wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 03:34 AM EDT

PDB ID	:	5JK7
Title	:	The X-ray structure of the DDB1-DCAF1-Vpr-UNG2 complex
Authors	:	Calero, G.; Ahn, J.; Wu, Y.
Deposited on	:	2016-04-26
Resolution	:	3.49 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 (i) 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 (i)) were used in the production of this report:

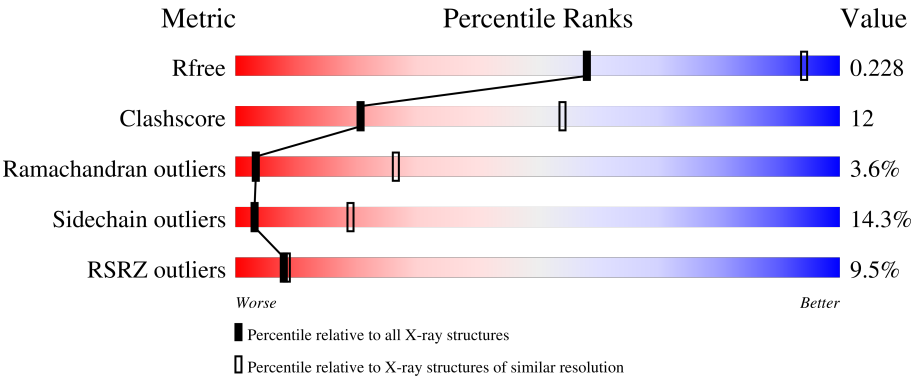
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.49 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.



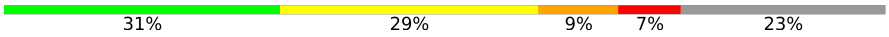
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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 <=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	1140	3% 66% 29% 5%
1	B	1140	24% 70% 26%
2	C	361	% 49% 33% 9% 8%
2	E	361	2% 53% 31% 7% 8%
3	D	222	3% 69% 26% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	222	
4	F	96	
4	H	96	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 27684 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 DNA damage-binding protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1133	Total	C	N	O	S	0	0	0
			8865	5619	1486	1711	49			
1	B	1133	Total	C	N	O	S	0	0	0
			8703	5541	1450	1663	49			

- Molecule 2 is a protein called Protein VPRBP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	331	Total	C	N	O	S	0	0	0
			2635	1666	457	494	18			
2	E	331	Total	C	N	O	S	0	0	0
			2635	1666	457	494	18			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1044	ALA	-	expression tag	UNP Q9Y4B6
C	1397	LEU	-	expression tag	UNP Q9Y4B6
C	1398	GLU	-	expression tag	UNP Q9Y4B6
C	1399	HIS	-	expression tag	UNP Q9Y4B6
C	1400	HIS	-	expression tag	UNP Q9Y4B6
C	1401	HIS	-	expression tag	UNP Q9Y4B6
C	1402	HIS	-	expression tag	UNP Q9Y4B6
C	1403	HIS	-	expression tag	UNP Q9Y4B6
C	1404	HIS	-	expression tag	UNP Q9Y4B6
E	1044	ALA	-	expression tag	UNP Q9Y4B6
E	1397	LEU	-	expression tag	UNP Q9Y4B6
E	1398	GLU	-	expression tag	UNP Q9Y4B6
E	1399	HIS	-	expression tag	UNP Q9Y4B6
E	1400	HIS	-	expression tag	UNP Q9Y4B6
E	1401	HIS	-	expression tag	UNP Q9Y4B6
E	1402	HIS	-	expression tag	UNP Q9Y4B6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	1403	HIS	-	expression tag	UNP Q9Y4B6
E	1404	HIS	-	expression tag	UNP Q9Y4B6

- Molecule 3 is a protein called Uracil-DNA glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	221	Total	C	N	O	S	0	0	0
			1791	1158	316	312	5			
3	G	221	Total	C	N	O	S	0	0	0
			1791	1158	316	312	5			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	83	VAL	-	expression tag	UNP P13051
D	84	PHE	-	expression tag	UNP P13051
G	83	VAL	-	expression tag	UNP P13051
G	84	PHE	-	expression tag	UNP P13051

- Molecule 4 is a protein called Protein Vpr.

