



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 6, 2016 – 09:52 AM EDT

PDB ID : 5GS1
Title : Crystal structure of homo-specific diabody
Authors : Kim, J.H.; Song, D.H.; Youn, S.J.; Kim, J.W.; Cho, G.; Lee, H.; Lee, J.O.
Deposited on : 2016-08-13
Resolution : 2.00 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

<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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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-20027939

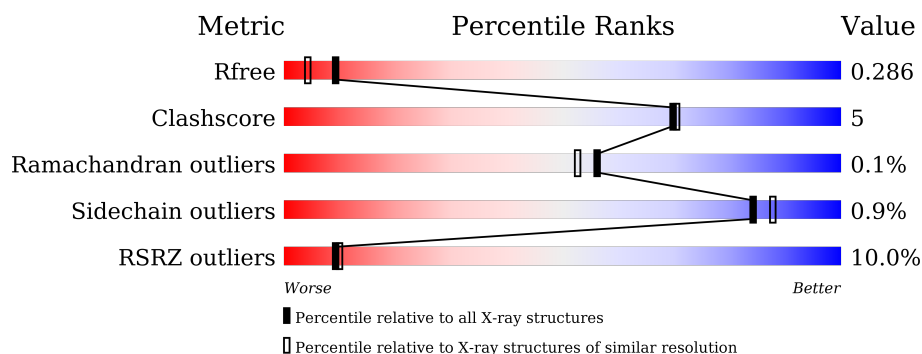
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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}	91344	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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. 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	108	<div> <div>90%</div> <div>8%</div> <div>.</div> </div>
1	M	108	<div> <div>6%</div> <div>88%</div> <div>10%</div> <div>.</div> </div>
1	O	108	<div> <div>9%</div> <div>86%</div> <div>12%</div> <div>.</div> </div>
1	Q	108	<div> <div>%</div> <div>91%</div> <div>7%</div> <div>.</div> </div>
2	B	124	<div> <div>4%</div> <div>90%</div> <div>5%</div> <div>5%</div> </div>
2	N	124	<div> <div>26%</div> <div>83%</div> <div>12%</div> <div>5%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	P	124	 % 90% 6% 5%
2	R	124	 87% 7% 5%
3	C	232	 2% 93% 6%
3	D	232	 5% 91% 8%
3	E	232	 7% 89% 9%
3	F	232	 7% 87% 12%
3	G	232	 41% 80% 16%
3	H	232	 37% 79% 16%
3	I	232	 3% 88% 10%
3	J	232	 5% 86% 13%
3	K	232	 2% 92% 6%
3	L	232	 5% 87% 11%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 25820 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 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	106	Total	C	N	O	S	0	0	0
			803	502	134	165	2			
1	M	106	Total	C	N	O	S	0	0	0
			803	502	134	165	2			
1	O	106	Total	C	N	O	S	0	0	0
			803	502	134	165	2			
1	Q	106	Total	C	N	O	S	0	0	0
			803	502	134	165	2			

- Molecule 2 is a protein called heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	118	Total	C	N	O	S	0	0	0
			901	569	156	172	4			
2	N	118	Total	C	N	O	S	0	0	0
			901	569	156	172	4			
2	P	118	Total	C	N	O	S	0	0	0
			901	569	156	172	4			
2	R	118	Total	C	N	O	S	0	0	0
			901	569	156	172	4			

- Molecule 3 is a protein called diabody.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			
3	D	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			
3	E	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			
3	F	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	223	Total	C	N	O	S	0	0	0
			1698	1068	289	335	6			
3	H	224	Total	C	N	O	S	0	0	0
			1704	1071	290	337	6			
3	I	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			
3	J	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			
3	K	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			
3	L	229	Total	C	N	O	S	0	0	0
			1726	1082	295	343	6			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	61	Total	O	0	0
			61	61		
4	B	84	Total	O	0	0
			84	84		
4	C	183	Total	O	0	0
			183	183		
4	D	145	Total	O	0	0
			145	145		
4	E	110	Total	O	0	0
			110	110		
4	F	101	Total	O	0	0
			101	101		
4	G	50	Total	O	0	0
			50	50		
4	H	58	Total	O	0	0
			58	58		
4	I	157	Total	O	0	0
			157	157		
4	J	168	Total	O	0	0
			168	168		
4	K	191	Total	O	0	0
			191	191		
4	L	142	Total	O	0	0
			142	142		
4	M	44	Total	O	0	0
			44	44		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	N	38	Total 38	O 38	0	0
4	O	37	Total 37	O 37	0	0
4	P	67	Total 67	O 67	0	0
4	Q	51	Total 51	O 51	0	0
4	R	107	Total 107	O 107	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 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.

