



wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Feb 24, 2018 – 10:48 AM EST

PDB ID : 6C23
EMDB ID: : EMD-7334
Title : Cryo-EM structure of PRC2 bound to cofactors AEBP2 and JARID2 in the Compact Active State
Authors : Kasinath, V.; Faini, M.; Poepsel, S.; Reif, D.; Feng, A.; Stjepanovic, G.; Aebersold, R.; Nogales, E.
Deposited on : 2018-01-05
Resolution : 3.90 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<http://wwpdb.org/validation/2016/EMValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20030736

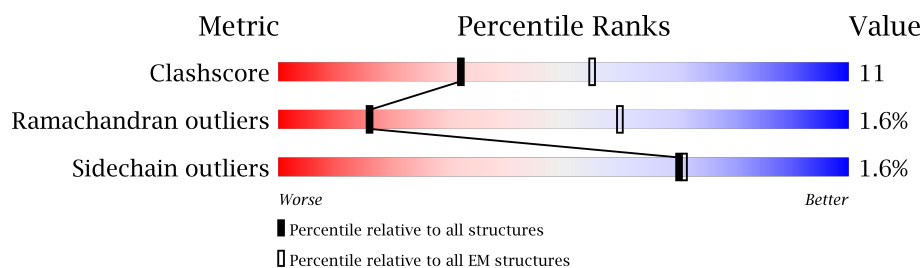
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 3.90 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)	EM structures (#Entries)
Clashscore	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	739	
1	M	739	
1	Q	739	
2	E	348	
3	C	746	
3	K	746	
4	L	441	
5	N	425	
6	O	7	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	P	295	<div><div></div><div>19% .. 78%</div></div>
8	Z	135	<div><div></div><div>93% 7%</div></div>
9	B	345	<div><div></div><div>.. 96%</div></div>

2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 13440 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 Polycomb protein SUZ12.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	123	Total	C	N	O	S	0	0
			860	549	157	148	6		
1	M	122	Total	C	N	O	S	0	0
			962	609	169	175	9		
1	Q	66	Total	C	N	O	S	0	0
			496	315	94	85	2		

- Molecule 2 is a protein called Protein Jumonji.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	E	27	Total	C	N	O	S	0	0
			181	111	32	37	1		

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	103	SER	-	expression tag	UNP Q92833
E	104	ASN	-	expression tag	UNP Q92833
E	105	ALA	-	expression tag	UNP Q92833

- Molecule 3 is a protein called Histone-lysine N-methyltransferase EZH2.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	K	325	Total	C	N	O	S	0	0
			2406	1514	426	436	30		
3	C	193	Total	C	N	O	S	0	0
			1307	820	230	253	4		

- Molecule 4 is a protein called Polycomb protein EED.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	L	362	Total	C	N	O	S	0	0
			2877	1826	505	526	20		

- Molecule 5 is a protein called Histone-binding protein RBBP4.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	N	386	Total	C	N	O	S	0	0
			3058	1931	524	593	10		

- Molecule 6 is a protein called JARID2-substrate.

Mol	Chain	Residues	Atoms				AltConf	Trace
6	O	7	Total	C	N	O	0	0
			51	33	11	7		

- Molecule 7 is a protein called Zinc finger protein AEBP2.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	P	64	Total	C	N	O	S	0	0
			484	304	91	87	2		

- Molecule 8 is a protein called SUZ12.

Mol	Chain	Residues	Atoms				AltConf	Trace
8	Z	135	Total	C	N	O	0	0
			674	404	135	135		

- Molecule 9 is a protein called Protein Jumonji.

