



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 11:27 AM GMT

PDB ID : 3P1Y
Title : Crystal structure of the chimeric *Archaeoglobus fulgidus* RNA splicing endonuclease with the broadest substrate specificity
Authors : Hirata, A.
Deposited on : 2010-10-01
Resolution : 2.05 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
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

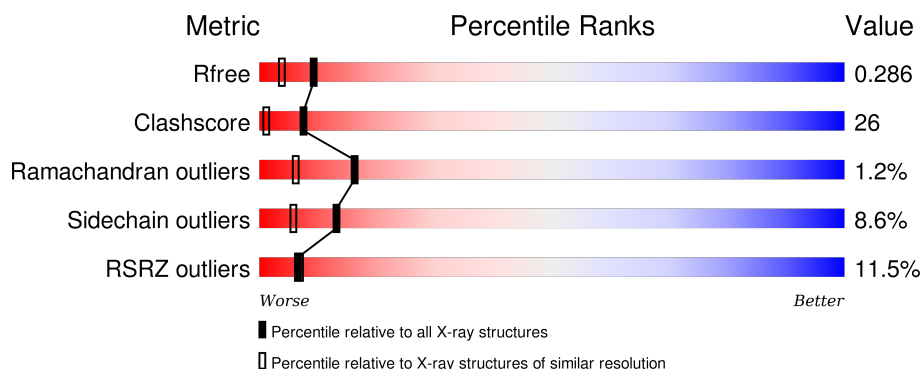
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.05 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}	91344	1192 (2.04-2.04)
Clashscore	102246	1269 (2.04-2.04)
Ramachandran outliers	100387	1258 (2.04-2.04)
Sidechain outliers	100360	1258 (2.04-2.04)
RSRZ outliers	91569	1194 (2.04-2.04)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	315	<div> <div>13%</div> <div>60%</div> <div>32%</div> <div>7%</div> </div>
1	B	315	<div> <div>8%</div> <div>60%</div> <div>34%</div> <div>6%</div> </div>
1	C	315	<div> <div>9%</div> <div>61%</div> <div>33%</div> <div>5%</div> </div>
1	D	315	<div> <div>16%</div> <div>55%</div> <div>40%</div> <div>5%</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10726 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 tRNA-splicing endonuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	314	Total	C	N	O	S	0	0	0
			2606	1668	446	486	6			
1	B	314	Total	C	N	O	S	0	0	0
			2606	1668	446	486	6			
1	C	314	Total	C	N	O	S	0	0	0
			2606	1668	446	486	6			
1	D	314	Total	C	N	O	S	0	0	0
			2606	1668	446	486	6			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP O29362
B	1	MET	-	EXPRESSION TAG	UNP O29362
C	1	MET	-	EXPRESSION TAG	UNP O29362
D	1	MET	-	EXPRESSION TAG	UNP O29362

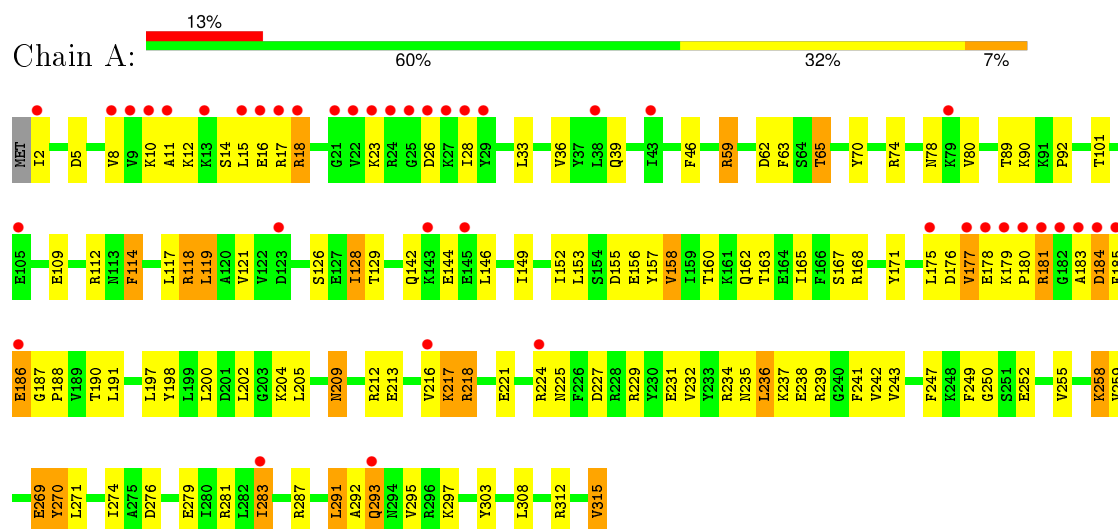
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	77	Total	O	0	0
			77	77		
2	B	84	Total	O	0	0
			84	84		
2	C	78	Total	O	0	0
			78	78		
2	D	63	Total	O	0	0
			63	63		