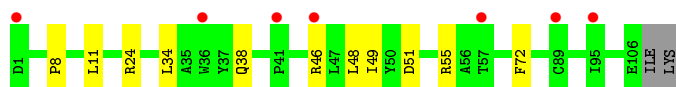
- Molecule 1: light chain

Chain A:  90% 8% .




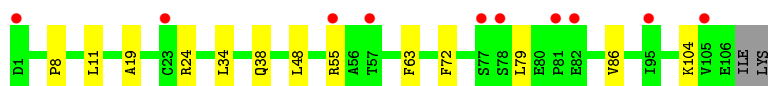
- Molecule 1: light chain

Chain M:  6% 88% 10% .

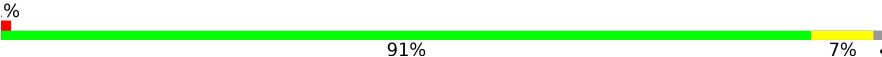


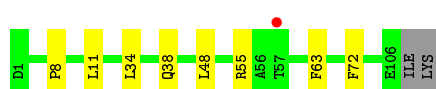
- Molecule 1: light chain

Chain O:  9% 86% 12% .



- Molecule 1: light chain

Chain Q:  0% 91% 7% .

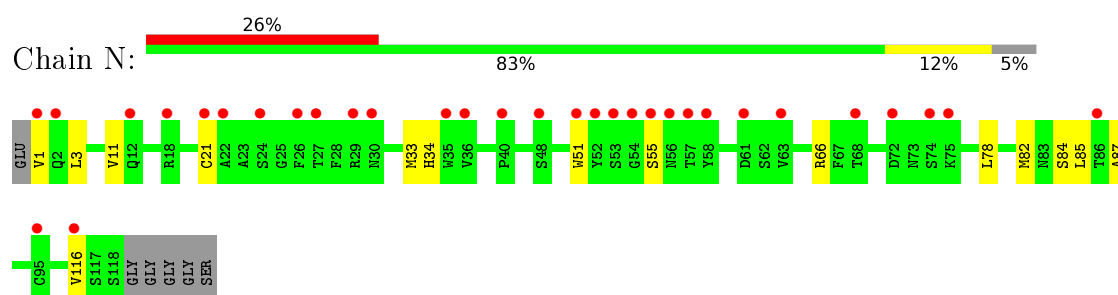


- Molecule 2: heavy chain

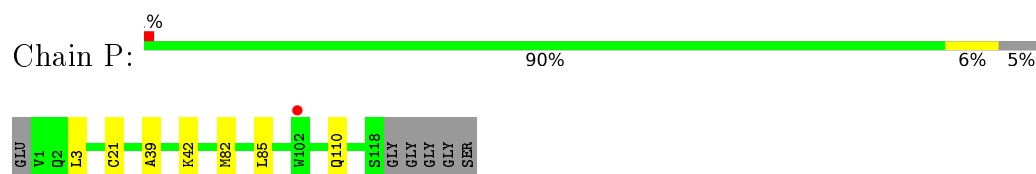
Chain B:  4% 90% 5% 5%



