



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2017 – 08:14 AM EDT

PDB ID : 3GUT
Title : Crystal structure of a higher-order complex of p50:RelA bound to the HIV-1 LTR
Authors : Stroud, J.C.; Oltman, A.J.; Han, A.; Bates, D.L.; Chen, L.
Deposited on : unknown
Resolution : 3.59 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

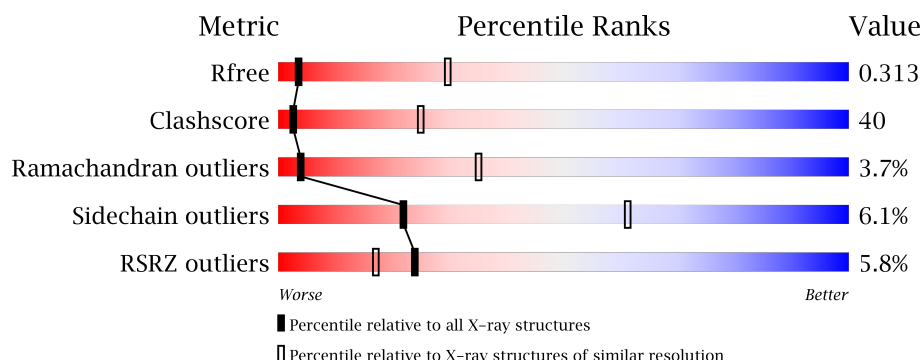
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.59 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}	100719	1026 (3.74-3.46)
Clashscore	112137	1036 (3.70-3.50)
Ramachandran outliers	110173	1030 (3.72-3.48)
Sidechain outliers	110143	1030 (3.72-3.48)
RSRZ outliers	101464	1051 (3.74-3.46)

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. 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	273	<div> <div>7%</div> <div>44%</div> <div>51%</div> <div>.</div> </div>
1	C	273	<div> <div>%</div> <div>42%</div> <div>49%</div> <div>8%</div> </div>
1	E	273	<div> <div>9%</div> <div>41%</div> <div>52%</div> <div>7%</div> </div>
1	G	273	<div> <div>%</div> <div>44%</div> <div>48%</div> <div>8%</div> </div>
2	B	312	<div> <div>6%</div> <div>50%</div> <div>46%</div> <div>..</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	312	
2	F	312	
2	H	312	
3	I	26	
3	X	26	
4	J	26	
4	Y	26	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 20640 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 Transcription factor p65.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	273	Total	C	N	O	S	0	0	0
			2176	1356	401	408	11			
1	C	273	Total	C	N	O	S	0	0	0
			2176	1356	401	408	11			
1	E	273	Total	C	N	O	S	0	0	0
			2176	1356	401	408	11			
1	G	273	Total	C	N	O	S	0	0	0
			2176	1356	401	408	11			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	19	ALA	-	EXPRESSION TAG	UNP Q04206
A	99	TYR	PHE	SEE REMARK 999	UNP Q04206
A	103	ASP	GLU	SEE REMARK 999	UNP Q04206
A	109	SER	CYS	SEE REMARK 999	UNP Q04206
A	142	HIS	GLN	SEE REMARK 999	UNP Q04206
A	169	ALA	SER	SEE REMARK 999	UNP Q04206
A	174	LEU	ARG	SEE REMARK 999	UNP Q04206
A	176	THR	PRO	SEE REMARK 999	UNP Q04206
C	19	ALA	-	EXPRESSION TAG	UNP Q04206
C	99	TYR	PHE	SEE REMARK 999	UNP Q04206
C	103	ASP	GLU	SEE REMARK 999	UNP Q04206
C	109	SER	CYS	SEE REMARK 999	UNP Q04206
C	142	HIS	GLN	SEE REMARK 999	UNP Q04206
C	169	ALA	SER	SEE REMARK 999	UNP Q04206
C	174	LEU	ARG	SEE REMARK 999	UNP Q04206
C	176	THR	PRO	SEE REMARK 999	UNP Q04206
E	19	ALA	-	EXPRESSION TAG	UNP Q04206
E	99	TYR	PHE	SEE REMARK 999	UNP Q04206
E	103	ASP	GLU	SEE REMARK 999	UNP Q04206
E	109	SER	CYS	SEE REMARK 999	UNP Q04206
E	142	HIS	GLN	SEE REMARK 999	UNP Q04206