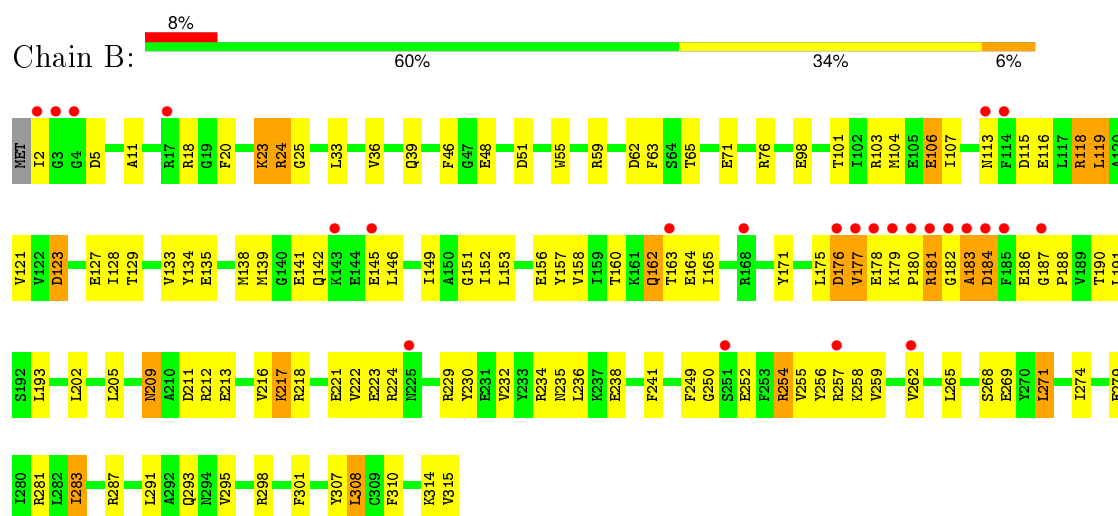
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: tRNA-splicing endonuclease