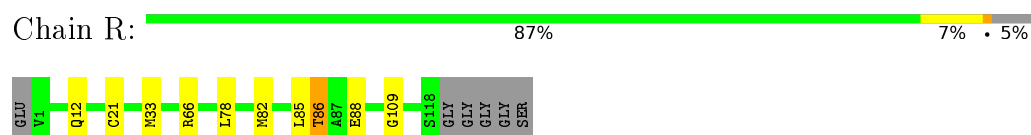
- Molecule 2: heavy chain



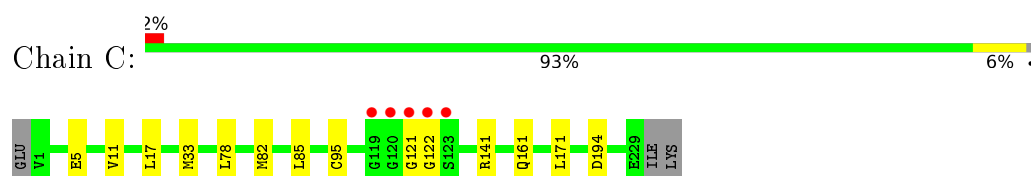
- Molecule 2: heavy chain



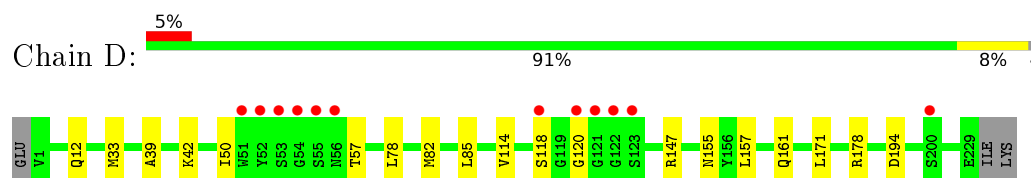
- Molecule 2: heavy chain



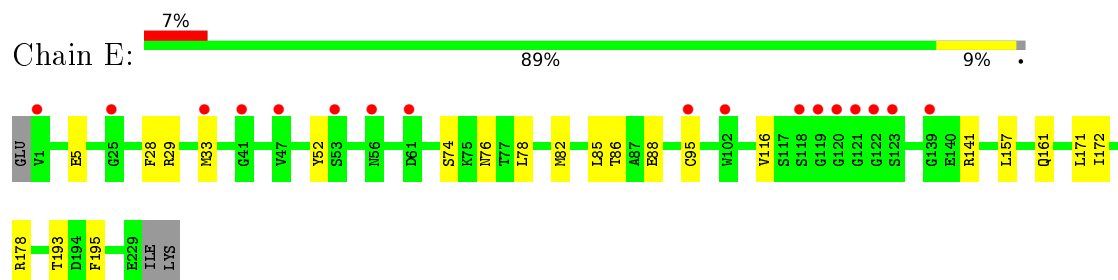
- Molecule 3: diabody



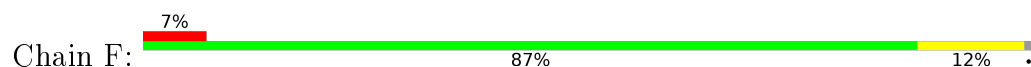
- Molecule 3: diabody

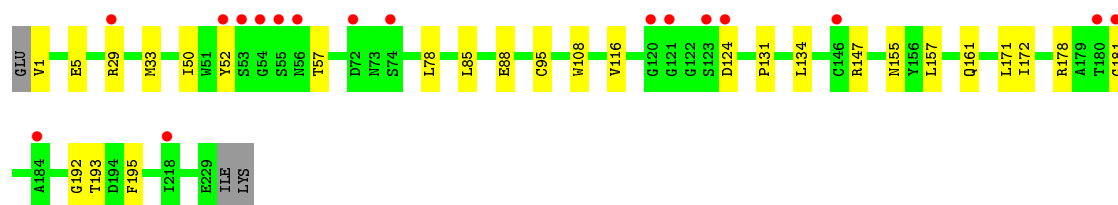


- Molecule 3: diabody

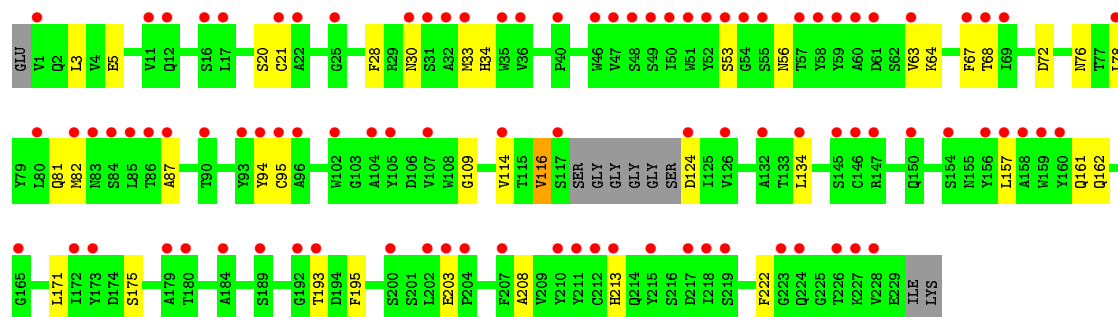
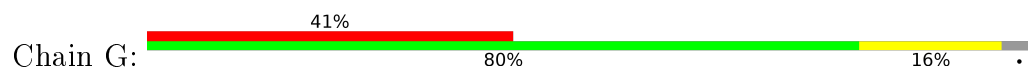


