0.69	7/6124 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	592	SER	N-CA	7.09	1.60	1.46
1	B	591	TRP	CA-C	7.07	1.71	1.52
1	B	591	TRP	CB-CG	6.54	1.62	1.50
1	B	591	TRP	N-CA	5.67	1.57	1.46

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	447	ARG	CG-CD-NE	-10.18	90.42	111.80
1	B	597	MET	C-N-CA	7.17	139.62	121.70
1	B	591	TRP	N-CA-C	6.85	129.49	111.00
1	B	399	MET	CA-CB-CG	6.22	123.88	113.30
1	A	470	ARG	NE-CZ-NH1	-6.02	117.29	120.30

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	2170	0	2066	19	1
1	B	2170	0	2066	56	1
2	C	52	0	39	1	0
3	A	179	0	0	0	0
3	B	96	0	0	5	0
3	C	7	0	0	0	0
All	All	4674	0	4171	75	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 9.

The worst 5 of 75 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:352:TRP:HE1	1:B:595:THR:HG1	1.17	0.91
1:B:335:PHE:HD2	3:B:783:HOH:O	1.58	0.85
1:B:594:VAL:HG23	1:B:595:THR:HG22	1.62	0.79
1:B:347:PRO:HG2	1:B:562:HIS:CE1	2.19	0.78
1:A:515:LEU:HD22	1:A:566:ILE:HG13	1.74	0.69

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:517:ASN:OD1	1:B:470:ARG:NH1[2_555]	1.55	0.65

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	281/336 (84%)	271 (96%)	10 (4%)	0	100	100
1	B	281/336 (84%)	262 (93%)	17 (6%)	2 (1%)	22	17
2	C	7/9 (78%)	7 (100%)	0	0	100	100
All	All	569/681 (84%)	540 (95%)	27 (5%)	2 (0%)	34	31

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	597	MET
1	B	598	THR

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	228/275 (83%)	223 (98%)	5 (2%)	52	55
1	B	228/275 (83%)	219 (96%)	9 (4%)	32	32
2	C	5/5 (100%)	5 (100%)	0	100	100
All	All	461/555 (83%)	447 (97%)	14 (3%)	41	43

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	354	ARG
1	B	494	ARG
1	B	596	ARG
1	B	591	TRP
1	B	592	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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](#)

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 [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	283/336 (84%)	0.41	22 (7%) 13 16	20, 28, 50, 76	0
1	B	283/336 (84%)	1.36	82 (28%) 0 0	19, 42, 80, 95	0
2	C	7/9 (77%)	0.01	0 100 100	29, 32, 51, 53	0
All	All	573/681 (84%)	0.87	104 (18%) 1 1	19, 33, 70, 95	0

The worst 5 of 104 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	399	MET	8.7
1	B	594	VAL	8.1
1	B	335	PHE	6.8
1	B	329	TYR	5.6
1	A	385	ASP	5.5

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.