



wwPDB EM Validation Summary Report ⓘ

Feb 24, 2025 – 12:06 pm GMT

PDB ID : 9I8I
EMDB ID : EMD-52724
Title : cryoEM structure of HIV-1 KAKA/G225R mature CA hexamer
Authors : Zhu, Y.; Shen, J.; Shen, Y.; Xu, J.; Zhang, P.
Deposited on : 2025-02-05
Resolution : 2.75 Å(reported)

This is a wwPDB EM 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/EMValidationReportHelp>

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:

EMDB validation analysis : **FAILED**
MolProbity : 4.02b-467
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
MapQ : **FAILED**
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.41

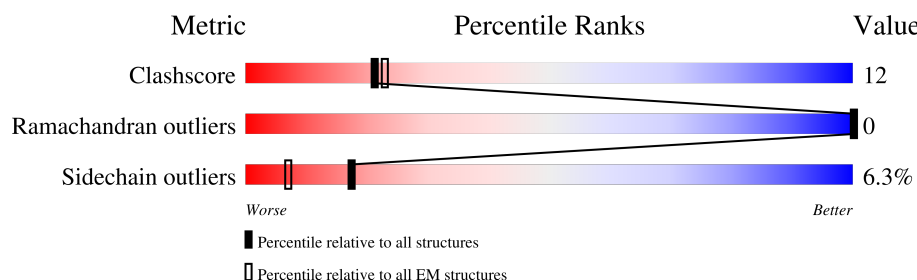
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

ELECTRON MICROSCOPY

The reported resolution of this entry is 2.75 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	210492	15764
Ramachandran outliers	207382	16835
Sidechain outliers	206894	16415

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	231	<div> <div>65%</div> <div>30%</div> <div>..</div> </div>
1	B	231	<div> <div>68%</div> <div>27%</div> <div>..</div> </div>
1	C	231	<div> <div>67%</div> <div>29%</div> <div>..</div> </div>
1	D	231	<div> <div>70%</div> <div>24%</div> <div>..</div> </div>
1	E	231	<div> <div>68%</div> <div>28%</div> <div>..</div> </div>
1	F	231	<div> <div>72%</div> <div>23%</div> <div>..</div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 10374 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.

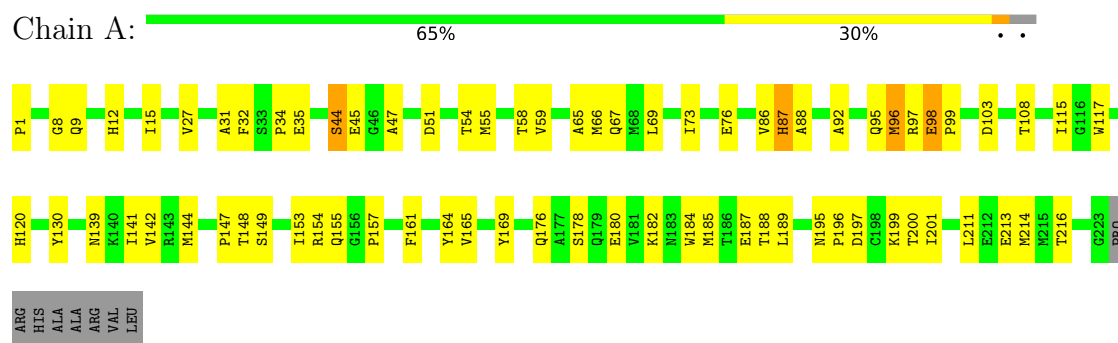
- Molecule 1 is a protein called HIV-1 KAKA/G225R CA hexamer.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	223	Total	C	N	O	S	0	0
			1729	1089	302	325	13		
1	F	223	Total	C	N	O	S	0	0
			1729	1089	302	325	13		
1	E	223	Total	C	N	O	S	0	0
			1729	1089	302	325	13		
1	D	223	Total	C	N	O	S	0	0
			1729	1089	302	325	13		
1	C	223	Total	C	N	O	S	0	0
			1729	1089	302	325	13		
1	B	223	Total	C	N	O	S	0	0
			1729	1089	302	325	13		

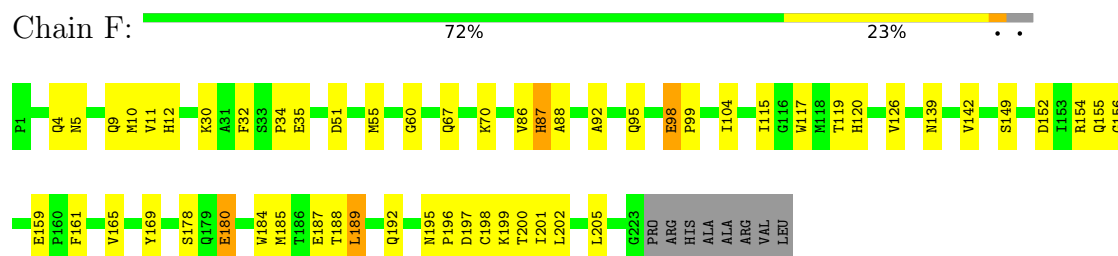
3 Residue-property plots

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. 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.

