



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 7, 2024 – 07:22 PM EDT

PDB ID : 4J80
Title : Thermus thermophilus DnaJ
Authors : Barends, T.R.M.; Brosi, R.W.; Steinmetz, A.; Scherer, A.; Hartmann, E.; Eschenbach, J.; Lorenz, T.; Seidel, R.; Shoeman, R.; Zimmermann, S.; Bittl, R.; Schlichting, I.; Reinstein, J.
Deposited on : 2013-02-14
Resolution : 2.90 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 ⓘ 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
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 1.20.1
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.003 (Gargrove)
Density-Fitness : 1.0.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.39

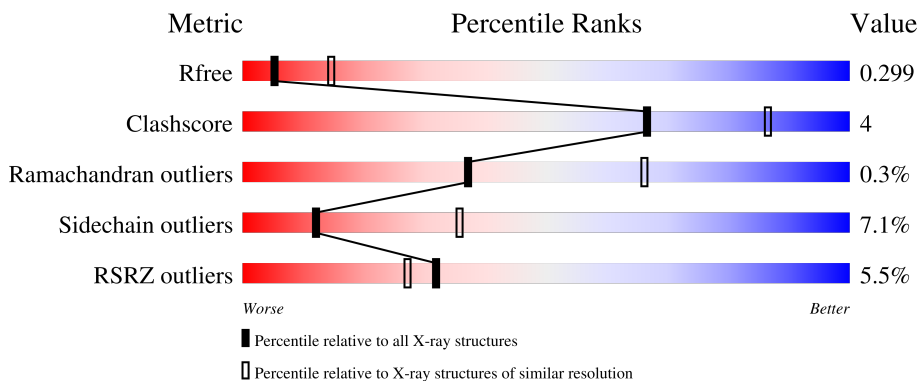
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION, EPR

The reported resolution of this entry is 2.90 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}	164625	2335 (2.90-2.90)
Clashscore	180529	2564 (2.90-2.90)
Ramachandran outliers	177936	2514 (2.90-2.90)
Sidechain outliers	177891	2516 (2.90-2.90)
RSRZ outliers	164620	2337 (2.90-2.90)

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 $\leq 5\%$. 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	284	
1	B	284	
1	C	284	
1	D	284	

2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 8460 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 Chaperone protein DnaJ 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	271	Total	C	N	O	Se	0	0	0
			2115	1348	374	387	6			
1	B	271	Total	C	N	O	Se	0	0	0
			2115	1348	374	387	6			
1	C	271	Total	C	N	O	Se	0	0	0
			2115	1348	374	387	6			
1	D	271	Total	C	N	O	Se	0	0	0
			2115	1348	374	387	6			

There are 92 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	57	MSE	LEU	engineered mutation	UNP Q56237
A	?	-	GLY	deletion	UNP Q56237
A	?	-	PHE	deletion	UNP Q56237
A	?	-	GLY	deletion	UNP Q56237
A	?	-	ARG	deletion	UNP Q56237
A	?	-	ARG	deletion	UNP Q56237
A	?	-	SER	deletion	UNP Q56237
A	?	-	ARG	deletion	UNP Q56237
A	142	MSE	ILE	engineered mutation	UNP Q56237
A	173	MSE	LEU	engineered mutation	UNP Q56237
A	226	MSE	LEU	engineered mutation	UNP Q56237
A	231	GLY	LYS	engineered mutation	UNP Q56237
A	274	ALA	-	expression tag	UNP Q56237
A	275	ALA	-	expression tag	UNP Q56237
A	276	ALA	-	expression tag	UNP Q56237
A	277	LEU	-	expression tag	UNP Q56237
A	278	GLU	-	expression tag	UNP Q56237
A	279	HIS	-	expression tag	UNP Q56237
A	280	HIS	-	expression tag	UNP Q56237
A	281	HIS	-	expression tag	UNP Q56237
A	282	HIS	-	expression tag	UNP Q56237

