Full wwPDB X-ray Structure Validation Report ⓘ

Sep 24, 2025 – 06:26 pm BST

PDB ID	:	9HB2 / pdb_00009hb2
Title	:	Structure of the truncated version of IdeC protease C94S from Streptococcus canis
Authors	:	Batuecas, M.T.; Miguel-Ruano, V.; Hermoso, J.A.
Deposited on	:	2024-11-05
Resolution	:	2.25 Å(reported)

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

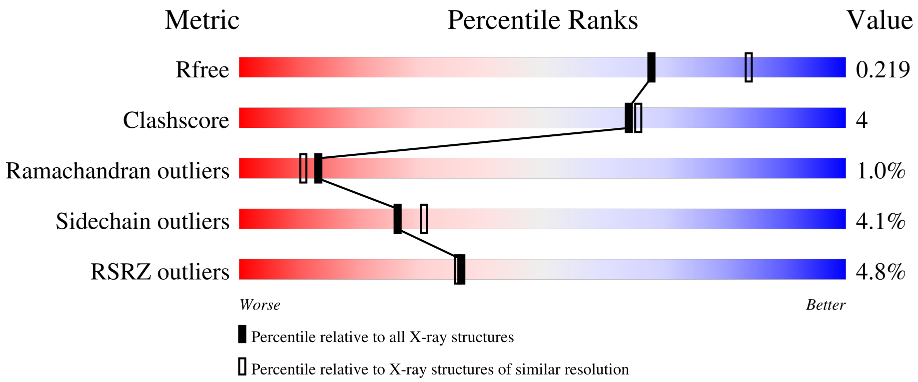
MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.46

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

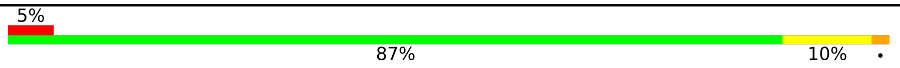
The reported resolution of this entry is 2.25 Å.

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	1763 (2.26-2.26)
Clashscore	180529	1919 (2.26-2.26)
Ramachandran outliers	177936	1884 (2.26-2.26)
Sidechain outliers	177891	1885 (2.26-2.26)
RSRZ outliers	164620	1763 (2.26-2.26)

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	291	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2457 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 IgM protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	291	2297	1458	393	442	4	0	1	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	46	GLN	-	expression tag	UNP A0A3P5YAY8
A	47	GLY	-	expression tag	UNP A0A3P5YAY8
A	48	PRO	-	expression tag	UNP A0A3P5YAY8
A	94	SER	CYS	engineered mutation	UNP A0A3P5YAY8
A	133	GLN	LYS	variant	UNP A0A3P5YAY8

- Molecule 2 is water.

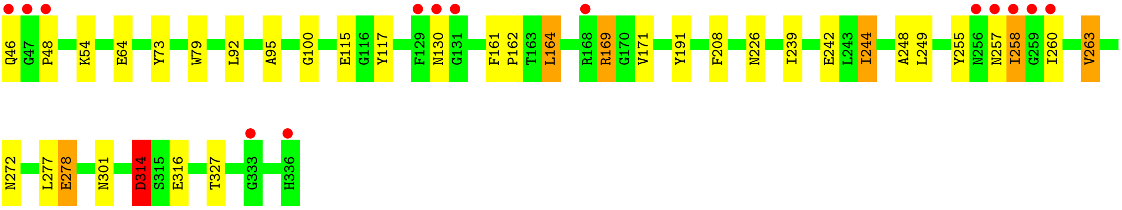
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	160	160	160	0	0

3 Residue-property plots

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: IgM protease

Chain A:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	109.38Å 109.38Å 109.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.96 – 2.25 48.96 – 2.25	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.96-2.25) 100.0 (48.96-2.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.42 (at 2.24Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.178 , 0.214 0.182 , 0.219	Depositor DCC
R_{free} test set	1057 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	36.1	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 25.5	EDS
L-test for twinning[2]	$< \lvert L \rvert > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.041 for l,-k,h	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2457	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.23% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< \lvert L \rvert >$, $< L^2 >$ 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	1.02	1/2352 (0.0%)	1.39	9/3184 (0.3%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	100	GLY	C-O	5.52	1.30	1.23