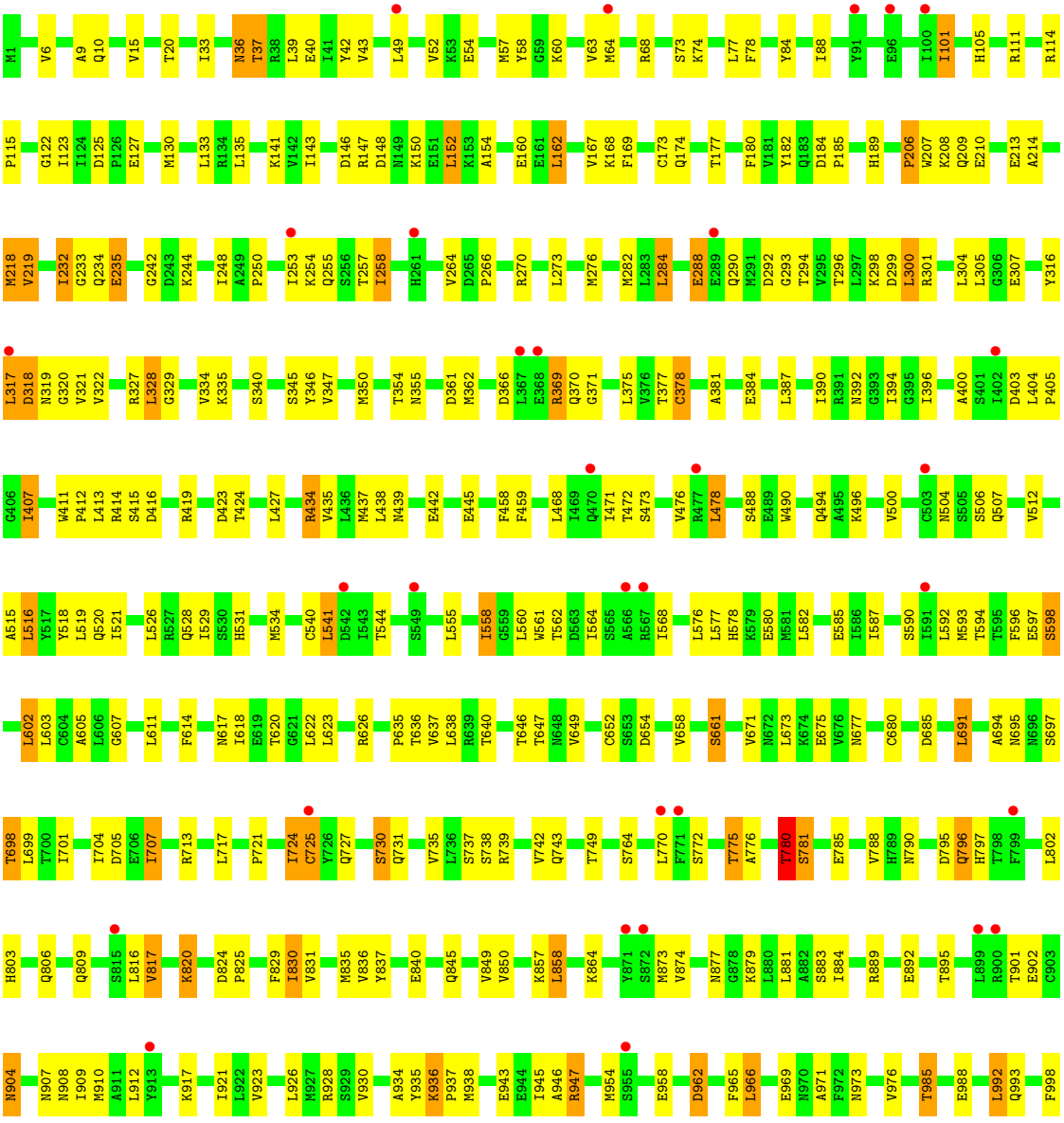
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	74	Total	C	N	O	S	0	0	0
			632	409	109	113	1			
4	H	74	Total	C	N	O	S	0	0	0
			632	409	109	113	1			