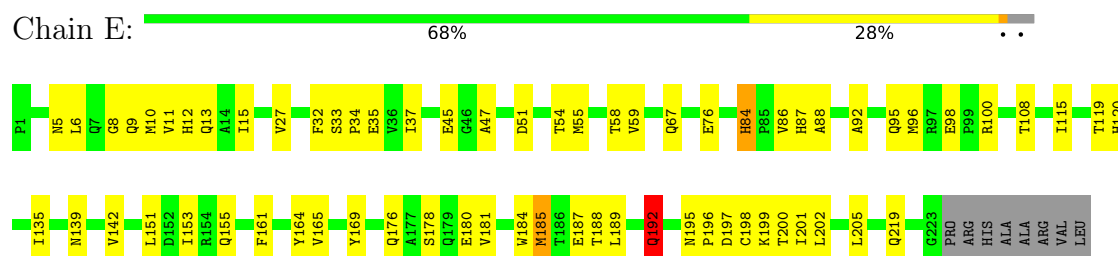
- Molecule 1: HIV-1 KAKA/G225R CA hexamer



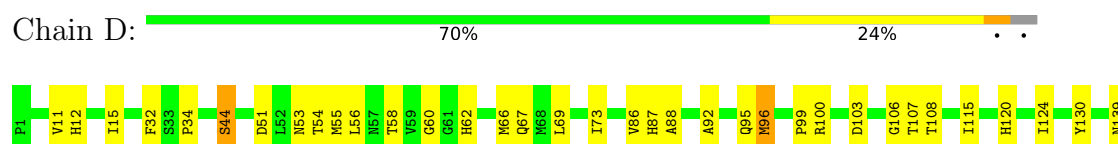
- Molecule 1: HIV-1 KAKA/G225R CA hexamer

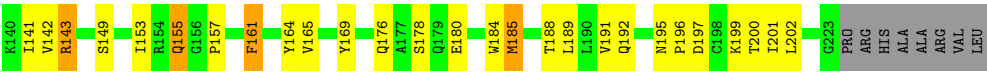


- Molecule 1: HIV-1 KAKA/G225R CA hexamer

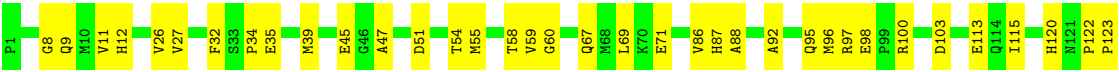


- Molecule 1: HIV-1 KAKA/G225R CA hexamer

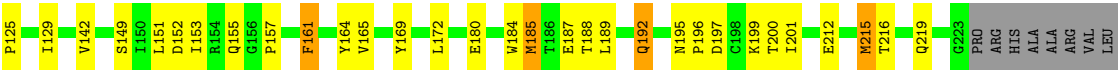
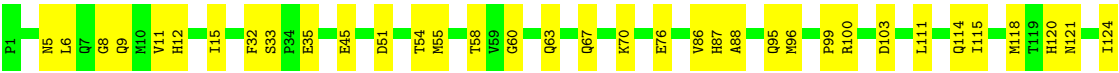




• Molecule 1: HIV-1 KAKA/G225R CA hexamer



• Molecule 1: HIV-1 KAKA/G225R CA hexamer



4 Experimental information

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C6	Depositor
Number of particles used	1646193	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	TFS KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ($e^-/\text{\AA}^2$)	40	Depositor
Minimum defocus (nm)	800	Depositor
Maximum defocus (nm)	2000	Depositor
Magnification	Not provided	
Image detector	GATAN K3 BIOQUANTUM (6k x 4k)	Depositor

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	0.25	0/1768	0.51	0/2404
1	B	0.26	0/1768	0.53	1/2404 (0.0%)
1	C	0.26	0/1768	0.53	0/2404
1	D	0.25	0/1768	0.50	0/2404
1	E	0.26	0/1768	0.50	1/2404 (0.0%)
1	F	0.25	0/1768	0.51	0/2404
All	All	0.25	0/10608	0.51	2/14424 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	192	GLN	CA-CB-CG	5.09	124.60	113.40
1	B	172	LEU	CA-CB-CG	5.07	126.96	115.30

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	1729	0	1725	48	0
1	B	1729	0	1725	38	0
1	C	1729	0	1725	49	0
1	D	1729	0	1725	40	0
1	E	1729	0	1725	40	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	F	1729	0	1725	36	0
All	All	10374	0	10350	240	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 12.

The worst 5 of 240 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:C:155:GLN:NE2	1:C:156:GLY:O	2.16	0.79
1:F:195:ASN:ND2	1:F:197:ASP:OD2	2.18	0.74
1:E:10:MET:HG3	1:E:119:THR:HG21	1.70	0.74
1:A:184:TRP:O	1:A:188:THR:OG1	2.05	0.74
1:C:192:GLN:O	1:C:192:GLN:NE2	2.20	0.72

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	221/231 (96%)	206 (93%)	15 (7%)	0	100	100
1	B	221/231 (96%)	211 (96%)	10 (4%)	0	100	100
1	C	221/231 (96%)	208 (94%)	13 (6%)	0	100	100
1	D	221/231 (96%)	210 (95%)	11 (5%)	0	100	100
1	E	221/231 (96%)	210 (95%)	11 (5%)	0	100	100
1	F	221/231 (96%)	208 (94%)	13 (6%)	0	100	100
All	All	1326/1386 (96%)	1253 (94%)	73 (6%)	0	100	100

There are no Ramachandran outliers to report.

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	187/193 (97%)	176 (94%)	11 (6%)	16	30
1	B	187/193 (97%)	172 (92%)	15 (8%)	10	18
1	C	187/193 (97%)	176 (94%)	11 (6%)	16	30
1	D	187/193 (97%)	175 (94%)	12 (6%)	14	27
1	E	187/193 (97%)	175 (94%)	12 (6%)	14	27
1	F	187/193 (97%)	177 (95%)	10 (5%)	19	34
All	All	1122/1158 (97%)	1051 (94%)	71 (6%)	17	27

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	63	GLN
1	B	76	GLU
1	B	180	GLU
1	E	84	HIS
1	E	76	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	9	GLN
1	E	155	GLN
1	D	63	GLN
1	B	63	GLN

5.3.3 RNA ⓘ

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](#)

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.