



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 8, 2026 – 01:15 pm BST

PDB ID : 9HK2 / pdb_00009hk2
Title : Crystal structure of cathepsin D from Schistosoma mansoni
Authors : Busa, M.; Houstecka, R.; Hanova, I.; Mares, M.
Deposited on : 2024-12-02
Resolution : 3.29 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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

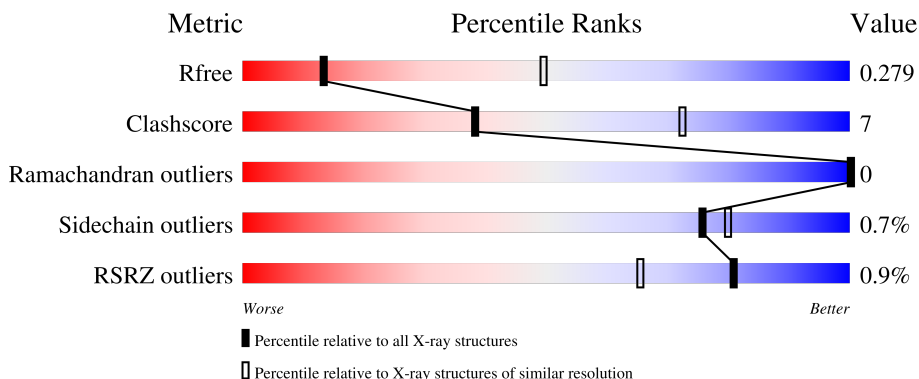
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.29 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}	180053	1169 (3.32-3.28)
Clashscore	190562	1209 (3.32-3.28)
Ramachandran outliers	187476	1188 (3.32-3.28)
Sidechain outliers	187428	1187 (3.32-3.28)
RSRZ outliers	180081	1169 (3.32-3.28)

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	351	<div> <div></div> <div>79%16%5%</div> </div>
1	B	351	<div> <div></div> <div>77%16%6%</div> </div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 5097 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 Cathepsin D (A01 family).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	331	Total	C	N	O	S	0	0	0
			2535	1639	397	483	16			
1	A	334	Total	C	N	O	S	0	0	0
			2562	1654	401	490	17			

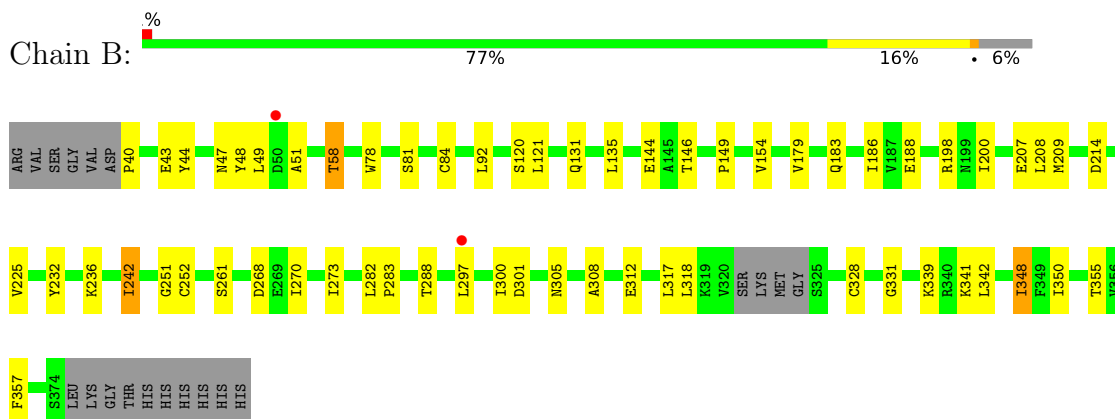
There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	373	ALA	-	expression tag	UNP G4VEV6
B	374	SER	-	expression tag	UNP G4VEV6
B	375	LEU	-	expression tag	UNP G4VEV6
B	376	LYS	-	expression tag	UNP G4VEV6
B	377	GLY	-	expression tag	UNP G4VEV6
B	378	THR	-	expression tag	UNP G4VEV6
B	379	HIS	-	expression tag	UNP G4VEV6
B	380	HIS	-	expression tag	UNP G4VEV6
B	381	HIS	-	expression tag	UNP G4VEV6
B	382	HIS	-	expression tag	UNP G4VEV6
B	383	HIS	-	expression tag	UNP G4VEV6
B	384	HIS	-	expression tag	UNP G4VEV6
A	373	ALA	-	expression tag	UNP G4VEV6
A	374	SER	-	expression tag	UNP G4VEV6
A	375	LEU	-	expression tag	UNP G4VEV6
A	376	LYS	-	expression tag	UNP G4VEV6
A	377	GLY	-	expression tag	UNP G4VEV6
A	378	THR	-	expression tag	UNP G4VEV6
A	379	HIS	-	expression tag	UNP G4VEV6
A	380	HIS	-	expression tag	UNP G4VEV6
A	381	HIS	-	expression tag	UNP G4VEV6
A	382	HIS	-	expression tag	UNP G4VEV6
A	383	HIS	-	expression tag	UNP G4VEV6
A	384	HIS	-	expression tag	UNP G4VEV6

