wwPDB EM Map/Model Validation Report (i)

Apr 10, 2016 – 01:56 PM BST

PDB ID	:	3J4U
EMDB ID:	:	EMD-5764
Title	:	A new topology of the HK97-like fold revealed in Bordetella bacteriophage: non-covalent chainmail secured by jellyrolls
Authors	:	Zhang, X.; Guo, H.; Jin, L.; Czornyj, E.; Hodes, A.; Hui, W.H.; Nieh, A.W.; Miller, J.F.; Zhou, Z.H.
Deposited on	:	2013-10-09
Resolution	:	3.50 Å(reported)

This is a wwPDB EM Map/Model Validation Report for a publicly released PDB/EMDB entry.
For rigid body fitted models, validation errors reported here could
stem from errors in the original structure(s) used in the fitting.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/EMValidationReportHelp

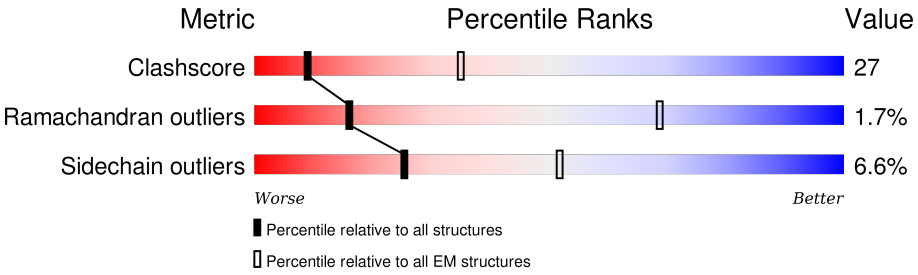
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk27241

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 3.50 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	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

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 <=5%

Mol	Chain	Length	Quality of chain
1	A	331	45% 48% 6% •
1	B	331	51% 40% 7% ••
1	C	331	44% 49% 5% ••
1	D	331	47% 47% 5% ••
1	E	331	47% 47% 5% ••
1	F	331	41% 50% 6% ••
1	G	331	49% 42% • • 6%
2	H	140	50% 47% •
2	I	140	59% 39% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	J	140	56% 39% 5%
2	K	140	61% 37%
2	L	140	55% 40% 5%
2	M	140	50% 48%
2	N	140	55% 41%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 24653 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 major capsid protein.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	327	Total	C	N	O	S	0	0
			2525	1574	444	495	12		
1	B	327	Total	C	N	O	S	0	0
			2525	1574	444	495	12		
1	C	327	Total	C	N	O	S	0	0
			2525	1574	444	495	12		
1	D	327	Total	C	N	O	S	0	0
			2525	1574	444	495	12		
1	E	327	Total	C	N	O	S	0	0
			2525	1574	444	495	12		
1	F	327	Total	C	N	O	S	0	0
			2525	1574	444	495	12		
1	G	312	Total	C	N	O	S	0	0
			2419	1508	427	473	11		

- Molecule 2 is a protein called cementing protein.

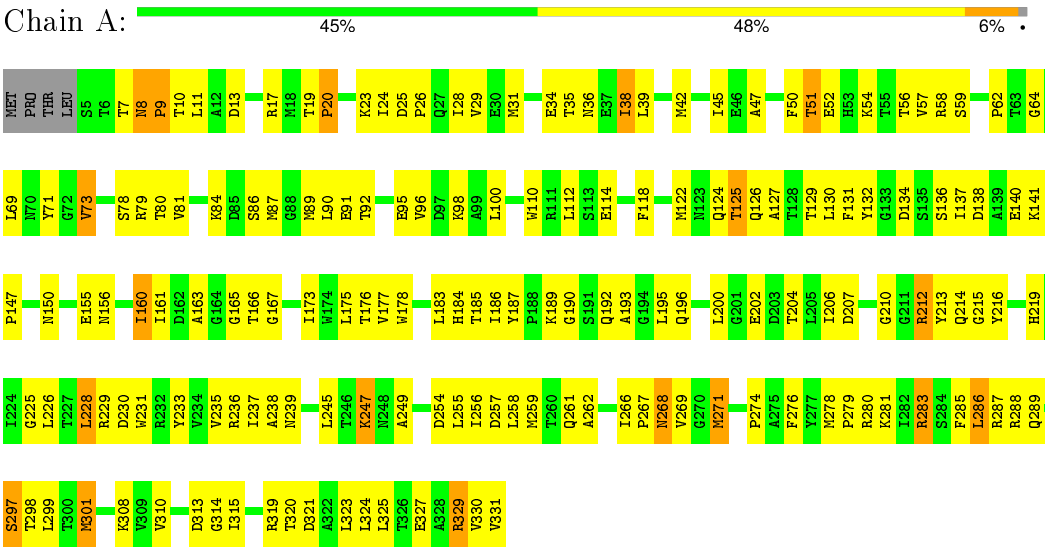
Mol	Chain	Residues	Atoms					AltConf	Trace
2	H	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		
2	I	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		
2	J	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		
2	K	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		
2	L	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		
2	M	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		
2	N	140	Total	C	N	O	S	0	0
			1012	628	174	205	5		

