WORLDWIDE
wwPDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

May 1, 2025 – 01:11 pm BST

PDB ID	:	9HTH / pdb_00009hth
Title	:	Crystal structure of the fused, cysteine-less Aes123 PolB1 split intein (with S1A, N159A mutations)
Authors	:	Fitzian, K.; Yilmaz, Z.; Humberg, C.; Mootz, H.D.; Kuemmel, D.
Deposited on	:	2024-12-19
Resolution	:	1.38 Å(reported)

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

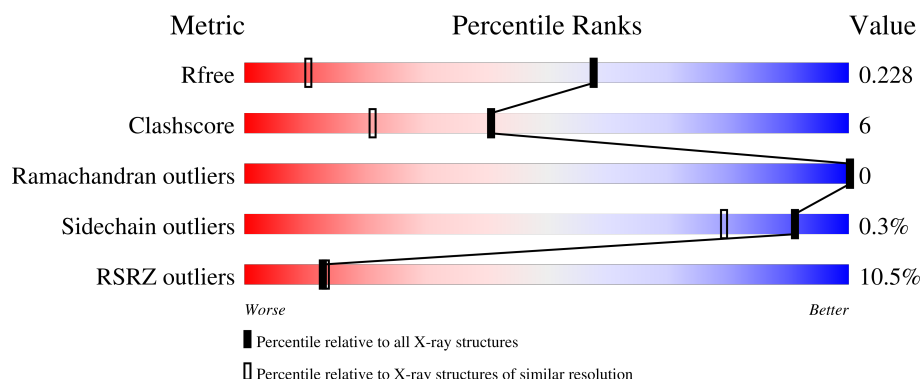|		
MolProbity	:	4-5-2 with Phenix2.0rc1
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.43.1

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 1.38 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	Percentile Ranks	Value
Rfree		0.228
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		0.3%
RSRZ outliers		10.5%

Worse — *Better*

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	3869 (1.40-1.36)
Clashscore	180529	4183 (1.40-1.36)
Ramachandran outliers	177936	4116 (1.40-1.36)
Sidechain outliers	177891	4115 (1.40-1.36)
RSRZ outliers	164620	3867 (1.40-1.36)

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	173	12% / 84% / 13% / ..
1	B	173	9% / 87% / 13%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2943 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-directed DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	170	Total 1362	C 863	N 232	O 258	S 9	0	0	0
1	B	173	Total 1385	C 877	N 235	O 264	S 9	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	ALA	SER	engineered mutation	UNP I6YMS9
A	120A	GLY	-	linker	UNP I6YMS9
A	120B	SER	-	linker	UNP I6YMS9
A	120C	HIS	-	linker	UNP I6YMS9
B	1	ALA	SER	engineered mutation	UNP I6YMS9
B	120A	GLY	-	linker	UNP I6YMS9
B	120B	SER	-	linker	UNP I6YMS9
B	120C	HIS	-	linker	UNP I6YMS9

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	81	Total 81	O 81	0	0
2	B	115	Total 115	O 115	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: DNA-directed DNA polymerase
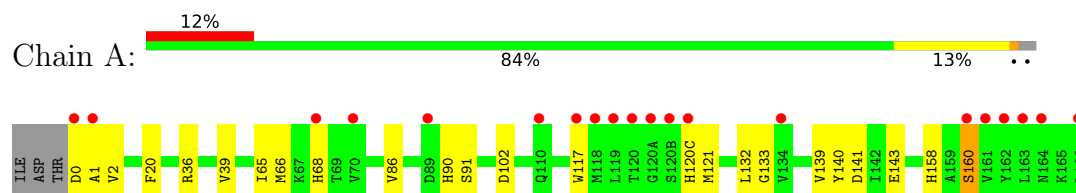
• Molecule 1: DNA-directed DNA polymerase
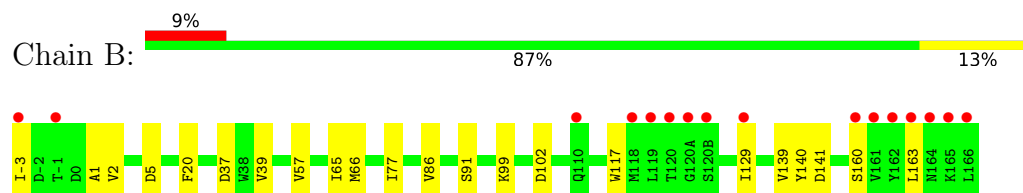
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	30.13Å　　63.01Å　　162.40Å 90.00°　　90.00°　　90.00°	Depositor
Resolution (Å)	40.60　–　1.38 40.60　–　1.38	Depositor EDS
% Data completeness (in resolution range)	96.7 (40.60-1.38) 96.1 (40.60-1.38)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.83 (at 1.38Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, R_{free}	0.193　,　0.226 0.194　,　0.228	Depositor DCC
R_{free} test set	63010 reflections (3.08%)	wwPDB-VP
Wilson B-factor (Å^2)	19.1	Xtriage
Anisotropy	0.347	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 28.6	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2943	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.83% 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 (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	0.31	0/1384	0.59	0/1860
1	B	0.32	0/1407	0.57	0/1892
All	All	0.32	0/2791	0.58	0/3752

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	1362	0	1373	21	0
1	B	1385	0	1395	17	0
2	A	81	0	0	4	0
2	B	115	0	0	2	0
All	All	2943	0	2768	35	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 6.

