



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 19, 2023 – 03:25 AM EDT

PDB ID : 4YC2
Title : Crystal structure of the stabilized inner domain of clade A/E HIV-1 gp120 from *E. coli* in complex with the antibody A32.
Authors : Tolbert, W.D.; Gohain, N.; Pazgier, M.
Deposited on : 2015-02-19
Resolution : 3.02 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 ⓘ 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:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.35.1
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) : 2.35.1

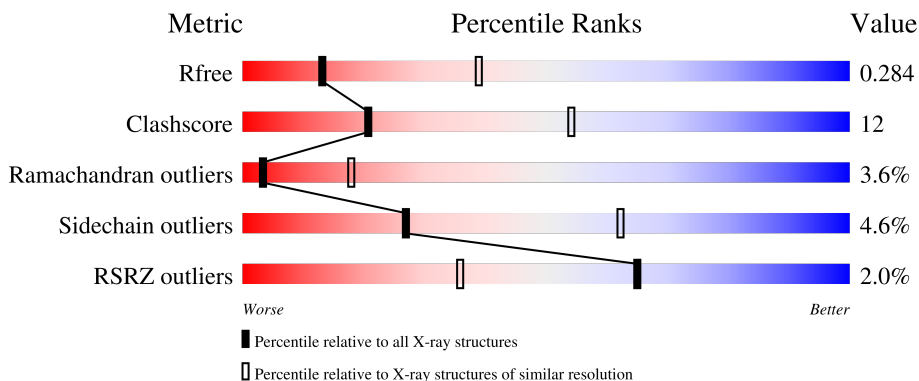
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.02 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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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 $\leq 5\%$. 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	156	
1	G	156	
2	B	224	
2	H	224	
3	C	210	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	L	210	 88% 12%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8604 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 The stabilized inner domain of clade A/E HIV-1 gp120 from E. coli.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	G	133	Total 1071	C 676	N 181	O 203	S 11	0	0	0
1	A	131	Total 1054	C 666	N 177	O 200	S 11	0	0	0

- Molecule 2 is a protein called The antibody A32 Fab heavy chain..

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	220	Total 1666	C 1052	N 286	O 323	S 5	0	0	0
2	B	223	Total 1685	C 1062	N 289	O 329	S 5	0	0	0

- Molecule 3 is a protein called The antibody A32 Fab light chain..

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	L	210	Total 1562	C 977	N 261	O 320	S 4	0	0	0
3	C	210	Total 1562	C 977	N 261	O 320	S 4	0	0	0

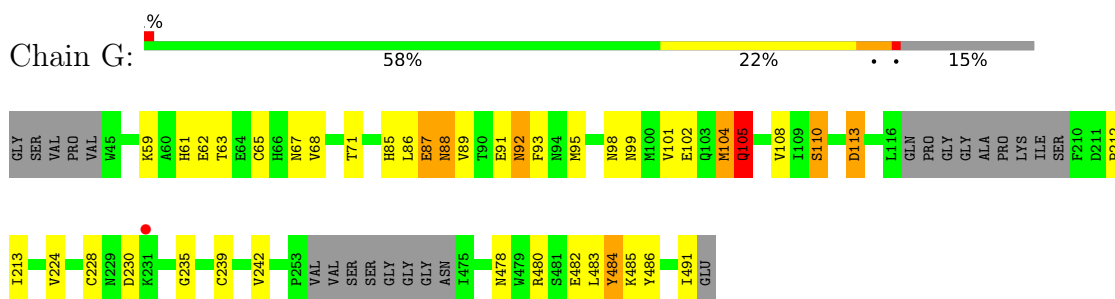
- Molecule 4 is water.

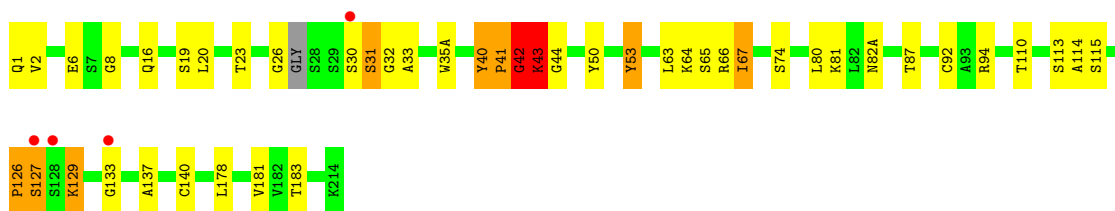
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	H	1	Total 1 O 1	0	0
4	B	2	Total 2 O 2	0	0
4	C	1	Total 1 O 1	0	0

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: The stabilized inner domain of clade A/E HIV-1 gp120 from *E. coli*





- Molecule 3: The antibody A32 Fab light chain.

Chain L: 88% 12%



- Molecule 3: The antibody A32 Fab light chain.