- Molecule 3: diabody

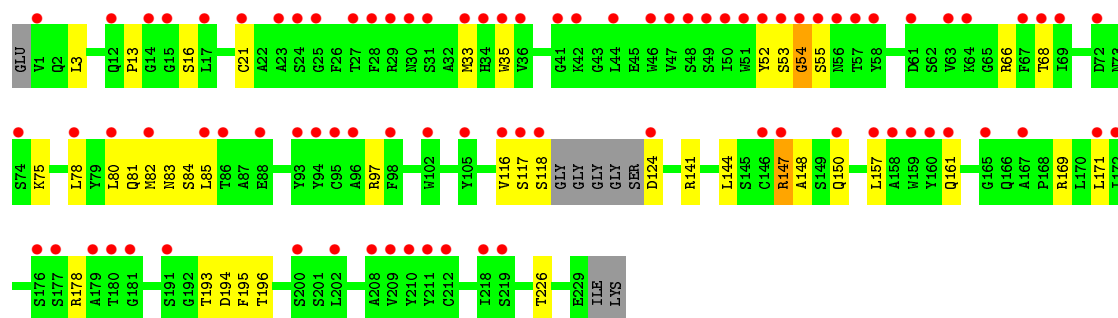
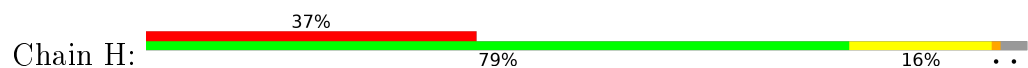




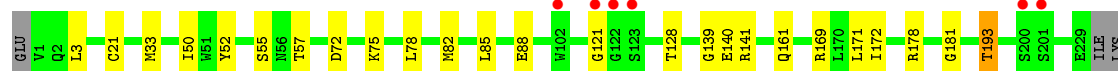
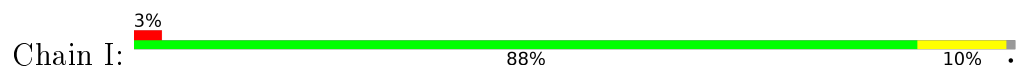
• Molecule 3: diabody



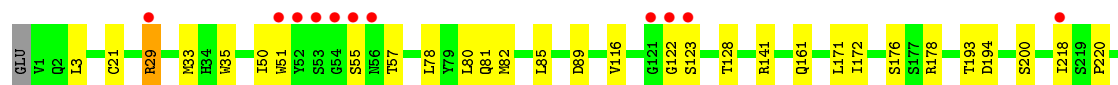
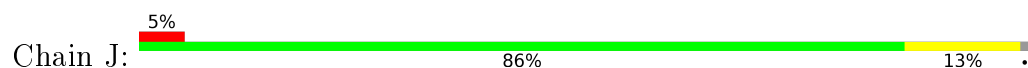
• Molecule 3: diabody



• Molecule 3: diabody

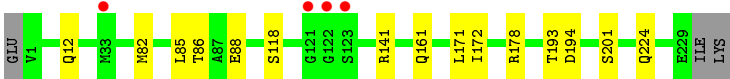
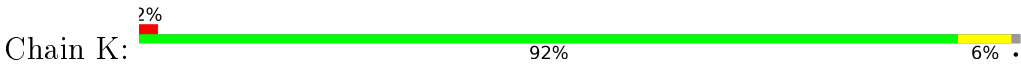


• Molecule 3: diabody

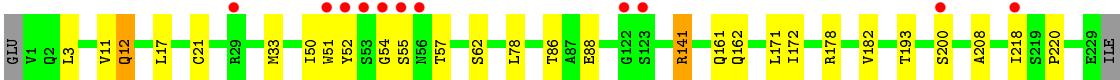
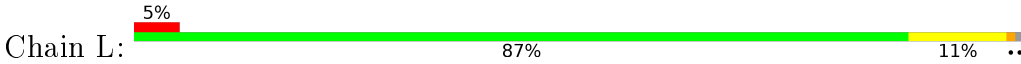