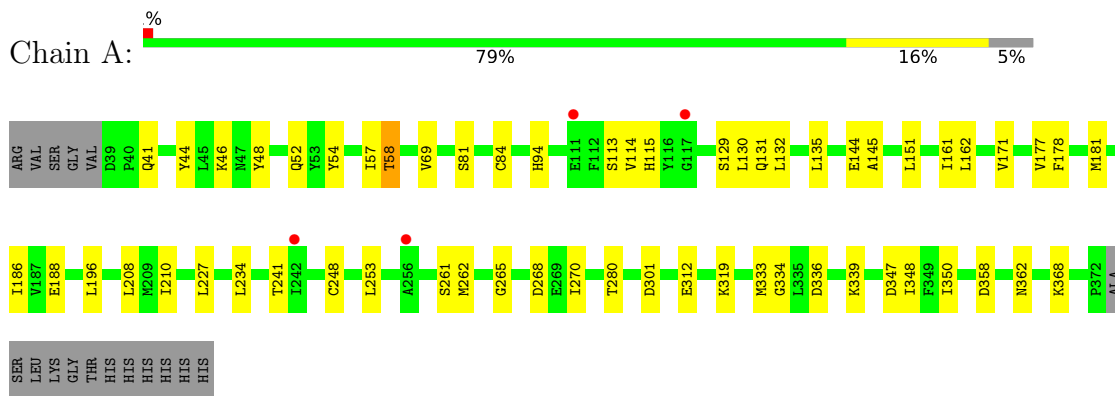
3 Residue-property plots [i](#)

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: Cathepsin D (A01 family)



- Molecule 1: Cathepsin D (A01 family)



4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, α , β , γ	202.49Å 202.49Å 104.27Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.85 – 3.29 44.85 – 3.29	Depositor EDS
% Data completeness (in resolution range)	99.4 (44.85-3.29) 99.3 (44.85-3.29)	Depositor EDS
R_{merge}	0.54	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.67 (at 3.32Å)	Xtriage
Refinement program	REFMAC 5.8.0430 (refmacat 0.4.88)	Depositor
R, R_{free}	0.269 , 0.281 0.268 , 0.279	Depositor DCC
R_{free} test set	601 reflections (4.80%)	wwPDB-VP
Wilson B-factor (Å ²)	82.2	Xtriage
Anisotropy	0.220	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 58.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	5097	wwPDB-VP
Average B, all atoms (Å ²)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.01% 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	0.76	0/2623	1.28	11/3558 (0.3%)
1	B	0.75	0/2595	1.31	11/3520 (0.3%)
All	All	0.76	0/5218	1.29	22/7078 (0.3%)

There are no bond length outliers.

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	44	TYR	CB-CA-C	8.36	123.95	109.65
1	B	242	ILE	N-CA-C	-8.02	104.75	112.29
1	B	232	TYR	CB-CA-C	-7.70	99.12	111.06
1	A	312	GLU	CB-CG-CD	7.19	124.82	112.60
1	B	312	GLU	CB-CG-CD	6.58	123.78	112.60

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	2562	0	2534	39	0
1	B	2535	0	2508	31	0
All	All	5097	0	5042	70	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 70 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:162:LEU:HD21	1:A:177:VAL:HG21	1.44	0.96
1:A:58:THR:HG22	1:A:131:GLN:HB2	1.61	0.83
1:B:282:LEU:HG	1:B:283:PRO:HD2	1.60	0.81
1:A:265:GLY:O	1:A:334:GLY:HA2	1.90	0.71
1:A:162:LEU:HD21	1:A:177:VAL:CG2	2.21	0.71

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	332/351 (95%)	315 (95%)	17 (5%)	0	100	100
1	B	327/351 (93%)	319 (98%)	8 (2%)	0	100	100
All	All	659/702 (94%)	634 (96%)	25 (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	284/298 (95%)	282 (99%)	2 (1%)	76	80

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	280/298 (94%)	278 (99%)	2 (1%)	76	80
All	All	564/596 (95%)	560 (99%)	4 (1%)	76	80

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	92	LEU
1	B	207	GLU
1	A	130	LEU
1	A	336	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	362	ASN
1	A	131	GLN
1	B	305	ASN
1	B	271	GLN
1	B	362	ASN

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 ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	334/351 (95%)	0.12	4 (1%)	76 58	64, 90, 130, 158	0
1	B	331/351 (94%)	0.05	2 (0%)	85 73	56, 85, 127, 149	0
All	All	665/702 (94%)	0.08	6 (0%)	81 65	56, 88, 128, 158	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	297	LEU	2.9
1	B	50	ASP	2.8
1	A	111	GLU	2.8
1	A	256	ALA	2.4
1	A	117	GLY	2.3

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 oligosaccharides 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.