WORLDWIDE

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 27, 2014 – 04:17 PM GMT

PDB ID : 2H3R

Title : Crystal structure of ORF52 from Murid herpesvirus 4 (MuHV-4) (Murine gammaherpesvirus 68). Northeast Structural Genomics Consortium target MhR28B.

Authors : Benach, J.; Chen, Y.; Seetharaman, J.; Janjua, H.; Xiao, R.; Cunningham, K.; Ma, L.-C.; Ho, C.K.; Acton, T.B.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)

Deposited on : 2006-05-23

Resolution : 2.70 Å(reported)

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

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
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)	:	stable22683

1 Overall quality at a glance ⓘ

The reported resolution of this entry is 2.70 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	Percentile Ranks	Value
Rfree		0.315
Clashscore		27
Ramachandran outliers		0.3%
Sidechain outliers		5.4%
RSRZ outliers		4.2%

Worse   Better

█ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
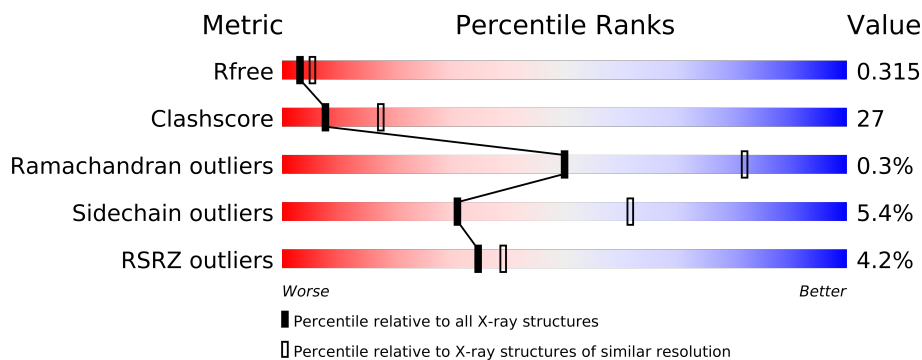
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1557 (2.70-2.70)
Clashscore	79885	1939 (2.70-2.70)
Ramachandran outliers	78287	1905 (2.70-2.70)
Sidechain outliers	78261	1905 (2.70-2.70)
RSRZ outliers	66119	1559 (2.70-2.70)

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. 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	110	
1	B	110	
1	C	110	
1	D	110	

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2779 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 Hypothetical protein BQLF2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	90	Total	C	N	O	Se	0	0	0
			703	434	125	141	3			
1	B	88	Total	C	N	O	Se	0	0	0
			686	425	124	134	3			
1	C	89	Total	C	N	O	Se	0	0	0
			695	430	124	138	3			
1	D	88	Total	C	N	O	Se	0	0	0
			685	424	124	134	3			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP P88989
A	13	MSE	MET	MODIFIED RESIDUE	UNP P88989
A	65	MSE	MET	MODIFIED RESIDUE	UNP P88989
A	100	MSE	MET	MODIFIED RESIDUE	UNP P88989
A	103	LEU	-	CLONING ARTIFACT	UNP P88989
A	104	GLU	-	CLONING ARTIFACT	UNP P88989
A	105	HIS	-	CLONING ARTIFACT	UNP P88989
A	106	HIS	-	CLONING ARTIFACT	UNP P88989
A	107	HIS	-	CLONING ARTIFACT	UNP P88989
A	108	HIS	-	CLONING ARTIFACT	UNP P88989
A	109	HIS	-	CLONING ARTIFACT	UNP P88989
A	110	HIS	-	CLONING ARTIFACT	UNP P88989
B	1	MSE	MET	MODIFIED RESIDUE	UNP P88989
B	13	MSE	MET	MODIFIED RESIDUE	UNP P88989
B	65	MSE	MET	MODIFIED RESIDUE	UNP P88989
B	100	MSE	MET	MODIFIED RESIDUE	UNP P88989
B	103	LEU	-	CLONING ARTIFACT	UNP P88989
B	104	GLU	-	CLONING ARTIFACT	UNP P88989
B	105	HIS	-	CLONING ARTIFACT	UNP P88989
B	106	HIS	-	CLONING ARTIFACT	UNP P88989
B	107	HIS	-	CLONING ARTIFACT	UNP P88989

