



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 10, 2023 – 03:56 PM EDT

PDB ID : 4JPO  
Title : 5A resolution structure of Proteasome Assembly Chaperone Hsm3 in complex with a C-terminal fragment of Rpt1  
Authors : Lovell, S.; Battaile, K.P.; Singh, R.; Roelofs, J.  
Deposited on : 2013-03-19  
Resolution : 5.00 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](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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35.1  
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) : 2.35.1

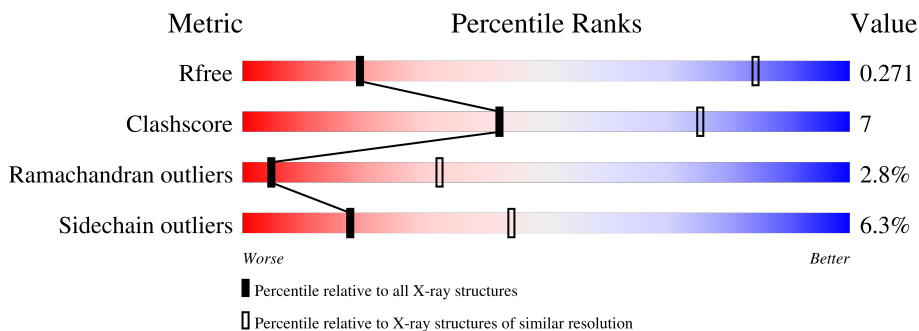
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 5.00 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))
$R_{free}$	130704	1140 (6.20-3.80)
Clashscore	141614	1000 (6.16-3.82)
Ramachandran outliers	138981	1146 (6.20-3.80)
Sidechain outliers	138945	1122 (6.20-3.80)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	491	71% (green), 20% (yellow), 8% (orange), 1% (red), 0% (grey)
1	B	491	68% (green), 21% (yellow), 9% (orange), 2% (red), 0% (grey)
2	C	100	54% (green), 17% (yellow), 26% (orange), 3% (red), 0% (grey)
2	D	100	60% (green), 11% (yellow), 26% (orange), 3% (red), 0% (grey)

## 2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 8391 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 DNA mismatch repair protein HSM3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	452	3629	2346	578	693	12	0	0	0
1	B	447	3600	2327	573	688	12	0	0	0

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P38348
A	2	PRO	-	expression tag	UNP P38348
A	3	LEU	-	expression tag	UNP P38348
A	4	THR	-	expression tag	UNP P38348
A	5	ARG	-	expression tag	UNP P38348
A	6	ARG	-	expression tag	UNP P38348
A	7	ALA	-	expression tag	UNP P38348
A	8	SER	-	expression tag	UNP P38348
A	9	VAL	-	expression tag	UNP P38348
A	10	GLY	-	expression tag	UNP P38348
A	11	SER	-	expression tag	UNP P38348
B	1	GLY	-	expression tag	UNP P38348
B	2	PRO	-	expression tag	UNP P38348
B	3	LEU	-	expression tag	UNP P38348
B	4	THR	-	expression tag	UNP P38348
B	5	ARG	-	expression tag	UNP P38348
B	6	ARG	-	expression tag	UNP P38348
B	7	ALA	-	expression tag	UNP P38348
B	8	SER	-	expression tag	UNP P38348
B	9	VAL	-	expression tag	UNP P38348
B	10	GLY	-	expression tag	UNP P38348
B	11	SER	-	expression tag	UNP P38348

- Molecule 2 is a protein called 26S protease regulatory subunit 7 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	74	Total	C	N	O	S	0	0	0
			584	365	112	103	4			
2	D	74	Total	C	N	O	S	0	0	0
			578	359	111	104	4			

There are 22 discrepancies between the modelled and reference sequences:

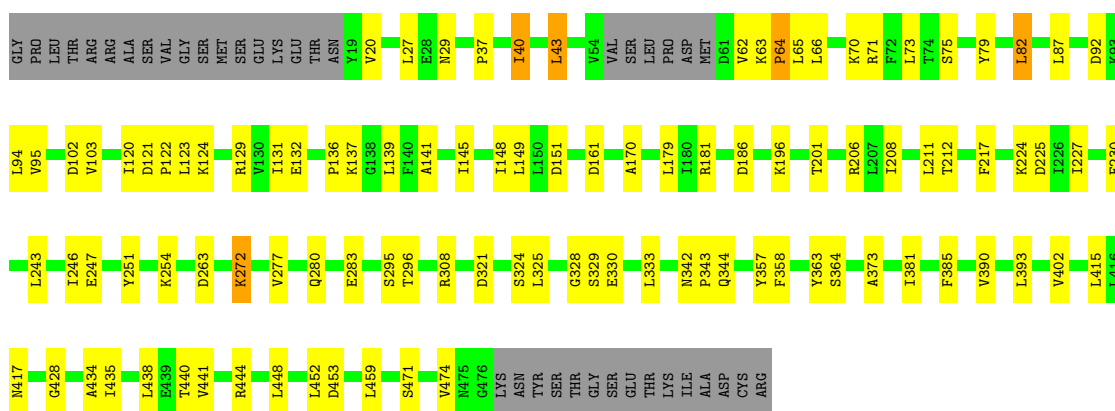
Chain	Residue	Modelled	Actual	Comment	Reference
C	368	MET	-	expression tag	UNP P33299
C	369	HIS	-	expression tag	UNP P33299
C	370	HIS	-	expression tag	UNP P33299
C	371	HIS	-	expression tag	UNP P33299
C	372	HIS	-	expression tag	UNP P33299
C	373	HIS	-	expression tag	UNP P33299
C	374	HIS	-	expression tag	UNP P33299
C	375	SER	-	expression tag	UNP P33299
C	376	GLN	-	expression tag	UNP P33299
C	377	HIS	-	expression tag	UNP P33299
C	378	MET	-	expression tag	UNP P33299
D	368	MET	-	expression tag	UNP P33299
D	369	HIS	-	expression tag	UNP P33299
D	370	HIS	-	expression tag	UNP P33299
D	371	HIS	-	expression tag	UNP P33299
D	372	HIS	-	expression tag	UNP P33299
D	373	HIS	-	expression tag	UNP P33299
D	374	HIS	-	expression tag	UNP P33299
D	375	SER	-	expression tag	UNP P33299
D	376	GLN	-	expression tag	UNP P33299
D	377	HIS	-	expression tag	UNP P33299
D	378	MET	-	expression tag	UNP P33299