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	283	HIS	-	expression tag	UNP Q56237
A	284	HIS	-	expression tag	UNP Q56237
B	57	MSE	LEU	engineered mutation	UNP Q56237
B	?	-	GLY	deletion	UNP Q56237
B	?	-	PHE	deletion	UNP Q56237
B	?	-	GLY	deletion	UNP Q56237
B	?	-	ARG	deletion	UNP Q56237
B	?	-	ARG	deletion	UNP Q56237
B	?	-	SER	deletion	UNP Q56237
B	?	-	ARG	deletion	UNP Q56237
B	142	MSE	ILE	engineered mutation	UNP Q56237
B	173	MSE	LEU	engineered mutation	UNP Q56237
B	226	MSE	LEU	engineered mutation	UNP Q56237
B	231	GLY	LYS	engineered mutation	UNP Q56237
B	274	ALA	-	expression tag	UNP Q56237
B	275	ALA	-	expression tag	UNP Q56237
B	276	ALA	-	expression tag	UNP Q56237
B	277	LEU	-	expression tag	UNP Q56237
B	278	GLU	-	expression tag	UNP Q56237
B	279	HIS	-	expression tag	UNP Q56237
B	280	HIS	-	expression tag	UNP Q56237
B	281	HIS	-	expression tag	UNP Q56237
B	282	HIS	-	expression tag	UNP Q56237
B	283	HIS	-	expression tag	UNP Q56237
B	284	HIS	-	expression tag	UNP Q56237
C	57	MSE	LEU	engineered mutation	UNP Q56237
C	?	-	GLY	deletion	UNP Q56237
C	?	-	PHE	deletion	UNP Q56237
C	?	-	GLY	deletion	UNP Q56237
C	?	-	ARG	deletion	UNP Q56237
C	?	-	ARG	deletion	UNP Q56237
C	?	-	SER	deletion	UNP Q56237
C	?	-	ARG	deletion	UNP Q56237
C	142	MSE	ILE	engineered mutation	UNP Q56237
C	173	MSE	LEU	engineered mutation	UNP Q56237
C	226	MSE	LEU	engineered mutation	UNP Q56237
C	231	GLY	LYS	engineered mutation	UNP Q56237
C	274	ALA	-	expression tag	UNP Q56237
C	275	ALA	-	expression tag	UNP Q56237
C	276	ALA	-	expression tag	UNP Q56237
C	277	LEU	-	expression tag	UNP Q56237
C	278	GLU	-	expression tag	UNP Q56237

Continued on next page...

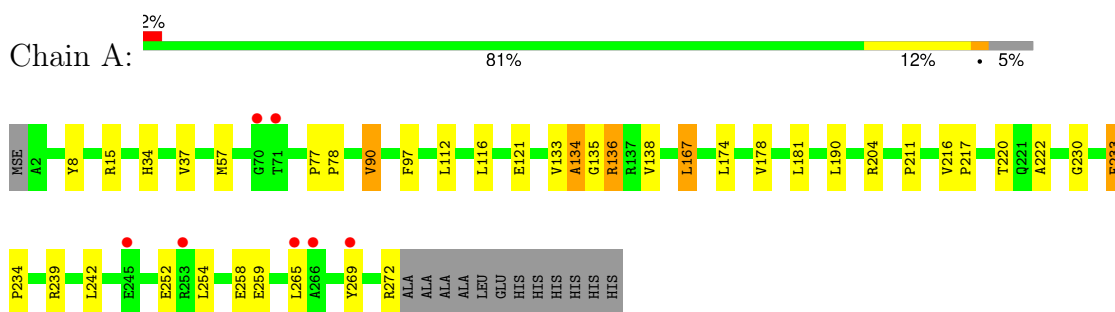
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	279	HIS	-	expression tag	UNP Q56237
C	280	HIS	-	expression tag	UNP Q56237
C	281	HIS	-	expression tag	UNP Q56237
C	282	HIS	-	expression tag	UNP Q56237
C	283	HIS	-	expression tag	UNP Q56237
C	284	HIS	-	expression tag	UNP Q56237
D	57	MSE	LEU	engineered mutation	UNP Q56237
D	?	-	GLY	deletion	UNP Q56237
D	?	-	PHE	deletion	UNP Q56237
D	?	-	GLY	deletion	UNP Q56237
D	?	-	ARG	deletion	UNP Q56237
D	?	-	ARG	deletion	UNP Q56237
D	?	-	SER	deletion	UNP Q56237
D	?	-	ARG	deletion	UNP Q56237
D	142	MSE	ILE	engineered mutation	UNP Q56237
D	173	MSE	LEU	engineered mutation	UNP Q56237
D	226	MSE	LEU	engineered mutation	UNP Q56237
D	231	GLY	LYS	engineered mutation	UNP Q56237
D	274	ALA	-	expression tag	UNP Q56237
D	275	ALA	-	expression tag	UNP Q56237
D	276	ALA	-	expression tag	UNP Q56237
D	277	LEU	-	expression tag	UNP Q56237
D	278	GLU	-	expression tag	UNP Q56237
D	279	HIS	-	expression tag	UNP Q56237
D	280	HIS	-	expression tag	UNP Q56237
D	281	HIS	-	expression tag	UNP Q56237
D	282	HIS	-	expression tag	UNP Q56237
D	283	HIS	-	expression tag	UNP Q56237
D	284	HIS	-	expression tag	UNP Q56237