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	169	ALA	SER	SEE REMARK 999	UNP Q04206
E	174	LEU	ARG	SEE REMARK 999	UNP Q04206
E	176	THR	PRO	SEE REMARK 999	UNP Q04206
G	19	ALA	-	EXPRESSION TAG	UNP Q04206
G	99	TYR	PHE	SEE REMARK 999	UNP Q04206
G	103	ASP	GLU	SEE REMARK 999	UNP Q04206
G	109	SER	CYS	SEE REMARK 999	UNP Q04206
G	142	HIS	GLN	SEE REMARK 999	UNP Q04206
G	169	ALA	SER	SEE REMARK 999	UNP Q04206
G	174	LEU	ARG	SEE REMARK 999	UNP Q04206
G	176	THR	PRO	SEE REMARK 999	UNP Q04206

- Molecule 2 is a protein called Nuclear factor NF-kappa-B p105 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	312	Total	C	N	O	S	0	0	0
			2454	1554	428	460	12			
2	D	312	Total	C	N	O	S	0	0	0
			2454	1554	428	460	12			
2	F	312	Total	C	N	O	S	0	0	0
			2454	1554	428	460	12			
2	H	312	Total	C	N	O	S	0	0	0
			2454	1554	428	460	12			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	420	VAL	ILE	SEE REMARK 999	UNP P19838
B	471	SER	PRO	SEE REMARK 999	UNP P19838
B	487	THR	GLY	SEE REMARK 999	UNP P19838
B	493	ILE	LEU	SEE REMARK 999	UNP P19838
B	499	VAL	LEU	SEE REMARK 999	UNP P19838
B	619	VAL	ILE	SEE REMARK 999	UNP P19838
D	420	VAL	ILE	SEE REMARK 999	UNP P19838
D	471	SER	PRO	SEE REMARK 999	UNP P19838
D	487	THR	GLY	SEE REMARK 999	UNP P19838
D	493	ILE	LEU	SEE REMARK 999	UNP P19838
D	499	VAL	LEU	SEE REMARK 999	UNP P19838
D	619	VAL	ILE	SEE REMARK 999	UNP P19838
F	420	VAL	ILE	SEE REMARK 999	UNP P19838
F	471	SER	PRO	SEE REMARK 999	UNP P19838
F	487	THR	GLY	SEE REMARK 999	UNP P19838

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	493	ILE	LEU	SEE REMARK 999	UNP P19838
F	499	VAL	LEU	SEE REMARK 999	UNP P19838
F	619	VAL	ILE	SEE REMARK 999	UNP P19838
H	420	VAL	ILE	SEE REMARK 999	UNP P19838
H	471	SER	PRO	SEE REMARK 999	UNP P19838
H	487	THR	GLY	SEE REMARK 999	UNP P19838
H	493	ILE	LEU	SEE REMARK 999	UNP P19838
H	499	VAL	LEU	SEE REMARK 999	UNP P19838
H	619	VAL	ILE	SEE REMARK 999	UNP P19838

- Molecule 3 is a DNA chain called HIV-LTR Core Forward Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	X	26	Total	C	N	O	P	0	0	0
			530	253	95	157	25			
3	I	26	Total	C	N	O	P	0	0	0
			530	253	95	157	25			