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 errors 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.

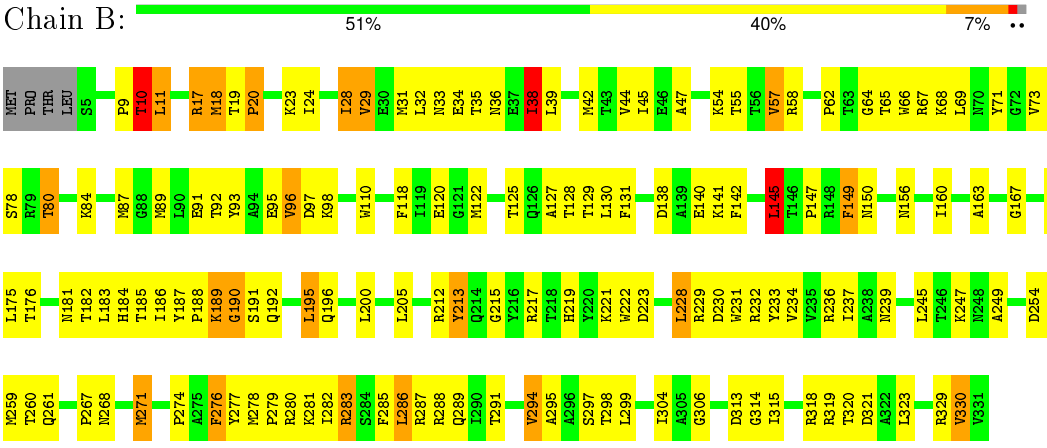
- Molecule 1: major capsid protein

Chain A:



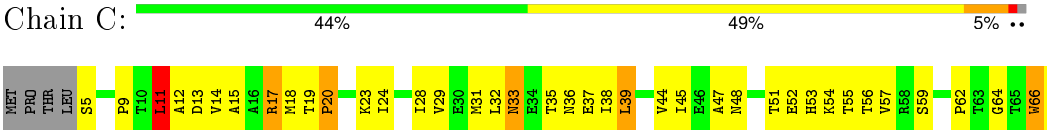
- Molecule 1: major capsid protein

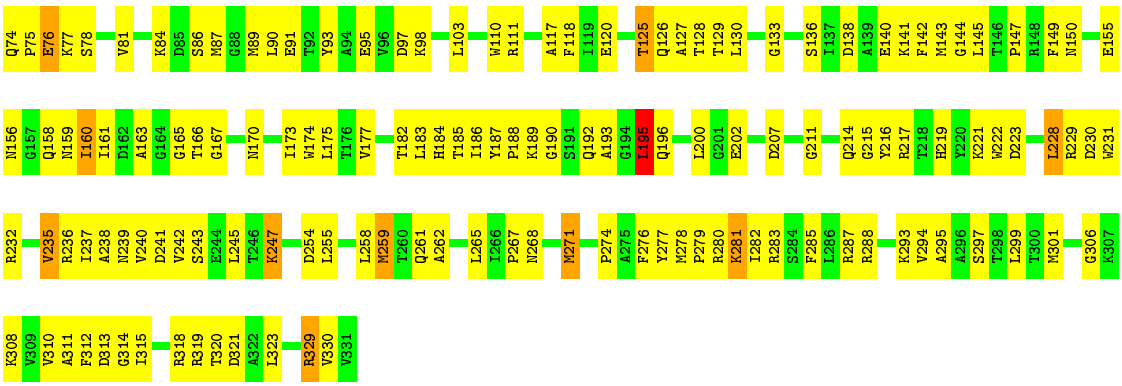
Chain B:



- Molecule 1: major capsid protein

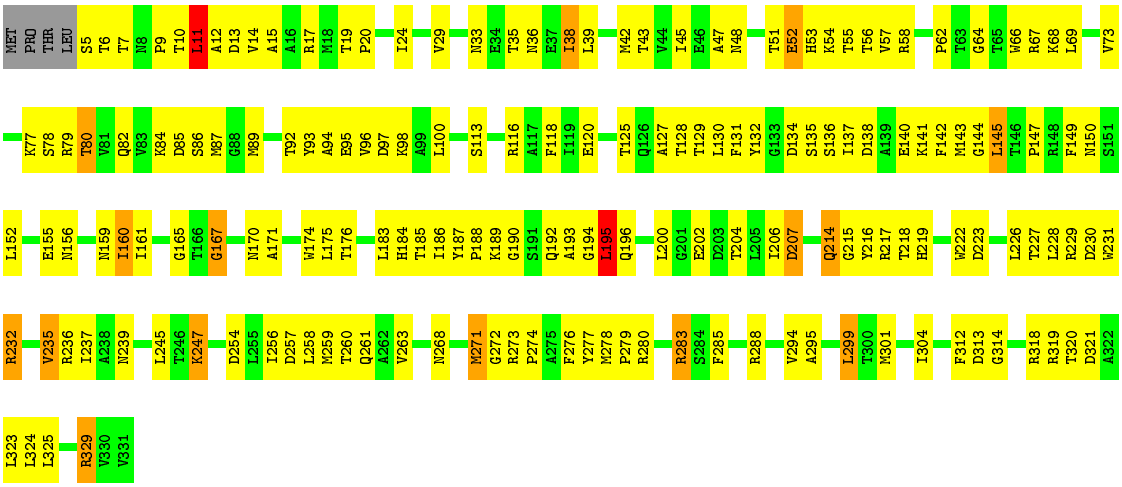
Chain C:





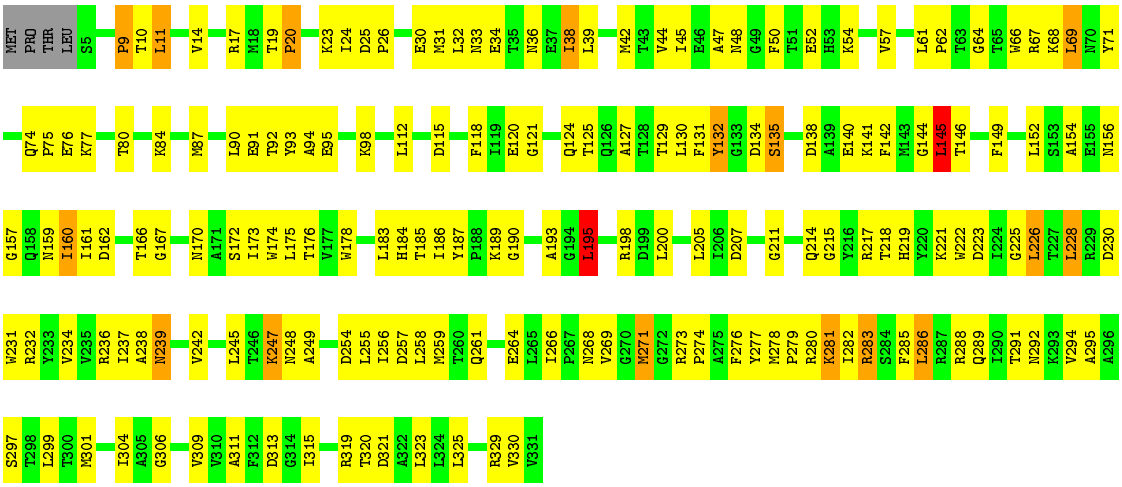
• Molecule 1: major capsid protein

Chain D:



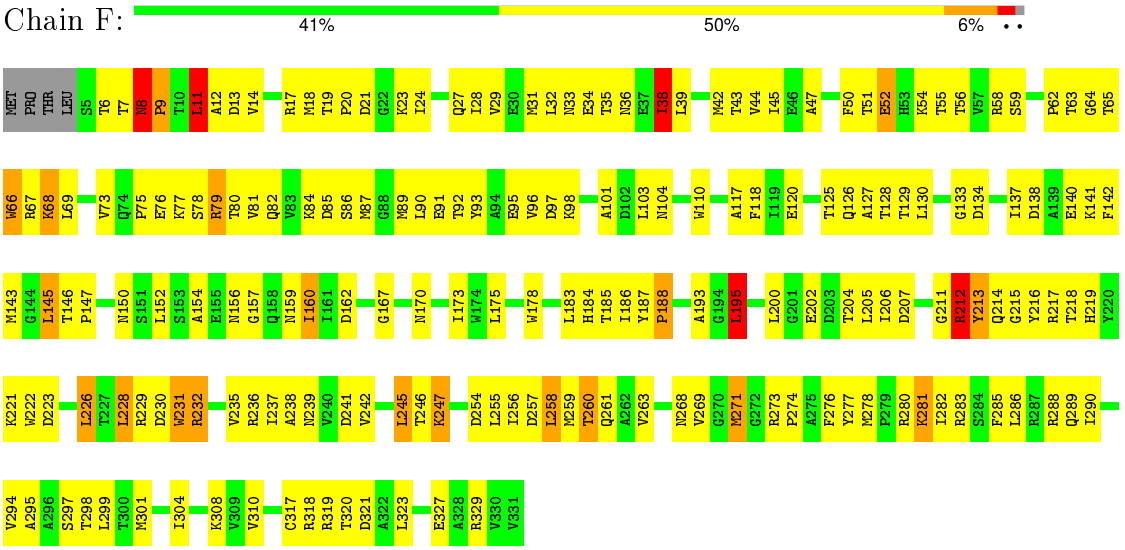
• Molecule 1: major capsid protein

Chain E:



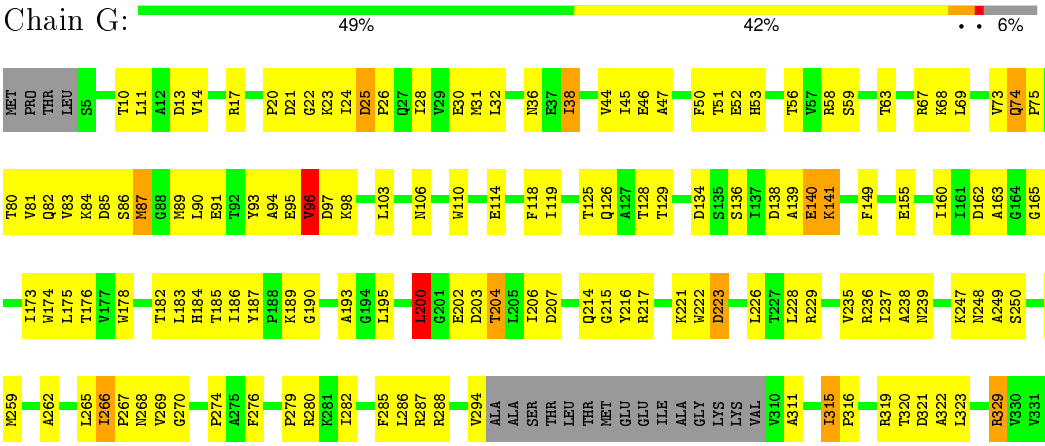
• Molecule 1: major capsid protein

Chain F:



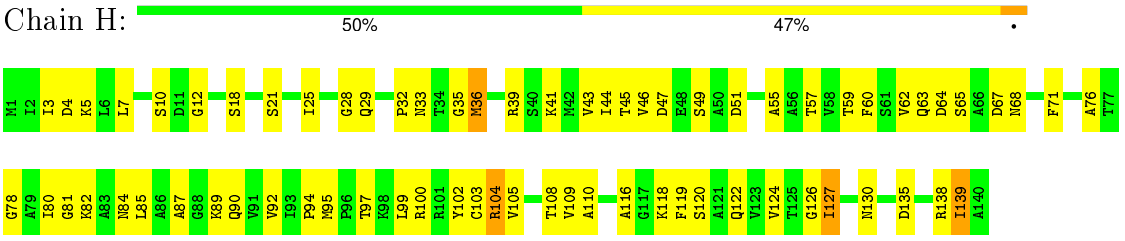
• Molecule 1: major capsid protein

Chain G:



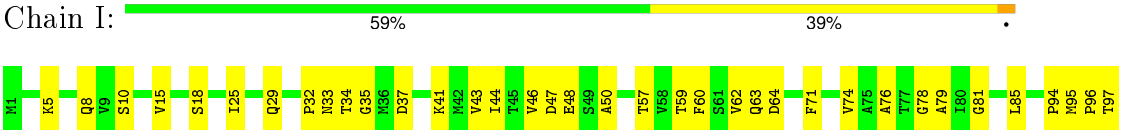
• Molecule 2: cementing protein

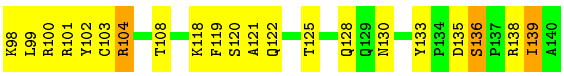
Chain H:



• Molecule 2: cementing protein

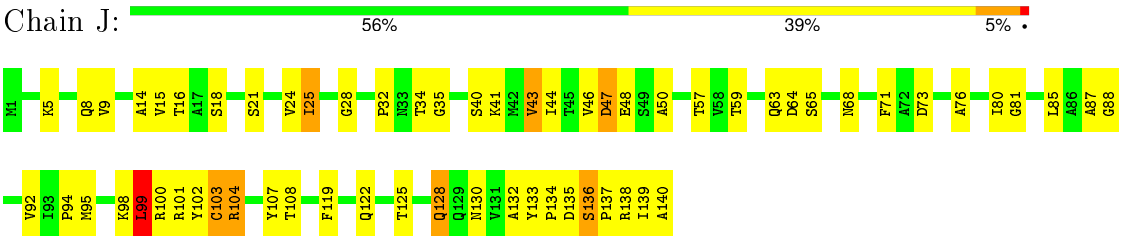
Chain I:





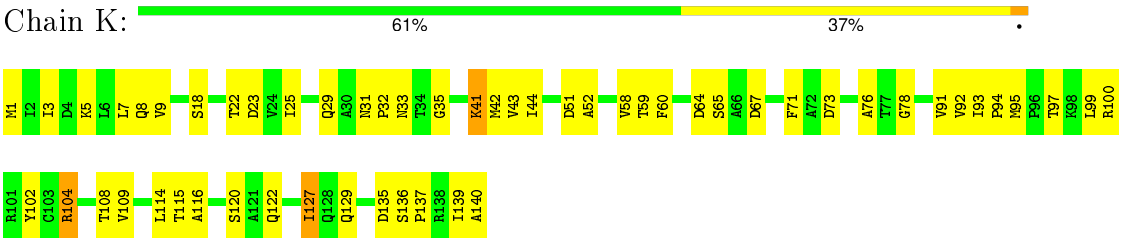
• Molecule 2: cementing protein

Chain J:



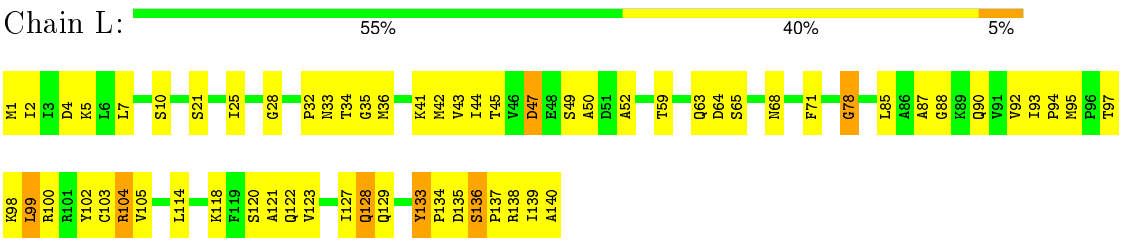
• Molecule 2: cementing protein

Chain K:



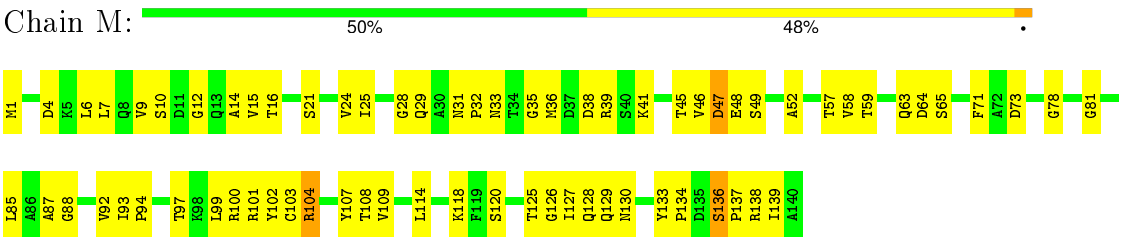
• Molecule 2: cementing protein

Chain L:



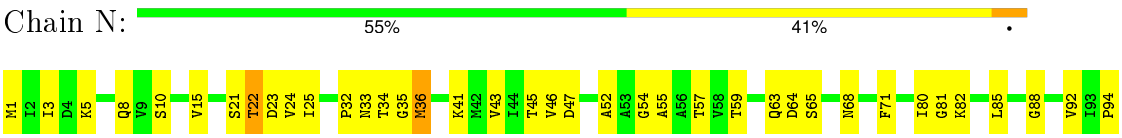
• Molecule 2: cementing protein

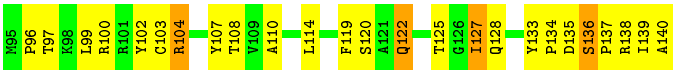
Chain M:



• Molecule 2: cementing protein

Chain N:




M95
P96
T97
K98
L99
R100
R101
Y102
C103
R104
Y107
T108
Y109
A110
L114
F119
S120
A121
Q122
T125
G126
I127
Q128
Y133
P134
D135
S136
P137
R138
I139
A140

4 Experimental information (i)

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of images	39549	Depositor
Resolution determination method	FSC at 0.143 cut-off, feature-based method and R-factor in model refinement	Depositor
CTF correction method	Each particle	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/Å^2$)	25	Depositor
Minimum defocus (nm)	800	Depositor
Maximum defocus (nm)	2300	Depositor
Magnification	59000	Depositor
Image detector	Kodak SO163 film	Depositor

5 Model quality

5.1 Standard geometry

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	0.81	0/2568	1.07	6/3484 (0.2%)
1	B	0.86	5/2568 (0.2%)	1.09	9/3484 (0.3%)
1	C	0.79	3/2568 (0.1%)	1.04	5/3484 (0.1%)
1	D	0.86	4/2568 (0.2%)	1.09	6/3484 (0.2%)
1	E	0.86	4/2568 (0.2%)	1.12	8/3484 (0.2%)
1	F	0.84	5/2568 (0.2%)	1.10	9/3484 (0.3%)
1	G	0.87	5/2461 (0.2%)	1.07	5/3339 (0.1%)
2	H	0.72	0/1025	0.97	0/1397
2	I	0.72	1/1025 (0.1%)	0.93	0/1397
2	J	0.66	0/1025	0.93	1/1397 (0.1%)
2	K	0.68	0/1025	0.91	1/1397 (0.1%)
2	L	0.61	0/1025	0.87	1/1397 (0.1%)
2	M	0.65	0/1025	0.91	0/1397
2	N	0.65	0/1025	0.90	1/1397 (0.1%)
All	All	0.80	27/25044 (0.1%)	1.04	52/34022 (0.2%)

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	G	0	1

The worst 5 of 27 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	214	GLN	CB-CG	8.11	1.74	1.52
2	I	37	ASP	CB-CG	7.70	1.68	1.51
1	G	140	GLU	CB-CG	7.64	1.66	1.52
1	G	140	GLU	CG-CD	7.56	1.63	1.51
1	E	214	GLN	CG-CD	7.50	1.68	1.51

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	8	ASN	C-N-CD	-9.91	98.79	120.60
1	B	212	ARG	NE-CZ-NH2	9.72	125.16	120.30
1	E	11	LEU	CA-CB-CG	-8.29	96.23	115.30
1	D	195	LEU	CA-CB-CG	-8.18	96.48	115.30
1	C	195	LEU	CA-CB-CG	-7.74	97.50	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	21	ASP	Peptide