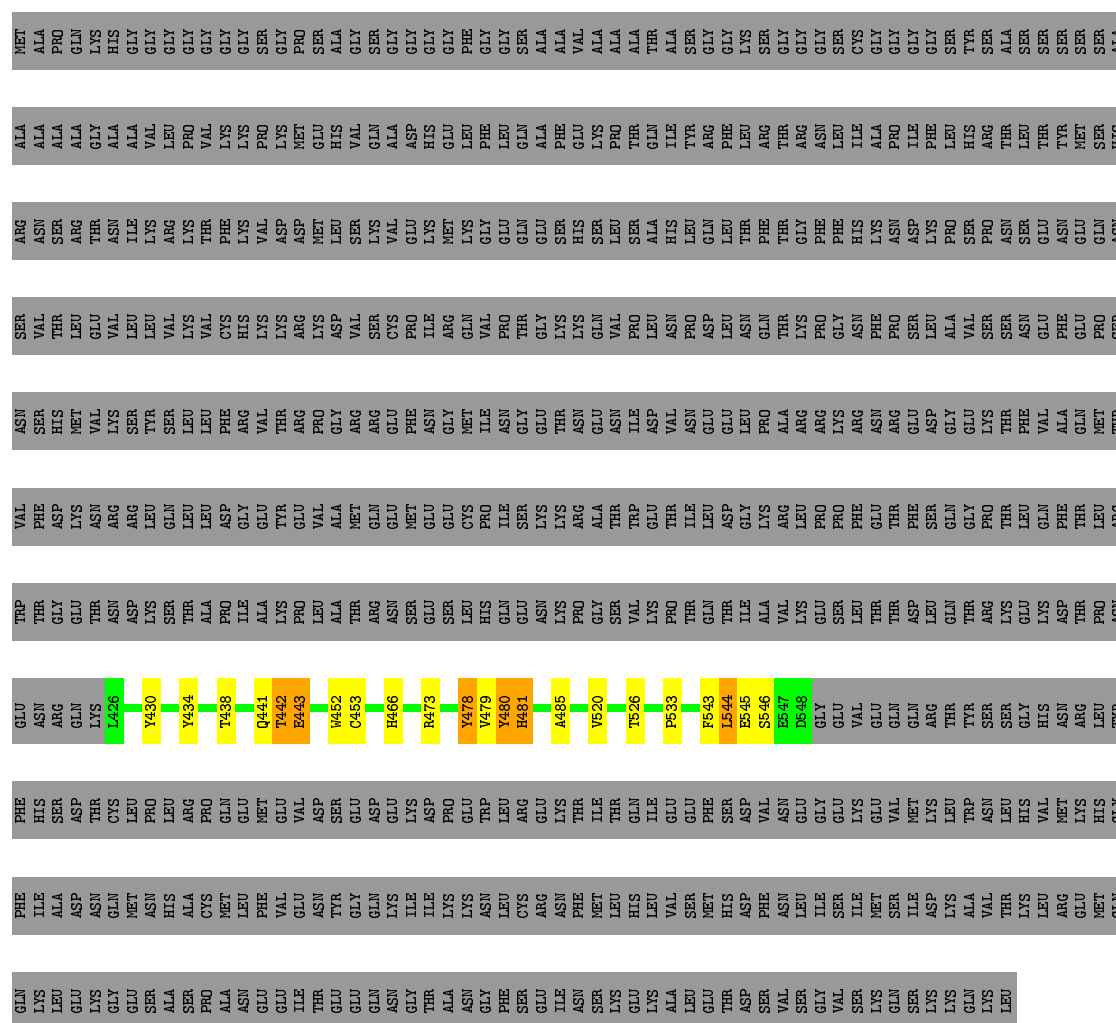
Mol	Chain	Residues	Atoms				AltConf	Trace
9	B	13	Total	C	N	O	0	0
			84	55	15	14		

3 Residue-property plots

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.

