



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 11:31 pm GMT

PDB ID : 4NHU  
Title : The M33 TCR p3M33l/H-2 Ld Complex  
Authors : Birnbaum, M.E.; Adams, J.J.; Garcia, K.C.  
Deposited on : 2013-11-05  
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

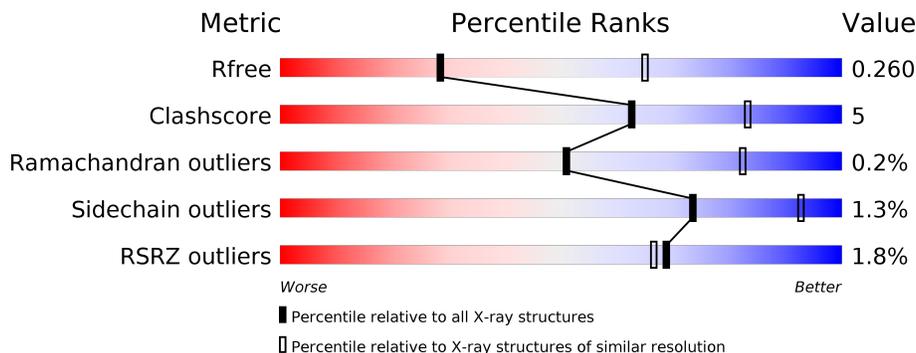
# 1 Overall quality at a glance

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	A	219	 4% (poor fit), 76% (0-1 outliers), 14% (2 outliers), 10% (3+ outliers or not modelled)
1	C	219	 3% (poor fit), 79% (0-1 outliers), 11% (2 outliers), 9% (3+ outliers or not modelled)
2	B	254	 78% (0-1 outliers), 15% (2 outliers), 6% (3+ outliers or not modelled)
2	D	254	 2% (poor fit), 82% (0-1 outliers), 11% (2 outliers), 6% (3+ outliers or not modelled)
3	E	204	 % (poor fit), 82% (0-1 outliers), 9% (2 outliers), 9% (3+ outliers or not modelled)
3	G	204	 77% (0-1 outliers), 14% (2 outliers), 9% (3+ outliers or not modelled)

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 9829 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 2C m33 alpha VmCh chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	197	1539	979	251	302	7	0	0	0
1	C	199	1563	991	260	305	7	0	0	0

- Molecule 2 is a protein called 2C m33 beta VmCh chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	238	1856	1165	329	356	6	0	0	0
2	D	238	1848	1161	328	353	6	0	0	0

- Molecule 3 is a protein called H-2 class I histocompatibility antigen, L-D alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	E	186	1514	953	266	286	9	0	0	0
3	G	185	1509	951	265	284	9	0	0	0

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	0	MET	-	INITIATING METHIONINE	UNP P01897
E	30	ASP	ASN	ENGINEERED MUTATION	UNP P01897
E	49	VAL	ALA	ENGINEERED MUTATION	UNP P01897
E	66	VAL	ILE	ENGINEERED MUTATION	UNP P01897
E	97	ARG	TRP	ENGINEERED MUTATION	UNP P01897
E	131	ARG	LYS	ENGINEERED MUTATION	UNP P01897
E	167	ALA	TRP	ENGINEERED MUTATION	UNP P01897
E	181	GLY	-	LINKER	UNP P01897

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	182	SER	-	LINKER	UNP P01897
E	183	LEU	-	LINKER	UNP P01897
E	184	VAL	-	LINKER	UNP P01897
E	185	PRO	-	LINKER	UNP P01897
E	186	ARG	-	LINKER	UNP P01897
E	187	GLY	-	LINKER	UNP P01897
E	188	SER	-	LINKER	UNP P01897
E	189	GLY	-	LINKER	UNP P01897
E	190	GLY	-	LINKER	UNP P01897
E	191	GLY	-	LINKER	UNP P01897
E	192	GLY	-	LINKER	UNP P01897
E	193	GLY	-	LINKER	UNP P01897
E	194	GLY	-	LINKER	UNP P01897
E	195	ALA	-	SEE REMARK 999	UNP P01897
E	196	PRO	-	SEE REMARK 999	UNP P01897
E	197	TRP	-	SEE REMARK 999	UNP P01897
E	198	ASN	-	SEE REMARK 999	UNP P01897
E	199	PRO	-	SEE REMARK 999	UNP P01897
E	200	ALA	-	SEE REMARK 999	UNP P01897
E	201	MET	-	SEE REMARK 999	UNP P01897
E	202	MET	-	SEE REMARK 999	UNP P01897
E	203	ILE	-	SEE REMARK 999	UNP P01897
G	0	MET	-	INITIATING METHIONINE	UNP P01897
G	30	ASP	ASN	ENGINEERED MUTATION	UNP P01897
G	49	VAL	ALA	ENGINEERED MUTATION	UNP P01897
G	66	VAL	ILE	ENGINEERED MUTATION	UNP P01897
G	97	ARG	TRP	ENGINEERED MUTATION	UNP P01897
G	131	ARG	LYS	ENGINEERED MUTATION	UNP P01897
G	167	ALA	TRP	ENGINEERED MUTATION	UNP P01897
G	181	GLY	-	LINKER	UNP P01897
G	182	SER	-	LINKER	UNP P01897
G	183	LEU	-	LINKER	UNP P01897
G	184	VAL	-	LINKER	UNP P01897
G	185	PRO	-	LINKER	UNP P01897
G	186	ARG	-	LINKER	UNP P01897
G	187	GLY	-	LINKER	UNP P01897
G	188	SER	-	LINKER	UNP P01897
G	189	GLY	-	LINKER	UNP P01897
G	190	GLY	-	LINKER	UNP P01897
G	191	GLY	-	LINKER	UNP P01897
G	192	GLY	-	LINKER	UNP P01897
G	193	GLY	-	LINKER	UNP P01897