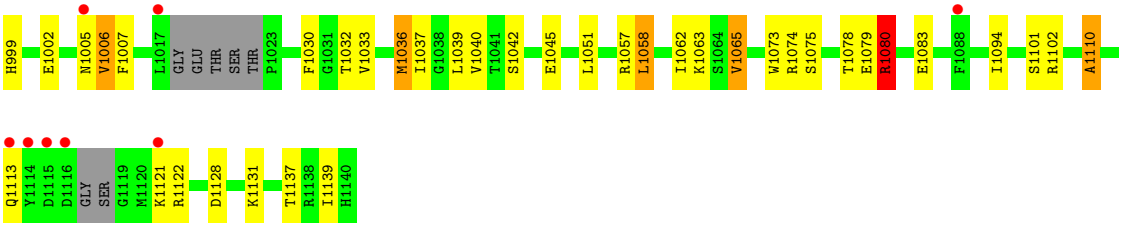
3 Residue-property plots (i)

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: DNA damage-binding protein 1

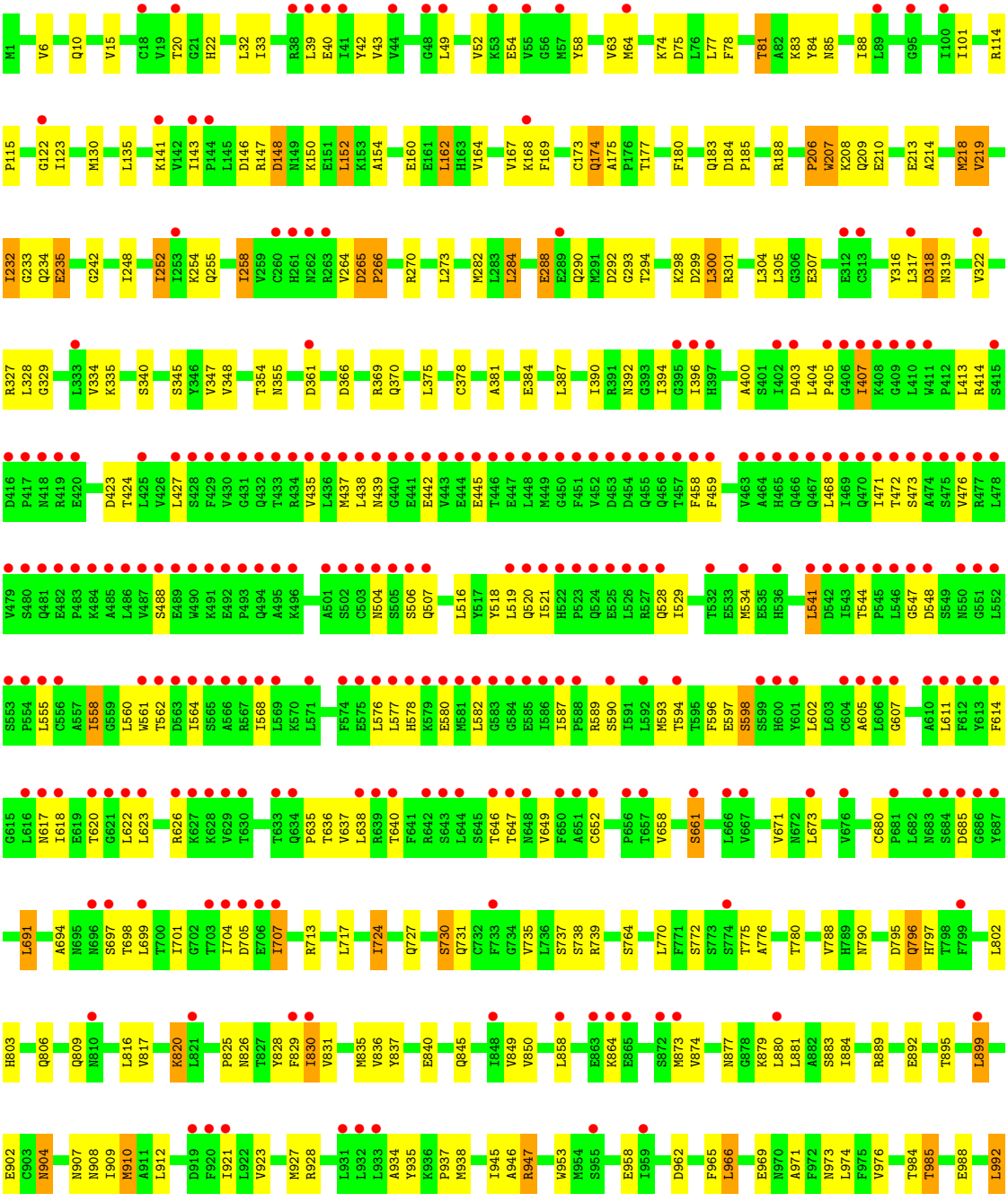
Chain A:

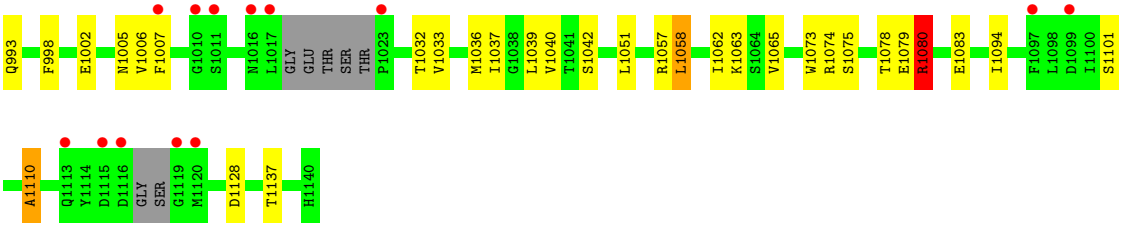




• Molecule 1: DNA damage-binding protein 1

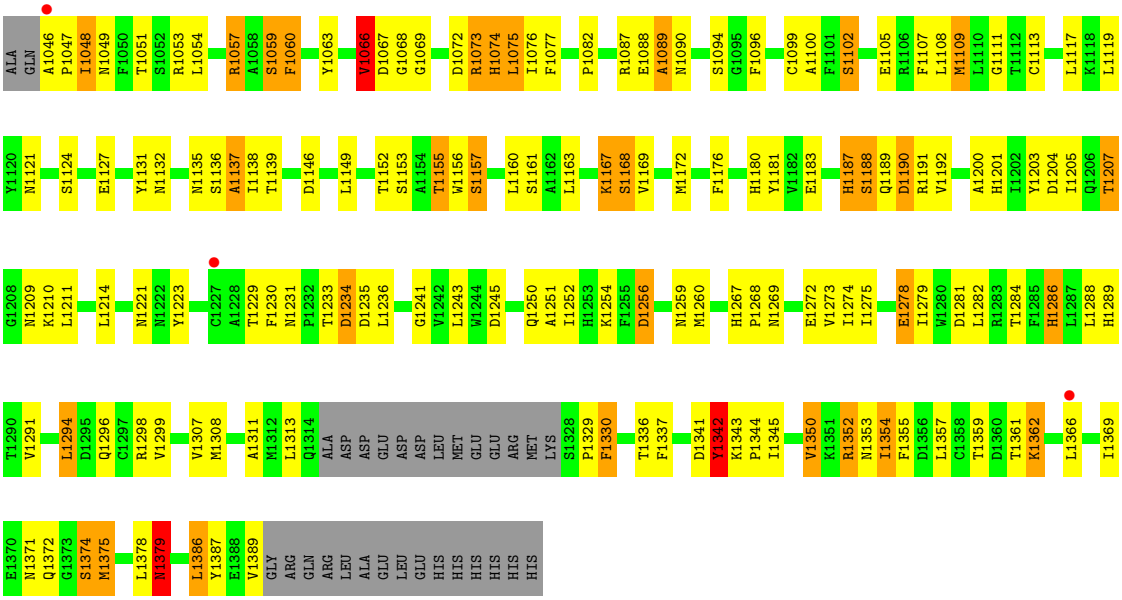
Chain B:





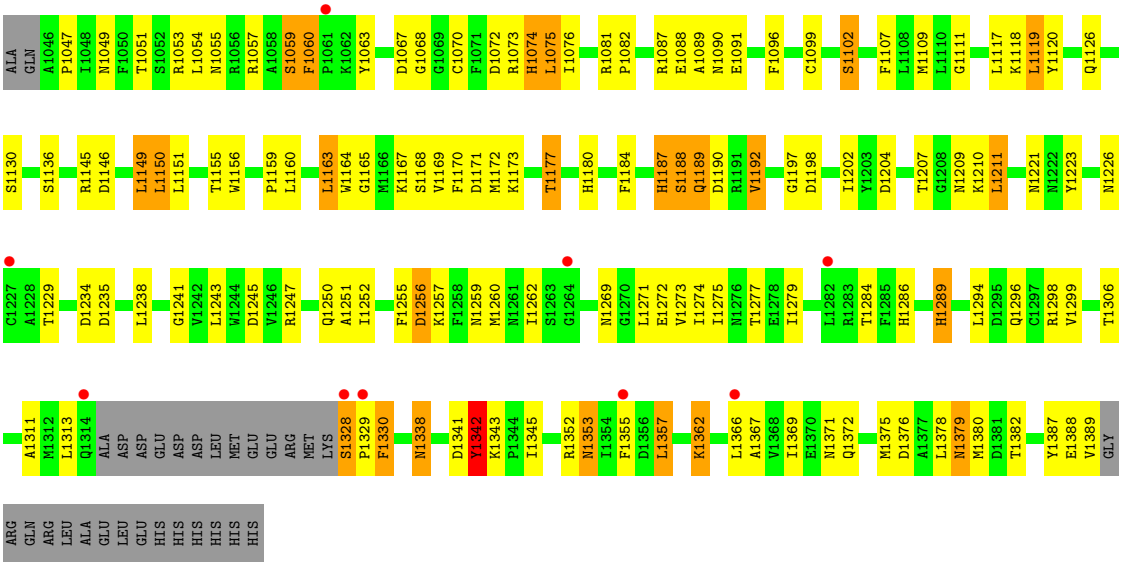
• Molecule 2: Protein VPRBP

Chain C:



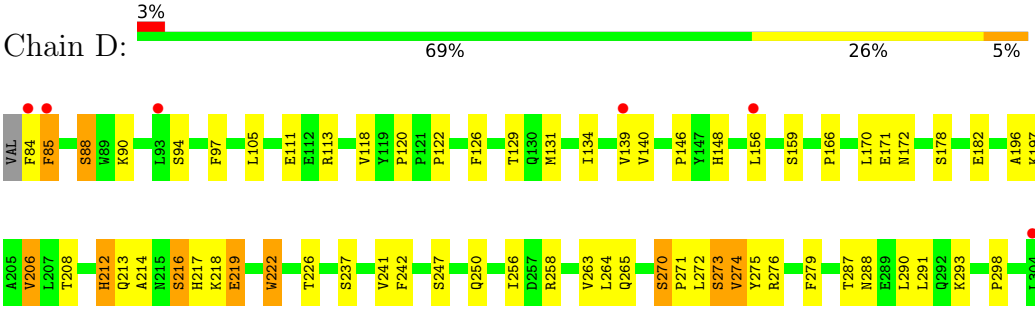
• Molecule 2: Protein VPRBP

Chain E:



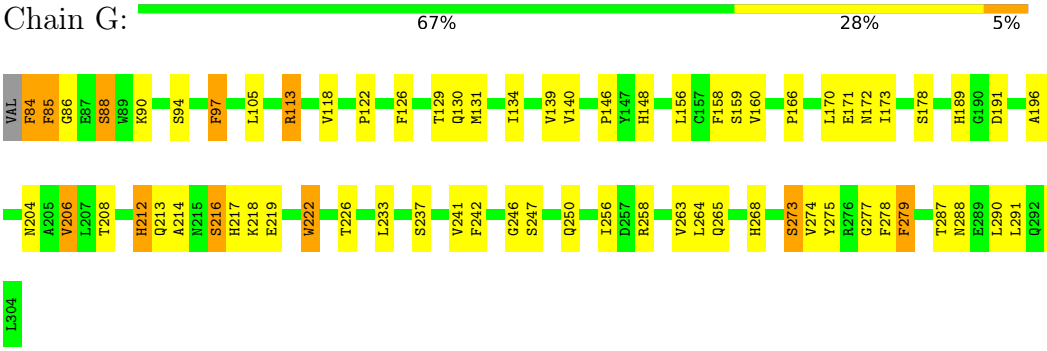
• Molecule 3: Uracil-DNA glycosylase

Chain D:



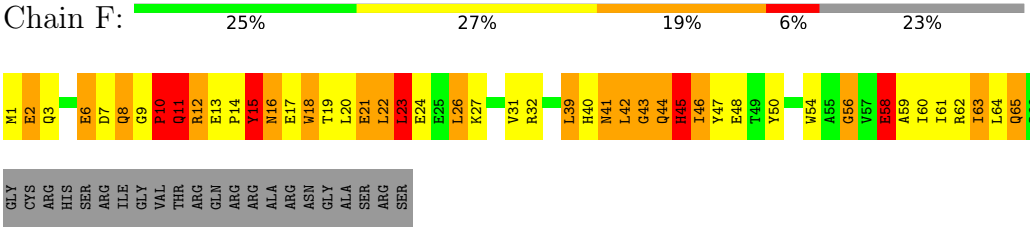
• Molecule 3: Uracil-DNA glycosylase

Chain G:



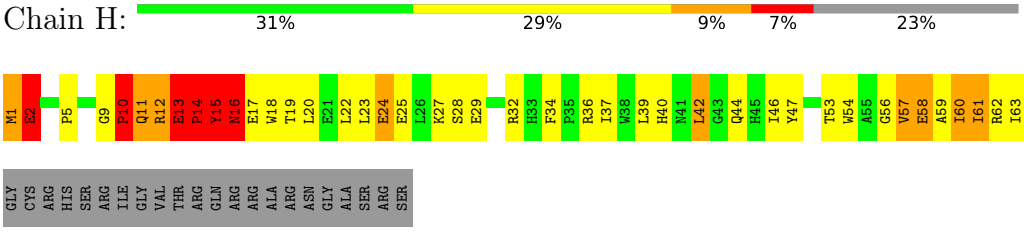
• Molecule 4: Protein Vpr

Chain F:



• Molecule 4: Protein Vpr

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	119.90Å 128.20Å 129.20Å 75.11° 89.44° 65.37°	Depositor
Resolution (Å)	40.30 – 3.49 39.69 – 3.49	Depositor EDS
% Data completeness (in resolution range)	90.6 (40.30-3.49) 90.6 (39.69-3.49)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.98 (at 3.48Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.176 , 0.206 0.198 , 0.228	Depositor DCC
R_{free} test set	2463 reflections (3.19%)	wwPDB-VP
Wilson B-factor (Å^2)	144.6	Xtriage
Anisotropy	0.404	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 154.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	27684	wwPDB-VP
Average B, all atoms (Å^2)	193.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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)

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	0/9028	0.79	0/12226
1	B	0.49	0/8864	0.78	2/12030 (0.0%)
2	C	0.61	1/2698 (0.0%)	0.92	1/3655 (0.0%)
2	E	0.56	0/2698	0.88	1/3655 (0.0%)
3	D	0.54	0/1852	0.80	1/2511 (0.0%)
3	G	0.51	0/1852	0.77	0/2511
4	F	1.71	13/651 (2.0%)	1.59	15/885 (1.7%)
4	H	1.76	19/651 (2.9%)	1.54	14/885 (1.6%)
All	All	0.63	33/28294 (0.1%)	0.86	34/38358 (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
4	F	0	2
4	H	0	3
All	All	0	5

The worst 5 of 33 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	H	25	GLU	C-O	-10.61	1.03	1.23
4	F	43	GLY	C-O	-10.24	1.07	1.23
4	H	58	GLU	C-O	-8.86	1.06	1.23
4	H	2	GLU	CD-OE2	-8.84	1.16	1.25
4	F	18	TRP	CE3-CZ3	-7.51	1.25	1.38