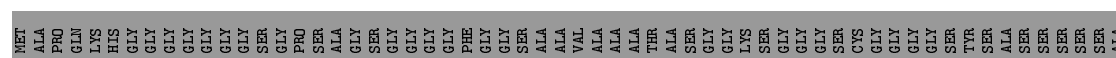
- Molecule 1: Polycomb protein SUZ12

Chain A:  14% .. 83%



- Molecule 1: Polycomb protein SUZ12

Chain M: 14% 83%





[illegible]

- Molecule 2: Protein Jumonji

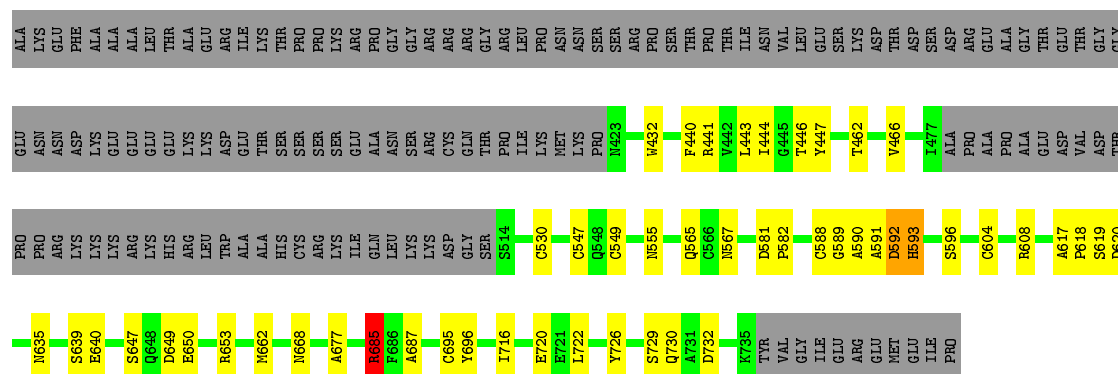
Chain E: 92%

VAL	GLN	PRO	GLY	SER
VAL	GLN	VAL	SER	ASN
ASP	LEU	GLN	SER	ALA
THR	ARG	LVS	GLN	ARG
LVS	LVS	HIS	ASP	LVS
PRO	GLN	LVS	GLU	ARG
ASN	VAL	SER	GLU	PRO
HIS	SER	LVS	GLU	LEU
HIS	LVS	GLU	VAL	GLN
PRO	ASN	THR	GLU	ALA
SER	GLY	PRO	ASP	GLN
SER	VAL	ALA	ASP	ARG
VAL	THR	LVS	GLU	LVS
VAL	ARG	GLU	GLU	PHE
ASN	MET	LVS	THR	ALA
HIS	SER	HIS	GLU	GLN
THR	SER	SER	ASP	SER
ILE	LEU	ASP	VAL	GLN
GLY	GLY	HIS	LVS	PRO
GLY	ALA	ARG	THR	ASN
LVS	VAL	ALA	SER	LVS
THR	MET	ALA	SER	ILE
ARG	ARG	SER	CYS	VAL
ARG	GLU	ALA	GLN	GLU
GLN	VAL	ASN	SER	PRO
LEU	PRO	PRO	THR	PRO
LEU	VAL	ALA	ARG	LEU
ARG	SER	ALA	LVS	PRO
GLU	SER	ALA	GLY	PRO
GLY	THR	ALA	LVS	P140
LEU	VAL	SER	THR	L147
GLN	THR	THR	HIS	K151
GLY	THR	SER	HIS	P152
ARG	THR	ALA	VAL	K153
PRO	GLY	ALA	HIS	T154
ALA	VAL	GLY	ASN	E155
VAL	THR	LEU	GLY	L158
ASN	LVS	ALA	VAL	L158
GLY	GLY	ALA	PHE	S166
GLY	ALA	THR	ASN	PRO
LEU	VAL	HIS	GLY	ALA
LVS	THR	HIS	SER	LEU
VAL	THR	HIS	SER	ALA
SER	THR	PRO	ARG	LEU
GLY	LVS	PRO	SER	PRO
ARG	ALA	LEU	THR	ASN
LEU	LVS	HIS	ARG	SER
ASN	ARG	THR	VAL	MET
PRO	GLU	SER	LVS	VAL
THR	THR	THR	THR	THR

