



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 15, 2018 – 11:21 AM EST

PDB ID : 3MJI
Title : Activation of catalytic cysteine without a base in a Mutant Penicillin Acylase Precursor
Authors : Pathak, M.C.; Suresh, C.G.; Dodson, G.G.; Murshudov, G.N.
Deposited on : 2010-04-12
Resolution : 2.50 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030736

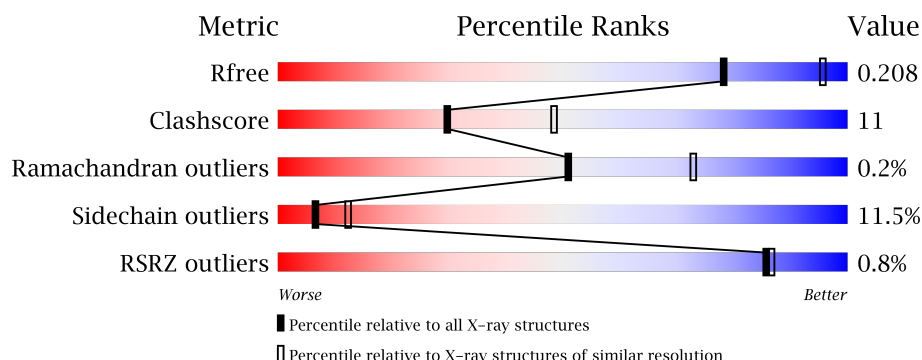
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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}	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	338	<div> <div>0.2%</div> <div> <div></div> <div>78%</div> <div>16%</div> <div>0.2%</div> </div> </div>
1	B	338	<div> <div>0.2%</div> <div> <div></div> <div>74%</div> <div>17%</div> <div>9%</div> </div> </div>
1	C	338	<div> <div>0.2%</div> <div> <div></div> <div>75%</div> <div>17%</div> <div>6%</div> </div> </div>
1	D	338	<div> <div>0.2%</div> <div> <div></div> <div>72%</div> <div>20%</div> <div>6%</div> </div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 10729 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 Penicillin acylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	334	Total	C	N	O	S	0	0	0
			2606	1659	427	507	13			
1	B	334	Total	C	N	O	S	0	0	0
			2606	1659	427	507	13			
1	C	333	Total	C	N	O	S	0	0	0
			2599	1654	426	506	13			
1	D	333	Total	C	N	O	S	0	0	0
			2599	1654	426	506	13			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	SER	CYS	CONFLICT	UNP P12256
A	101	ARG	THR	CONFLICT	UNP P12256
B	4	SER	CYS	CONFLICT	UNP P12256
B	101	ARG	THR	CONFLICT	UNP P12256
C	4	SER	CYS	CONFLICT	UNP P12256
C	101	ARG	THR	CONFLICT	UNP P12256
D	4	SER	CYS	CONFLICT	UNP P12256
D	101	ARG	THR	CONFLICT	UNP P12256

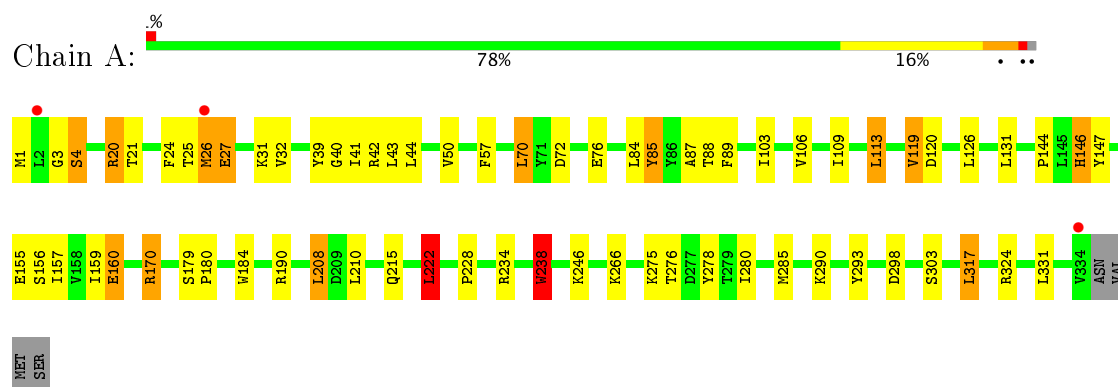
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	90	Total	O	0	0
			90	90		
2	B	69	Total	O	0	0
			69	69		
2	C	73	Total	O	0	0
			73	73		
2	D	87	Total	O	0	0
			87	87		