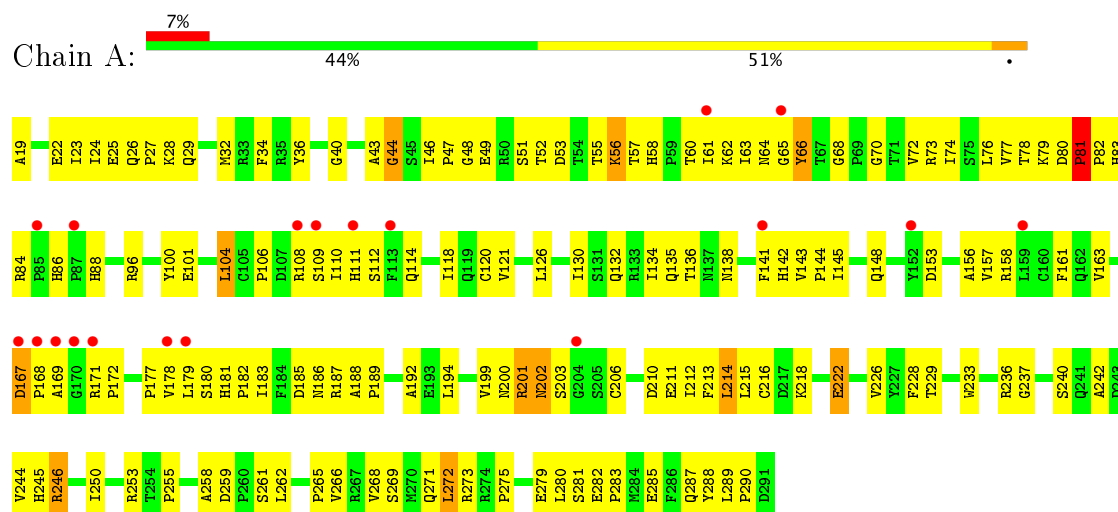
- Molecule 4 is a DNA chain called HIV-LTR Core Reverse Strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	Y	26	Total	C	N	O	P	0	0	0
			530	252	102	151	25			
4	J	26	Total	C	N	O	P	0	0	0
			530	252	102	151	25			

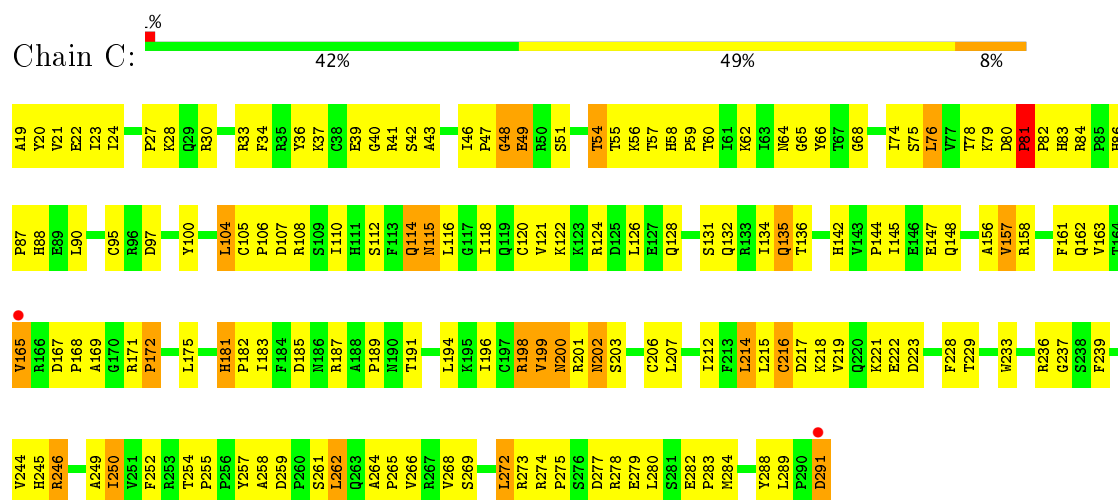
3 Residue-property plots

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 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: Transcription factor p65