E229
ILE
LYS

● Molecule 3: diabody



● Molecule 3: diabody



LYS

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	114.12Å 128.81Å 259.83Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.05 – 2.00 20.05 – 1.99	Depositor EDS
% Data completeness (in resolution range)	96.0 (20.05-2.00) 95.6 (20.05-1.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	7.25 (at 1.99Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.267 , 0.286 0.266 , 0.286	Depositor DCC
R_{free} test set	1999 reflections (0.81%)	DCC
Wilson B-factor (Å ²)	19.1	Xtriage
Anisotropy	0.213	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 61.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	25820	wwPDB-VP
Average B, all atoms (Å ²)	30.0	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 75.78 % 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 1.1822e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² 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	0.26	0/822	0.48	0/1120
1	M	0.26	0/822	0.48	0/1120
1	O	0.26	0/822	0.49	0/1120
1	Q	0.25	0/822	0.48	0/1120
2	B	0.25	0/924	0.48	0/1255
2	N	0.27	0/924	0.50	0/1255
2	P	0.26	0/924	0.48	0/1255
2	R	0.25	0/924	0.48	0/1255
3	C	0.25	0/1769	0.48	0/2406
3	D	0.25	0/1769	0.48	0/2406
3	E	0.26	0/1769	0.48	0/2406
3	F	0.26	0/1769	0.50	0/2406
3	G	0.27	0/1740	0.52	0/2367
3	H	0.26	0/1746	0.54	1/2375 (0.0%)
3	I	0.26	0/1769	0.50	0/2406
3	J	0.26	0/1769	0.50	0/2406
3	K	0.26	0/1769	0.49	0/2406
3	L	0.26	0/1769	0.49	0/2406
All	All	0.26	0/24622	0.49	1/33490 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	H	54	GLY	N-CA-C	5.32	126.41	113.10

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	803	0	768	10	0
1	M	803	0	768	9	0
1	O	803	0	768	8	0
1	Q	803	0	768	4	0
2	B	901	0	856	4	0
2	N	901	0	856	11	0
2	P	901	0	856	5	0
2	R	901	0	856	9	0
3	C	1726	0	1639	8	0
3	D	1726	0	1639	15	0
3	E	1726	0	1639	13	0
3	F	1726	0	1639	15	0
3	G	1698	0	1616	24	1
3	H	1704	0	1621	36	0
3	I	1726	0	1639	16	0
3	J	1726	0	1639	18	0
3	K	1726	0	1639	11	0
3	L	1726	0	1639	18	0
4	A	61	0	0	2	0
4	B	84	0	0	1	0
4	C	183	0	0	0	0
4	D	145	0	0	3	0
4	E	110	0	0	2	0
4	F	101	0	0	3	0
4	G	50	0	0	8	1
4	H	58	0	0	10	0
4	I	157	0	0	5	0
4	J	168	0	0	4	0
4	K	191	0	0	4	0
4	L	142	0	0	2	0
4	M	44	0	0	2	0
4	N	38	0	0	3	0
4	O	37	0	0	1	0
4	P	67	0	0	1	0
4	Q	51	0	0	0	0
4	R	107	0	0	4	0
All	All	25820	0	22845	213	1

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 5.

The worst 5 of 213 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:I:139:GLY:O	1:O:24:ARG:NH1	2.04	0.89
1:A:24:ARG:NH1	4:A:201:HOH:O	2.12	0.81
3:G:72:ASP:O	4:G:301:HOH:O	2.04	0.75
3:J:81:GLN:NE2	4:J:302:HOH:O	2.18	0.74
3:D:120:GLY:HA2	2:P:110:GLN:HB3	1.70	0.73

All (1) 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)
3:G:53:SER:O	4:G:304:HOH:O[3_454]	2.19	0.01

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	104/108 (96%)	101 (97%)	3 (3%)	0	100	100
1	M	104/108 (96%)	101 (97%)	3 (3%)	0	100	100
1	O	104/108 (96%)	102 (98%)	2 (2%)	0	100	100
1	Q	104/108 (96%)	101 (97%)	3 (3%)	0	100	100
2	B	116/124 (94%)	113 (97%)	3 (3%)	0	100	100
2	N	116/124 (94%)	112 (97%)	4 (3%)	0	100	100
2	P	116/124 (94%)	113 (97%)	3 (3%)	0	100	100
2	R	116/124 (94%)	113 (97%)	3 (3%)	0	100	100
3	C	227/232 (98%)	221 (97%)	6 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	D	227/232 (98%)	217 (96%)	10 (4%)	0	100	100
3	E	227/232 (98%)	223 (98%)	4 (2%)	0	100	100
3	F	227/232 (98%)	217 (96%)	9 (4%)	1 (0%)	39	33
3	G	219/232 (94%)	213 (97%)	6 (3%)	0	100	100
3	H	220/232 (95%)	213 (97%)	7 (3%)	0	100	100
3	I	227/232 (98%)	221 (97%)	6 (3%)	0	100	100
3	J	227/232 (98%)	221 (97%)	5 (2%)	1 (0%)	39	33
3	K	227/232 (98%)	223 (98%)	4 (2%)	0	100	100
3	L	227/232 (98%)	223 (98%)	4 (2%)	0	100	100
All	All	3135/3248 (96%)	3048 (97%)	85 (3%)	2 (0%)	56	53

