



# wwPDB X-ray Structure Validation Summary Report

May 30, 2014 – 02:18 AM EDT

PDB ID : 4CSG  
Title : Structural insights into Toscana virus RNA encapsidation  
Authors : Olal, D.; Daumke, O.  
Deposited on : 2014-03-07  
Resolution : 3.32 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

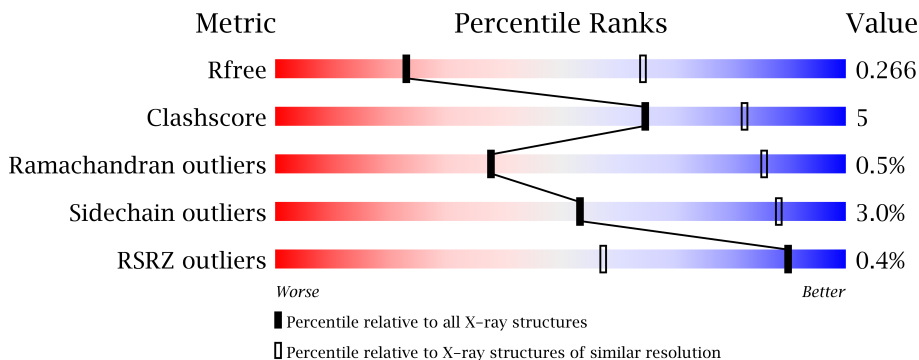
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : **FAILED**  
Xtriage (Phenix) : dev-1439  
EDS : stable23161  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable23161

# 1 Overall quality at a glance

The reported resolution of this entry is 3.32 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))
$R_{free}$	66092	1372 (3.44-3.20)
Clashscore	79885	1016 (3.42-3.22)
Ramachandran outliers	78287	1699 (3.44-3.20)
Sidechain outliers	78261	1697 (3.44-3.20)
RSRZ outliers	66119	1373 (3.44-3.20)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	253	
1	B	253	
1	C	253	
1	D	253	
1	E	253	
1	F	253	
1	H	253	
1	I	253	
1	J	253	
1	K	253	
1	L	253	
2	G	253	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 22844 atoms, of which 0 are hydrogen and 0 are deuterium.

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 NUCLEOPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	247	Total	C	N	O	S	0	0	1
			1889	1199	330	349	11			
1	B	248	Total	C	N	O	S	0	0	0
			1905	1208	332	354	11			
1	C	248	Total	C	N	O	S	0	0	1
			1901	1205	332	353	11			
1	D	249	Total	C	N	O	S	0	0	0
			1902	1207	332	352	11			
1	E	248	Total	C	N	O	S	0	0	0
			1905	1208	332	354	11			
1	F	248	Total	C	N	O	S	0	0	1
			1901	1205	332	353	11			
1	H	251	Total	C	N	O	S	0	0	0
			1927	1220	335	360	12			
1	I	248	Total	C	N	O	S	0	0	1
			1897	1203	332	351	11			
1	J	248	Total	C	N	O	S	0	0	0
			1901	1206	332	352	11			
1	K	249	Total	C	N	O	S	0	0	0
			1909	1210	333	355	11			
1	L	248	Total	C	N	O	S	0	0	1
			1901	1205	332	353	11			

- Molecule 2 is a protein called NUCLEOPROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	248	Total	C	N	O	S	0	0	0
			1906	1209	332	354	11			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	41	ILE	VAL	CONFLICT	UNP P21701



Sequence logo for the R213 site. The y-axis represents information content in bits, ranging from 0 to 1.5. The x-axis shows amino acid positions from M150 to V251. The R213 position is highlighted in orange. A red dot is placed above the A167 position, indicating a specific mutation or site of interest.

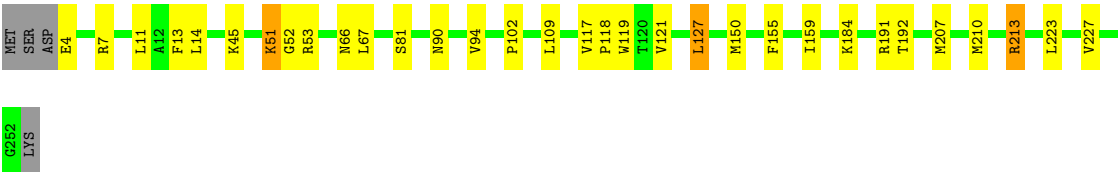
Diagram illustrating the protein structure with residues M210, R213, V227, D233, and V251. Residue D233 is highlighted in red, indicating a mutation site.