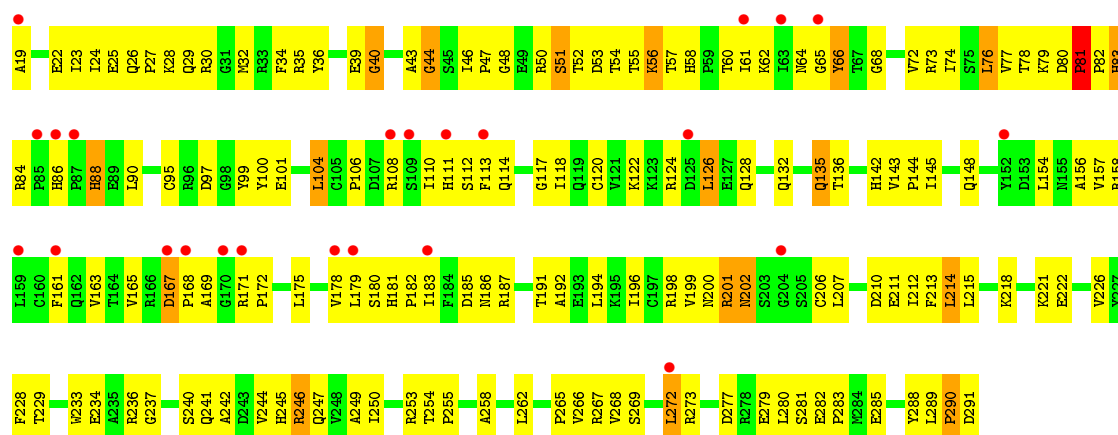


• Molecule 1: Transcription factor p65

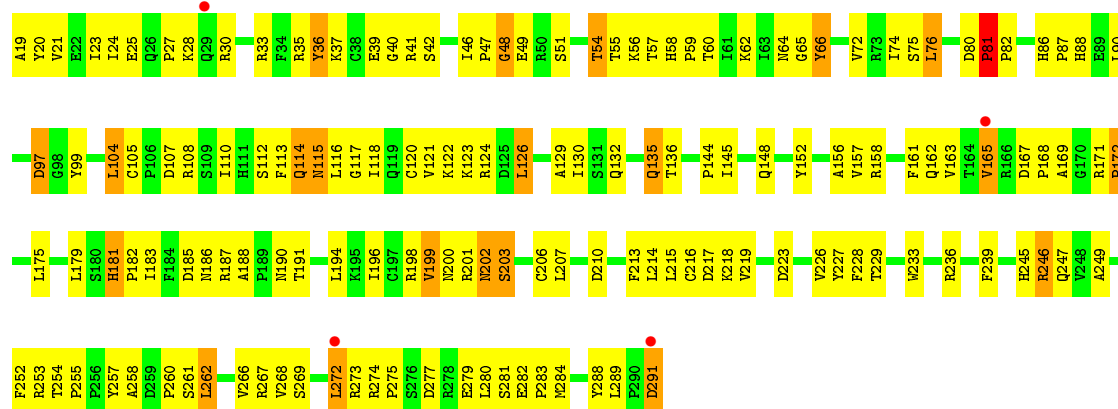
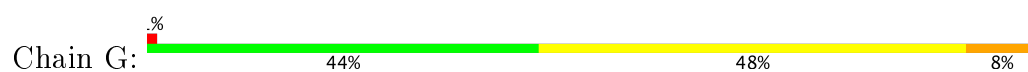