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	108	HIS	-	CLONING ARTIFACT	UNP P88989
B	109	HIS	-	CLONING ARTIFACT	UNP P88989
B	110	HIS	-	CLONING ARTIFACT	UNP P88989
C	1	MSE	MET	MODIFIED RESIDUE	UNP P88989
C	13	MSE	MET	MODIFIED RESIDUE	UNP P88989
C	65	MSE	MET	MODIFIED RESIDUE	UNP P88989
C	100	MSE	MET	MODIFIED RESIDUE	UNP P88989
C	103	LEU	-	CLONING ARTIFACT	UNP P88989
C	104	GLU	-	CLONING ARTIFACT	UNP P88989
C	105	HIS	-	CLONING ARTIFACT	UNP P88989
C	106	HIS	-	CLONING ARTIFACT	UNP P88989
C	107	HIS	-	CLONING ARTIFACT	UNP P88989
C	108	HIS	-	CLONING ARTIFACT	UNP P88989
C	109	HIS	-	CLONING ARTIFACT	UNP P88989
C	110	HIS	-	CLONING ARTIFACT	UNP P88989
D	1	MSE	MET	MODIFIED RESIDUE	UNP P88989
D	13	MSE	MET	MODIFIED RESIDUE	UNP P88989
D	65	MSE	MET	MODIFIED RESIDUE	UNP P88989
D	100	MSE	MET	MODIFIED RESIDUE	UNP P88989
D	103	LEU	-	CLONING ARTIFACT	UNP P88989
D	104	GLU	-	CLONING ARTIFACT	UNP P88989
D	105	HIS	-	CLONING ARTIFACT	UNP P88989
D	106	HIS	-	CLONING ARTIFACT	UNP P88989
D	107	HIS	-	CLONING ARTIFACT	UNP P88989
D	108	HIS	-	CLONING ARTIFACT	UNP P88989
D	109	HIS	-	CLONING ARTIFACT	UNP P88989
D	110	HIS	-	CLONING ARTIFACT	UNP P88989

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total 3	O 3	0	0
2	B	1	Total 1	O 1	0	0
2	C	3	Total 3	O 3	0	0
2	D	3	Total 3	O 3	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.
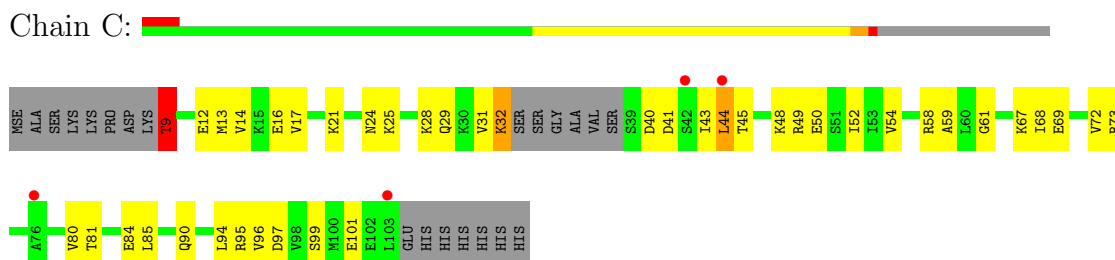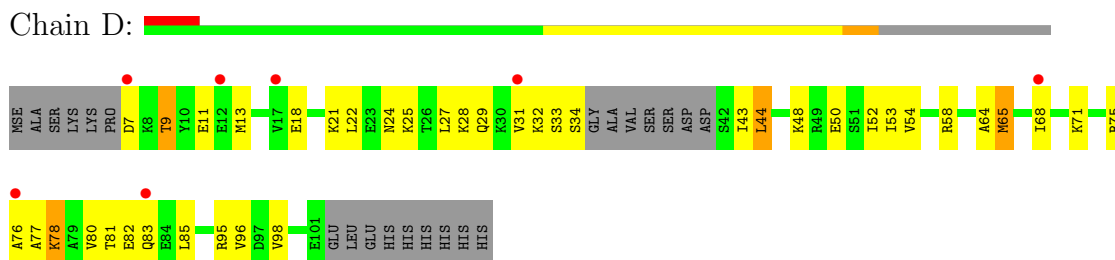
- Molecule 1: Hypothetical protein BQLF2
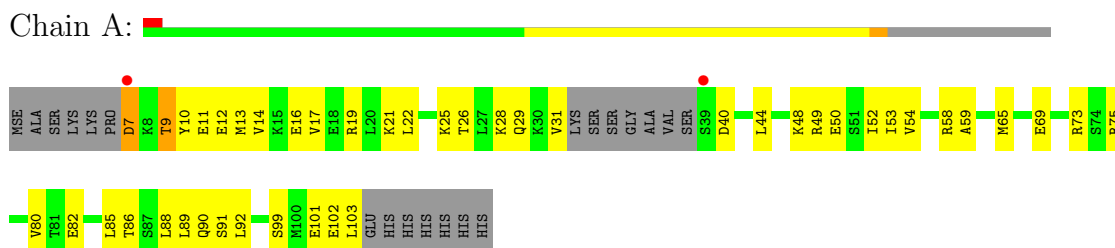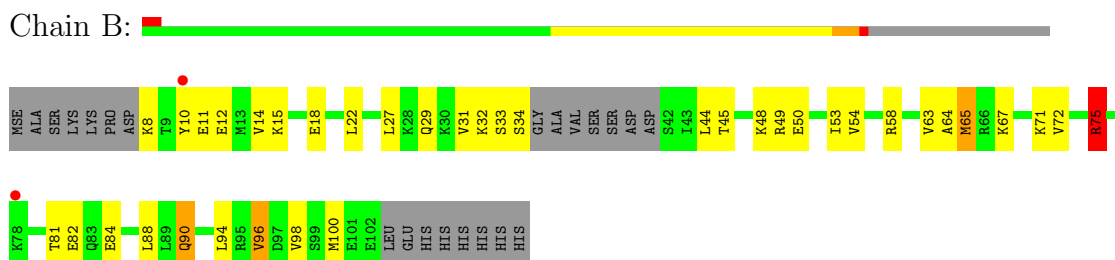
Chain A:

- Molecule 1: Hypothetical protein BQLF2

Chain B:

- Molecule 1: Hypothetical protein BQLF2

Chain C:

- Molecule 1: Hypothetical protein BQLF2

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	54.93Å 49.24Å 88.92Å 90.00° 105.61° 90.00°	Depositor
Resolution (Å)	20.00 – 2.70 49.24 – 2.68	Depositor EDS
% Data completeness (in resolution range)	83.4 (20.00-2.70) 95.5 (49.24-2.68)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.25 (at 2.69Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.275 , 0.295 0.291 , 0.315	Depositor DCC
R_{free} test set	1102 reflections (8.93%)	DCC
Wilson B-factor (Å^2)	45.3	Xtriage
Anisotropy	0.360	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 40.3	EDS
Estimated twinning fraction	0.014 for h,-k,-h-l	Xtriage
L-test for twinning	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Outliers	0 of 23521 reflections	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	2779	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 25.11 % 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 3.3731e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[1]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	0.37	0/699	0.65	0/927
1	B	0.39	0/682	0.66	1/902 (0.1%)
1	C	0.48	1/691 (0.1%)	0.65	1/916 (0.1%)
1	D	0.36	0/681	0.62	0/901
All	All	0.40	1/2753 (0.0%)	0.64	2/3646 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	9	THR	CB-CG2	-7.08	1.28	1.52

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	9	THR	OG1-CB-CG2	-5.42	97.52	110.00
1	B	75	ARG	NE-CZ-NH1	5.12	122.86	120.30

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	703	0	753	48	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	686	0	748	44	0
1	C	695	0	749	55	0
1	D	685	0	746	50	0
2	A	3	0	0	0	0
2	B	1	0	0	0	0
2	C	3	0	0	0	0
2	D	3	0	0	1	0
All	All	2779	0	2996	158	0

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 27.

The worst 5 of 158 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:29:GLN:O	1:B:32:LYS:HG2	1.65	0.97
1:A:85:LEU:HD22	1:B:44:LEU:HD21	1.50	0.91
1:A:22:LEU:HD11	1:B:90:GLN:HG2	1.53	0.89
1:C:44:LEU:HD21	1:D:85:LEU:HD22	1.56	0.87
1:C:25:LYS:HG2	1:C:29:GLN:HE21	1.40	0.87

There are no symmetry-related clashes.

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	86/110 (78%)	79 (92%)	7 (8%)	0	100	100
1	B	84/110 (76%)	79 (94%)	5 (6%)	0	100	100
1	C	85/110 (77%)	82 (96%)	2 (2%)	1 (1%)	19	45
1	D	84/110 (76%)	78 (93%)	6 (7%)	0	100	100
All	All	339/440 (77%)	318 (94%)	20 (6%)	1 (0%)	50	82

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	31	VAL

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	80/93 (86%)	74 (92%)	6 (8%)	19	43
1	B	78/93 (84%)	74 (95%)	4 (5%)	33	64
1	C	79/93 (85%)	76 (96%)	3 (4%)	44	76
1	D	78/93 (84%)	74 (95%)	4 (5%)	33	64
All	All	315/372 (85%)	298 (95%)	17 (5%)	31	61

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	75	ARG
1	B	90	GLN
1	D	9	THR
1	B	65	MSE
1	D	44	LEU

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

Mol	Chain	Res	Type
1	A	90	GLN
1	C	29	GLN
1	C	90	GLN
1	D	90	GLN

5.3.3 RNA (i)

There are no RNA chains 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 carbohydrates 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, 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	A	90/110 (81%)	0.59	2 (2%)	59	65	21, 37, 56, 71	0
1	B	88/110 (80%)	0.72	2 (2%)	57	64	30, 41, 54, 60	0
1	C	89/110 (80%)	0.65	4 (4%)	32	36	20, 37, 55, 67	0
1	D	88/110 (80%)	0.75	7 (7%)	12	13	32, 44, 52, 59	0
All	All	355/440 (80%)	0.68	15 (4%)	35	39	20, 40, 55, 71	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	39	SER	4.4
1	B	10	TYR	3.5
1	D	7	ASP	3.2
1	D	76	ALA	3.2
1	C	44	LEU	3.1

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.