



# Full wwPDB X-ray Structure Validation Report ⓘ

May 8, 2017 – 05:20 PM EDT

PDB ID : 5J9I  
Title : Crystal structure of the HigA2 antitoxin C-terminal domain  
Authors : Hadzi, S.; Loris, R.  
Deposited on : 2016-04-10  
Resolution : 1.80 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](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 : rb-20029077  
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) : rb-20029077

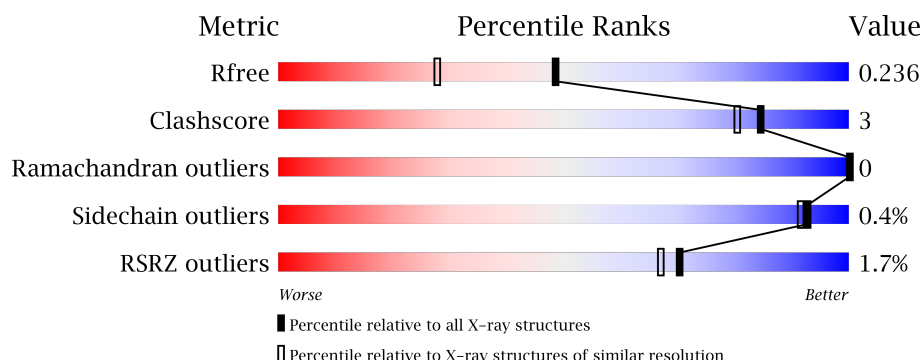
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	4827 (1.80-1.80)
Clashscore	112137	5742 (1.80-1.80)
Ramachandran outliers	110173	5676 (1.80-1.80)
Sidechain outliers	110143	5675 (1.80-1.80)
RSRZ outliers	101464	4906 (1.80-1.80)

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	104	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between;"> <span>64%</span> <span>36%</span> </div> </div>
1	B	104	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between;"> <span>62%</span> <span>35%</span> </div> </div>
1	C	104	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between;"> <span>64%</span> <span>34%</span> </div> </div>
1	D	104	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between;"> <span>64%</span> <span>35%</span> </div> </div>
1	E	104	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between;"> <span>56%</span> <span>37%</span> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	104	
1	G	104	
1	H	104	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8751 atoms, of which 4244 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 Antitoxin igA-2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	67	Total	C	H	N	O	S	0	0	0
			1077	334	541	101	100	1			
1	B	68	Total	C	H	N	O	S	0	0	0
			1066	334	531	99	101	1			
1	C	69	Total	C	H	N	O	S	0	1	0
			1094	341	543	105	104	1			
1	D	68	Total	C	H	N	O	S	0	0	0
			1074	336	535	99	103	1			
1	E	66	Total	C	H	N	O	S	0	1	0
			1047	328	520	98	100	1			
1	F	68	Total	C	H	N	O	S	0	0	0
			1066	334	531	99	101	1			
1	G	65	Total	C	H	N	O	S	0	0	0
			1026	321	512	95	97	1			
1	H	68	Total	C	H	N	O	S	0	0	0
			1066	334	531	99	101	1			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	35	Total	O	0	0
			35	35		
2	B	32	Total	O	0	0
			32	32		
2	C	43	Total	O	0	0
			43	43		
2	D	37	Total	O	0	0
			37	37		
2	E	23	Total	O	0	0
			23	23		
2	F	17	Total	O	0	0
			17	17		

*Continued on next page...*

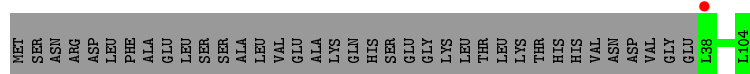
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	18	Total	O	0	0
			18	18		
2	H	30	Total	O	0	0
			30	30		

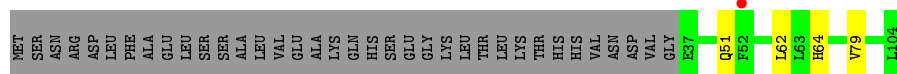
### 3 Residue-property plots [i](#)

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.

