Full wwPDB X-ray Structure Validation Report (i)

May 29, 2020 – 09:18 am BST

PDB ID	:	6AVZ
Title	:	Crystal structure of the HopQ-CEACAM3 WT complex
Authors	:	Bonsor, D.A.; Sundberg, E.J.
Deposited on	:	2017-09-05
Resolution	:	2.05 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references (1)) were used in the production of this report:

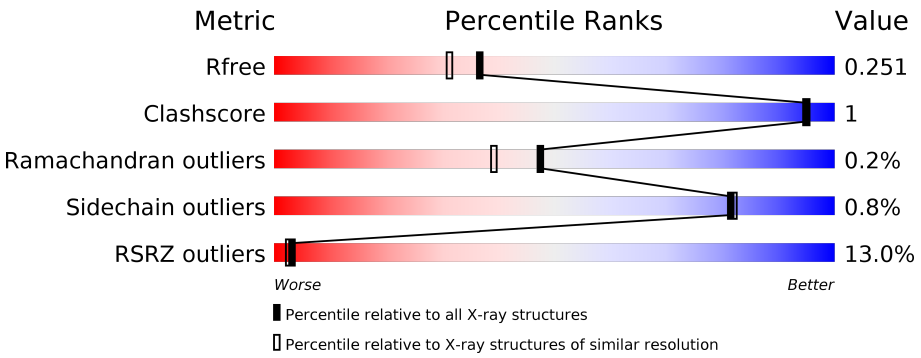
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.05 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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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 >=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	B	439	11% 77% • 21%
2	A	109	9% 94% 5% •
3	E	3	67% 33%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	CA	B	502	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3513 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 HopQ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	347	2591	1580	460	541	10	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	4	MET	-	initiating methionine	UNP H6A3H4
B	5	GLY	-	expression tag	UNP H6A3H4
B	6	SER	-	expression tag	UNP H6A3H4
B	7	SER	-	expression tag	UNP H6A3H4
B	8	HIS	-	expression tag	UNP H6A3H4
B	9	HIS	-	expression tag	UNP H6A3H4
B	10	HIS	-	expression tag	UNP H6A3H4
B	11	HIS	-	expression tag	UNP H6A3H4
B	12	HIS	-	expression tag	UNP H6A3H4
B	13	HIS	-	expression tag	UNP H6A3H4
B	14	SER	-	expression tag	UNP H6A3H4
B	15	GLN	-	expression tag	UNP H6A3H4
B	16	ASP	-	expression tag	UNP H6A3H4
B	17	PRO	-	expression tag	UNP H6A3H4

- Molecule 2 is a protein called Carcinoembryonic antigen-related cell adhesion molecule 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	108	842	539	138	164	1	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	initiating methionine	UNP P40198

- Molecule 3 is a protein called part of HopQ loop.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	E	3	15	9	3	3	0	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
4	B	2	2	2	0	0

- Molecule 5 is water.

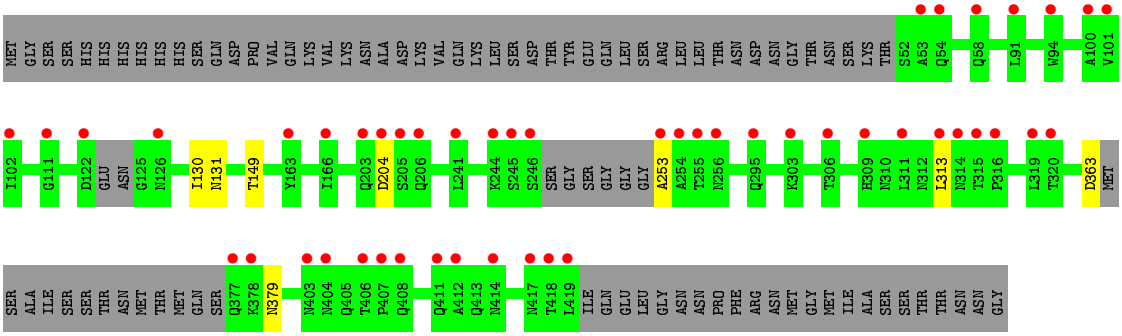
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	B	41	41	41	0	0
			Total	O		
5	A	21	21	21	0	0
			Total	O		
5	E	1	1	1	0	0

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.