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	301	ASN	CA-C-N	6.42	129.75	120.38
1	A	301	ASN	C-N-CA	6.42	129.75	120.38
1	A	314	ASP	CB-CA-C	-6.25	98.92	111.17
1	A	327	THR	CA-CB-OG1	-6.19	100.31	109.60
1	A	278	GLU	N-CA-C	-5.81	106.11	113.43
1	A	272	ASN	CA-C-N	5.25	127.57	120.38
1	A	272	ASN	C-N-CA	5.25	127.57	120.38
1	A	260	ILE	N-CA-C	-5.19	108.45	113.53
1	A	115	GLU	CB-CA-C	5.09	119.50	110.85

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	2297	0	2237	17	0
2	A	160	0	0	6	2
All	All	2457	0	2237	17	2

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.

All (17) 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:169:ARG:NH2	1:A:191:TYR:OH	2.12	0.83
1:A:242:GLU:OE1	2:A:401:HOH:O	2.05	0.72
1:A:48:PRO:HB2	2:A:423:HOH:O	1.95	0.67
1:A:226:ASN:HB3	2:A:402:HOH:O	2.06	0.56
1:A:226:ASN:CG	2:A:402:HOH:O	2.49	0.55
1:A:263[A]:VAL:CG2	2:A:551:HOH:O	2.58	0.52
1:A:314:ASP:HB3	1:A:316:GLU:H	1.79	0.48
1:A:161:PHE:N	1:A:162:PRO:CD	2.78	0.47
1:A:244:ILE:C	1:A:244:ILE:HD12	2.40	0.47
1:A:161:PHE:HB3	1:A:164:LEU:HD22	1.97	0.46
1:A:73:TYR:HA	1:A:79:TRP:CH2	2.53	0.44
1:A:239:ILE:HD12	1:A:277:LEU:HD11	2.00	0.43
1:A:95:ALA:HB3	1:A:171:VAL:HG23	2.00	0.43
1:A:117:TYR:CZ	1:A:208:PHE:HA	2.54	0.42
1:A:226:ASN:CB	2:A:402:HOH:O	2.65	0.42
1:A:92:LEU:CD2	1:A:258:ILE:CG2	2.98	0.42
1:A:248:ALA:C	1:A:249:LEU:HD12	2.46	0.40

All (2) 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)
2:A:409:HOH:O	2:A:411:HOH:O[7_555]	1.51	0.69
2:A:509:HOH:O	2:A:537:HOH:O[10_545]	1.54	0.66

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	290/291 (100%)	278 (96%)	9 (3%)	3 (1%)	13 10

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	130	ASN
1	A	257	ASN
1	A	258	ILE

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	244/243 (100%)	233 (96%)	11 (4%)	23 26

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46	GLN
1	A	54	LYS
1	A	64	GLU
1	A	164	LEU
1	A	169	ARG
1	A	244	ILE
1	A	255	TYR
1	A	263[A]	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	263[B]	VAL
1	A	278	GLU
1	A	314	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	125	GLN
1	A	233	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 ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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	291/291 (100%)	-0.11	14 (4%)	36	36	18, 38, 79, 148	1 (0%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	258	ILE	8.5
1	A	129	PHE	5.0
1	A	336	HIS	4.6
1	A	259	GLY	4.6
1	A	260	ILE	3.8
1	A	130	ASN	3.2
1	A	46	GLN	3.0
1	A	48	PRO	3.0
1	A	256	ASN	2.9
1	A	257	ASN	2.9
1	A	47	GLY	2.8
1	A	131	GLY	2.7
1	A	168	ARG	2.3
1	A	333	GLY	2.1

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 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.