



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 11:55 am GMT

PDB ID : 4EMZ  
Title : HIV-1 Nef in complex with MHC-I cytoplasmic domain and Mu1 adaptin subunit of AP1 adaptor (second domain)  
Authors : Jia, X.; Xiong, Y.  
Deposited on : 2012-04-12  
Resolution : 2.90 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

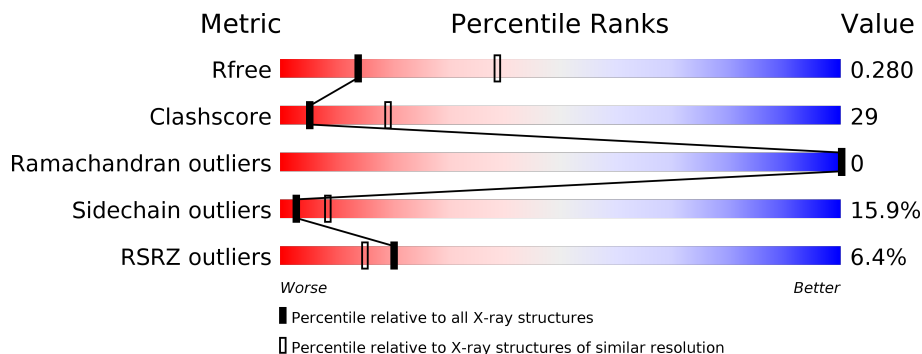
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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	B	206	<div> <div>11%</div> <div> <div>35%</div> <div>27%</div> <div>5%</div> <div>33%</div> </div> </div>
1	C	206	<div> <div>3%</div> <div> <div>41%</div> <div>25%</div> <div>•</div> <div>31%</div> </div> </div>
2	D	28	<div> <div>4%</div> <div> <div>43%</div> <div>21%</div> <div>•</div> <div>32%</div> </div> </div>
2	E	28	<div> <div>25%</div> <div>25%</div> <div>11%</div> <div>39%</div> </div>
3	A	266	<div> <div>6%</div> <div> <div>50%</div> <div>38%</div> <div>8%</div> <div>•</div> </div> </div>
3	M	266	<div> <div>3%</div> <div> <div>51%</div> <div>33%</div> <div>11%</div> <div>•</div> </div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6789 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 Protein Nef.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	C	142	Total	C	N	O	S	Se	0	0	0
			1193	775	210	204	2	2			
1	B	138	Total	C	N	O	S	Se	0	0	0
			1163	756	205	199	1	2			

- Molecule 2 is a protein called MHC-I.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	19	Total	C	N	O	0	0	0
			127	71	25	31			
2	E	17	Total	C	N	O	0	0	0
			113	64	23	26			

- Molecule 3 is a protein called AP-1 complex subunit mu-1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	A	255	Total	C	N	O	S	Se	0	1	0
			2088	1341	367	374	1	5			
3	M	255	Total	C	N	O	S	Se	0	1	0
			2087	1341	367	373	1	5			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	2	Total	O	0	0
			2	2		
4	D	1	Total	O	0	0
			1	1		
4	B	2	Total	O	0	0
			2	2		
4	A	8	Total	O	0	0
			8	8		

*Continued on next page...*

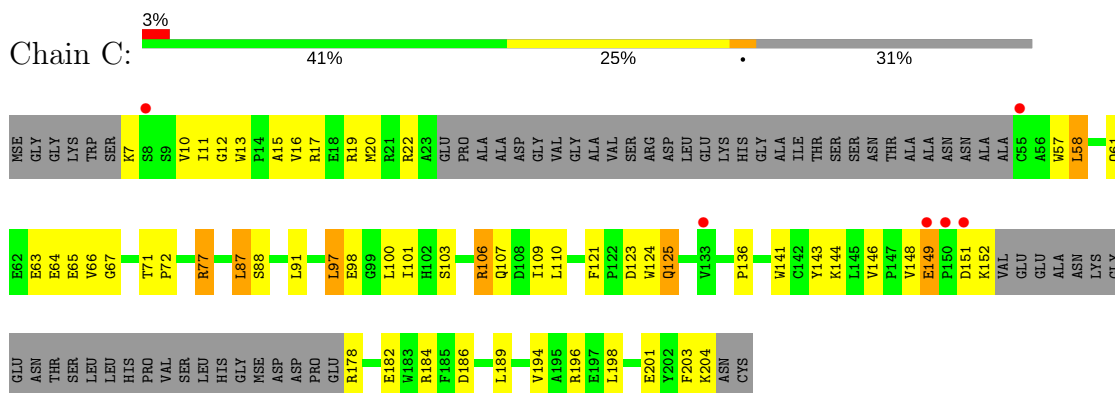
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	M	5	Total	O	0	0
			5	5		