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	J	122	GLY
3	F	52	TYR

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	90/92 (98%)	90 (100%)	0	100	100
1	M	90/92 (98%)	90 (100%)	0	100	100
1	O	90/92 (98%)	90 (100%)	0	100	100
1	Q	90/92 (98%)	90 (100%)	0	100	100
2	B	93/95 (98%)	93 (100%)	0	100	100
2	N	93/95 (98%)	93 (100%)	0	100	100
2	P	93/95 (98%)	93 (100%)	0	100	100
2	R	93/95 (98%)	92 (99%)	1 (1%)	80	83
3	C	184/187 (98%)	184 (100%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	D	184/187 (98%)	183 (100%)	1 (0%)	92	94
3	E	184/187 (98%)	181 (98%)	3 (2%)	70	73
3	F	184/187 (98%)	182 (99%)	2 (1%)	80	83
3	G	182/187 (97%)	178 (98%)	4 (2%)	60	62
3	H	183/187 (98%)	181 (99%)	2 (1%)	80	83
3	I	184/187 (98%)	182 (99%)	2 (1%)	80	83
3	J	184/187 (98%)	181 (98%)	3 (2%)	70	73
3	K	184/187 (98%)	183 (100%)	1 (0%)	92	94
3	L	184/187 (98%)	181 (98%)	3 (2%)	70	73
All	All	2569/2618 (98%)	2547 (99%)	22 (1%)	84	88

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	193	THR
3	I	128	THR
3	L	193	THR
3	H	75	LYS
3	H	147	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	12	GLN
3	F	155	ASN
3	F	224	GLN
3	H	161	GLN
3	K	2	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, 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	106/108 (98%)	-0.04	0 100 100	8, 20, 36, 58	0
1	M	106/108 (98%)	0.99	7 (6%) 22 22	30, 37, 51, 70	0
1	O	106/108 (98%)	0.98	10 (9%) 11 11	21, 37, 54, 68	0
1	Q	106/108 (98%)	0.24	1 (0%) 85 86	14, 30, 41, 64	0
2	B	118/124 (95%)	0.21	5 (4%) 40 41	8, 16, 45, 68	0
2	N	118/124 (95%)	1.58	32 (27%) 1 1	29, 40, 58, 73	0
2	P	118/124 (95%)	0.17	1 (0%) 87 88	12, 22, 37, 54	0
2	R	118/124 (95%)	-0.17	0 100 100	7, 15, 34, 48	0
3	C	229/232 (98%)	0.10	5 (2%) 65 66	8, 19, 39, 113	0
3	D	229/232 (98%)	0.28	12 (5%) 31 33	10, 23, 44, 112	0
3	E	229/232 (98%)	0.76	17 (7%) 17 18	13, 32, 52, 110	0
3	F	229/232 (98%)	0.72	17 (7%) 17 18	12, 33, 58, 118	0
3	G	223/232 (96%)	1.94	94 (42%) 0 1	37, 51, 66, 93	0
3	H	224/232 (96%)	2.01	86 (38%) 0 1	37, 51, 71, 100	0
3	I	229/232 (98%)	0.17	6 (2%) 59 60	8, 19, 46, 70	0
3	J	229/232 (98%)	0.29	11 (4%) 34 36	7, 19, 46, 69	0
3	K	229/232 (98%)	0.07	4 (1%) 73 73	7, 16, 39, 66	0
3	L	229/232 (98%)	0.34	11 (4%) 34 36	7, 21, 51, 69	0
All	All	3175/3248 (97%)	0.61	319 (10%) 9 10	7, 28, 59, 118	0

The worst 5 of 319 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	121	GLY	14.2
3	F	53	SER	13.1
3	E	121	GLY	11.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	E	120	GLY	11.5
3	D	53	SER	9.9

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.