Chain C: 90% 9%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	75.80Å 211.80Å 72.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.02 47.07 – 3.02	Depositor EDS
% Data completeness (in resolution range)	86.2 (50.00-3.02) 86.3 (47.07-3.02)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.31 (at 3.01Å)	Xtrriage
Refinement program	REFMAC 5.8.0103, PHENIX, CNS	Depositor
R, R_{free}	0.227 , 0.288 0.229 , 0.284	Depositor DCC
R_{free} test set	1087 reflections (5.27%)	wwPDB-VP
Wilson B-factor (Å ²)	85.4	Xtrriage
Anisotropy	0.260	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 42.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.034 for l,-k,h	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	8604	wwPDB-VP
Average B, all atoms (Å ²)	90.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.58% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [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	0.52	0/1082	0.89	4/1473 (0.3%)
1	G	0.40	0/1100	0.71	1/1498 (0.1%)
2	B	0.41	0/1727	0.74	3/2355 (0.1%)
2	H	0.38	0/1708	0.70	0/2329
3	C	0.37	0/1603	0.58	0/2191
3	L	0.35	0/1603	0.57	0/2191
All	All	0.40	0/8823	0.69	8/12037 (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	4
2	B	0	2
All	All	0	6

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	228	CYS	CA-CB-SG	-7.95	99.69	114.00
2	B	40	TYR	C-N-CD	-7.36	104.41	120.60
2	B	40	TYR	N-CA-C	7.24	130.54	111.00
1	A	90	THR	CA-CB-CG2	-6.10	103.86	112.40
1	A	86	LEU	CB-CG-CD1	5.55	120.44	111.00

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	241	ASN	Peptide
1	A	77	THR	Peptide
1	A	89	VAL	Peptide
1	A	90	THR	Peptide
2	B	35(A)	TRP	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	1054	0	982	96	0
1	G	1071	0	1002	41	0
2	B	1685	0	1659	28	0
2	H	1666	0	1642	33	0
3	C	1562	0	1499	8	1
3	L	1562	0	1499	9	0
4	B	2	0	0	0	0
4	C	1	0	0	0	0
4	H	1	0	0	0	0
All	All	8604	0	8283	208	1

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 208 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:A:249:HIS:ND1	1:A:486:TYR:OH	1.69	1.25
1:A:230:ASP:CG	1:A:233:PHE:HB2	1.58	1.21
1:A:230:ASP:OD2	1:A:233:PHE:HB2	1.45	1.14
1:A:233:PHE:HE1	1:A:239:CYS:HB2	1.08	1.10
1:A:233:PHE:CE1	1:A:239:CYS:HB2	1.89	1.07

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:93:ASP:OD2	3:C:151:ASP:OD2[2_455]	2.17	0.03

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	123/156 (79%)	103 (84%)	13 (11%)	7 (6%)	1	9
1	G	127/156 (81%)	118 (93%)	5 (4%)	4 (3%)	4	21
2	B	219/224 (98%)	186 (85%)	22 (10%)	11 (5%)	2	12
2	H	216/224 (96%)	184 (85%)	21 (10%)	11 (5%)	2	11
3	C	208/210 (99%)	187 (90%)	17 (8%)	4 (2%)	8	34
3	L	208/210 (99%)	180 (86%)	25 (12%)	3 (1%)	11	41
All	All	1101/1180 (93%)	958 (87%)	103 (9%)	40 (4%)	3	18

5 of 40 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	88	ASN
1	G	104	MET
1	G	105	GLN
2	H	64	LYS
2	H	126	PRO

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	120/138 (87%)	111 (92%)	9 (8%)	13	42
1	G	122/138 (88%)	115 (94%)	7 (6%)	20	54
2	B	193/193 (100%)	182 (94%)	11 (6%)	20	54
2	H	190/193 (98%)	181 (95%)	9 (5%)	26	61
3	C	177/177 (100%)	174 (98%)	3 (2%)	60	85
3	L	177/177 (100%)	171 (97%)	6 (3%)	37	72
All	All	979/1016 (96%)	934 (95%)	45 (5%)	27	62

5 of 45 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	476	LYS
2	B	74	SER
1	A	480	ARG
2	B	50	TYR
2	B	113	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	G	92	ASN
1	G	232	ASN
1	A	82	GLN
1	A	241	ASN
2	B	39	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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	131/156 (83%)	0.56	14 (10%) 6 2	85, 130, 178, 187	0
1	G	133/156 (85%)	0.03	1 (0%) 86 65	86, 120, 150, 156	0
2	B	223/224 (99%)	-0.19	4 (1%) 68 39	54, 75, 113, 129	0
2	H	220/224 (98%)	-0.24	3 (1%) 75 48	59, 76, 102, 127	0
3	C	210/210 (100%)	-0.34	0 100 100	55, 72, 94, 104	0
3	L	210/210 (100%)	-0.32	0 100 100	63, 78, 101, 110	0
All	All	1127/1180 (95%)	-0.14	22 (1%) 65 36	54, 82, 146, 187	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	229	ASN	3.7
1	A	93	PHE	3.6
1	A	240	LYS	3.6
1	G	231	LYS	3.3
1	A	230	ASP	3.2

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides 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.