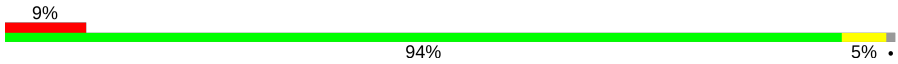
- Molecule 1: HopQ

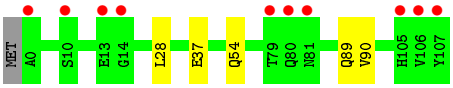
Chain B:



- Molecule 2: Carcinoembryonic antigen-related cell adhesion molecule 3

Chain A:





- Molecule 3: part of HopQ loop

Chain E:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	39.48Å 103.26Å 112.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	103.26 – 2.05 27.40 – 2.05	Depositor EDS
% Data completeness (in resolution range)	99.1 (103.26-2.05) 99.2 (27.40-2.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	1.28 (at 2.04Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.203 , 0.250 0.213 , 0.251	Depositor DCC
R_{free} test set	1483 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	39.5	Xtriage
Anisotropy	0.378	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 40.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	3513	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.02% of the height of the origin peak. No significant pseudotranslation is detected.*

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	0.36	0/2622	0.51	0/3553
2	A	0.38	0/859	0.60	0/1168
All	All	0.37	0/3481	0.54	0/4721

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	B	2591	0	2522	5	0
2	A	842	0	832	4	0
3	E	15	0	5	3	0
4	B	2	0	0	0	0
5	A	21	0	0	1	0
5	B	41	0	0	0	0
5	E	1	0	0	0	0
All	All	3513	0	3359	8	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 1.

All (8) 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:B:131:ASN:H	3:E:3:UNK:CB	1.91	0.83
2:A:37:GLU:O	5:A:201:HOH:O	2.11	0.67
1:B:131:ASN:N	3:E:3:UNK:CB	2.66	0.55
2:A:28:LEU:HD22	2:A:90:VAL:HG13	1.93	0.51
1:B:130:ILE:HA	3:E:3:UNK:CB	2.43	0.48
2:A:28:LEU:HD22	2:A:90:VAL:CG1	2.48	0.43
1:B:253:ALA:O	1:B:379:ASN:ND2	2.48	0.41
1:B:149:THR:HG21	2:A:89:GLN:HE22	1.85	0.41

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	339/439 (77%)	328 (97%)	10 (3%)	1 (0%)	41	31
2	A	106/109 (97%)	104 (98%)	2 (2%)	0	100	100
All	All	445/548 (81%)	432 (97%)	12 (3%)	1 (0%)	47	39

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	204	ASP

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	B	286/368 (78%)	284 (99%)	2 (1%)	84	84
2	A	91/92 (99%)	90 (99%)	1 (1%)	73	73
All	All	377/460 (82%)	374 (99%)	3 (1%)	81	82

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	313	LEU
1	B	363	ASP
2	A	54	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	167	HIS
1	B	221	GLN
1	B	285	ASN
1	B	301	GLN
1	B	410	ASN
2	A	89	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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95th 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	B	347/439 (79%)	0.72	49 (14%) 2 2	31, 44, 81, 108	0
2	A	108/109 (99%)	0.48	10 (9%) 8 9	32, 42, 68, 90	0
3	E	0/3	-	-	-	-
All	All	455/551 (82%)	0.66	59 (12%) 3 2	31, 44, 81, 108	0

All (59) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	204	ASP	7.4
1	B	319	LEU	5.7
1	B	205	SER	5.6
2	A	107	TYR	5.4
1	B	256	ASN	5.2
1	B	246	SER	4.9
1	B	306	THR	4.5
1	B	203	GLN	4.5
1	B	311	LEU	4.3
1	B	309	HIS	4.2
1	B	320	THR	4.2
1	B	316	PRO	3.9
1	B	53	ALA	3.9
1	B	253	ALA	3.8
1	B	255	THR	3.8
1	B	378	LYS	3.6
1	B	407	PRO	3.6
1	B	54	GLN	3.5
2	A	14	GLY	3.5
1	B	377	GLN	3.4
1	B	417	ASN	3.3
1	B	404	ASN	3.2
1	B	91	LEU	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	313	LEU	3.1
1	B	126	ASN	3.0
1	B	414	ASN	3.0
1	B	94	TRP	2.9
2	A	106	VAL	2.9
1	B	315	THR	2.9
1	B	419	LEU	2.9
2	A	80	GLN	2.8
1	B	111	GLY	2.8
2	A	10	SER	2.8
1	B	122	ASP	2.8
2	A	13	GLU	2.8
1	B	101	VAL	2.7
1	B	166	ILE	2.7
1	B	163	TYR	2.7
1	B	241	LEU	2.6
1	B	58	GLN	2.6
2	A	105	HIS	2.5
1	B	303	LYS	2.5
1	B	403	ASN	2.5
2	A	79	THR	2.5
1	B	206	GLN	2.5
1	B	102	ILE	2.4
1	B	408	GLN	2.4
1	B	314	ASN	2.3
1	B	406	THR	2.3
2	A	81	ASN	2.2
1	B	100	ALA	2.2
1	B	418	THR	2.2
1	B	244	LYS	2.1
1	B	245	SER	2.1
1	B	411	GLN	2.1
1	B	254	ALA	2.1
1	B	295	GLN	2.0
1	B	412	ALA	2.0
2	A	0	ALA	2.0

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	CA	B	502	1/1	0.30	0.56	134,134,134,134	1
4	CA	B	501	1/1	0.88	0.09	84,84,84,84	0

6.5 Other polymers (i)

There are no such residues in this entry.