wwPDB X-ray Structure Validation Summary Report (i)

Nov 23, 2023 – 12:49 AM JST

PDB ID	:	8GTJ
Title	:	Crystal structure of IpaH7.8-LRR and GSDMB isoform-4 complex
Authors	:	Zhong, X.; Hou, Y.J.; Ding, J.
Deposited on	:	2022-09-08
Resolution	:	2.70 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

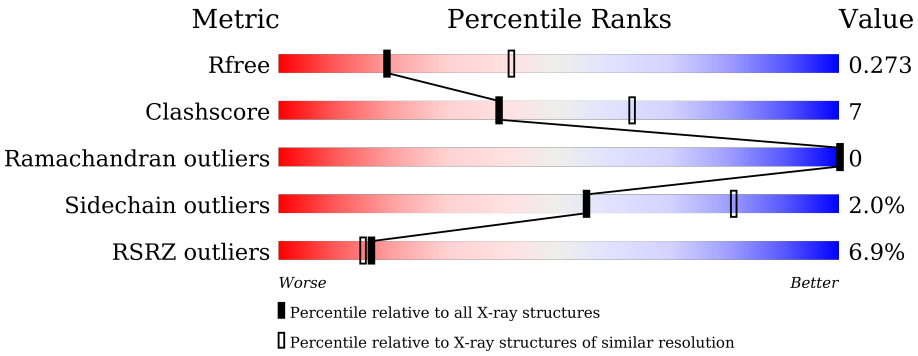
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.70 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (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 of 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	A	414	2% 77% 14% 8%
1	C	414	9% 60% 15% 23%
2	B	241	8% 85% 12%
2	D	241	7% 73% 24%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 9607 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 Isoform 4 of Gasdermin-B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	379	Total	C	N	O	S	0	0	0
			3038	1915	520	589	14			
1	C	318	Total	C	N	O	S	0	0	0
			2573	1629	439	493	12			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	GLY	-	expression tag	UNP Q8TAX9
A	-5	PRO	-	expression tag	UNP Q8TAX9
A	-4	GLY	-	expression tag	UNP Q8TAX9
A	-3	SER	-	expression tag	UNP Q8TAX9
A	-2	GLY	-	expression tag	UNP Q8TAX9
A	-1	ARG	-	expression tag	UNP Q8TAX9
A	0	PRO	-	expression tag	UNP Q8TAX9
A	?	-	GLU	deletion	UNP Q8TAX9
A	?	-	LYS	deletion	UNP Q8TAX9
A	?	-	ASP	deletion	UNP Q8TAX9
A	?	-	GLY	deletion	UNP Q8TAX9
A	?	-	ALA	deletion	UNP Q8TAX9
A	?	-	SER	deletion	UNP Q8TAX9
A	?	-	SER	deletion	UNP Q8TAX9
A	?	-	CYS	deletion	UNP Q8TAX9
A	?	-	LEU	deletion	UNP Q8TAX9
C	-6	GLY	-	expression tag	UNP Q8TAX9
C	-5	PRO	-	expression tag	UNP Q8TAX9
C	-4	GLY	-	expression tag	UNP Q8TAX9
C	-3	SER	-	expression tag	UNP Q8TAX9
C	-2	GLY	-	expression tag	UNP Q8TAX9
C	-1	ARG	-	expression tag	UNP Q8TAX9
C	0	PRO	-	expression tag	UNP Q8TAX9
C	?	-	GLU	deletion	UNP Q8TAX9
C	?	-	LYS	deletion	UNP Q8TAX9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	?	-	ASP	deletion	UNP Q8TAX9
C	?	-	GLY	deletion	UNP Q8TAX9
C	?	-	ALA	deletion	UNP Q8TAX9
C	?	-	SER	deletion	UNP Q8TAX9
C	?	-	SER	deletion	UNP Q8TAX9
C	?	-	CYS	deletion	UNP Q8TAX9
C	?	-	LEU	deletion	UNP Q8TAX9

- Molecule 2 is a protein called Probable E3 ubiquitin-protein ligase ipaH7.8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	236	Total	C	N	O	S	0	0	0
			1922	1225	331	363	3			
2	D	236	Total	C	N	O	S	0	0	0
			1922	1225	331	363	3			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	22	MET	-	initiating methionine	UNP P18014
D	22	MET	-	initiating methionine	UNP P18014

- Molecule 3 is water.