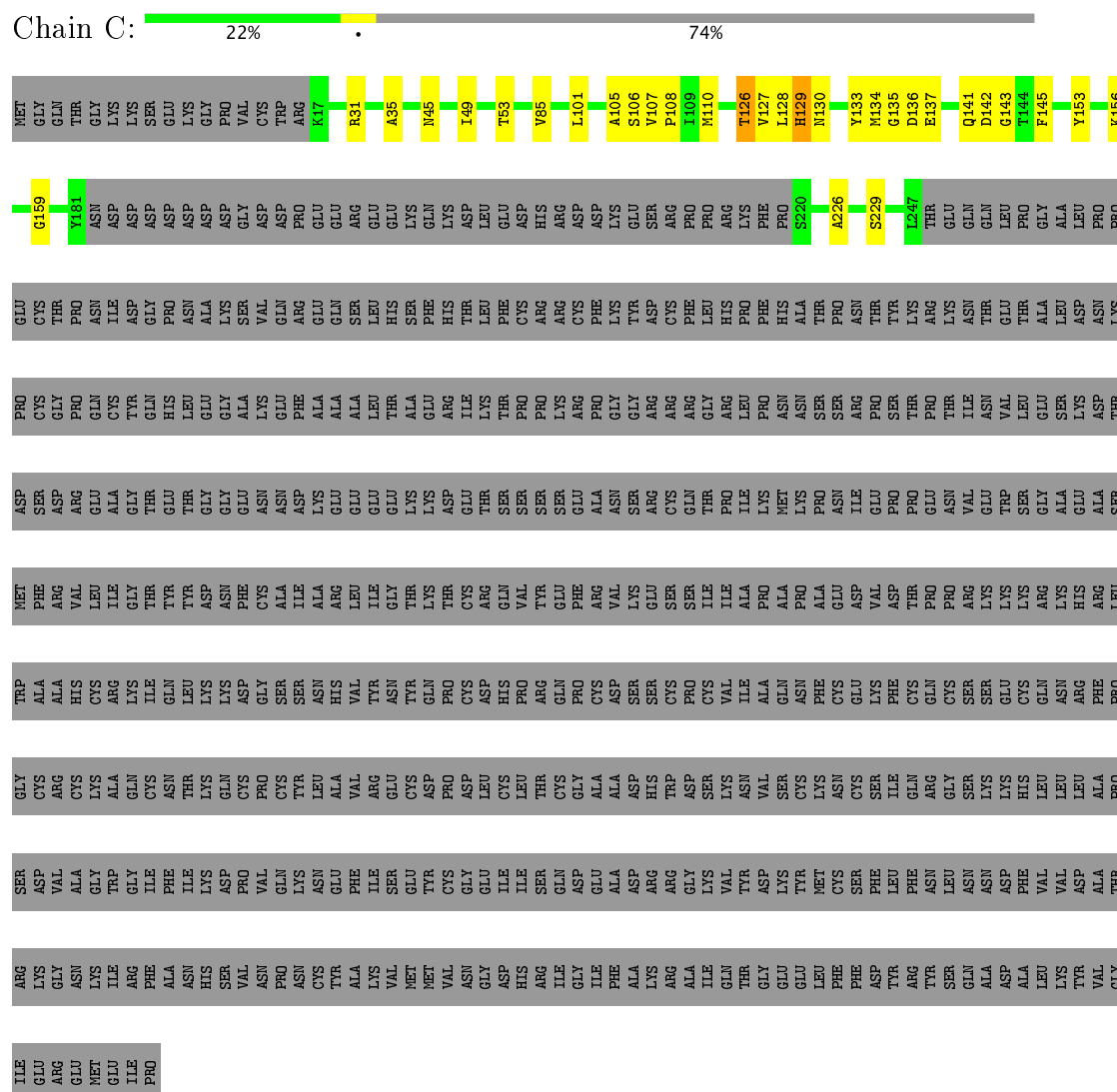
- Molecule 3: Histone-lysine N-methyltransferase EZH2