Q194  
A195  
N196  
A199  
F202  
M207  
M210  
R213  
F214  
F215  
V227  
A239  
R242  
V251  
GLY  
LYS

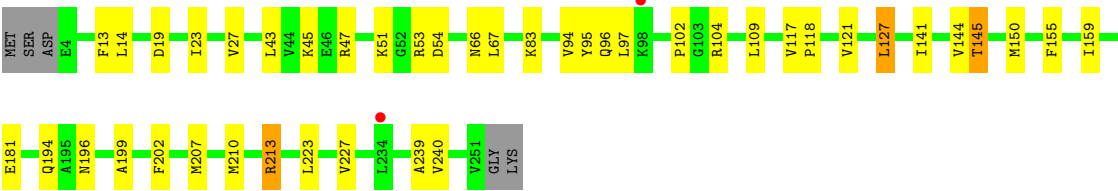
Sequence logo for the 1000th position. The y-axis represents information content in bits, ranging from 0 to 0.4. The x-axis shows amino acids: M207, M210, R213, F214, V227, V251, GLY, and LYS. The bars for M207, M210, V227, and V251 are yellow, while the bars for R213 and F214 are orange. The bars for GLY and LYS are grey.

V227 V251  
GLY LYS

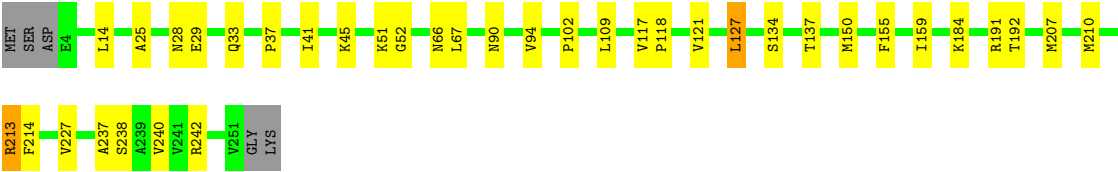
Chain K:



● Molecule 1: NUCLEOPROTEIN



● Molecule 2: NUCLEOPROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	104.48Å 104.48Å 510.76Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	34.12 – 3.32 34.12 – 3.32	Depositor EDS
% Data completeness (in resolution range)	99.1 (34.12-3.32) 99.2 (34.12-3.32)	Depositor EDS
$R_{merge}$	0.01	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.44 (at 3.32Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.240 , 0.268 0.238 , 0.266	Depositor DCC
$R_{free}$ test set	2293 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	66.8	Xtriage
Anisotropy	0.124	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 20.1	EDS
Estimated twinning fraction	0.095 for h,-h-k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.25$	Xtriage
Outliers	0 of 45888 reflections	Xtriage
$F_o, F_c$ correlation	0.88	EDS
Total number of atoms	22844	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.90% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality

### 5.1 Standard geometry

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	0.26	0/1921	0.43	0/2593
1	B	0.24	0/1937	0.42	0/2614
1	C	0.25	0/1933	0.41	0/2609
1	D	0.23	0/1934	0.42	0/2610
1	E	0.23	0/1937	0.41	0/2614
1	F	0.25	0/1933	0.43	0/2609
1	H	0.23	0/1959	0.41	0/2643
1	I	0.25	0/1929	0.45	0/2604
1	J	0.23	0/1933	0.42	0/2609
1	K	0.22	0/1941	0.41	0/2619
1	L	0.25	0/1933	0.43	0/2609
2	G	0.25	0/1938	0.42	0/2615
All	All	0.24	0/23228	0.42	0/31348

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1889	0	1937	32	0
1	B	1905	0	1949	30	1

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1901	0	1947	22	0
1	D	1902	0	1944	24	0
1	E	1905	0	1949	23	1
1	F	1901	0	1947	20	0
1	H	1927	0	1970	28	0
1	I	1897	0	1943	25	0
1	J	1901	0	1945	19	0
1	K	1909	0	1952	22	0
1	L	1901	0	1947	24	0
2	G	1906	0	1951	31	0
All	All	22844	0	23381	246	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 5.

