W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

May 23, 2020 – 03:34 am BST

PDB ID	:	3VIE
Title	:	HIV-gp41 fusion inhibitor Sifuvirtide
Authors	:	Yao, X.; Waltersperger, S.; Wang, M.T.; Cui, S.
Deposited on	:	2011-09-29
Resolution	:	1.80 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
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)	:	2.11

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 1.80 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.
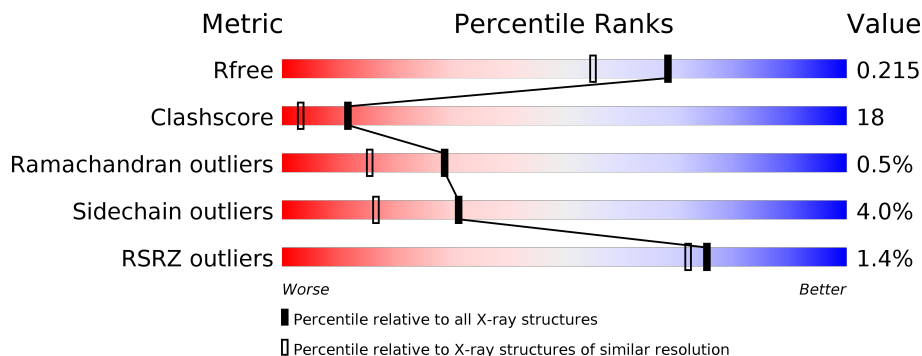
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 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	37	70% 24% · ·
1	C	37	5% 70% 24% · ·
1	E	37	3% 73% 27%
2	B	37	68% 27% 5%
2	D	37	65% 35%
2	F	37	68% 27% 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4378 atoms, of which 2153 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 gp160.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	36	Total	C	H	N	O	0	2	0
			627	190	326	61	50			
1	C	37	Total	C	H	N	O	0	3	0
			647	199	335	60	53			
1	E	37	Total	C	H	N	O	0	8	0
			729	221	383	68	57			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	34	ACE	-	ACETYLATION	UNP Q9YP39
C	34	ACE	-	ACETYLATION	UNP Q9YP39
E	34	ACE	-	ACETYLATION	UNP Q9YP39

- Molecule 2 is a protein called Sifuvirtide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	37	Total	C	H	N	O	0	10	0
			781	248	381	72	80			
2	D	37	Total	C	H	N	O	21	8	0
			758	242	369	69	78			
2	F	37	Total	C	H	N	O	0	7	0
			743	239	359	67	78			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	23	Total	O	0	0
			23	23		
3	C	19	Total	O	0	0
			19	19		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	E	13	Total 13	O 13	0	0
3	B	13	Total 13	O 13	0	0
3	D	10	Total 10	O 10	0	0
3	F	15	Total 15	O 15	0	0

3 Residue-property plots (i)

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: Envelope glycoprotein gp160

Chain A:

- Molecule 1: Envelope glycoprotein gp160
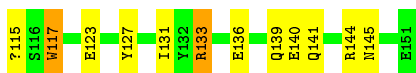
Chain C:
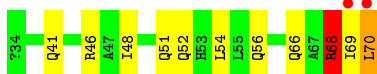
- Molecule 1: Envelope glycoprotein gp160

Chain E:

- Molecule 2: Sifuvirtide

Chain B:

- Molecule 2: Sifuvirtide

Chain D:

- Molecule 2: Sifuvirtide

Chain F:

?115 S116 W117 E118 T119 W120 E121 Y132 E136 E137 S138 Q141 Q142 D143 R144 E151

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	88.79Å　　49.18Å　　55.82Å 90.00°　　90.65°　　90.00°	Depositor
Resolution (Å)	34.94　–　1.80 44.39　–　1.70	Depositor EDS
% Data completeness (in resolution range)	98.8 (34.94-1.80) 93.3 (44.39-1.70)	Depositor EDS
R_{merge}	0.03	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.85 (at 1.70Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.1_357)	Depositor
R, R_{free}	0.182　,　0.219 0.180　,　0.215	Depositor DCC
R_{free} test set	1243 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	23.6	Xtriage
Anisotropy	0.498	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.44 , 54.5	EDS
L-test for twinning [2]	$<|L|> = 0.50$, $<L^2> = 0.34$	Xtriage
Estimated twinning fraction	0.013 for -1/2*h-3/2*k,-1/2*h+1/2*k,-l 0.011 for -1/2*h+3/2*k,1/2*h+1/2*k,-l 0.090 for 1/2*h-3/2*k,-1/2*h-1/2*k,-l 0.043 for 1/2*h+3/2*k,1/2*h-1/2*k,-l 0.023 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4378	wwPDB-VP
Average B, all atoms (Å^2)	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 19.47% 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: ACE