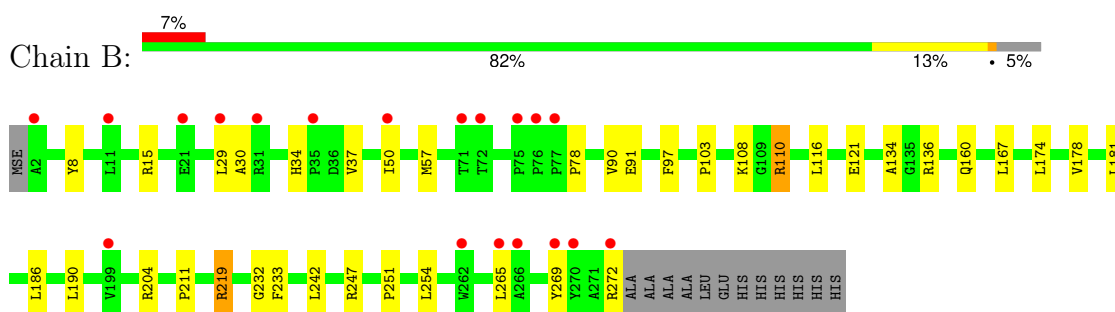
3 Residue-property plots [i](#)

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.

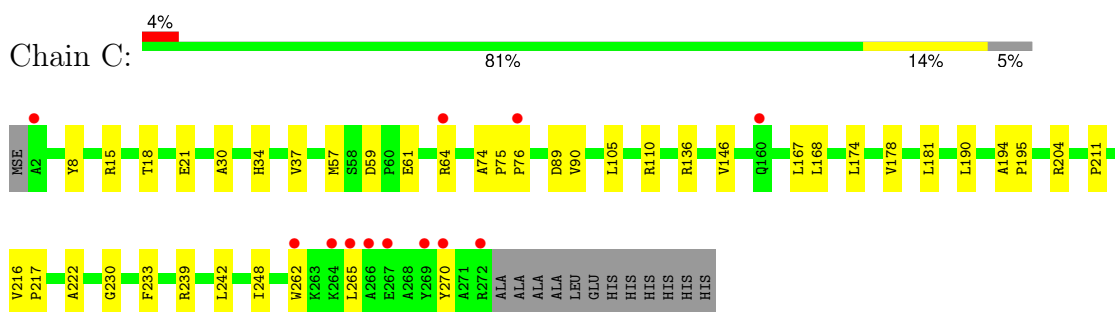
- Molecule 1: Chaperone protein DnaJ 2



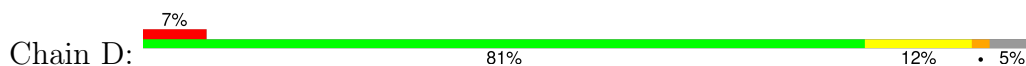
- Molecule 1: Chaperone protein DnaJ 2

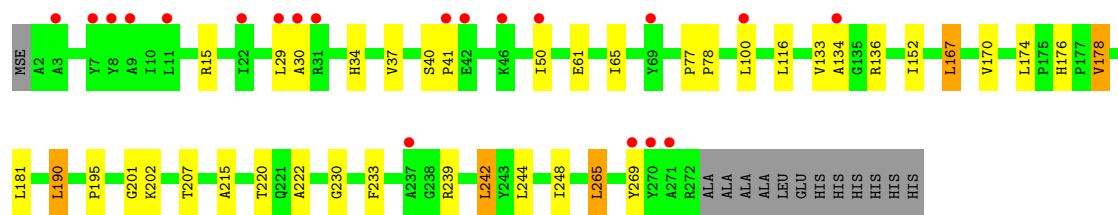


- Molecule 1: Chaperone protein DnaJ 2



- Molecule 1: Chaperone protein DnaJ 2





4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	102.09Å 105.00Å 130.03Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.90 20.00 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.9 (20.00-2.90) 98.9 (20.00-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.43 (at 2.88Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.265 , 0.307 0.259 , 0.299	Depositor DCC
R_{free} test set	1591 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	66.2	Xtrriage
Anisotropy	0.154	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 30.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.024 for k,h,-l	Xtrriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	8460	wwPDB-VP
Average B, all atoms (Å ²)	95.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.47% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.32	0/2164	0.48	0/2923
1	B	0.32	0/2164	0.47	0/2923
1	C	0.32	0/2164	0.49	0/2923
1	D	0.32	0/2164	0.49	0/2923
All	All	0.32	0/8656	0.48	0/11692

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	2115	0	2110	17	0
1	B	2115	0	2110	14	0
1	C	2115	0	2110	17	0
1	D	2115	0	2110	17	0
All	All	8460	0	8440	61	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.

The worst 5 of 61 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:78:PRO:HG2	1:B:103:PRO:HB2	1.67	0.76
