



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 12, 2017 – 08:59 pm GMT

PDB ID : 5BN0  
Title : A new HIV fusion peptide inhibitor  
Authors : Xue, Y.  
Deposited on : 2015-05-25  
Resolution : 2.80 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](mailto: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	:	trunk28620
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)	:	recalc28949

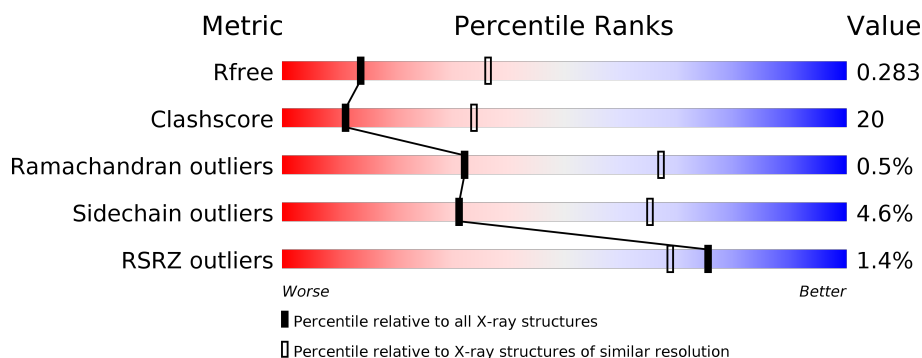
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	100719	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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  $\geq 3$ , 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	C	37	<div> <div>89%</div> <div>11%</div> </div>
1	D	37	<div> <div>65%</div> <div>30%</div> <div>5%</div> </div>
2	B	36	<div> <div>56%</div> <div>39%</div> <div>• •</div> </div>
2	E	36	<div> <div>6%</div> <div>67%</div> <div>33%</div> </div>
2	N	36	<div> <div>3%</div> <div>53%</div> <div>39%</div> <div>• 6%</div> </div>
3	A	36	<div> <div>78%</div> <div>14%</div> <div>6%</div> <div>•</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 1804 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 gp160.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	37	Total	C	N	O	S	0	0	0
			316	196	52	67	1			
1	D	35	Total	C	N	O	S	0	0	0
			300	184	50	65	1			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	625	ACE	-	expression tag	UNP B2CPZ5
C	626	LEU	-	expression tag	UNP B2CPZ5
D	625	ACE	-	expression tag	UNP B2CPZ5
D	626	LEU	-	expression tag	UNP B2CPZ5

- Molecule 2 is a protein called Envelope glycoprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	N	34	Total	C	N	O	0	0	0
			274	172	54	48			
2	B	35	Total	C	N	O	0	0	0
			282	178	55	49			
2	E	36	Total	C	N	O	0	0	0
			290	184	56	50			

- Molecule 3 is a protein called Envelope glycoprotein gp160.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	36	Total	C	N	O	S	0	0	0
			313	194	52	66	1			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	626	LEU	-	expression tag	UNP B2CPZ5


- Molecule 4 is water.

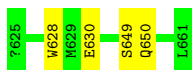
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	N	3	Total O 3 3	0	0
4	A	10	Total O 10 10	0	0
4	B	3	Total O 3 3	0	0
4	D	6	Total O 6 6	0	0
4	E	7	Total O 7 7	0	0

### 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 C: 



- Molecule 1: Envelope glycoprotein gp160

Chain D: 



- Molecule 2: Envelope glycoprotein

Chain N: 



- Molecule 2: Envelope glycoprotein

Chain B: 



- Molecule 2: Envelope glycoprotein

Chain E: 



- Molecule 3: Envelope glycoprotein gp160

Chain A: 