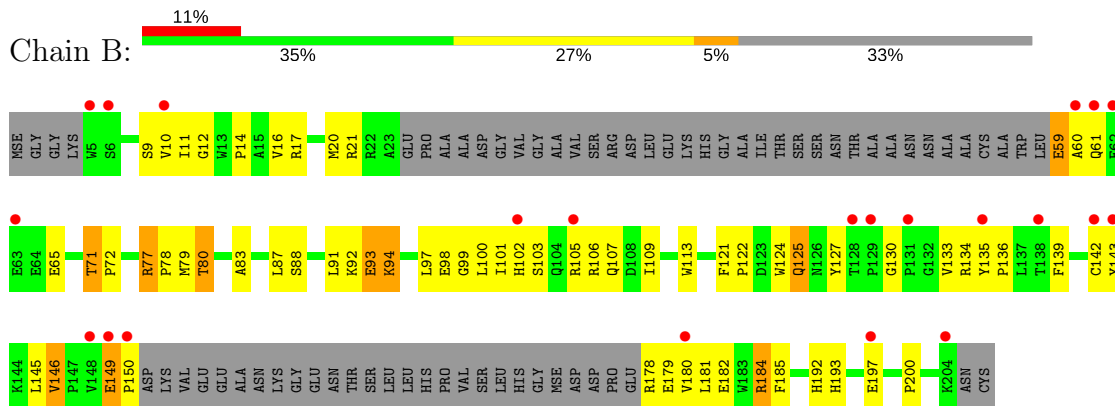
### 3 Residue-property plots [i](#)

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.

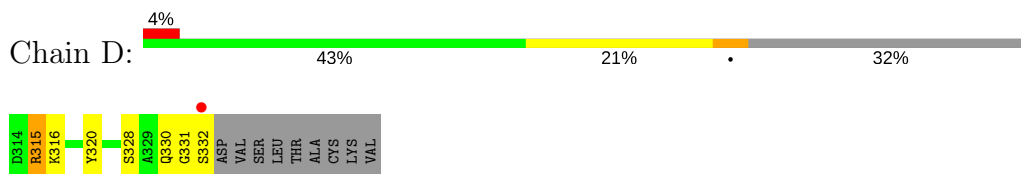
#### • Molecule 1: Protein Nef



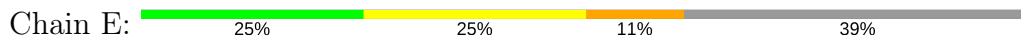
#### • Molecule 1: Protein Nef

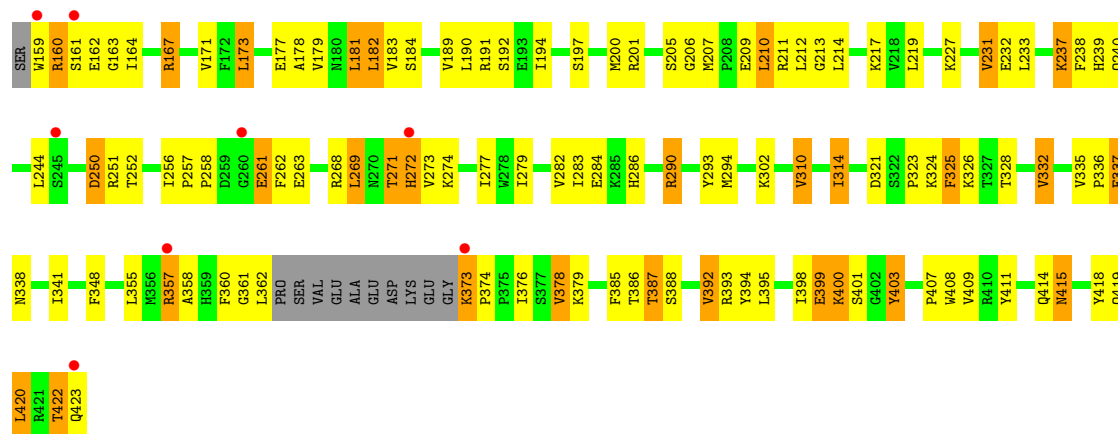


#### • Molecule 2: MHC-I



#### • Molecule 2: MHC-I





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.32Å 112.47Å 114.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.90 29.84 – 2.90	Depositor EDS
% Data completeness (in resolution range)	99.4 (30.00-2.90) 99.5 (29.84-2.90)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.66 (at 2.90Å)	Xtriage
Refinement program	REFMAC 5.7.0025	Depositor
R, $R_{free}$	0.223 , 0.278 0.225 , 0.280	Depositor DCC
$R_{free}$ test set	1317 reflections (5.35%)	DCC
Wilson B-factor (Å <sup>2</sup> )	77.2	Xtriage
Anisotropy	0.102	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 85.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.015 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6789	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	117.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.84% 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](#)