### 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. 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. 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.

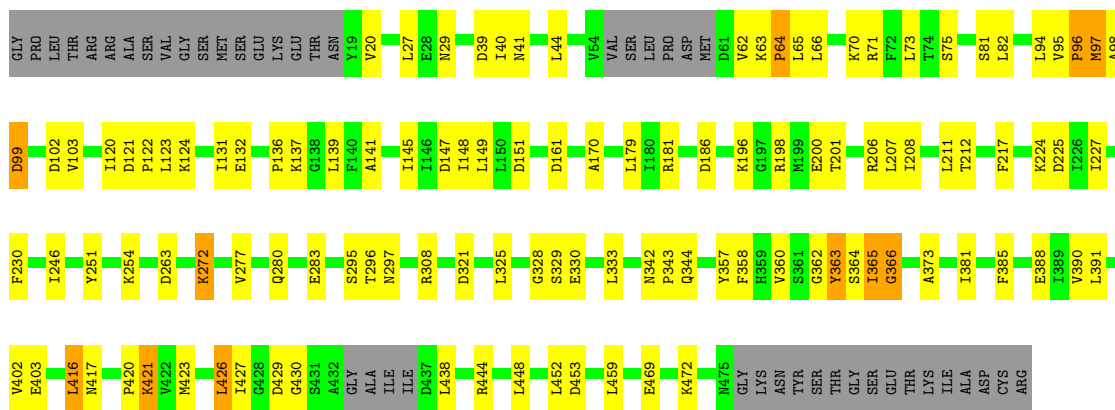
- Molecule 1: DNA mismatch repair protein HSM3

Chain A:  71% 20% 8%



- Molecule 1: DNA mismatch repair protein HSM3

Chain B:  68% 21% 9%



- Molecule 2: 26S protease regulatory subunit 7 homolog

Chain C:  54% 17% 26%



- Molecule 2: 26S protease regulatory subunit 7 homolog

Chain D:  60% 11% 26%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	185.30Å 185.30Å 357.38Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.87 – 5.00 46.32 – 5.00	Depositor EDS
% Data completeness (in resolution range)	99.9 (46.87-5.00) 99.9 (46.32-5.00)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.05 (at 5.10Å)	Xtriage
Refinement program	BUSTER-TNT, BUSTER 2.10.0	Depositor
R, $R_{free}$	0.253 , 0.273 0.262 , 0.271	Depositor DCC
$R_{free}$ test set	829 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	300.2	Xtriage
Anisotropy	0.036	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 262.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	8391	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	273.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.80% 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	0.43	0/3693	0.62	0/5012
1	B	0.44	0/3663	0.63	0/4970
2	C	0.38	0/591	0.60	0/788
2	D	0.43	0/585	0.63	0/781
All	All	0.43	0/8532	0.63	0/11551

There are no bond length outliers.

There are no bond angle outliers.

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	3629	0	3610	49	0
1	B	3600	0	3576	55	0
2	C	584	0	610	10	0
2	D	578	0	600	7	0
All	All	8391	0	8396	115	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 115 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:B:40:ILE:O	1:B:41:ASN:HB2	1.62	0.97
1:B:363:TYR:N	1:B:364:SER:HA	1.93	0.81
1:A:63:LYS:HD3	1:A:102:ASP:O	1.79	0.80
1:B:63:LYS:HD3	1:B:102:ASP:O	1.83	0.78
2:C:382:LEU:HA	2:C:385:ARG:HD3	1.67	0.74

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	448/491 (91%)	395 (88%)	41 (9%)	12 (3%)	5	33
1	B	441/491 (90%)	395 (90%)	32 (7%)	14 (3%)	4	30
2	C	72/100 (72%)	64 (89%)	6 (8%)	2 (3%)	5	32
2	D	72/100 (72%)	63 (88%)	8 (11%)	1 (1%)	11	46
All	All	1033/1182 (87%)	917 (89%)	87 (8%)	29 (3%)	5	32

5 of 29 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	272	LYS
1	A	329	SER
1	B	97	MET
1	B	272	LYS
1	B	329	SER

### 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	402/453 (89%)	382 (95%)	20 (5%)	24	50
1	B	400/453 (88%)	374 (94%)	26 (6%)	17	43
2	C	62/88 (70%)	56 (90%)	6 (10%)	8	29
2	D	62/88 (70%)	56 (90%)	6 (10%)	8	29
All	All	926/1082 (86%)	868 (94%)	58 (6%)	18	44

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	207	LEU
2	D	404	TRP
1	B	363	TYR
2	D	399	GLU
2	C	424	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	450	ASN
1	B	280	GLN
1	B	344	GLN
1	B	342	ASN
1	A	344	GLN

### 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

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.