• Molecule 1: Transcription factor p65

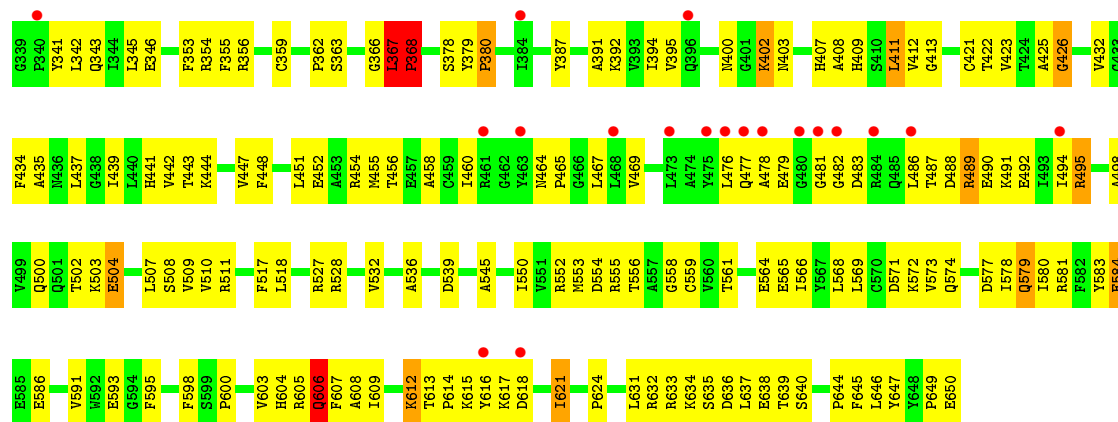




• Molecule 1: Transcription factor p65

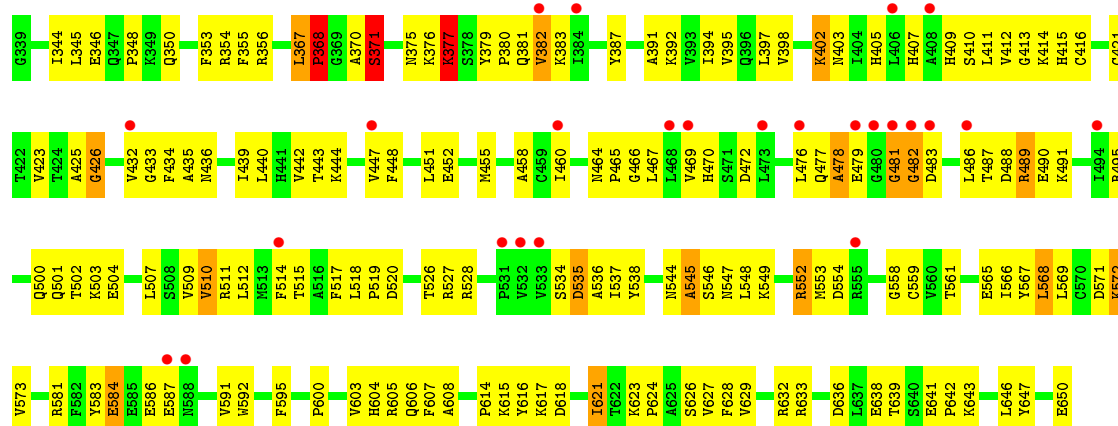


• Molecule 2: Nuclear factor NF-kappa-B p105 subunit

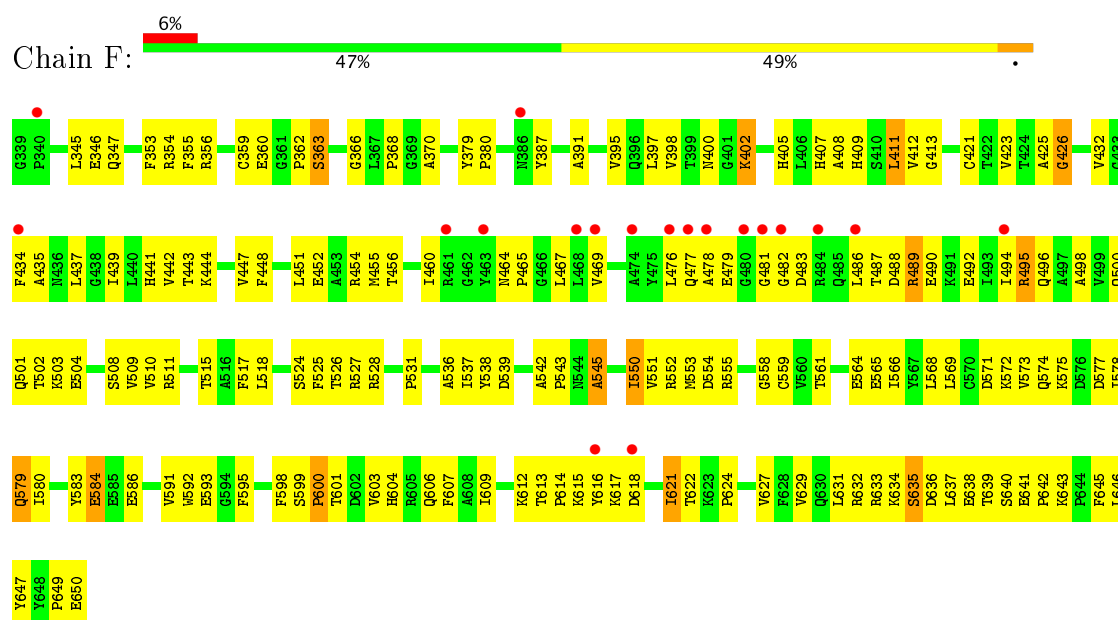