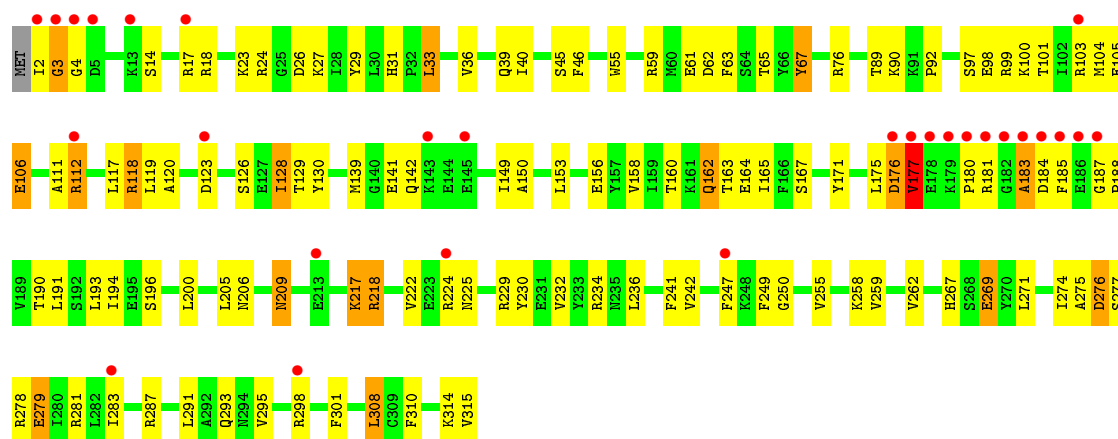


- Molecule 1: tRNA-splicing endonuclease

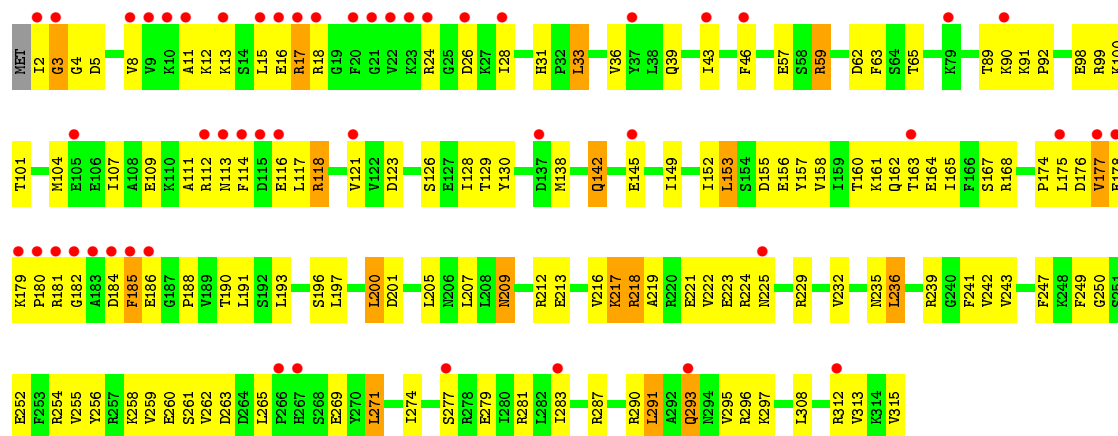


- Molecule 1: tRNA-splicing endonuclease





• Molecule 1: tRNA-splicing endonuclease



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	81.77Å 104.64Å 165.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.89 – 2.05 38.89 – 2.05	Depositor EDS
% Data completeness (in resolution range)	93.9 (38.89-2.05) 94.2 (38.89-2.05)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.81 (at 2.05Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.247 , 0.287 0.247 , 0.286	Depositor DCC
R_{free} test set	8397 reflections (9.98%)	DCC
Wilson B-factor (Å ²)	30.1	Xtriage
Anisotropy	0.388	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 47.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	5 of 88812 reflections (0.006%)	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	10726	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 45.47 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.3152e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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	0.38	1/2654 (0.0%)	0.67	3/3563 (0.1%)
1	B	0.47	2/2654 (0.1%)	0.64	0/3563
1	C	0.38	0/2654	0.66	1/3563 (0.0%)
1	D	0.44	2/2654 (0.1%)	0.63	1/3563 (0.0%)
All	All	0.42	5/10616 (0.0%)	0.65	5/14252 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	151	GLY	C-O	-11.51	1.05	1.23
1	B	152	ILE	CB-CG2	-7.39	1.29	1.52
1	D	152	ILE	C-O	-6.73	1.10	1.23
1	D	153	LEU	C-O	-6.29	1.11	1.23
1	A	152	ILE	C-N	-5.77	1.20	1.34

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	152	ILE	O-C-N	-9.58	107.38	122.70
1	A	152	ILE	CA-C-N	6.66	131.85	117.20
1	C	242	VAL	N-CA-C	-5.49	96.19	111.00
1	D	242	VAL	N-CA-C	-5.28	96.75	111.00
1	A	242	VAL	N-CA-C	-5.18	97.03	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	67	TYR	Sidechain

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	2606	0	2616	141	0
1	B	2606	0	2617	137	0
1	C	2606	0	2616	135	0
1	D	2606	0	2617	155	0
2	A	77	0	0	1	0
2	B	84	0	0	4	0
2	C	78	0	0	5	0
2	D	63	0	0	4	0
All	All	10726	0	10466	533	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 26.

The worst 5 of 533 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:149:ILE:HB	1:B:162:GLN:HG3	1.38	1.05
1:A:149:ILE:HB	1:A:162:GLN:HG3	1.39	1.03
1:C:149:ILE:HB	1:C:162:GLN:HG3	1.42	1.02
1:A:218:ARG:HH11	1:A:218:ARG:HB2	1.20	1.02
1:B:283:ILE:HD12	1:B:283:ILE:H	1.30	0.97

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	312/315 (99%)	292 (94%)	19 (6%)	1 (0%)	46	36
1	B	312/315 (99%)	295 (95%)	12 (4%)	5 (2%)	12	3
1	C	312/315 (99%)	292 (94%)	14 (4%)	6 (2%)	10	2
1	D	312/315 (99%)	291 (93%)	18 (6%)	3 (1%)	19	8
All	All	1248/1260 (99%)	1170 (94%)	63 (5%)	15 (1%)	16	6

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	177	VAL
1	B	176	ASP
1	B	177	VAL
1	B	183	ALA
1	C	177	VAL

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	280/281 (100%)	252 (90%)	28 (10%)	9	3
1	B	280/281 (100%)	254 (91%)	26 (9%)	11	4
1	C	280/281 (100%)	257 (92%)	23 (8%)	14	6
1	D	280/281 (100%)	261 (93%)	19 (7%)	20	11
All	All	1120/1124 (100%)	1024 (91%)	96 (9%)	13	6

5 of 96 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	181	ARG
1	B	308	LEU
1	D	236	LEU
1	B	193	LEU
1	B	254	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	39	GLN
1	C	206	ASN
1	D	267	HIS
1	C	142	GLN
1	C	209	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 carbohydrates 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 [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	314/315 (99%)	0.89	41 (13%) 5 5	19, 32, 59, 89	0
1	B	314/315 (99%)	0.78	25 (7%) 15 17	18, 33, 52, 87	0
1	C	314/315 (99%)	0.80	28 (8%) 12 13	20, 32, 53, 91	0
1	D	314/315 (99%)	1.08	51 (16%) 3 2	22, 35, 61, 92	0
All	All	1256/1260 (99%)	0.89	145 (11%) 6 7	18, 33, 59, 92	0

The worst 5 of 145 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	2	ILE	16.0
1	C	180	PRO	16.0
1	B	2	ILE	14.0
1	D	183	ALA	13.7
1	A	183	ALA	12.4

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 carbohydrates 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.