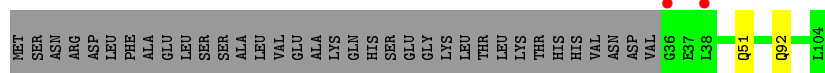
- Molecule 1: Antitoxin igA-2



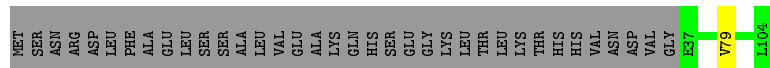
- Molecule 1: Antitoxin igA-2



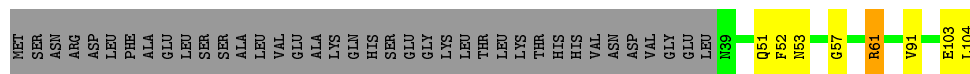
- Molecule 1: Antitoxin igA-2



- Molecule 1: Antitoxin igA-2

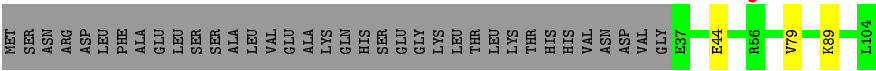


- Molecule 1: Antitoxin igA-2

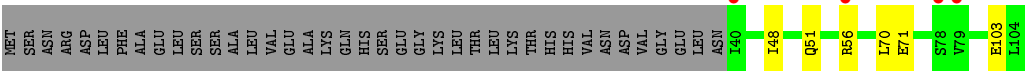


- Molecule 1: Antitoxin igA-2

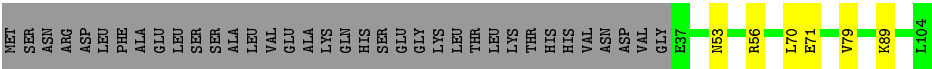




● Molecule 1: Antitoxin igA-2



● Molecule 1: Antitoxin igA-2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	115.36Å 61.16Å 73.85Å 90.00° 106.69° 90.00°	Depositor
Resolution (Å)	38.51 – 1.80 40.08 – 1.80	Depositor EDS
% Data completeness (in resolution range)	97.4 (38.51-1.80) 97.6 (40.08-1.80)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.23 (at 1.79Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.194 , 0.236 0.194 , 0.236	Depositor DCC
$R_{free}$ test set	2242 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	26.6	Xtriage
Anisotropy	0.574	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 52.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	8751	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	37.0	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 79.11 % 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 5.9975e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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

### 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	0.68	0/545	0.75	0/738
1	B	0.71	1/544 (0.2%)	0.88	3/738 (0.4%)
1	C	0.65	0/560	0.73	0/758
1	D	0.58	0/548	0.69	0/743
1	E	0.63	0/536	0.72	2/727 (0.3%)
1	F	0.58	0/544	0.65	0/738
1	G	0.59	0/523	0.65	0/709
1	H	0.62	0/544	0.64	0/738
All	All	0.63	1/4344 (0.0%)	0.72	5/5889 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	62	LEU	C-N	-10.22	1.10	1.34

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	62	LEU	C-N-CA	9.79	146.18	121.70
1	B	62	LEU	O-C-N	-9.31	107.81	122.70
1	B	62	LEU	CA-C-N	6.56	131.63	117.20
1	E	61	ARG	CG-CD-NE	-5.90	99.42	111.80
1	E	61	ARG	NE-CZ-NH1	-5.08	117.76	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	536	541	540	0	0
1	B	535	531	530	5	0
1	C	551	543	543	3	0
1	D	539	535	535	2	0
1	E	527	520	519	6	0
1	F	535	531	531	3	0
1	G	514	512	512	4	0
1	H	535	531	531	6	0
2	A	35	0	0	0	2
2	B	32	0	0	2	0
2	C	43	0	0	2	0
2	D	37	0	0	0	1
2	E	23	0	0	4	0
2	F	17	0	0	0	0
2	G	18	0	0	0	1
2	H	30	0	0	1	0
All	All	4507	4244	4241	23	2

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 3.