• Molecule 2: Nuclear factor NF-kappa-B p105 subunit

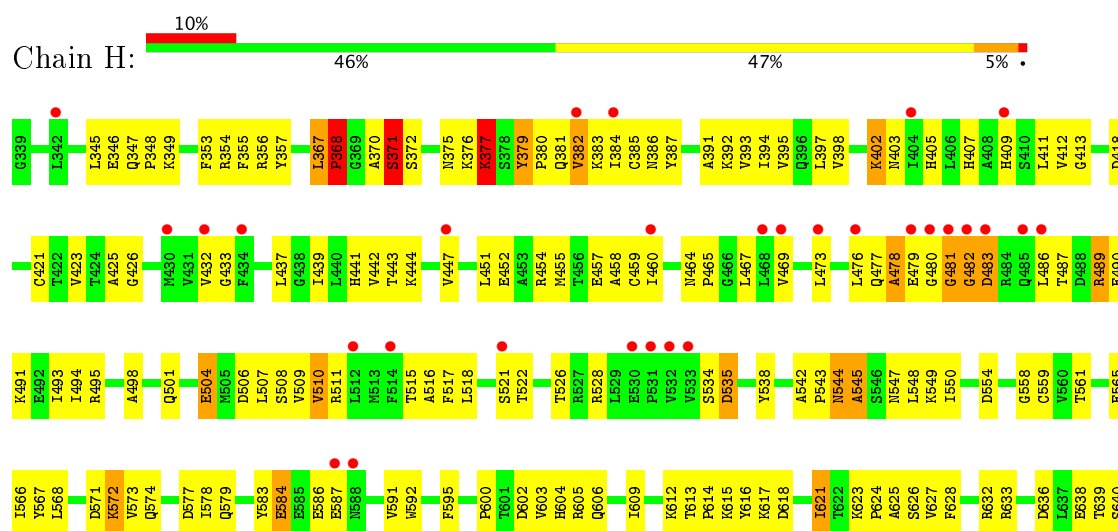




• Molecule 2: Nuclear factor NF-kappa-B p105 subunit



• Molecule 2: Nuclear factor NF-kappa-B p105 subunit





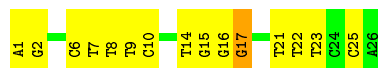
- Molecule 3: HIV-LTR Core Forward Strand

