wwPDB X-ray Structure Validation Summary Report (i)

Oct 11, 2023 – 11:24 AM EDT

PDB ID	:	7RK2
Title	:	Crystal structure of the human astrovirus serotype 8 capsid spike in complex with scFv 2D9, an astrovirus-neutralizing antibody, at 2.65-A resolution
Authors	:	Meyer, L.; Cuellar, C.; DuBois, R.M.
Deposited on	:	2021-07-21
Resolution	:	2.65 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (1)) were used in the production of this report:

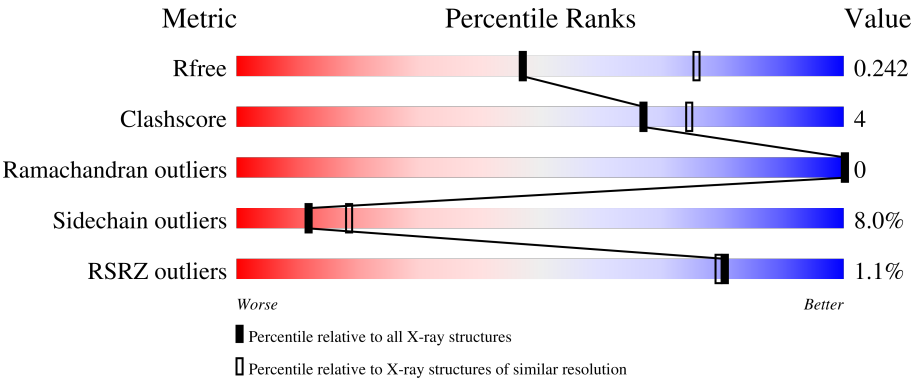
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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 of 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	230	81% 13% 6%
1	B	230	80% 15% 5%
2	C	251	% 72% 17% 11%
2	D	251	3% 71% 18% • 11%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 7106 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 Capsid protein VP25.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	217	Total	C	N	O	S	0	0	0
			1740	1113	292	329	6			
1	B	218	Total	C	N	O	S	0	1	0
			1754	1121	296	331	6			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	427	MET	-	initiating methionine	UNP Q9IFX1
A	428	GLY	-	expression tag	UNP Q9IFX1
A	648	ALA	-	expression tag	UNP Q9IFX1
A	649	ALA	-	expression tag	UNP Q9IFX1
A	650	GLU	-	expression tag	UNP Q9IFX1
A	651	LEU	-	expression tag	UNP Q9IFX1
A	652	ALA	-	expression tag	UNP Q9IFX1
A	653	LEU	-	expression tag	UNP Q9IFX1
A	654	VAL	-	expression tag	UNP Q9IFX1
A	655	PRO	-	expression tag	UNP Q9IFX1
A	656	ARG	-	expression tag	UNP Q9IFX1
B	427	MET	-	initiating methionine	UNP Q9IFX1
B	428	GLY	-	expression tag	UNP Q9IFX1
B	648	ALA	-	expression tag	UNP Q9IFX1
B	649	ALA	-	expression tag	UNP Q9IFX1
B	650	GLU	-	expression tag	UNP Q9IFX1
B	651	LEU	-	expression tag	UNP Q9IFX1
B	652	ALA	-	expression tag	UNP Q9IFX1
B	653	LEU	-	expression tag	UNP Q9IFX1
B	654	VAL	-	expression tag	UNP Q9IFX1
B	655	PRO	-	expression tag	UNP Q9IFX1
B	656	ARG	-	expression tag	UNP Q9IFX1

- Molecule 2 is a protein called scFv 2D9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	224	Total	C	N	O	S	0	1	0
			1731	1101	285	338	7			
2	D	224	Total	C	N	O	S	0	1	0
			1731	1101	285	338	7			

- Molecule 3 is water.