Chain K: 36% 8% 56%

LYS	LYS	TTR	MET	LYS	MET
GLU	ASN	ASN	VAL	GLN	GLY
LYS	LYS	ASP	GLU	ARG	GLN
TTR	ASP	ASP	ASP	ARG	THR
LYS	ASP	ASP	GLU	ILE	GLY
GLU	ASP	ASP	THR	GLN	LYS
LEU	ASP	ASP	VAL	PRO	LYS
THR	ASP	ASP	LEU	VAL	SER
GLU	ASP	ASP	HIS	HIS	GLU
GLN	GLY	GLY	ASN	ILE	LYS
LEU	ASP	ASP	ILE	LEU	GLY
PRO	ASP	ASP	PRO	THR	PRO
PRO	PRO	PRO	TTR	SER	VAL
E229	GLU	GLU	MET	VAL	CYS
P262	GLU	GLU	GLY	THR	TRP
V272	ARG	ARG	ASP	LEU	LYS
L278	GLN	GLY	VAL	ARG	ARG
S280	LYS	LYS	LEU	GLY	VAL
F281	ASP	ASP	GLY	THR	ARG
H282	ASP	ASP	ILE	ASP	LEU
T283	LYS	LYS	ILE	ASP	LYS
L284	GLU	GLU	LYS	PHE	ARG
L296	SER	ARG	ASN	PRO	PHE
L296	ARG	TTR	TTR	THR	ARG
T305	PRO	PRO	GLY	GLN	ALA
Y306	ARG	ARG	LYS	VAL	ASP
LYS	LYS	LYS	VAL	PRO	GLU
ARG	PHE	PHE	HIS	LEU	VAL
LYS	PRO	PRO	GLY	LYS	LYS
ASN	SER	SER	ASP	THR	SER
THR	ASP	ASP	ARG	LEU	MET
GLU	LYS	LYS	GLU	ASN	PHE
THR	ILE	PHE	CYS	ALA	SER
ALA	PHE	GLU	GLY	VAL	SER
ASP	ALA	ILE	ILE	SER	ASN
ASN	ILE	ASN	ASN	VAL	ARG
LYS	SER	SER	ASP	PRO	GLN
PRO	SER	SER	GLU	ILE	LYS
CYS	MET	MET	ILE	ILE	ILE
GLY	PHE	PHE	PHE	TTR	LEU
PRO	PRO	PRO	VAL	SER	ARG
GLN	ASP	GLU	GLU	TRP	THR
LYS	LYS	LYS	LEU	SER	GLU
TTR	GLY	VAL	VAL	PRO	ILE
GLN	THR	THR	ASN	LEU	LEU
HIS	ALA	ALA	ALA	GLN	ASN
LEU	GLU	GLU	LEU	GLN	GLN
GLU	GLU	GLU	GLY	ASN	GLU