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	B	0.80	1/1202 (0.1%)	0.91	2/1629 (0.1%)
1	C	0.80	1/1232 (0.1%)	0.89	0/1669
2	D	0.85	0/127	0.81	0/167
2	E	0.80	0/113	0.78	0/148
3	A	0.71	0/2135	0.90	0/2871
3	M	0.72	0/2134	0.88	0/2871
All	All	0.75	2/6943 (0.0%)	0.89	2/9355 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	149	GLU	CD-OE1	5.46	1.31	1.25
1	B	113	TRP	CD2-CE2	5.25	1.47	1.41

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	65	GLU	N-CA-C	-5.36	96.53	111.00
1	B	149	GLU	C-N-CD	5.06	139.02	128.40

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	B	1163	0	1120	72	0
1	C	1193	0	1153	65	0
2	D	127	0	110	5	0
2	E	113	0	101	24	0
3	A	2088	0	2117	129	0
3	M	2087	0	2119	114	0
4	A	8	0	0	1	0
4	B	2	0	0	0	0
4	C	2	0	0	0	0
4	D	1	0	0	0	0
4	M	5	0	0	0	0
All	All	6789	0	6720	394	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 29.

The worst 5 of 394 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:A:283:ILE:HG22	3:A:293:TYR:CD1	1.79	1.17
1:B:79:MSE:HE2	1:B:83:ALA:HB3	1.28	1.11
1:C:20:MSE:HA	1:C:109:ILE:HD11	1.31	1.09
3:M:286:HIS:CD2	3:M:290:ARG:CZ	2.38	1.07
3:M:286:HIS:HD2	3:M:290:ARG:NH1	1.54	1.03

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	B	132/206 (64%)	127 (96%)	5 (4%)	0	100	100
1	C	136/206 (66%)	132 (97%)	4 (3%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	17/28 (61%)	16 (94%)	1 (6%)	0	100	100
2	E	15/28 (54%)	14 (93%)	1 (7%)	0	100	100
3	A	252/266 (95%)	241 (96%)	11 (4%)	0	100	100
3	M	252/266 (95%)	246 (98%)	6 (2%)	0	100	100
All	All	804/1000 (80%)	776 (96%)	28 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	B	122/169 (72%)	108 (88%)	14 (12%)	6	20
1	C	125/169 (74%)	105 (84%)	20 (16%)	3	8
2	D	12/20 (60%)	9 (75%)	3 (25%)	1	2
2	E	10/20 (50%)	6 (60%)	4 (40%)	0	0
3	A	234/237 (99%)	200 (86%)	34 (14%)	4	11
3	M	234/237 (99%)	192 (82%)	42 (18%)	2	6
All	All	737/852 (86%)	620 (84%)	117 (16%)	3	8

5 of 117 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	256	ILE
3	A	321	ASP
3	M	392	VAL
3	A	263	GLU
3	A	287	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	186	ASN
3	M	286	HIS
3	A	288	HIS
1	B	193	HIS
3	A	286	HIS

### 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	B	136/206 (66%)	0.67	22 (16%) 2 1	86, 122, 195, 256	0
1	C	140/206 (67%)	0.49	6 (4%) 36 31	72, 113, 176, 201	0
2	D	19/28 (67%)	0.14	1 (5%) 27 23	71, 94, 139, 149	0
2	E	17/28 (60%)	0.11	0 100 100	86, 102, 148, 151	0
3	A	250/266 (93%)	0.34	15 (6%) 23 17	81, 115, 170, 236	0
3	M	250/266 (93%)	0.07	8 (3%) 48 42	74, 105, 148, 182	0
All	All	812/1000 (81%)	0.33	52 (6%) 20 15	71, 113, 168, 256	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	149	GLU	6.3
1	C	151	ASP	6.0
1	B	5	TRP	5.5
1	C	149	GLU	5.0
3	A	401	SER	4.6

### 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.