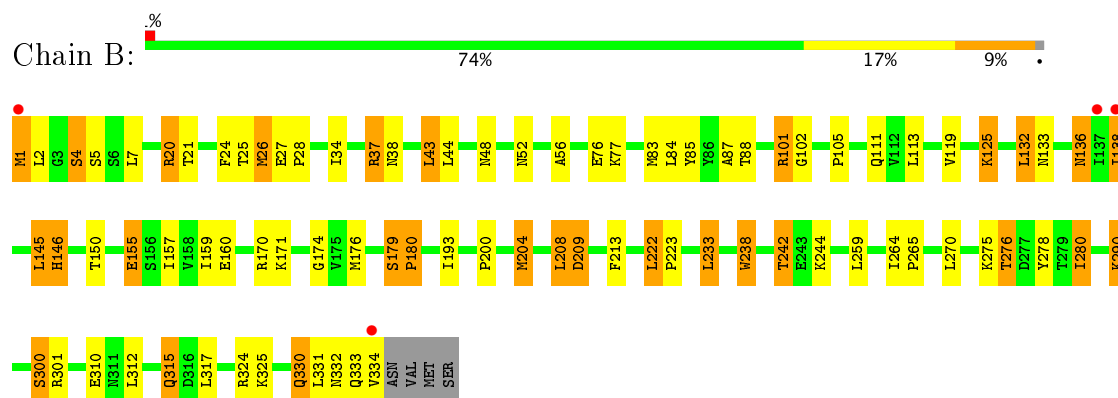
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