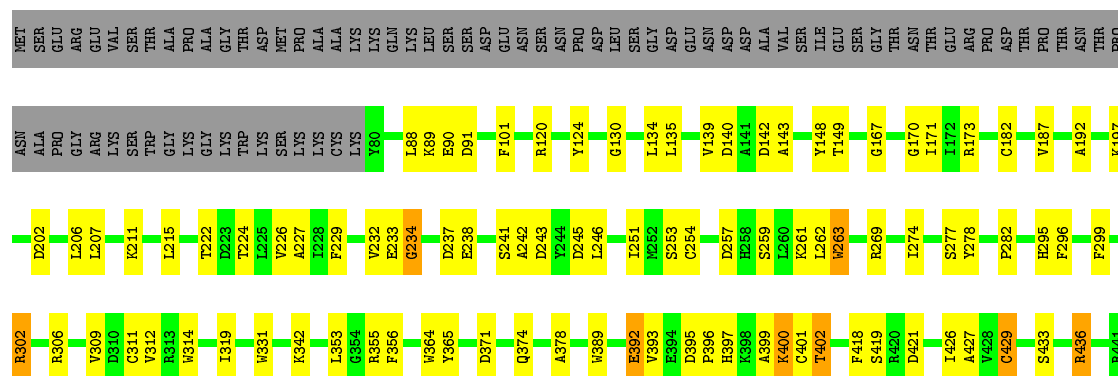


- Molecule 3: Histone-lysine N-methyltransferase EZH2



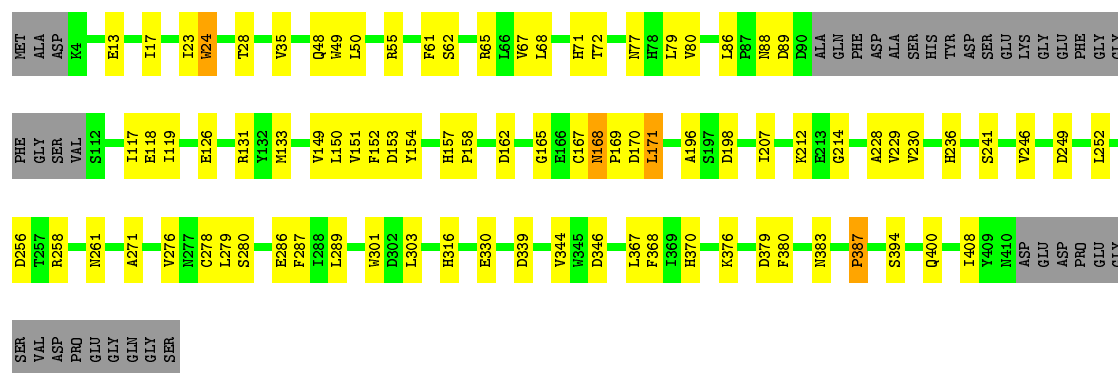
- Molecule 4: Polycomb protein EED





• Molecule 5: Histone-binding protein RBBP4

Chain N: 71% 19% 9%



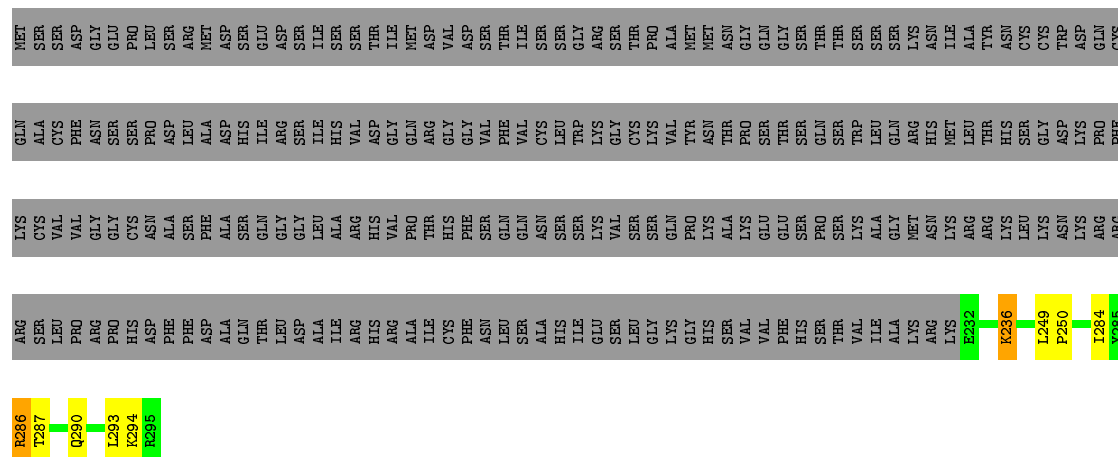
• Molecule 6: JARID2-substrate

Chain O: 100%

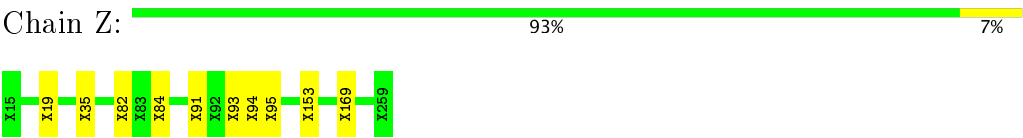
There are no outlier residues recorded for this chain.