The worst 5 of 246 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:G:29:GLU:OE2	1:L:83:LYS:NZ	2.16	0.78
1:A:106:THR:HG23	2:G:242:ARG:HH12	1.49	0.76
1:H:2:SER:HB2	1:H:5:ASN:HB2	1.67	0.76
1:A:122:GLN:NE2	1:B:14:LEU:O	2.19	0.75
1:A:34:GLY:HA2	1:F:79:LYS:HD2	1.69	0.75

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:20:SER:OG	1:E:19:ASP:OD1[1.545]	2.16	0.04

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	245/253 (97%)	232 (95%)	12 (5%)	1 (0%)	43 89

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	246/253 (97%)	236 (96%)	9 (4%)	1 (0%)	43	89
1	C	246/253 (97%)	230 (94%)	15 (6%)	1 (0%)	43	89
1	D	247/253 (98%)	236 (96%)	10 (4%)	1 (0%)	43	89
1	E	246/253 (97%)	235 (96%)	10 (4%)	1 (0%)	43	89
1	F	246/253 (97%)	235 (96%)	10 (4%)	1 (0%)	43	89
1	H	249/253 (98%)	237 (95%)	11 (4%)	1 (0%)	43	89
1	I	246/253 (97%)	234 (95%)	11 (4%)	1 (0%)	43	89
1	J	246/253 (97%)	232 (94%)	12 (5%)	2 (1%)	27	78
1	K	247/253 (98%)	235 (95%)	11 (4%)	1 (0%)	43	89
1	L	246/253 (97%)	232 (94%)	12 (5%)	2 (1%)	27	78
2	G	246/253 (97%)	235 (96%)	9 (4%)	2 (1%)	27	78
All	All	2956/3036 (97%)	2809 (95%)	132 (4%)	15 (0%)	38	86

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	L	96	GLN
1	A	102	PRO
1	B	102	PRO
1	C	102	PRO
1	D	102	PRO

### 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	A	199/206 (97%)	193 (97%)	6 (3%)	53	88
1	B	201/206 (98%)	195 (97%)	6 (3%)	53	88
1	C	201/206 (98%)	194 (96%)	7 (4%)	48	86
1	D	199/206 (97%)	194 (98%)	5 (2%)	60	90
1	E	201/206 (98%)	194 (96%)	7 (4%)	48	86
1	F	201/206 (98%)	195 (97%)	6 (3%)	53	88

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	H	204/206 (99%)	196 (96%)	8 (4%)	43	84
1	I	200/206 (97%)	193 (96%)	7 (4%)	48	86
1	J	200/206 (97%)	196 (98%)	4 (2%)	68	92
1	K	201/206 (98%)	196 (98%)	5 (2%)	60	90
1	L	201/206 (98%)	195 (97%)	6 (3%)	53	88
2	G	201/206 (98%)	196 (98%)	5 (2%)	60	90
All	All	2409/2472 (98%)	2337 (97%)	72 (3%)	53	88

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	51	LYS
2	G	207	MET
1	L	19	ASP
1	F	53	ARG
1	F	213	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	5	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

Mogul failed to run properly - this section will therefore be empty.

## 5.5 Carbohydrates ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.6 Ligand geometry ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.7 Other polymers ⓘ

Mogul failed to run properly - this section will therefore be empty.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	247/253 (97%)	-0.11	1 (0%) 90 57	29, 58, 100, 127	0
1	B	248/253 (98%)	-0.17	0 100 100	35, 56, 85, 106	0
1	C	248/253 (98%)	-0.08	1 (0%) 90 57	35, 62, 101, 136	0
1	D	249/253 (98%)	0.01	1 (0%) 90 57	32, 74, 130, 150	0
1	E	248/253 (98%)	0.45	6 (2%) 56 15	44, 105, 128, 144	0
1	F	248/253 (98%)	0.31	1 (0%) 90 57	52, 105, 138, 158	0
1	H	251/253 (99%)	-0.05	1 (0%) 90 57	35, 72, 101, 133	0
1	I	248/253 (98%)	0.10	0 100 100	52, 86, 113, 128	0
1	J	248/253 (98%)	-0.14	0 100 100	36, 64, 93, 110	0
1	K	249/253 (98%)	-0.02	0 100 100	35, 75, 121, 149	0
1	L	248/253 (98%)	-0.00	2 (0%) 83 39	37, 74, 123, 161	0
2	G	248/253 (98%)	-0.13	0 100 100	33, 57, 87, 98	0
All	All	2980/3036 (98%)	0.01	13 (0%) 90 57	29, 72, 121, 161	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	46	GLU	3.1
1	F	233	ASP	2.5
1	E	230	ILE	2.5
1	D	230	ILE	2.3
1	L	98	LYS	2.3

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

There are no ligands in this entry.

### 6.5 Other polymers ⓘ

There are no such residues in this entry.