*Continued on next page...*

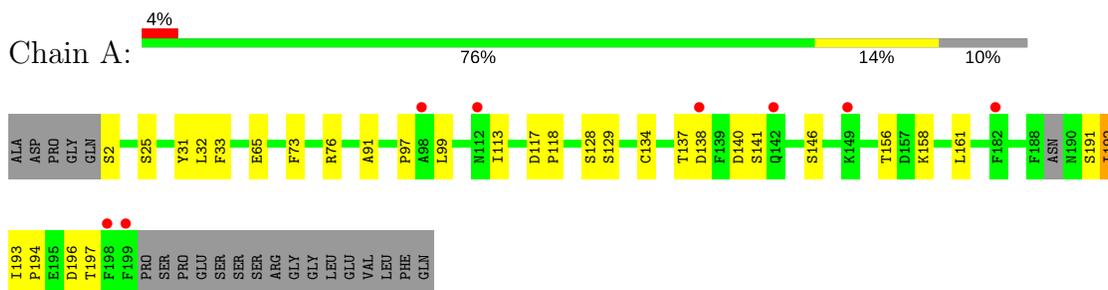
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	194	GLY	-	LINKER	UNP P01897
G	195	ALA	-	SEE REMARK 999	UNP P01897
G	196	PRO	-	SEE REMARK 999	UNP P01897
G	197	TRP	-	SEE REMARK 999	UNP P01897
G	198	ASN	-	SEE REMARK 999	UNP P01897
G	199	PRO	-	SEE REMARK 999	UNP P01897
G	200	ALA	-	SEE REMARK 999	UNP P01897
G	201	MET	-	SEE REMARK 999	UNP P01897
G	202	MET	-	SEE REMARK 999	UNP P01897
G	203	ILE	-	SEE REMARK 999	UNP P01897

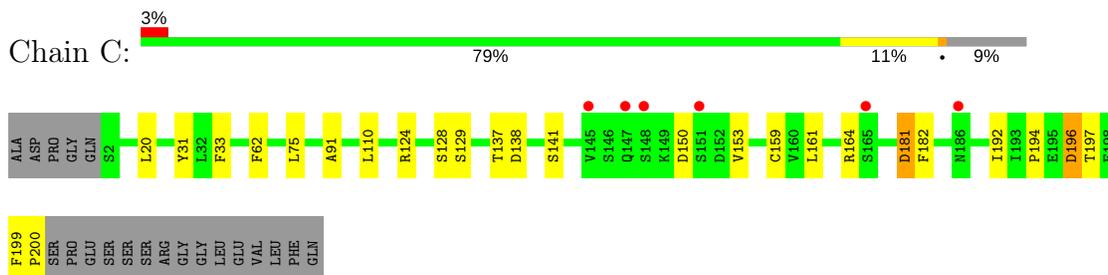
### 3 Residue-property plots [i](#)

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.