1:C:64:ARG:HH22	1:C:76:PRO:HD2	1.52	0.75
1:A:204:ARG:HH12	1:A:211:PRO:HB2	1.52	0.73
1:D:34:HIS:HB3	1:D:37:VAL:HG22	1.73	0.71
1:A:8:TYR:CE1	1:A:57:MSE:HE1	2.30	0.66

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	269/284 (95%)	263 (98%)	5 (2%)	1 (0%)	30	60
1	B	269/284 (95%)	259 (96%)	9 (3%)	1 (0%)	30	60
1	C	269/284 (95%)	263 (98%)	6 (2%)	0	100	100
1	D	269/284 (95%)	262 (97%)	6 (2%)	1 (0%)	30	60
All	All	1076/1136 (95%)	1047 (97%)	26 (2%)	3 (0%)	37	66

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	134	ALA
1	B	134	ALA
1	D	134	ALA

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	217/219 (99%)	202 (93%)	15 (7%)	13	37
1	B	217/219 (99%)	201 (93%)	16 (7%)	11	34
1	C	217/219 (99%)	201 (93%)	16 (7%)	11	34
1	D	217/219 (99%)	202 (93%)	15 (7%)	13	37
All	All	868/876 (99%)	806 (93%)	62 (7%)	12	36

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	219	ARG
1	D	178	VAL
1	C	110	ARG
1	D	174	LEU
1	D	233	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	34	HIS

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 oligosaccharides 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 [i](#)

6.1 Protein, DNA and RNA chains [i](#)

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	265/284 (93%)	0.02	7 (2%) 57 51	36, 64, 130, 202	0
1	B	265/284 (93%)	0.36	19 (7%) 23 20	44, 95, 236, 329	0
1	C	265/284 (93%)	0.10	12 (4%) 39 32	38, 73, 136, 200	0
1	D	265/284 (93%)	0.38	20 (7%) 22 18	37, 87, 246, 323	0
All	All	1060/1136 (93%)	0.21	58 (5%) 32 27	36, 77, 215, 329	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	266	ALA	5.1
1	B	269	TYR	4.9
1	D	3	ALA	4.3
1	B	266	ALA	4.1
1	C	269	TYR	3.7

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 monosaccharides 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.