The worst 5 of 34 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	H	61	ILE	CG1-CB-CG2	-10.76	87.73	111.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	F	39	LEU	CB-CG-CD1	-10.14	93.76	111.00
4	F	42	LEU	CB-CG-CD1	-10.13	93.77	111.00
4	H	42	LEU	CA-CB-CG	9.92	138.12	115.30
4	H	12	ARG	NE-CZ-NH2	-9.14	115.73	120.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	F	11	GLN	Peptide
4	F	21	GLU	Peptide
4	H	1	MET	Peptide
4	H	10	PRO	Mainchain
4	H	14	PRO	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	8865	0	8824	173	0
1	B	8703	0	8508	140	0
2	C	2635	0	2533	99	0
2	E	2635	0	2533	70	0
3	D	1791	0	1747	48	0
3	G	1791	0	1745	58	0
4	F	632	0	609	58	0
4	H	632	0	609	50	0
All	All	27684	0	27108	671	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 671 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)
3:G:84:PHE:CE1	3:G:130:GLN:NE2	1.71	1.55

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:G:84:PHE:CD1	3:G:130:GLN:NE2	1.84	1.45
4:F:26:LEU:HD13	4:F:65:GLN:NE2	1.53	1.23
2:E:1170:PHE:O	4:F:8:GLN:O	1.60	1.17
4:H:1:MET:HB2	4:H:2:GLU:HG3	1.17	1.11

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 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	1127/1140 (99%)	991 (88%)	106 (9%)	30 (3%)	5 33
1	B	1127/1140 (99%)	997 (88%)	104 (9%)	26 (2%)	6 36
2	C	327/361 (91%)	278 (85%)	31 (10%)	18 (6%)	2 17
2	E	327/361 (91%)	279 (85%)	35 (11%)	13 (4%)	3 24
3	D	219/222 (99%)	191 (87%)	17 (8%)	11 (5%)	2 19
3	G	219/222 (99%)	186 (85%)	22 (10%)	11 (5%)	2 19
4	F	72/96 (75%)	50 (69%)	12 (17%)	10 (14%)	0 3
4	H	72/96 (75%)	57 (79%)	8 (11%)	7 (10%)	0 7
All	All	3490/3638 (96%)	3029 (87%)	335 (10%)	126 (4%)	3 26

5 of 126 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	162	LEU
1	A	206	PRO
1	A	290	GLN
1	A	576	LEU
1	A	770	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 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	994/999 (100%)	852 (86%)	142 (14%)	3 19
1	B	934/999 (94%)	801 (86%)	133 (14%)	3 19
2	C	291/317 (92%)	239 (82%)	52 (18%)	2 9
2	E	291/317 (92%)	238 (82%)	53 (18%)	1 9
3	D	193/194 (100%)	181 (94%)	12 (6%)	18 51
3	G	193/194 (100%)	179 (93%)	14 (7%)	14 45
4	F	66/83 (80%)	50 (76%)	16 (24%)	0 4
4	H	66/83 (80%)	55 (83%)	11 (17%)	2 12
All	All	3028/3186 (95%)	2595 (86%)	433 (14%)	3 19

5 of 433 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	797	HIS
2	C	1109	MET
4	F	21	GLU
1	B	849	VAL
1	B	1007	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 69 such sidechains are listed below:

Mol	Chain	Res	Type
4	F	16	ASN
4	F	41	ASN
4	H	33	HIS
1	B	261	HIS
1	B	234	GLN

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

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, 95^{th} 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	1133/1140 (99%)	0.03	39 (3%) 45 40	110, 180, 236, 272	0
1	B	1133/1140 (99%)	1.14	278 (24%) 0 0	157, 230, 290, 296	0
2	C	331/361 (91%)	-0.04	3 (0%) 84 79	107, 149, 194, 240	0
2	E	331/361 (91%)	0.03	9 (2%) 54 48	121, 174, 216, 242	0
3	D	221/222 (99%)	-0.16	6 (2%) 54 48	121, 178, 222, 248	0
3	G	221/222 (99%)	-0.27	0 100 100	129, 172, 221, 253	0
4	F	74/96 (77%)	-0.25	0 100 100	125, 161, 196, 206	0
4	H	74/96 (77%)	-0.20	0 100 100	120, 149, 184, 200	0
All	All	3518/3638 (96%)	0.34	335 (9%) 8 8	107, 190, 284, 296	0

The worst 5 of 335 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	470	GLN	18.8
1	B	520	GLN	11.9
1	B	562	THR	10.8
1	B	521	ILE	10.5
1	B	526	LEU	9.9

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

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.