Chain X: 35% 58% 8%



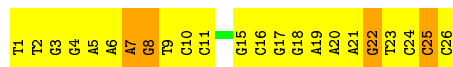
- Molecule 3: HIV-LTR Core Forward Strand

Chain I: 42% 54%



- Molecule 4: HIV-LTR Core Reverse Strand

Chain Y: 12% 73% 15%



- Molecule 4: HIV-LTR Core Reverse Strand

Chain J: 12% 73% 15%



4 Data and refinement statistics

Property	Value	Source
Space group	P 42	Depositor
Cell constants a, b, c, α , β , γ	167.49 Å 167.49 Å 172.76 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.45 – 3.59 46.89 – 3.59	Depositor EDS
% Data completeness (in resolution range)	81.2 (46.45-3.59) 91.9 (46.89-3.59)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.26 (at 3.57 Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.245 , 0.301 0.259 , 0.313	Depositor DCC
R_{free} test set	5197 reflections (10.15%)	DCC
Wilson B-factor (Å ²)	118.8	Xtriage
Anisotropy	0.196	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 85.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.011 for -h,-l,-k 0.005 for -h,l,k 0.006 for l,-k,h 0.012 for -l,-k,-h 0.438 for h,-k,-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	20640	wwPDB-VP
Average B, all atoms (Å ²)	150.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 40.03 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.9062e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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

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 > 5$	RMSZ	# $ Z > 5$
1	A	0.50	0/2228	0.75	1/3021 (0.0%)
1	C	0.54	0/2228	0.79	0/3021
1	E	0.50	0/2228	0.76	0/3021
1	G	0.52	0/2228	0.79	0/3021
2	B	0.67	3/2506 (0.1%)	0.94	6/3384 (0.2%)
2	D	0.63	3/2506 (0.1%)	0.97	11/3384 (0.3%)
2	F	0.54	0/2506	0.72	0/3384
2	H	0.65	3/2506 (0.1%)	0.98	12/3384 (0.4%)
3	I	0.87	0/593	0.97	0/914
3	X	0.89	0/593	1.00	0/914
4	J	0.91	1/595 (0.2%)	0.98	0/916
4	Y	0.95	2/595 (0.3%)	0.98	0/916
All	All	0.62	12/21312 (0.1%)	0.86	30/29280 (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	G	0	1
2	B	0	1
2	D	0	1
2	H	0	1
3	I	0	2
3	X	0	2
4	J	0	3
4	Y	0	2
All	All	0	13

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	367	LEU	CG-CD2	-10.36	1.13	1.51
2	B	368	PRO	N-CD	8.14	1.59	1.47
2	H	370	ALA	C-O	6.68	1.36	1.23
2	D	370	ALA	CA-C	-6.58	1.35	1.52
2	H	370	ALA	CA-C	-6.19	1.36	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	367	LEU	C-N-CD	-25.69	64.09	120.60
2	H	371	SER	N-CA-CB	17.03	136.04	110.50
2	D	371	SER	N-CA-CB	17.00	136.00	110.50
2	H	370	ALA	CA-C-N	-12.45	89.82	117.20
2	D	370	ALA	CA-C-N	-12.13	90.51	117.20

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	367	LEU	Mainchain
2	D	371	SER	Mainchain
1	G	227	TYR	Sidechain
2	H	371	SER	Mainchain
3	X	16	DG	Sidechain

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	2176	0	2137	194	0
1	C	2176	0	2137	183	0
1	E	2176	0	2137	222	0
1	G	2176	0	2137	184	0
2	B	2454	0	2451	164	0
2	D	2454	0	2450	172	0
2	F	2454	0	2451	176	0
2	H	2454	0	2450	166	0
3	I	530	0	295	28	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	X	530	0	295	39	0
4	J	530	0	290	74	0
4	Y	530	0	288	71	0
All	All	20640	0	19518	1596	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 40.

