



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 8, 2018 – 08:10 pm GMT

PDB ID : 5GRU  
Title : Structure of mono-specific diabody  
Authors : Kim, J.H.; Song, D.H.; Youn, S.J.; Cho, G.; Lee, H.; Lee, J.O.  
Deposited on : 2016-08-12  
Resolution : 1.96 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

<https://www.wwpdb.org/validation/2017/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) : 1.13  
EDS : trunk30967  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30967

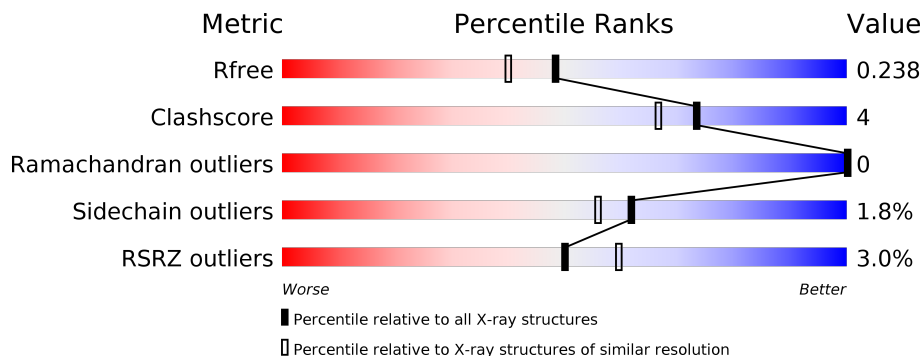
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.96 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))
$R_{free}$	111664	2220 (1.96-1.96)
Clashscore	122126	2333 (1.96-1.96)
Ramachandran outliers	120053	2314 (1.96-1.96)
Sidechain outliers	120020	2314 (1.96-1.96)
RSRZ outliers	108989	2174 (1.96-1.96)

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	383	 % 86% 7% 6%
2	H	249	 7% 87% 8% . .
3	L	228	 % 93% 7%

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 7153 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 Maltose-binding periplasmic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	360	2782	1792	450	534	6	0	0	0

There are 17 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	367	LEU	-	expression tag	UNP P0AEX9
A	368	VAL	-	expression tag	UNP P0AEX9
A	369	PRO	-	expression tag	UNP P0AEX9
A	370	ARG	-	expression tag	UNP P0AEX9
A	371	GLY	-	expression tag	UNP P0AEX9
A	372	SER	-	expression tag	UNP P0AEX9
A	373	ALA	-	expression tag	UNP P0AEX9
A	374	ALA	-	expression tag	UNP P0AEX9
A	375	ALA	-	expression tag	UNP P0AEX9
A	376	LEU	-	expression tag	UNP P0AEX9
A	377	GLU	-	expression tag	UNP P0AEX9
A	378	HIS	-	expression tag	UNP P0AEX9
A	379	HIS	-	expression tag	UNP P0AEX9
A	380	HIS	-	expression tag	UNP P0AEX9
A	381	HIS	-	expression tag	UNP P0AEX9
A	382	HIS	-	expression tag	UNP P0AEX9
A	383	HIS	-	expression tag	UNP P0AEX9

- Molecule 2 is a protein called diabody protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	H	240	1816	1140	303	366	7	0	0	0

- Molecule 3 is a protein called diabody protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	L	228	1746	1100	292	348	6	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	320	Total 320	O 320	0	0
4	H	211	Total 211	O 211	0	0
4	L	278	Total 278	O 278	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	197.27Å 50.82Å 105.87Å 90.00° 111.28° 90.00°	Depositor
Resolution (Å)	33.78 – 1.96 33.78 – 1.95	Depositor EDS
% Data completeness (in resolution range)	93.1 (33.78-1.96) 89.0 (33.78-1.95)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.23 (at 1.95Å)	Xtrriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, $R_{free}$	0.197 , 0.237 0.198 , 0.238	Depositor DCC
$R_{free}$ test set	2000 reflections (3.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	18.8	Xtrriage
Anisotropy	0.517	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 49.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	7153	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.05% of the height of the origin peak. No significant pseudotranslation is detected.*

<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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.42	0/2851	0.55	1/3878 (0.0%)
2	H	0.41	0/1860	0.64	1/2526 (0.0%)
3	L	0.42	0/1787	0.64	0/2426
All	All	0.42	0/6498	0.60	2/8830 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	H	98	ARG	NE-CZ-NH1	-6.83	116.89	120.30
1	A	43	LEU	CA-CB-CG	6.51	130.28	115.30

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	A	2782	0	2727	20	0
2	H	1816	0	1718	17	0
3	L	1746	0	1693	10	0
4	A	320	0	0	12	0
4	H	211	0	0	2	0
4	L	278	0	0	2	0
All	All	7153	0	6138	45	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.