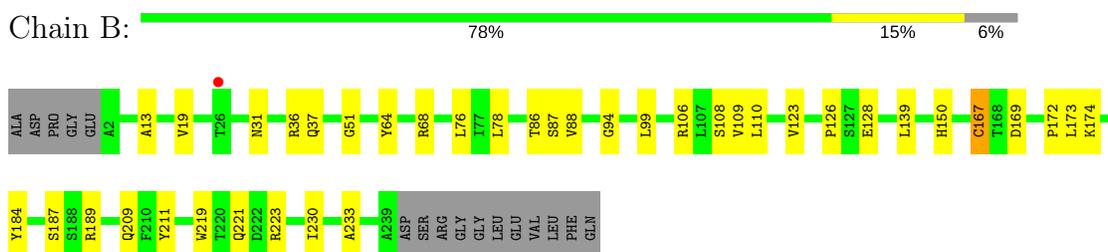
- Molecule 1: 2C m33 alpha VmCh chimera



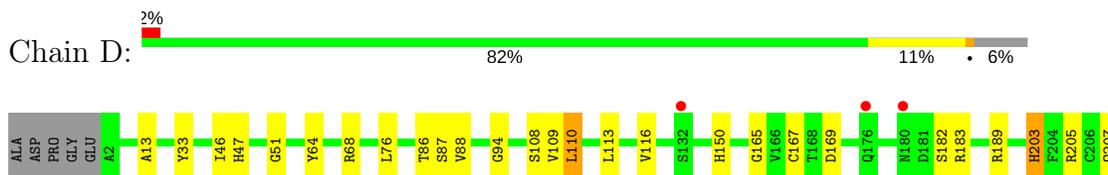
- Molecule 1: 2C m33 alpha VmCh chimera

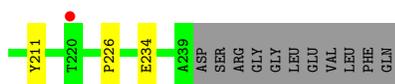


- Molecule 2: 2C m33 beta VmCh chimera

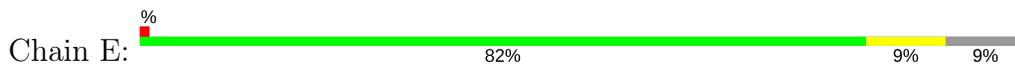


- Molecule 2: 2C m33 beta VmCh chimera





• Molecule 3: H-2 class I histocompatibility antigen, L-D alpha chain



• Molecule 3: H-2 class I histocompatibility antigen, L-D alpha chain



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	49.76Å 92.80Å 94.95Å 112.29° 103.57° 102.37°	Depositor
Resolution (Å)	47.86 – 2.90 77.94 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.8 (47.86-2.90) 82.5 (77.94-2.90)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.55 (at 2.91Å)	Xtrriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.229 , 0.260 0.227 , 0.260	Depositor DCC
$R_{free}$ test set	1567 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	53.2	Xtrriage
Anisotropy	0.049	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 34.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.003 for -h,-l,-k	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	9829	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.90% 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	A	0.23	0/1578	0.38	0/2145
1	C	0.22	0/1604	0.37	0/2181
2	B	0.22	0/1906	0.40	0/2597
2	D	0.21	0/1898	0.40	0/2586
3	E	0.24	0/1556	0.39	0/2106
3	G	0.21	0/1551	0.36	0/2101
All	All	0.22	0/10093	0.38	0/13716

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	1539	0	1457	22	0
1	C	1563	0	1484	14	0
2	B	1856	0	1761	24	0
2	D	1848	0	1748	17	0
3	E	1514	0	1406	10	0
3	G	1509	0	1403	15	0
All	All	9829	0	9259	96	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 5.

The worst 5 of 96 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:G:111:ARG:NH1	3:G:112:GLY:O	2.25	0.70
1:A:99:LEU:HD13	2:B:99:LEU:HD13	1.73	0.69
3:E:111:ARG:NH1	3:E:112:GLY:O	2.26	0.68
1:A:146:SER:O	1:A:191:SER:OG	2.15	0.64
2:D:182:SER:OG	2:D:183:ARG:N	2.33	0.62

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	193/219 (88%)	175 (91%)	17 (9%)	1 (0%)	32	68
1	C	197/219 (90%)	183 (93%)	13 (7%)	1 (0%)	32	68
2	B	236/254 (93%)	227 (96%)	9 (4%)	0	100	100
2	D	236/254 (93%)	226 (96%)	10 (4%)	0	100	100
3	E	182/204 (89%)	173 (95%)	9 (5%)	0	100	100
3	G	181/204 (89%)	173 (96%)	8 (4%)	0	100	100
All	All	1225/1354 (90%)	1157 (94%)	66 (5%)	2 (0%)	51	82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	192	ILE
1	C	192	ILE

### 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	174/194 (90%)	172 (99%)	2 (1%)	78	94
1	C	177/194 (91%)	171 (97%)	6 (3%)	42	76
2	B	199/212 (94%)	197 (99%)	2 (1%)	80	95
2	D	196/212 (92%)	194 (99%)	2 (1%)	80	95
3	E	150/161 (93%)	149 (99%)	1 (1%)	87	97
3	G	150/161 (93%)	149 (99%)	1 (1%)	87	97
All	All	1046/1134 (92%)	1032 (99%)	14 (1%)	73	93

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	159	CYS
1	C	164	ARG
2	D	110	LEU
1	C	31	TYR
1	C	196	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	95	HIS
2	D	6	GLN
2	D	203	HIS
2	D	207	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	A	197/219 (89%)	0.18	8 (4%) 38 32	23, 63, 112, 156	0
1	C	199/219 (90%)	0.07	6 (3%) 51 44	20, 53, 119, 152	0
2	B	238/254 (93%)	-0.12	1 (0%) 92 92	23, 47, 89, 118	0
2	D	238/254 (93%)	-0.02	4 (1%) 70 68	20, 49, 99, 135	0
3	E	186/204 (91%)	-0.04	3 (1%) 72 70	23, 41, 78, 155	0
3	G	185/204 (90%)	-0.03	0 100 100	22, 48, 78, 126	0
All	All	1243/1354 (91%)	0.00	22 (1%) 69 66	20, 50, 105, 156	0

The worst 5 of 22 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	E	90	GLY	8.0
1	A	142	GLN	3.2
2	D	180	ASN	3.0
3	E	91	GLY	3.0
2	D	220	THR	2.9

### 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.