wwPDB X-ray Structure Validation Summary Report (i)

Mar 13, 2024 – 06:09 PM JST

PDB ID	:	5HFL
Title	:	Gp41-targeting HIV-1 fusion inhibitors with helical Ile-Asp-Leu tail
Authors	:	Zhu, Y.; Ye, S.; Zhang, R.
Deposited on	:	2016-01-07
Resolution	:	2.29 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 (1)) were used in the production of this report:

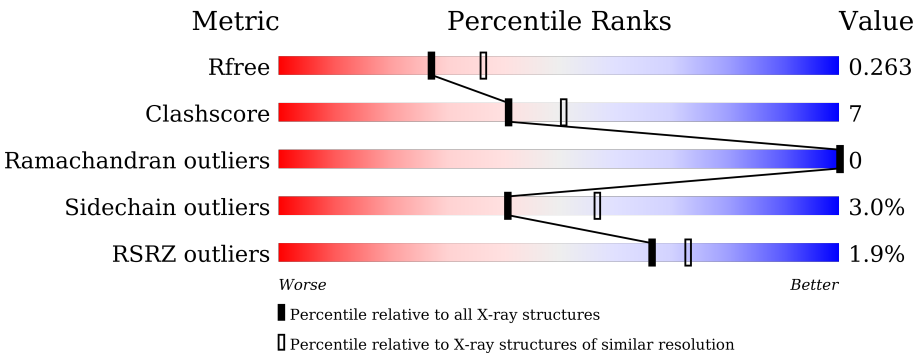
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.29 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)
RSRZ outliers	127900	4938 (2.30-2.30)

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	75	80% 11% • 8%
1	B	75	87% 8% 5%
1	C	75	4% 79% 13% • 7%
1	D	75	% 71% 20% • 8%
1	E	75	3% 71% 15% • 13%
1	F	75	3% 75% 15% • 8%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3587 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 Envelope glycoprotein,gp41 CHR region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	69	572	363	101	107	1	3	0	0
1	B	71	585	371	103	110	1	3	0	0
1	C	70	578	366	102	109	1	2	0	0
1	D	69	570	361	101	108		0	0	0
1	E	65	539	341	97	100	1	5	0	0
1	F	69	572	363	101	107	1	9	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	542	GLY	-	expression tag	UNP A1YNW7
A	543	PRO	-	expression tag	UNP A1YNW7
A	544	MET	-	expression tag	UNP A1YNW7
A	545	ALA	-	expression tag	UNP A1YNW7
B	542	GLY	-	expression tag	UNP A1YNW7
B	543	PRO	-	expression tag	UNP A1YNW7
B	544	MET	-	expression tag	UNP A1YNW7
B	545	ALA	-	expression tag	UNP A1YNW7
C	542	GLY	-	expression tag	UNP A1YNW7
C	543	PRO	-	expression tag	UNP A1YNW7
C	544	MET	-	expression tag	UNP A1YNW7
C	545	ALA	-	expression tag	UNP A1YNW7
D	542	GLY	-	expression tag	UNP A1YNW7
D	543	PRO	-	expression tag	UNP A1YNW7
D	544	MET	-	expression tag	UNP A1YNW7
D	545	ALA	-	expression tag	UNP A1YNW7
E	542	GLY	-	expression tag	UNP A1YNW7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	543	PRO	-	expression tag	UNP A1YNW7
E	544	MET	-	expression tag	UNP A1YNW7
E	545	ALA	-	expression tag	UNP A1YNW7
F	542	GLY	-	expression tag	UNP A1YNW7
F	543	PRO	-	expression tag	UNP A1YNW7
F	544	MET	-	expression tag	UNP A1YNW7
F	545	ALA	-	expression tag	UNP A1YNW7

- Molecule 2 is water.

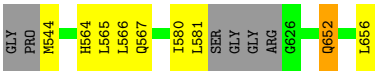
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	38	38	38	0	0
			Total	O		
2	B	23	23	23	0	0
			Total	O		
2	C	34	34	34	0	0
			Total	O		
2	D	29	29	29	0	0
			Total	O		
2	E	21	21	21	0	0
			Total	O		
2	F	26	26	26	0	0

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: Envelope glycoprotein,gp41 CHR region

Chain A:

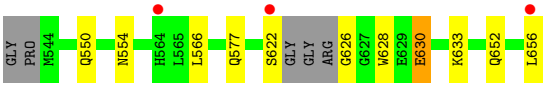


• Molecule 1: Envelope glycoprotein,gp41 CHR region

Chain B:

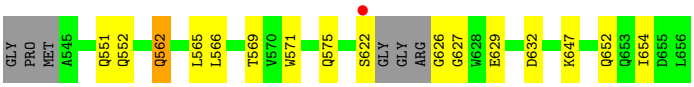
• Molecule 1: Envelope glycoprotein,gp41 CHR region

Chain C:



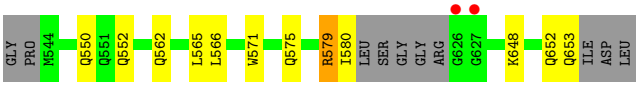
• Molecule 1: Envelope glycoprotein,gp41 CHR region

Chain D:



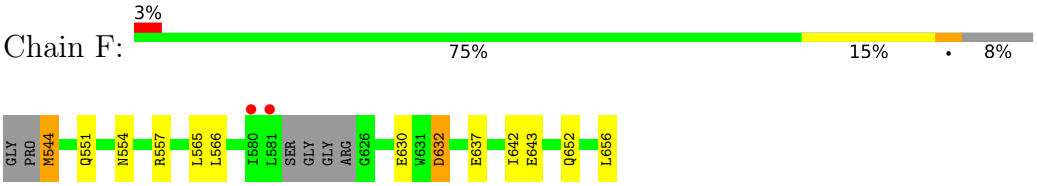
• Molecule 1: Envelope glycoprotein,gp41 CHR region

Chain E:



• Molecule 1: Envelope glycoprotein,gp41 CHR region

Chain F:


3%
75%
15%
8%
GLY
PRO
M544
Q551
N554
R557
L565
L566
I580
L581
SER
GLY
GLY
ARG
G626
E630
W631
D632
E637
I642
E643
Q652
L656

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	42.45Å 114.47Å 42.94Å 90.00° 91.80° 90.00°	Depositor
Resolution (Å)	42.91 – 2.29 42.91 – 2.29	Depositor EDS
% Data completeness (in resolution range)	93.8 (42.91-2.29) 93.5 (42.91-2.29)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.17 (at 2.29Å)	Xtriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, R_{free}	0.214 , 0.263 0.214 , 0.263	Depositor DCC
R_{free} test set	1816 reflections (9.92%)	wwPDB-VP
Wilson B-factor (Å^2)	29.5	Xtriage
Anisotropy	0.550	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 45.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.022 for l,k,-h 0.060 for h,-k,-l 0.046 for l,-k,h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3587	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.07% 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	0.21	0/578	0.34	0/773
1	B	0.20	0/592	0.34	0/792
1	C	0.20	0/584	0.34	0/781
1	D	0.20	0/576	0.33	0/771
1	E	0.20	0/545	0.34	0/729
1	F	0.20	0/578	0.36	0/773
All	All	0.20	0/3453	0.34	0/4619

There are no bond length outliers.

There are no bond angle outliers.

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	572	0	587	9	0
1	B	585	0	600	5	0
1	C	578	0	592	8	0
1	D	570	0	583	13	0
1	E	539	0	550	14	0
1	F	572	0	587	15	0
2	A	38	0	0	3	0
2	B	23	0	0	4	0
2	C	34	0	0	3	0
2	D	29	0	0	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	21	0	0	0	0
2	F	26	0	0	5	0
All	All	3587	0	3499	49	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 7.

The worst 5 of 49 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:D:629:GLU:OE2	2:D:701:HOH:O	1.87	0.91
1:A:564:HIS:ND1	2:A:701:HOH:O	2.08	0.85
1:B:554:ASN:ND2	2:B:701:HOH:O	2.17	0.78
1:F:632:ASP:OD2	2:F:701:HOH:O	2.00	0.77
1:F:557:ARG:NH2	2:F:704:HOH:O	2.19	0.75

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	65/75 (87%)	64 (98%)	1 (2%)	0	100	100
1	B	67/75 (89%)	67 (100%)	0	0	100	100
1	C	66/75 (88%)	65 (98%)	1 (2%)	0	100	100
1	D	65/75 (87%)	65 (100%)	0	0	100	100
1	E	61/75 (81%)	60 (98%)	1 (2%)	0	100	100
1	F	65/75 (87%)	64 (98%)	1 (2%)	0	100	100
All	All	389/450 (86%)	385 (99%)	4 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	61/64 (95%)	60 (98%)	1 (2%)	62	78
1	B	63/64 (98%)	63 (100%)	0	100	100
1	C	62/64 (97%)	60 (97%)	2 (3%)	39	54
1	D	61/64 (95%)	59 (97%)	2 (3%)	38	53
1	E	57/64 (89%)	54 (95%)	3 (5%)	22	31
1	F	61/64 (95%)	58 (95%)	3 (5%)	25	35
All	All	365/384 (95%)	354 (97%)	11 (3%)	41	57

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	580	ILE
1	F	544	MET
1	F	632	ASP
1	F	630	GLU
1	D	647	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	575	GLN
1	E	652	GLN
1	F	652	GLN
1	F	554	ASN
1	F	650	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	69/75 (92%)	-0.03	0	100	100	26, 36, 51, 69	1 (1%)
1	B	71/75 (94%)	0.01	0	100	100	25, 37, 50, 61	1 (1%)
1	C	70/75 (93%)	0.14	3 (4%)	35	42	28, 37, 52, 67	1 (1%)
1	D	69/75 (92%)	0.10	1 (1%)	75	80	26, 39, 57, 70	0
1	E	65/75 (86%)	0.12	2 (3%)	49	56	26, 40, 54, 62	2 (3%)
1	F	69/75 (92%)	0.11	2 (2%)	51	58	27, 39, 61, 69	2 (2%)
All	All	413/450 (91%)	0.08	8 (1%)	66	73	25, 38, 55, 70	7 (1%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	622	SER	2.8
1	F	580	ILE	2.6
1	F	581	LEU	2.5
1	E	626	GLY	2.3
1	C	622	SER	2.2

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.