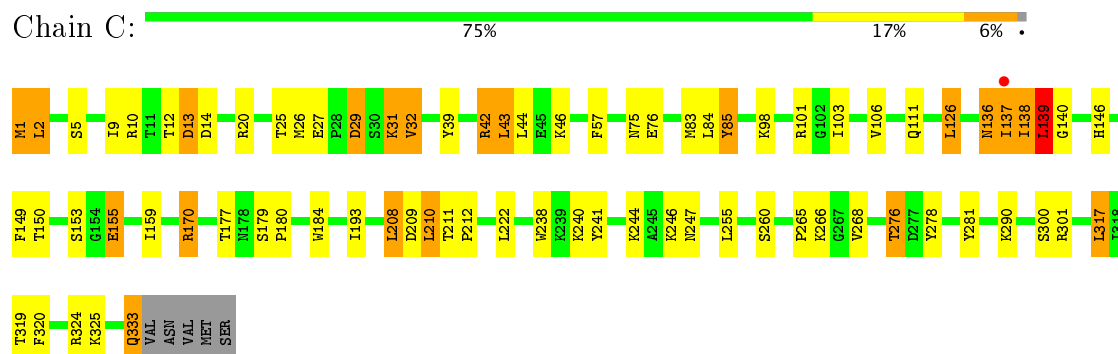
• Molecule 1: Penicillin acylase



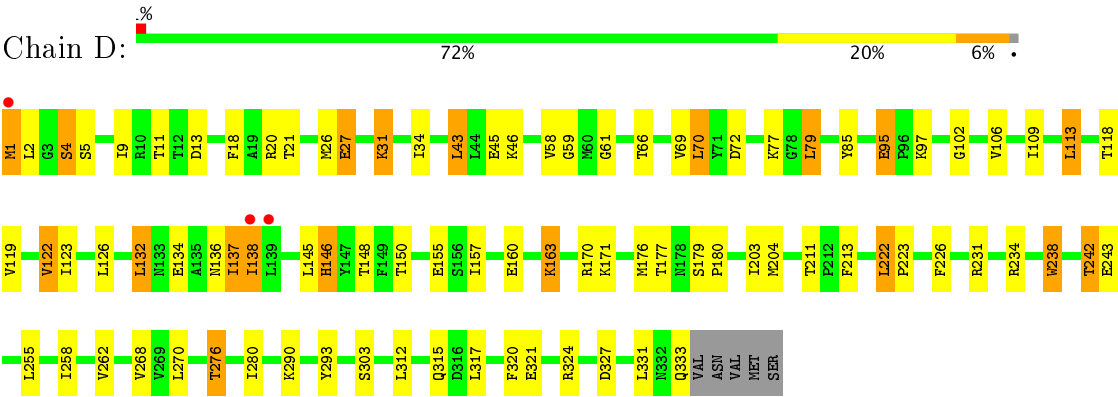
• Molecule 1: Penicillin acylase



• Molecule 1: Penicillin acylase



● Molecule 1: Penicillin acylase



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	102.64Å 90.09Å 102.27Å 90.00° 102.13° 90.00°	Depositor
Resolution (Å)	19.95 – 2.50 19.95 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.6 (19.95-2.50) 99.0 (19.95-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.69 (at 2.50Å)	Xtriage
Refinement program	REFMAC 5.6.0041	Depositor
R, R_{free}	0.166 , 0.211 0.172 , 0.208	Depositor DCC
R_{free} test set	3170 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	26.7	Xtriage
Anisotropy	0.473	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 12.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.124 for l,-k,h	Xtriage
Reported twinning fraction	0.901 for H, K, L 0.099 for L, -K, H	Depositor
Outliers	2 of 62981 reflections (0.003%)	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	10729	wwPDB-VP
Average B, all atoms (Å ²)	27.0	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 22.84 % 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 5.2485e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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

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	1.23	3/2661 (0.1%)	0.99	2/3616 (0.1%)
1	B	1.15	3/2661 (0.1%)	0.99	4/3616 (0.1%)
1	C	1.15	2/2654 (0.1%)	0.98	3/3606 (0.1%)
1	D	1.12	2/2654 (0.1%)	1.00	2/3606 (0.1%)
All	All	1.16	10/10630 (0.1%)	0.99	11/14444 (0.1%)

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	2

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	179	SER	CB-OG	-5.93	1.34	1.42
1	B	209	ASP	CB-CG	-5.92	1.39	1.51
1	D	226	PHE	CE1-CZ	5.62	1.48	1.37
1	C	179	SER	CB-OG	-5.49	1.35	1.42
1	A	160	GLU	CG-CD	5.44	1.60	1.51

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	2	LEU	CA-CB-CG	5.92	128.92	115.30
1	D	70	LEU	CA-CB-CG	-5.78	102.01	115.30
1	B	233	LEU	CA-CB-CG	5.74	128.50	115.30
1	A	190	ARG	NE-CZ-NH1	-5.46	117.57	120.30
1	C	126	LEU	CB-CG-CD1	5.33	120.07	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	136	ASN	Peptide
1	C	139	LEU	Peptide

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	2587	48	0
1	B	2606	0	2587	57	0
1	C	2599	0	2578	58	0
1	D	2599	0	2578	67	0
2	A	90	0	0	2	0
2	B	69	0	0	1	0
2	C	73	0	0	0	0
2	D	87	0	0	3	0
All	All	10729	0	10330	221	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 11.

The worst 5 of 221 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:C:137:ILE:O	1:C:140:GLY:HA2	1.34	1.27
1:A:24:PHE:HB3	1:A:280:ILE:CD1	1.66	1.26
1:C:137:ILE:HD12	1:C:137:ILE:H	1.07	1.13
1:C:137:ILE:H	1:C:137:ILE:CD1	1.63	1.11
1:C:136:ASN:HB3	1:C:137:ILE:C	1.73	1.09

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	332/338 (98%)	316 (95%)	15 (4%)	1 (0%)	44	66
1	B	332/338 (98%)	309 (93%)	22 (7%)	1 (0%)	44	66
1	C	331/338 (98%)	311 (94%)	20 (6%)	0	100	100
1	D	331/338 (98%)	313 (95%)	17 (5%)	1 (0%)	44	66
All	All	1326/1352 (98%)	1249 (94%)	74 (6%)	3 (0%)	51	73

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	163	LYS
1	A	119	VAL
1	B	180	PRO

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	290/294 (99%)	268 (92%)	22 (8%)	15	29
1	B	290/294 (99%)	250 (86%)	40 (14%)	4	8
1	C	289/294 (98%)	253 (88%)	36 (12%)	5	10
1	D	289/294 (98%)	254 (88%)	35 (12%)	6	11
All	All	1158/1176 (98%)	1025 (88%)	133 (12%)	6	12

5 of 133 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	331	LEU
1	C	85	TYR
1	D	242	THR
1	B	333	GLN
1	C	29	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	188	ASN
1	A	315	GLN
1	B	188	ASN

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 carbohydrates 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	334/338 (98%)	-0.42	3 (0%) 84 85	15, 24, 42, 71	0
1	B	334/338 (98%)	-0.37	4 (1%) 79 80	16, 26, 48, 67	0
1	C	333/338 (98%)	-0.43	1 (0%) 93 94	15, 26, 44, 66	0
1	D	333/338 (98%)	-0.40	3 (0%) 84 85	15, 26, 49, 62	0
All	All	1334/1352 (98%)	-0.41	11 (0%) 86 86	15, 25, 46, 71	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	137	ILE	5.9
1	B	138	ILE	4.0
1	A	26	MET	3.2
1	D	1	MET	3.0
1	A	2	LEU	2.8

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.