The worst 5 of 1596 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)
2:B:572:LYS:NZ	2:B:606:GLN:HG2	1.45	1.32
2:F:572:LYS:NZ	2:F:606:GLN:HG2	1.53	1.23
2:H:572:LYS:NZ	2:H:606:GLN:HG2	1.59	1.18
3:I:15:DG:H2"	3:I:16:DG:H5"	1.17	1.15
4:Y:9:DT:H2"	4:Y:10:DC:H5"	1.14	1.13

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	271/273 (99%)	231 (85%)	29 (11%)	11 (4%)	3	32
1	C	271/273 (99%)	231 (85%)	29 (11%)	11 (4%)	3	32
1	E	271/273 (99%)	229 (84%)	31 (11%)	11 (4%)	3	32
1	G	271/273 (99%)	231 (85%)	31 (11%)	9 (3%)	4	38
2	B	310/312 (99%)	274 (88%)	25 (8%)	11 (4%)	4	37
2	D	310/312 (99%)	264 (85%)	34 (11%)	12 (4%)	3	33
2	F	310/312 (99%)	272 (88%)	29 (9%)	9 (3%)	5	41

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	H	310/312 (99%)	261 (84%)	36 (12%)	13 (4%)	3	31
All	All	2324/2340 (99%)	1993 (86%)	244 (10%)	87 (4%)	4	35

5 of 87 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	81	PRO
1	A	83	HIS
1	A	192	ALA
1	A	246	ARG
2	B	363	SER

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	242/242 (100%)	228 (94%)	14 (6%)	23	62
1	C	242/242 (100%)	221 (91%)	21 (9%)	12	47
1	E	242/242 (100%)	226 (93%)	16 (7%)	19	59
1	G	242/242 (100%)	221 (91%)	21 (9%)	12	47
2	B	268/268 (100%)	250 (93%)	18 (7%)	19	58
2	D	268/268 (100%)	257 (96%)	11 (4%)	35	72
2	F	268/268 (100%)	256 (96%)	12 (4%)	32	70
2	H	268/268 (100%)	257 (96%)	11 (4%)	35	72
All	All	2040/2040 (100%)	1916 (94%)	124 (6%)	22	61

5 of 124 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	489	ARG
1	E	97	ASP
2	H	489	ARG
2	D	510	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	600	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 71 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	436	ASN
1	E	111	HIS
2	H	409	HIS
2	D	464	ASN
2	D	547	ASN

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 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](#)

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	273/273 (100%)	0.38	19 (6%) 17 13	76, 176, 199, 200	0
1	C	273/273 (100%)	-0.02	2 (0%) 87 78	67, 142, 189, 199	0
1	E	273/273 (100%)	0.36	24 (8%) 11 8	77, 177, 199, 200	0
1	G	273/273 (100%)	0.03	4 (1%) 74 61	66, 144, 190, 200	0
2	B	312/312 (100%)	0.17	19 (6%) 22 16	75, 155, 198, 200	0
2	D	312/312 (100%)	0.33	25 (8%) 13 10	75, 164, 199, 200	0
2	F	312/312 (100%)	0.17	19 (6%) 22 16	66, 156, 199, 200	0
2	H	312/312 (100%)	0.39	30 (9%) 9 7	74, 169, 199, 200	0
3	I	26/26 (100%)	-0.03	0 100 100	84, 106, 119, 122	0
3	X	26/26 (100%)	0.01	0 100 100	90, 103, 116, 130	0
4	J	26/26 (100%)	-0.01	0 100 100	85, 108, 118, 127	0
4	Y	26/26 (100%)	-0.05	0 100 100	79, 104, 115, 129	0
All	All	2444/2444 (100%)	0.22	142 (5%) 24 17	66, 157, 199, 200	0

The worst 5 of 142 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	179	LEU	6.0
1	A	167	ASP	5.9
1	A	179	LEU	5.6
2	D	532	VAL	5.4
1	A	168	PRO	5.3

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

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.