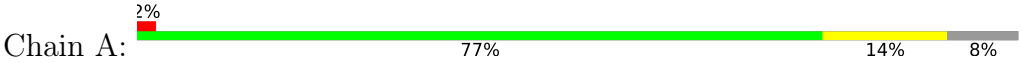
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	58	Total	O	0	0
			58	58		
3	B	35	Total	O	0	0
			35	35		
3	C	24	Total	O	0	0
			24	24		
3	D	35	Total	O	0	0
			35	35		

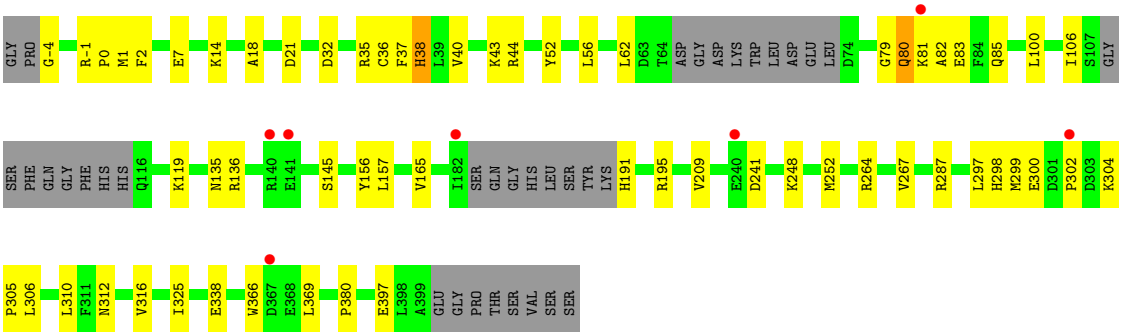
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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: Isoform 4 of Gasdermin-B

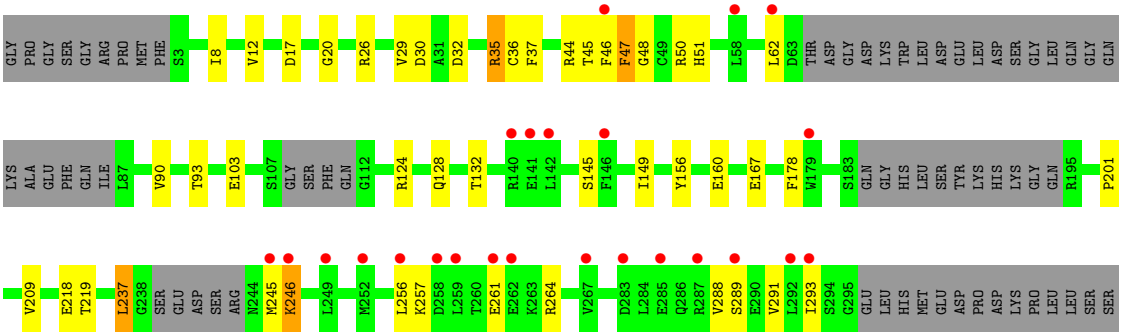
Chain A:





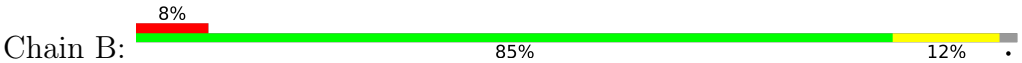
• Molecule 1: Isoform 4 of Gasdermin-B

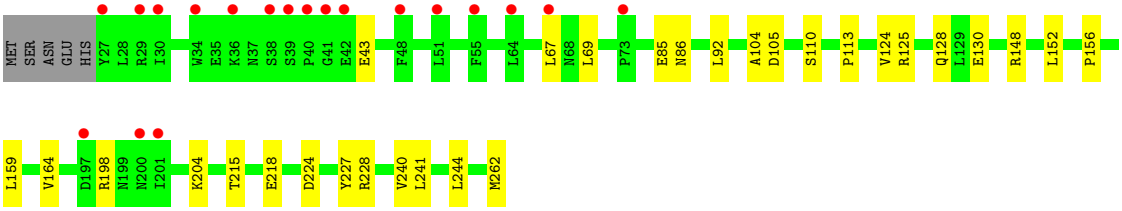
Chain C:



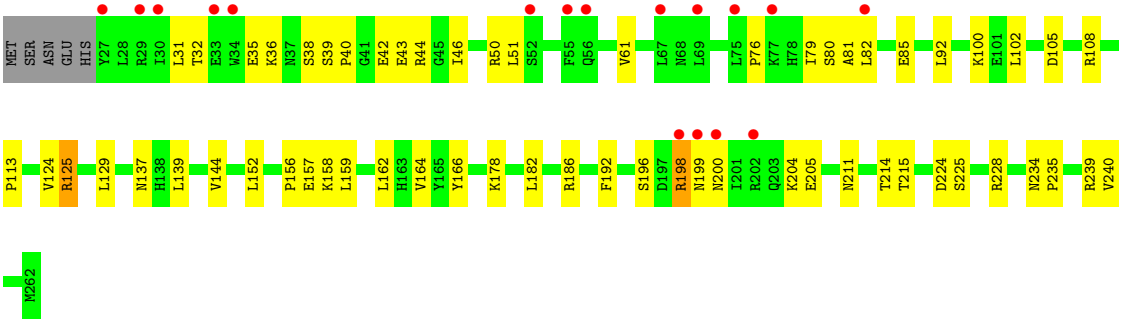
• Molecule 2: Probable E3 ubiquitin-protein ligase ipaH7.8

Chain B:





• Molecule 2: Probable E3 ubiquitin-protein ligase ipaH7.8



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	88.34Å 88.34Å 335.62Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	45.15 – 2.70 45.15 – 2.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (45.15-2.70) 97.6 (45.15-2.70)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.58 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, R_{free}	0.241 , 0.276 0.237 , 0.273	Depositor DCC
R_{free} test set	2005 reflections (4.67%)	wwPDB-VP
Wilson B-factor (Å^2)	58.3	Xtriage
Anisotropy	0.217	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 57.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.037 for -h,-k,l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	9607	wwPDB-VP
Average B, all atoms (Å^2)	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.09% 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

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	0.32	0/3079	0.52	0/4139
1	C	0.30	0/2602	0.51	0/3489
2	B	0.30	0/1966	0.50	0/2678
2	D	0.32	0/1966	0.52	0/2678
All	All	0.31	0/9613	0.51	0/12984

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1
2	D	0	2
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	35	ARG	Sidechain
2	D	125	ARG	Sidechain
2	D	198	ARG	Sidechain

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	3038	0	3084	45	0
1	C	2573	0	2629	47	0
2	B	1922	0	1925	21	0
2	D	1922	0	1925	33	0
3	A	58	0	0	0	0
3	B	35	0	0	1	0
3	C	24	0	0	0	0
3	D	35	0	0	0	0
All	All	9607	0	9563	141	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 7.

The worst 5 of 141 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:A:100:LEU:HD11	1:A:106:ILE:HD11	1.50	0.94
1:C:264:ARG:HD3	1:C:398:LEU:O	1.74	0.88
1:A:-4:GLY:HA2	1:A:52:TYR:CE1	2.20	0.77
1:A:100:LEU:CD1	1:A:106:ILE:HD11	2.16	0.76
1:A:1:MET:CE	1:A:380:PRO:HB3	2.16	0.76

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	371/414 (90%)	360 (97%)	11 (3%)	0	100	100
1	C	304/414 (73%)	296 (97%)	8 (3%)	0	100	100
2	B	234/241 (97%)	217 (93%)	17 (7%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	234/241 (97%)	221 (94%)	13 (6%)	0	100	100
All	All	1143/1310 (87%)	1094 (96%)	49 (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	343/372 (92%)	337 (98%)	6 (2%)	60	84
1	C	292/372 (78%)	282 (97%)	10 (3%)	37	66
2	B	225/230 (98%)	225 (100%)	0	100	100
2	D	225/230 (98%)	219 (97%)	6 (3%)	44	74
All	All	1085/1204 (90%)	1063 (98%)	22 (2%)	55	81

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

Mol	Chain	Res	Type
1	C	363	GLU
2	D	108	ARG
2	D	43	GLU
2	D	192	PHE
1	C	47	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 monosaccharides 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^{th} 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	379/414 (91%)	0.02	7 (1%) 68 70	26, 50, 98, 131	0
1	C	318/414 (76%)	0.60	38 (11%) 4 3	37, 77, 117, 138	0
2	B	236/241 (97%)	0.25	19 (8%) 12 10	28, 50, 104, 129	0
2	D	236/241 (97%)	0.34	17 (7%) 15 13	32, 60, 96, 123	0
All	All	1169/1310 (89%)	0.29	81 (6%) 16 15	26, 59, 109, 138	0

The worst 5 of 81 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	256	LEU	6.4
2	B	40	PRO	5.7
1	C	46	PHE	5.5
2	B	27	TYR	5.4
1	C	346	LEU	5.2

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.