All (23) 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:H:89:LYS:NZ	2:H:201:HOH:O	2.20	0.73
1:E:104:LEU:O	2:E:201:HOH:O	2.13	0.65
1:B:79:VAL:HB	1:H:79:VAL:CG2	2.26	0.65
1:C:51:GLN:OE1	2:C:201:HOH:O	2.14	0.65
1:D:79:VAL:CG2	1:F:79:VAL:HB	2.27	0.65
1:H:56:ARG:NE	1:H:71:GLU:OE1	2.33	0.61
1:C:92[A]:GLN:NE2	2:C:202:HOH:O	2.27	0.53
1:G:56:ARG:HD3	1:G:71:GLU:HB2	1.92	0.52
1:C:51:GLN:OE1	1:C:92[B]:GLN:NE2	2.43	0.51
1:E:51:GLN:NE2	2:E:204:HOH:O	2.44	0.51
1:B:79:VAL:HB	1:H:79:VAL:HG22	1.93	0.51
1:E:53[A]:ASN:OD1	2:E:202:HOH:O	2.20	0.50
1:D:79:VAL:HG22	1:F:79:VAL:HB	1.94	0.49
1:B:51:GLN:NE2	2:B:203:HOH:O	2.41	0.49
1:F:44:GLU:OE2	1:F:89:LYS:NZ	2.48	0.45

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:56:ARG:HG3	1:G:70:LEU:HD23	1.99	0.44
1:E:57:GLY:HA3	1:G:103:GLU:HG2	2.01	0.43
1:B:64:HIS:ND1	2:B:202:HOH:O	2.36	0.43
1:B:79:VAL:HB	1:H:79:VAL:HG23	2.02	0.42
1:H:56:ARG:HG3	1:H:70:LEU:HD23	2.02	0.42
1:E:103:GLU:OE1	2:E:203:HOH:O	2.21	0.42
1:E:52:PHE:CE2	1:E:91:VAL:HG12	2.55	0.41
1:G:48:ILE:O	1:G:51:GLN:HB2	2.21	0.40

All (2) 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	Interatomic distance (Å)	Clash overlap (Å)
2:A:205:HOH:O	2:D:234:HOH:O[2_556]	2.11	0.09
2:A:214:HOH:O	2:G:217:HOH:O[1_565]	2.15	0.05

## 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	65/104 (62%)	64 (98%)	1 (2%)	0	100	100
1	B	66/104 (64%)	66 (100%)	0	0	100	100
1	C	68/104 (65%)	68 (100%)	0	0	100	100
1	D	66/104 (64%)	66 (100%)	0	0	100	100
1	E	65/104 (62%)	65 (100%)	0	0	100	100
1	F	66/104 (64%)	66 (100%)	0	0	100	100
1	G	63/104 (61%)	63 (100%)	0	0	100	100
1	H	66/104 (64%)	66 (100%)	0	0	100	100
All	All	525/832 (63%)	524 (100%)	1 (0%)	0	100	100

There are no Ramachandran outliers to report.

### 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	61/93 (66%)	61 (100%)	0	100	100
1	B	60/93 (64%)	60 (100%)	0	100	100
1	C	61/93 (66%)	61 (100%)	0	100	100
1	D	61/93 (66%)	61 (100%)	0	100	100
1	E	59/93 (63%)	58 (98%)	1 (2%)	66	55
1	F	60/93 (64%)	60 (100%)	0	100	100
1	G	58/93 (62%)	58 (100%)	0	100	100
1	H	60/93 (64%)	59 (98%)	1 (2%)	66	55
All	All	480/744 (64%)	478 (100%)	2 (0%)	93	91

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	61	ARG
1	H	53	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	39	ASN
1	G	51	GLN

### 5.3.3 RNA ⓘ

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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	62:LEU	C	63:LEU	N	1.10

## 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	67/104 (64%)	0.22	1 (1%) 74 70	18, 29, 45, 64	0
1	B	68/104 (65%)	0.15	1 (1%) 74 70	20, 31, 48, 52	0
1	C	69/104 (66%)	0.23	2 (2%) 52 47	19, 29, 52, 80	0
1	D	68/104 (65%)	0.09	0 100 100	18, 31, 48, 56	0
1	E	66/104 (63%)	0.09	0 100 100	23, 33, 49, 73	0
1	F	68/104 (65%)	0.15	1 (1%) 74 70	23, 33, 54, 73	0
1	G	65/104 (62%)	0.19	4 (6%) 21 17	22, 33, 52, 63	0
1	H	68/104 (65%)	0.06	0 100 100	20, 31, 50, 72	0
All	All	539/832 (64%)	0.15	9 (1%) 70 67	18, 31, 51, 80	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	36	GLY	4.1
1	C	38	LEU	3.0
1	G	79	VAL	2.8
1	F	56	ARG	2.4
1	G	56	ARG	2.3
1	G	40	ILE	2.3
1	G	78	SER	2.1
1	B	52	PHE	2.0
1	A	38	LEU	2.0

### 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 [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.