• Molecule 7: Zinc finger protein AEBP2

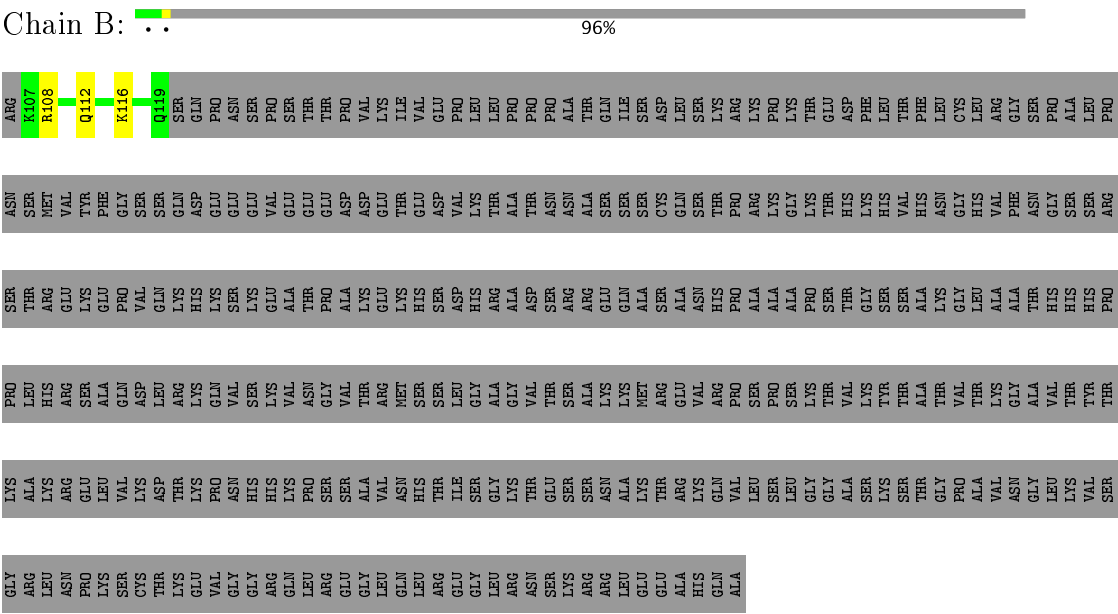
Chain P: 19% 78%



● Molecule 8: SUZ12



● Molecule 9: Protein Jumonji



4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	145592	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING ONLY	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	60	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: M3L

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 >2	RMSZ	# Z >2
1	A	0.58	0/882	0.82	1/1208 (0.1%)
1	M	0.58	0/983	0.75	0/1328
1	Q	0.61	0/506	0.86	0/687
2	E	0.40	0/182	0.75	0/246
3	C	0.49	0/1328	0.66	0/1815
3	K	0.49	0/2466	0.66	2/3352 (0.1%)
4	L	0.84	1/2952 (0.0%)	0.92	7/4008 (0.2%)
5	N	0.58	1/3142 (0.0%)	0.73	0/4286
6	O	0.54	0/51	0.74	0/66
7	P	0.53	0/492	0.81	0/666
9	B	0.44	0/73	0.43	0/100
All	All	0.62	2/13057 (0.0%)	0.77	10/17762 (0.1%)

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	A	0	3
1	M	0	1
1	Q	0	1
3	C	0	1
All	All	0	6

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	L	234	GLY	CA-C	-22.83	1.15	1.51
5	N	24	TRP	CB-CG	-6.19	1.39	1.50

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	L	234	GLY	CA-C-O	-12.45	98.19	120.60
4	L	234	GLY	CA-C-N	9.64	138.40	117.20
4	L	234	GLY	N-CA-C	8.16	133.51	113.10
4	L	302	ARG	NE-CZ-NH2	-7.22	116.69	120.30
3	K	685	ARG	NE-CZ-NH2	5.79	123.20	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	441	GLN	Peptide
1	A	478	TYR	Peptide
1	A	480	TYR	Peptide
1	M	579	MET	Peptide
1	Q	129	ARG	Peptide

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	860	0	729	20	0
1	M	962	0	898	13	0
1	Q	496	0	450	11	0
2	E	181	0	170	3	0
3	C	1307	0	1048	23	0
3	K	2406	0	2122	65	0
4	L	2877	0	2749	68	0
5	N	3058	0	2904	81	0
6	O	51	0	54	0	0
7	P	484	0	476	8	0
8	Z	674	0	167	5	0
9	B	84	0	65	3	0
All	All	13440	0	11832	274	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 11.

