



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 1, 2016 – 05:45 PM GMT

PDB ID : 4JEI  
Title : Nonglycosylated Yarrowia lipolytica LIP2 lipase  
Authors : Aloulou, A.; Benarouche, A.; Puccinelli, D.; Spinelli, S.; Cavalier, J.-F.; Cam-  
billau, C.; Carriere, F.  
Deposited on : 2013-02-27  
Resolution : 2.60 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto: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

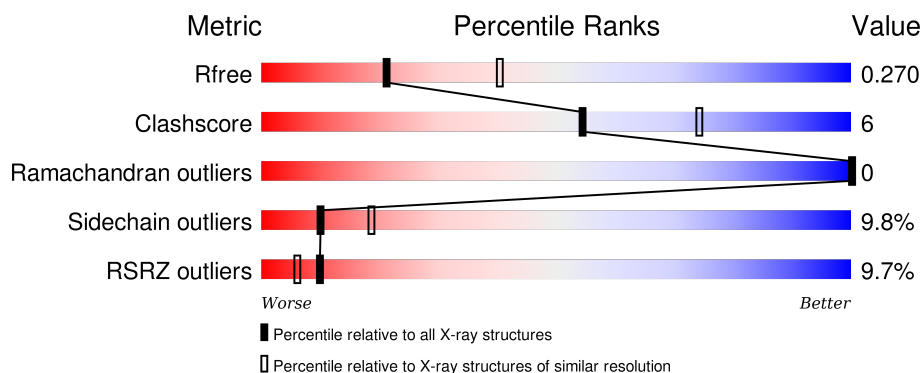
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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	2328 (2.60-2.60)
Clashscore	102246	2679 (2.60-2.60)
Ramachandran outliers	100387	2635 (2.60-2.60)
Sidechain outliers	100360	2635 (2.60-2.60)
RSRZ outliers	91569	2334 (2.60-2.60)

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  $\geq 3$ , 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	301	<div> <div>10%</div> <div>79%</div> <div>18%</div> <div>..</div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2444 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 Lipase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	299	Total	C	N	O	S	12	0	0
			2334	1491	394	438	11			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	88	THR	ILE	SEE REMARK 999	UNP E0A7J0
A	96	THR	SER	SEE REMARK 999	UNP E0A7J0
A	113	GLN	ASN	ENGINEERED MUTATION	UNP E0A7J0
A	134	GLN	ASN	ENGINEERED MUTATION	UNP E0A7J0

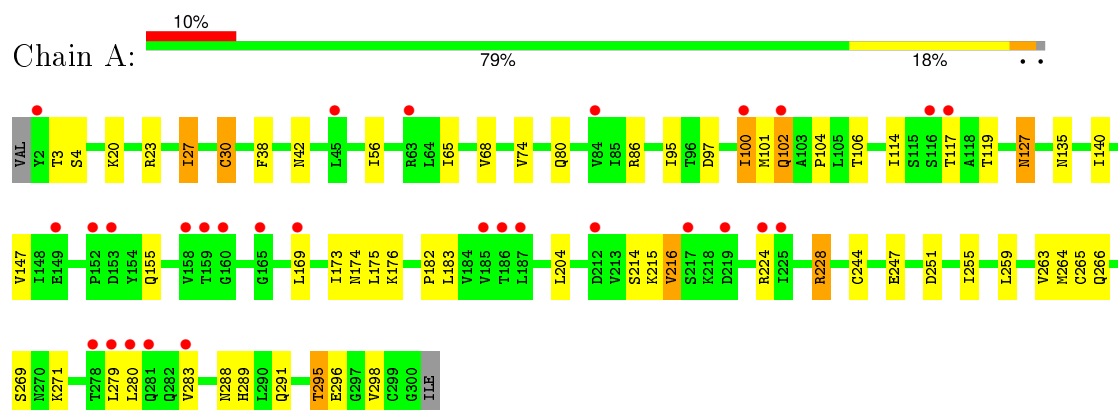
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	110	Total	O	0	0
			110	110		

### 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: Lipase 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	116.69 Å 116.69 Å 170.27 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	27.97 – 2.60 27.66 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.5 (27.97-2.60) 99.5 (27.66-2.60)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.82 (at 2.61 Å)	Xtriage
Refinement program	BUSTER 2.11.2	Depositor
R, $R_{free}$	0.224 , 0.250 0.239 , 0.270	Depositor DCC
$R_{free}$ test set	1112 reflections (5.42%)	DCC
Wilson B-factor (Å <sup>2</sup> )	56.9	Xtriage
Anisotropy	0.576	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 54.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 21618 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	2444	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [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	0.41	0/2395	0.69	0/3264