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	0.89	0/307	0.86	0/413
1	C	0.95	0/322	0.78	1/435 (0.2%)
1	E	0.84	0/371	0.80	0/499
2	B	0.80	0/437	0.96	1/588 (0.2%)
2	D	0.68	0/418	0.81	2/563 (0.4%)
2	F	0.79	1/410 (0.2%)	0.84	1/552 (0.2%)
All	All	0.82	1/2265 (0.0%)	0.85	5/3050 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	115	ACE	C-N	-5.80	1.20	1.34

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	115	ACE	C-N-CA	-13.97	86.77	121.70
2	D	115	ACE	O-C-N	9.93	138.58	122.70
2	D	115	ACE	C-N-CA	-5.26	108.54	121.70
1	C	68	ARG	NE-CZ-NH2	-5.24	117.68	120.30
2	F	143	ASP	CB-CG-OD1	5.11	122.90	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

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	301	326	328	16	0
1	C	312	335	336	26	0
1	E	346	383	384	19	0
2	B	400	381	379	22	0
2	D	389	369	372	15	0
2	F	384	359	363	10	0
3	A	23	0	0	0	0
3	B	13	0	0	3	0
3	C	19	0	0	3	0
3	D	10	0	0	1	0
3	E	13	0	0	0	0
3	F	15	0	0	0	0
All	All	2225	2153	2162	74	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 18.

The worst 5 of 74 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:41[B]:GLN:OE1	1:E:41[B]:GLN:NE2	2.05	0.89
1:A:66:GLN:NE2	2:B:117[A]:TRP:CD1	2.47	0.81
2:B:144[B]:ARG:HD2	2:B:145:ASN:OD1	1.86	0.75
2:B:123:GLU:OE1	3:B:53:HOH:O	2.05	0.74
1:C:56:GLN:OE1	3:C:91:HOH:O	2.06	0.73

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	36/37 (97%)	36 (100%)	0	0	100	100
1	C	38/37 (103%)	38 (100%)	0	0	100	100
1	E	43/37 (116%)	43 (100%)	0	0	100	100
2	B	45/37 (122%)	45 (100%)	0	0	100	100
2	D	43/37 (116%)	43 (100%)	0	0	100	100
2	F	42/37 (114%)	41 (98%)	0	1 (2%)	6	1
All	All	247/222 (111%)	246 (100%)	0	1 (0%)	29	21

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	F	116	SER

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	32/31 (103%)	31 (97%)	1 (3%)	40	25
1	C	34/31 (110%)	32 (94%)	2 (6%)	19	7
1	E	39/31 (126%)	39 (100%)	0	100	100
2	B	46/36 (128%)	42 (91%)	4 (9%)	10	3
2	D	44/36 (122%)	42 (96%)	2 (4%)	27	13
2	F	43/36 (119%)	40 (93%)	3 (7%)	15	5
All	All	238/201 (118%)	226 (95%)	12 (5%)	31	10

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	133[A]	ARG
2	B	133[B]	ARG
2	F	116	SER
2	B	117[B]	TRP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	117[B]	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	42	ASN
1	C	56	GLN
2	B	130	GLN
1	A	66	GLN
1	E	43	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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	35/37 (94%)	-0.21	0	100	100	17, 24, 41, 50	0
1	C	36/37 (97%)	0.21	2 (5%)	24	19	17, 25, 43, 56	0
1	E	36/37 (97%)	-0.09	1 (2%)	53	47	17, 22, 34, 40	0
2	B	36/37 (97%)	-0.00	0	100	100	17, 29, 42, 48	0
2	D	36/37 (97%)	0.06	0	100	100	20, 31, 48, 68	0
2	F	36/37 (97%)	0.05	0	100	100	20, 30, 44, 52	0
All	All	215/222 (96%)	0.00	3 (1%)	75	72	17, 26, 44, 68	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	70	LEU	5.3
1	C	69	ILE	5.2
1	E	70	LEU	2.1

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 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.