The worst 5 of 274 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)
5:N:152:PHE:CD1	5:N:170:ASP:OD2	1.71	1.43
3:K:443:LEU:HB3	3:K:447:TYR:CE2	1.61	1.36
3:K:443:LEU:CB	3:K:447:TYR:CE2	2.09	1.34
3:K:443:LEU:CB	3:K:447:TYR:HE2	1.41	1.32
3:K:662:MET:SD	3:K:732:ASP:OD1	1.92	1.28

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 PDB entries followed by that with respect to all EM entries.

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	121/739 (16%)	89 (74%)	28 (23%)	4 (3%)	4	39
1	M	120/739 (16%)	105 (88%)	13 (11%)	2 (2%)	11	52
1	Q	64/739 (9%)	48 (75%)	15 (23%)	1 (2%)	11	53
2	E	25/348 (7%)	15 (60%)	8 (32%)	2 (8%)	1	17
3	C	189/746 (25%)	160 (85%)	24 (13%)	5 (3%)	6	44
3	K	319/746 (43%)	270 (85%)	46 (14%)	3 (1%)	20	63
4	L	360/441 (82%)	305 (85%)	51 (14%)	4 (1%)	17	60
5	N	382/425 (90%)	335 (88%)	42 (11%)	5 (1%)	14	57
6	O	5/7 (71%)	4 (80%)	1 (20%)	0	100	100
7	P	62/295 (21%)	48 (77%)	14 (23%)	0	100	100
9	B	10/345 (3%)	9 (90%)	1 (10%)	0	100	100
All	All	1657/5570 (30%)	1388 (84%)	243 (15%)	26 (2%)	16	53

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	481	HIS
4	L	400	LYS
1	M	581	VAL
5	N	171	LEU
3	C	127	VAL

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 PDB entries followed by that with respect to all EM entries.

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	72/646 (11%)	72 (100%)	0	100	100
1	M	101/646 (16%)	100 (99%)	1 (1%)	80	90
1	Q	45/646 (7%)	45 (100%)	0	100	100
2	E	18/297 (6%)	17 (94%)	1 (6%)	25	61
3	C	100/667 (15%)	99 (99%)	1 (1%)	80	90
3	K	233/667 (35%)	227 (97%)	6 (3%)	51	78
4	L	310/392 (79%)	305 (98%)	5 (2%)	68	86
5	N	340/375 (91%)	335 (98%)	5 (2%)	70	87
6	O	3/3 (100%)	3 (100%)	0	100	100
7	P	50/263 (19%)	48 (96%)	2 (4%)	36	69
9	B	3/294 (1%)	3 (100%)	0	100	100
All	All	1275/4896 (26%)	1254 (98%)	21 (2%)	70	86

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

Mol	Chain	Res	Type
4	L	342	LYS
4	L	436	ARG
5	N	387	PRO
4	L	269	ARG
7	P	236	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
5	N	188	ASN
1	Q	83	HIS
5	N	261	ASN
4	L	213	HIS
5	N	226	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
9	M3L	B	116	9,3	11,11,12	0.62	0	11,14,16	0.82	1 (9%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
9	M3L	B	116	9,3	-	0/8/10/12	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
9	B	116	M3L	O-C-CA	-2.01	119.46	125.02

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
9	B	116	M3L	2	0

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
8	Z	6

The worst 5 of 6 chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	Z	182:UNK	C	247:UNK	N	39.79
1	Z	45:UNK	C	55:UNK	N	32.97
1	Z	67:UNK	C	77:UNK	N	29.83
1	Z	138:UNK	C	143:UNK	N	25.66
1	Z	99:UNK	C	120:UNK	N	10.57