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	2334	0	2238	29	0
2	A	110	0	0	0	0
All	All	2444	0	2238	29	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 (29) 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:216:VAL:HG23	1:A:264:MET:HE1	1.55	0.88
1:A:68:VAL:H	1:A:135:ASN:HD22	1.22	0.82
1:A:68:VAL:H	1:A:135:ASN:ND2	1.86	0.72
1:A:216:VAL:HG23	1:A:264:MET:CE	2.21	0.70

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:97:ASP:O	1:A:101:MET:HB2	1.94	0.67
1:A:56:ILE:HD11	1:A:74:VAL:HG22	1.83	0.61
1:A:56:ILE:HD11	1:A:74:VAL:CG2	2.34	0.58
1:A:247:GLU:HB3	1:A:265:CYS:HB2	1.89	0.55
1:A:100:ILE:HG12	1:A:100:ILE:O	2.07	0.55
1:A:68:VAL:N	1:A:135:ASN:HD22	2.00	0.55
1:A:176:LYS:HD2	1:A:204:LEU:HD22	1.88	0.55
1:A:27:ILE:O	1:A:30:CYS:SG	2.66	0.53
1:A:3:THR:HG22	1:A:269:SER:HB3	1.92	0.52
1:A:280:LEU:O	1:A:283:VAL:HG23	2.10	0.51
1:A:38:PHE:HB2	1:A:42:ASN:HB2	1.93	0.51
1:A:214:SER:HB2	1:A:266:GLN:HE22	1.76	0.50
1:A:114:ILE:HD11	1:A:174:ASN:HB2	1.95	0.48
1:A:4:SER:OG	1:A:266:GLN:HB3	2.13	0.48
1:A:23:ARG:O	1:A:27:ILE:HG12	2.14	0.47
1:A:20:LYS:NZ	1:A:80:GLN:HE22	2.13	0.47
1:A:228:ARG:HG3	1:A:251:ASP:HA	1.98	0.45
1:A:288:ASN:HA	1:A:291:GLN:HE21	1.83	0.43
1:A:95:ILE:HD11	1:A:289:HIS:CD2	2.54	0.42
1:A:255:ILE:HD13	1:A:295:THR:HG22	2.02	0.41
1:A:175:LEU:HB2	1:A:182:PRO:HG3	2.02	0.41
1:A:104:PRO:HD2	1:A:127:ASN:OD1	2.21	0.41
1:A:102:GLN:O	1:A:104:PRO:HD3	2.21	0.40
1:A:56:ILE:HD13	1:A:147:VAL:HG21	2.04	0.40
1:A:224:ARG:NH1	1:A:244:CYS:O	2.54	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	297/301 (99%)	280 (94%)	17 (6%)	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	256/260 (98%)	231 (90%)	25 (10%)	10	19

All (25) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	ILE
1	A	30	CYS
1	A	65	ILE
1	A	86	ARG
1	A	100	ILE
1	A	102	GLN
1	A	106	THR
1	A	117	THR
1	A	119	THR
1	A	127	ASN
1	A	140	ILE
1	A	155	GLN
1	A	169	LEU
1	A	173	ILE
1	A	183	LEU
1	A	215	LYS
1	A	216	VAL
1	A	228	ARG
1	A	259	LEU
1	A	263	VAL
1	A	271	LYS
1	A	279	LEU
1	A	295	THR
1	A	296	GLU
1	A	298	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:



Mol	Chain	Res	Type
1	A	12	GLN
1	A	80	GLN
1	A	113	GLN
1	A	135	ASN
1	A	139	GLN
1	A	150	GLN
1	A	155	GLN
1	A	174	ASN
1	A	210	ASN
1	A	291	GLN

### 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 ⓘ

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	299/301 (99%)	0.54	29 (9%) 10 6	43, 64, 86, 101	12 (4%)

All (29) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	281	GLN	8.8
1	A	280	LEU	7.6
1	A	100	ILE	5.5
1	A	186	THR	4.3
1	A	117	THR	3.2
1	A	63	ARG	3.0
1	A	2	TYR	3.0
1	A	219	ASP	2.9
1	A	279	LEU	2.9
1	A	116	SER	2.9
1	A	160	GLY	2.9
1	A	185	VAL	2.8
1	A	165	GLY	2.8
1	A	283	VAL	2.8
1	A	169	LEU	2.6
1	A	158	VAL	2.6
1	A	153	ASP	2.6
1	A	212	ASP	2.5
1	A	45	LEU	2.5
1	A	149	GLU	2.4
1	A	159	THR	2.4
1	A	187	LEU	2.3
1	A	84	VAL	2.2
1	A	225	ILE	2.2
1	A	278	THR	2.2
1	A	152	PRO	2.1
1	A	224	ARG	2.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	102	GLN	2.0
1	A	217	SER	2.0

## 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.