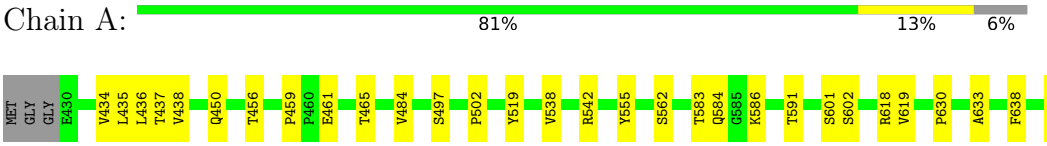
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	46	Total	O	0	0
			46	46		
3	B	39	Total	O	0	0
			39	39		
3	C	44	Total	O	0	0
			44	44		
3	D	21	Total	O	0	0
			21	21		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

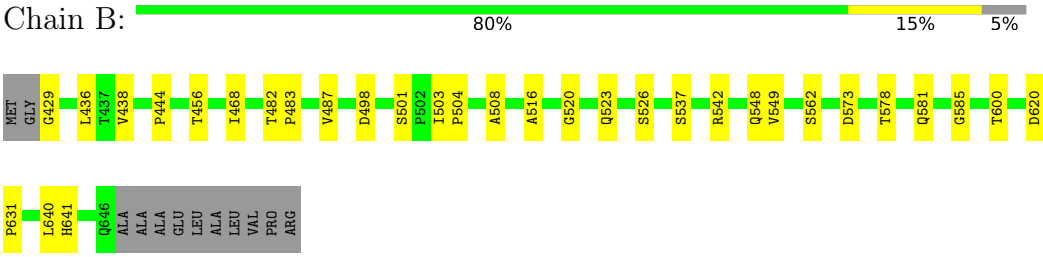
• Molecule 1: Capsid protein VP25

Chain A:



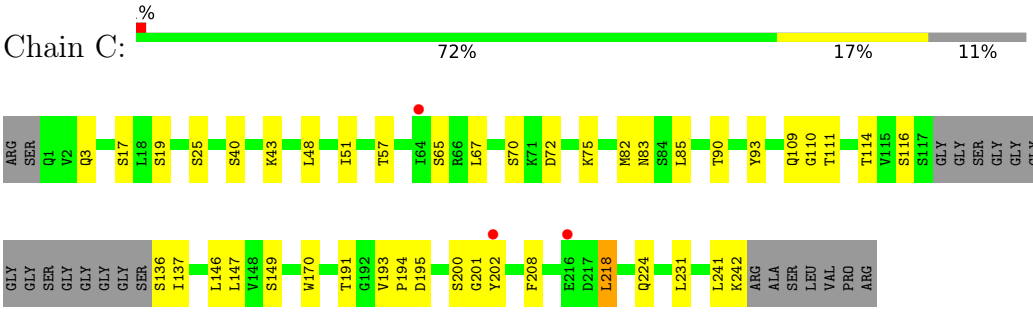
• Molecule 1: Capsid protein VP25

Chain B:



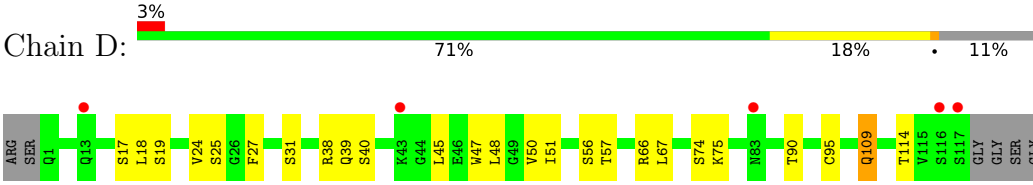
• Molecule 2: scFv 2D9

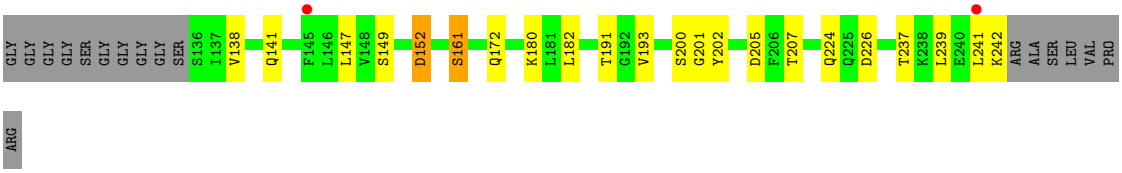
Chain C:



• Molecule 2: scFv 2D9

Chain D:




GLY GLY GLY GLY SER GLY GLY GLY GLY SER S136 I137 V138 Q141 F145 L146 L147 V148 S149 D152 S161 Q172 K180 L181 L182 T191 G192 V193 S200 G201 Y202 D205 F206 T207 Q224 Q225 D226 T237 K238 L239 E240 L241 K242 ARG ALA SER LEU VAL PRO

ARG

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	49.59Å 97.34Å 208.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	104.15 – 2.65 104.15 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.9 (104.15-2.65) 99.9 (104.15-2.65)	Depositor EDS
R_{merge}	0.45	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.93 (at 2.65Å)	Xtriage
Refinement program	PHENIX 1.19_4092	Depositor
R, R_{free}	0.218 , 0.247 0.212 , 0.242	Depositor DCC
R_{free} test set	1998 reflections (6.61%)	wwPDB-VP
Wilson B-factor (Å^2)	42.5	Xtriage
Anisotropy	0.440	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 42.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7106	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.75 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.4997e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

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	0.61	0/1793	0.76	0/2457
1	B	0.69	0/1808	0.73	0/2477
2	C	0.64	0/1771	0.75	0/2406
2	D	0.61	0/1771	0.69	0/2406
All	All	0.64	0/7143	0.73	0/9746

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	C	0	3
2	D	0	4
All	All	0	7

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	C	201	GLY	Mainchain
2	C	202[A]	TYR	Mainchain
2	D	201	GLY	Mainchain
2	D	202[A]	TYR	Mainchain
2	D	202[B]	TYR	Mainchain

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	1740	0	1680	14	0
1	B	1754	0	1689	13	0
2	C	1731	0	1685	12	0
2	D	1731	0	1685	15	0
3	A	46	0	0	1	0
3	B	39	0	0	1	0
3	C	44	0	0	1	0
3	D	21	0	0	0	0
All	All	7106	0	6739	53	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 4.

The worst 5 of 53 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:A:438:VAL:HG11	1:A:630:PRO:HG2	1.80	0.61
2:D:172:GLN:HB2	2:D:182:LEU:HD11	1.82	0.61
1:A:437:THR:HG23	1:A:633:ALA:HA	1.83	0.59
1:B:438:VAL:HG11	1:B:630:PRO:HG2	1.85	0.58
1:B:620:ASP:OD1	1:B:623:MET:HG3	2.04	0.58

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	215/230 (94%)	211 (98%)	4 (2%)	0	100	100
1	B	217/230 (94%)	211 (97%)	6 (3%)	0	100	100
2	C	221/251 (88%)	215 (97%)	6 (3%)	0	100	100
2	D	221/251 (88%)	214 (97%)	7 (3%)	0	100	100
All	All	874/962 (91%)	851 (97%)	23 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	197/204 (97%)	189 (96%)	8 (4%)	30	46
1	B	198/204 (97%)	188 (95%)	10 (5%)	24	37
2	C	192/203 (95%)	171 (89%)	21 (11%)	6	9
2	D	192/203 (95%)	169 (88%)	23 (12%)	5	7
All	All	779/814 (96%)	717 (92%)	62 (8%)	12	18

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	116	SER
2	D	180	LYS
2	C	218	LEU
2	D	161	SER
2	D	207	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	522	GLN
1	B	551	ASN
2	C	109	GLN

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 monosaccharides 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

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^{th} 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	217/230 (94%)	-0.23	0 100 100	22, 32, 46, 62	0
1	B	218/230 (94%)	-0.15	0 100 100	24, 35, 51, 64	0
2	C	224/251 (89%)	-0.03	3 (1%) 77 75	27, 40, 62, 83	0
2	D	224/251 (89%)	0.11	7 (3%) 49 45	27, 47, 69, 81	0
All	All	883/962 (91%)	-0.07	10 (1%) 80 79	22, 38, 62, 83	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	145	PHE	3.8
2	D	241	LEU	3.0
2	C	64	ILE	2.8
2	D	83	ASN	2.5
2	D	116	SER	2.5

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.