All (45) 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)
3:L:124:GLN:OE1	4:L:301:HOH:O	1.79	1.01
2:H:130:GLY:H	2:H:131:GLY:HA2	1.30	0.92
2:H:217:ILE:HD11	2:H:240:ILE:HG12	1.52	0.92
1:A:239:LYS:NZ	4:A:404:HOH:O	2.04	0.90
1:A:98:ARG:NH1	4:A:403:HOH:O	1.98	0.86
1:A:345:THR:OG1	4:A:402:HOH:O	1.97	0.83
2:H:135:ASP:O	4:H:301:HOH:O	2.03	0.75
1:A:256:LYS:NZ	4:A:412:HOH:O	2.22	0.73
1:A:83:LYS:NZ	4:A:401:HOH:O	1.76	0.70
3:L:44:ASN:OD1	4:L:302:HOH:O	2.11	0.68
2:H:130:GLY:N	2:H:131:GLY:HA2	2.04	0.68
2:H:144:SER:HA	2:H:237:LYS:O	1.94	0.67
1:A:7:LEU:N	4:A:414:HOH:O	2.28	0.66
1:A:359:GLU:OE1	4:A:405:HOH:O	2.14	0.66
2:H:142:THR:HG21	2:H:145:LEU:HD22	1.79	0.63
1:A:310:GLU:OE1	4:A:407:HOH:O	2.16	0.63
1:A:366:THR:O	4:A:406:HOH:O	2.15	0.60
1:A:287:ASP:OD2	4:A:408:HOH:O	2.18	0.56
1:A:102:LYS:HD2	1:A:102:LYS:N	2.20	0.55
1:A:41:ASP:OD1	4:A:409:HOH:O	2.18	0.55
3:L:158:GLN:HB2	3:L:168:LEU:HD11	1.89	0.54
1:A:22:GLU:OE1	4:A:411:HOH:O	2.20	0.50
3:L:145:ARG:NE	3:L:191:ASP:OD1	2.41	0.47
2:H:65:LYS:HD3	4:H:334:HOH:O	2.16	0.46
3:L:157:TYR:HE1	3:L:210:GLN:HE21	1.64	0.46
2:H:73:ASP:CG	2:H:76:LYS:HG3	2.37	0.45
3:L:157:TYR:OH	3:L:210:GLN:NE2	2.50	0.45
2:H:130:GLY:N	2:H:131:GLY:CA	2.78	0.45
1:A:211:SER:HB3	2:H:100:TYR:CE1	2.51	0.45
1:A:356:THR:OG1	1:A:359:GLU:HG3	2.17	0.45
3:L:48:TRP:CZ2	3:L:50:GLY:HA2	2.51	0.44
1:A:96:ALA:HB2	1:A:329:ILE:HD12	2.00	0.44
3:L:148:GLN:HB3	3:L:148:GLN:HE21	1.64	0.43
2:H:83:MET:HB3	2:H:86:LEU:HD21	2.00	0.43
2:H:228:LEU:HD13	3:L:59:ASN:OD1	2.19	0.43
1:A:196:VAL:HG12	1:A:200:LYS:HD3	2.01	0.42
1:A:296:ASP:O	1:A:297:LYS:HD2	2.19	0.42

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:H:32:SER:OG	2:H:98:ARG:HD2	2.19	0.42
1:A:109:ALA:HA	1:A:302:VAL:HA	2.02	0.42
1:A:278:GLU:HG3	1:A:282:ASN:HD22	1.85	0.41
2:H:240:ILE:HD13	2:H:240:ILE:HA	1.75	0.41
3:L:5:LYS:HE3	3:L:108:GLN:NE2	2.36	0.41
2:H:129:SER:HA	2:H:130:GLY:HA3	1.75	0.41
2:H:155:ILE:HD13	2:H:236:SER:HB2	2.03	0.41
2:H:142:THR:HG21	2:H:145:LEU:HD13	2.01	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	A	358/383 (94%)	347 (97%)	11 (3%)	0	100	100
2	H	238/249 (96%)	225 (94%)	13 (6%)	0	100	100
3	L	226/228 (99%)	216 (96%)	10 (4%)	0	100	100
All	All	822/860 (96%)	788 (96%)	34 (4%)	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	285/306 (93%)	281 (99%)	4 (1%)	69	65
2	H	196/205 (96%)	191 (97%)	5 (3%)	49	39
3	L	195/195 (100%)	192 (98%)	3 (2%)	67	63
All	All	676/706 (96%)	664 (98%)	12 (2%)	62	56

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	34	LYS
1	A	102	LYS
1	A	142	LYS
1	A	295	LYS
2	H	32	SER
2	H	48	VAL
2	H	74	THR
2	H	240	ILE
2	H	241	LYS
3	L	1	GLN
3	L	210	GLN
3	L	216	LEU

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

Mol	Chain	Res	Type
1	A	173	ASN
1	A	282	ASN
3	L	108	GLN
3	L	124	GLN
3	L	148	GLN
3	L	210	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 [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](#)

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	360/383 (93%)	0.07	5 (1%) 75 82	14, 24, 43, 77	0
2	H	240/249 (96%)	0.20	17 (7%) 16 24	11, 23, 43, 90	0
3	L	228/228 (100%)	-0.20	3 (1%) 77 83	9, 20, 32, 54	0
All	All	828/860 (96%)	0.03	25 (3%) 50 59	9, 23, 39, 90	0

All (25) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	143	SER	13.9
2	H	144	SER	8.7
3	L	117	GLY	8.1
2	H	133	GLY	7.6
2	H	145	LEU	7.0
3	L	118	GLY	6.2
2	H	130	GLY	6.1
2	H	146	PRO	5.6
1	A	341	TYR	5.6
2	H	132	GLY	5.1
2	H	131	GLY	4.1
1	A	173	ASN	4.0
2	H	129	SER	3.2
2	H	240	ILE	3.0
2	H	142	THR	2.8
2	H	134	SER	2.6
1	A	53	THR	2.5
2	H	105	LEU	2.5
3	L	119	GLY	2.3
2	H	149	LEU	2.3
2	H	241	LYS	2.3
1	A	99	TYR	2.2
2	H	103	THR	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	H	102	TYR	2.1
1	A	354	ARG	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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.