All (35) 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:120(C):HIS:HD2	1:B:99:LYS:HE2	1.56	0.71

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:120(C):HIS:CD2	1:B:99:LYS:HE2	2.26	0.70
1:A:1:ALA:HB1	1:A:140:TYR:O	1.92	0.70
1:A:1:ALA:HB3	1:A:139:VAL:HB	1.74	0.70
1:A:36:ARG:O	1:A:68:HIS:HD2	1.77	0.67
1:B:39:VAL:HG22	1:B:66:MET:HG2	1.78	0.66
1:B:77:ILE:HD12	1:B:129:ILE:HD12	1.81	0.62
1:A:132:LEU:HD21	2:A:212:HOH:O	2.00	0.62
1:A:39:VAL:HG22	1:A:66:MET:HG2	1.81	0.61
1:A:0:ASP:HB2	1:A:90:HIS:HB2	1.83	0.60
1:A:91:SER:HB3	1:A:102:ASP:HB2	1.87	0.56
1:B:5:ASP:HB2	2:B:308:HOH:O	2.09	0.51
1:B:2:VAL:HG13	2:B:279:HOH:O	2.10	0.51
1:A:90:HIS:CD2	2:A:216:HOH:O	2.64	0.50
1:A:0:ASP:OD1	1:A:90:HIS:ND1	2.27	0.50
1:B:-3:ILE:HD12	1:B:37:ASP:OD2	2.11	0.49
1:B:141:ASP:OD2	1:B:160:SER:HB3	2.11	0.49
1:A:143:GLU:HB2	2:A:235:HOH:O	2.12	0.48
1:B:1:ALA:HB3	1:B:139:VAL:HB	1.94	0.48
1:A:2:VAL:HG12	1:A:86:VAL:HG12	1.96	0.48
1:B:1:ALA:HB1	1:B:140:TYR:O	2.16	0.46
1:A:90:HIS:NE2	1:A:158:HIS:O	2.50	0.45
1:B:20:PHE:CD2	1:B:65:ILE:HD12	2.52	0.44
1:B:57:VAL:HG11	1:B:117:TRP:HZ3	1.83	0.44
1:A:141:ASP:OD2	1:A:160:SER:HB3	2.18	0.44
1:B:91:SER:HB3	1:B:102:ASP:HB2	2.01	0.43
1:B:163:LEU:HD12	1:B:163:LEU:HA	1.65	0.43
1:A:120(C):HIS:HD2	1:B:99:LYS:CE	2.25	0.42
1:B:2:VAL:HG12	1:B:86:VAL:HG12	2.00	0.42
1:A:133:GLY:HA2	2:A:203:HOH:O	2.18	0.42
1:A:117:TRP:CH2	1:A:121:MET:HE2	2.55	0.42
1:B:39:VAL:HG22	1:B:66:MET:CG	2.46	0.41
1:A:1:ALA:HB1	1:A:140:TYR:C	2.44	0.41
1:A:20:PHE:CD2	1:A:65:ILE:HD12	2.56	0.41
1:A:39:VAL:HG22	1:A:66:MET:CG	2.49	0.40

There are no symmetry-related clashes.

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	168/173 (97%)	165 (98%)	3 (2%)	0	100	100
1	B	171/173 (99%)	169 (99%)	2 (1%)	0	100	100
All	All	339/346 (98%)	334 (98%)	5 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	152/155 (98%)	151 (99%)	1 (1%)	81	61
1	B	155/155 (100%)	155 (100%)	0	100	100
All	All	307/310 (99%)	306 (100%)	1 (0%)	91	79

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	160	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	68	HIS
1	A	120(C)	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	148	HIS
1	B	158	HIS

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(Å²)	Q<0.9
1	A	170/173 (98%)	0.89	20 (11%)	10	11	15, 25, 55, 71	0
1	B	173/173 (100%)	0.66	16 (9%)	16	17	14, 21, 47, 73	0
All	All	343/346 (99%)	0.78	36 (10%)	13	13	14, 24, 49, 73	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	119	LEU	6.3
1	B	163	LEU	6.1
1	B	119	LEU	6.0
1	A	120(A)	GLY	5.9
1	A	166	LEU	5.7
1	A	162	TYR	5.7
1	B	166	LEU	5.5
1	A	1	ALA	4.7
1	B	120	THR	4.6
1	B	162	TYR	4.6
1	B	-3	ILE	4.5
1	A	161	VAL	4.4
1	A	163	LEU	4.4
1	A	120	THR	4.1
1	B	118	MET	4.1
1	A	120(B)	SER	4.0
1	B	161	VAL	3.6
1	B	129	ILE	3.6
1	A	120(C)	HIS	3.5
1	B	120(A)	GLY	3.5
1	B	160	SER	3.3
1	A	0	ASP	3.3
1	A	118	MET	3.3
1	B	164	ASN	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	165	LYS	2.9
1	A	160	SER	2.9
1	B	-1	THR	2.8
1	A	117	TRP	2.7
1	A	68	HIS	2.7
1	A	164	ASN	2.6
1	A	70	VAL	2.5
1	B	120(B)	SER	2.4
1	A	89	ASP	2.2
1	A	110	GLN	2.2
1	A	134	VAL	2.1
1	B	110	GLN	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 monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.