L626					
Y638					
T639					
S640					
L641					
T642					
Q658					
E659					
L660					
L661					

## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.15Å 52.34Å 60.26Å 90.00° 117.46° 90.00°	Depositor
Resolution (Å)	26.73 – 2.80 26.74 – 2.80	Depositor EDS
% Data completeness (in resolution range)	93.2 (26.73-2.80) 93.2 (26.74-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	26.84 (at 2.80Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, $R_{free}$	0.271 , 0.288 0.259 , 0.283	Depositor DCC
$R_{free}$ test set	246 reflections (4.92%)	DCC
Wilson B-factor (Å <sup>2</sup> )	6.1	Xtriage
Anisotropy	0.287	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 50.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.80	EDS
Total number of atoms	1804	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.74% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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 [i](#)

### 5.1 Standard geometry [i](#)

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	C	0.42	0/319	0.49	0/432
1	D	0.38	0/303	0.45	0/410
2	B	0.38	0/284	0.50	0/383
2	E	0.43	0/292	0.60	0/394
2	N	0.37	0/276	0.51	0/372
3	A	0.46	0/318	0.80	2/430 (0.5%)
All	All	0.41	0/1792	0.57	2/2421 (0.1%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	661	LEU	CA-CB-CG	-9.40	93.67	115.30
3	A	660	LEU	CA-CB-CG	6.55	130.35	115.30

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	C	316	0	293	4	0
1	D	300	0	271	10	0
2	B	282	0	298	23	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	290	0	309	26	0
2	N	274	0	287	25	0
3	A	313	0	289	15	0
4	A	10	0	0	0	0
4	B	3	0	0	0	0
4	D	6	0	0	1	0
4	E	7	0	0	2	0
4	N	3	0	0	0	0
All	All	1804	0	1747	70	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 20.

The worst 5 of 70 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:N:565:LEU:HD23	2:B:566:LEU:HD13	1.50	0.93
2:N:566:LEU:HD22	2:E:565:LEU:HD22	1.52	0.92
2:E:580:ILE:O	2:E:581:LEU:HB2	1.71	0.90
3:A:638:TYR:O	3:A:642:ILE:HG23	1.76	0.85
3:A:642:ILE:HG22	2:B:561:ALA:HB1	1.68	0.74

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	C	35/37 (95%)	34 (97%)	1 (3%)	0	100	100
1	D	33/37 (89%)	31 (94%)	2 (6%)	0	100	100
2	B	33/36 (92%)	32 (97%)	1 (3%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	34/36 (94%)	34 (100%)	0	0	100	100
2	N	32/36 (89%)	32 (100%)	0	0	100	100
3	A	34/36 (94%)	33 (97%)	0	1 (3%)	5	18
All	All	201/218 (92%)	196 (98%)	4 (2%)	1 (0%)	32	67

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	660	LEU

### 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	C	36/36 (100%)	36 (100%)	0	100	100
1	D	34/36 (94%)	33 (97%)	1 (3%)	48	81
2	B	30/31 (97%)	27 (90%)	3 (10%)	9	26
2	E	31/31 (100%)	31 (100%)	0	100	100
2	N	29/31 (94%)	27 (93%)	2 (7%)	18	46
3	A	36/36 (100%)	33 (92%)	3 (8%)	13	36
All	All	196/201 (98%)	187 (95%)	9 (5%)	31	65

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	660	LEU
1	D	626	LEU
2	B	577	GLN
3	A	640	SER
2	B	555	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	562	GLN
2	E	564	HIS
1	D	658	GLN
2	B	552	GLN
2	E	563	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 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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	C	36/37 (97%)	0.10	0	100 100	5, 17, 36, 49	0
1	D	34/37 (91%)	0.32	0	100 100	9, 28, 46, 53	0
2	B	35/36 (97%)	0.04	0	100 100	5, 17, 45, 54	0
2	E	36/36 (100%)	0.19	2 (5%)	25 16	3, 15, 37, 54	0
2	N	34/36 (94%)	0.22	1 (2%)	52 41	5, 19, 48, 51	0
3	A	36/36 (100%)	0.06	0	100 100	5, 19, 39, 44	0
All	All	211/218 (96%)	0.15	3 (1%)	75 69	3, 19, 45, 54	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	547	GLY	3.3
2	N	546	SER	2.8
2	E	546	SER	2.7

### 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.