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	2525	0	2501	173	0
1	B	2525	0	2501	167	0
1	C	2525	0	2501	177	0
1	D	2525	0	2501	185	0
1	E	2525	0	2501	164	0
1	F	2525	0	2501	180	0
1	G	2419	0	2385	106	0
2	H	1012	0	1012	55	0
2	I	1012	0	1012	48	0
2	J	1012	0	1012	65	0
2	K	1012	0	1012	51	0
2	L	1012	0	1012	57	0
2	M	1012	0	1012	65	0
2	N	1012	0	1012	56	0
All	All	24653	0	24475	1327	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 27.

The worst 5 of 1327 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:E:295:ALA:HA	1:E:299:LEU:HD22	1.44	0.99
1:F:295:ALA:HA	1:F:299:LEU:HD22	1.46	0.96
1:F:54:LYS:HD3	1:F:84:LYS:HB3	1.45	0.96
1:D:86:SER:H	2:J:128:GLN:HG2	1.28	0.95
2:J:64:ASP:HB3	2:J:100:ARG:HD3	1.48	0.95

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	325/331 (98%)	288 (89%)	31 (10%)	6 (2%)	11 53
1	B	325/331 (98%)	283 (87%)	35 (11%)	7 (2%)	8 49
1	C	325/331 (98%)	289 (89%)	29 (9%)	7 (2%)	8 49
1	D	325/331 (98%)	280 (86%)	39 (12%)	6 (2%)	11 53
1	E	325/331 (98%)	285 (88%)	33 (10%)	7 (2%)	8 49
1	F	325/331 (98%)	282 (87%)	36 (11%)	7 (2%)	8 49
1	G	308/331 (93%)	272 (88%)	29 (9%)	7 (2%)	8 48
2	H	138/140 (99%)	123 (89%)	14 (10%)	1 (1%)	26 72
2	I	138/140 (99%)	124 (90%)	12 (9%)	2 (1%)	14 58
2	J	138/140 (99%)	122 (88%)	15 (11%)	1 (1%)	26 72
2	K	138/140 (99%)	124 (90%)	14 (10%)	0	100 100
2	L	138/140 (99%)	119 (86%)	18 (13%)	1 (1%)	26 72
2	M	138/140 (99%)	123 (89%)	14 (10%)	1 (1%)	26 72
2	N	138/140 (99%)	122 (88%)	14 (10%)	2 (1%)	14 58
All	All	3224/3297 (98%)	2836 (88%)	333 (10%)	55 (2%)	16 54

5 of 55 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	9	PRO
1	A	268	ASN
1	B	10	THR
1	B	268	ASN
1	D	195	LEU

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 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	270/274 (98%)	253 (94%)	17 (6%)	22 63
1	B	270/274 (98%)	249 (92%)	21 (8%)	16 53
1	C	270/274 (98%)	250 (93%)	20 (7%)	17 56
1	D	270/274 (98%)	256 (95%)	14 (5%)	29 68
1	E	270/274 (98%)	258 (96%)	12 (4%)	35 73
1	F	270/274 (98%)	248 (92%)	22 (8%)	15 52
1	G	259/274 (94%)	228 (88%)	31 (12%)	6 30
2	H	106/106 (100%)	102 (96%)	4 (4%)	40 76
2	I	106/106 (100%)	103 (97%)	3 (3%)	51 82
2	J	106/106 (100%)	96 (91%)	10 (9%)	11 44
2	K	106/106 (100%)	103 (97%)	3 (3%)	51 82
2	L	106/106 (100%)	100 (94%)	6 (6%)	25 66
2	M	106/106 (100%)	102 (96%)	4 (4%)	40 76
2	N	106/106 (100%)	101 (95%)	5 (5%)	32 72
All	All	2621/2660 (98%)	2449 (93%)	172 (7%)	25 61

5 of 172 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	160	ILE
1	F	213	TYR
2	L	47	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	247	LYS
1	F	38	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	184	HIS
2	H	29	GLN
2	J	128	GLN
1	E	192	GLN
1	E	219	HIS

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.