



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 30, 2016 – 10:51 AM EST

PDB ID : 5L3C  
Title : Human LSD1/CoREST: LSD1 E379K mutation  
Authors : Pilotto, S.; Speranzini, V.; Marabelli, C.; Mattevi, A.  
Deposited on : 2016-04-06  
Resolution : 3.31 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 : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20028320  
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) : rb-20028320

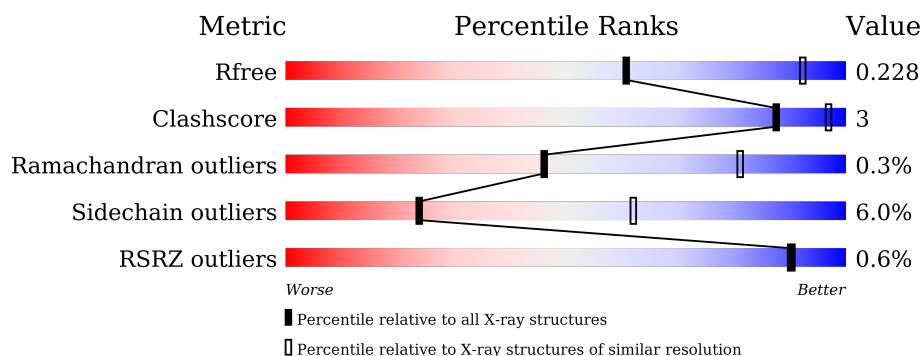
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.31 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	1198 (3.40-3.24)
Clashscore	102246	1280 (3.40-3.24)
Ramachandran outliers	100387	1260 (3.40-3.24)
Sidechain outliers	100360	1259 (3.40-3.24)
RSRZ outliers	91569	1203 (3.40-3.24)

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	852	
2	B	482	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 6356 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 Lysine-specific histone demethylase 1A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	666	5217	3325	907	965	20	0	0	0

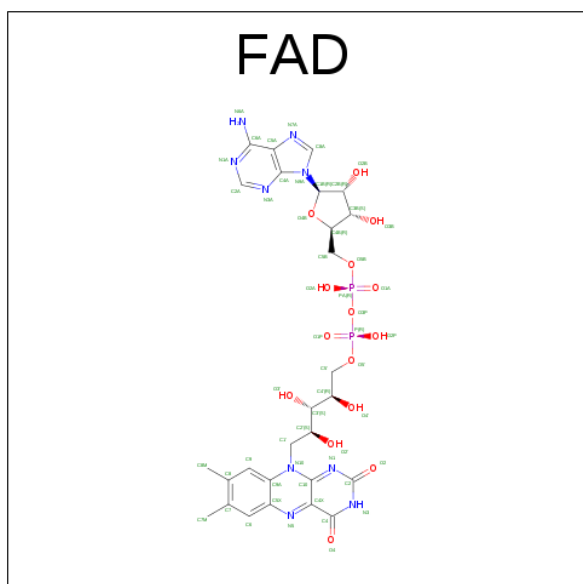
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	379	LYS	GLU	engineered mutation	UNP O60341

- Molecule 2 is a protein called REST corepressor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	134	1086	682	197	204	3	0	0	0

- Molecule 3 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula:  $C_{27}H_{33}N_9O_{15}P_2$ ).

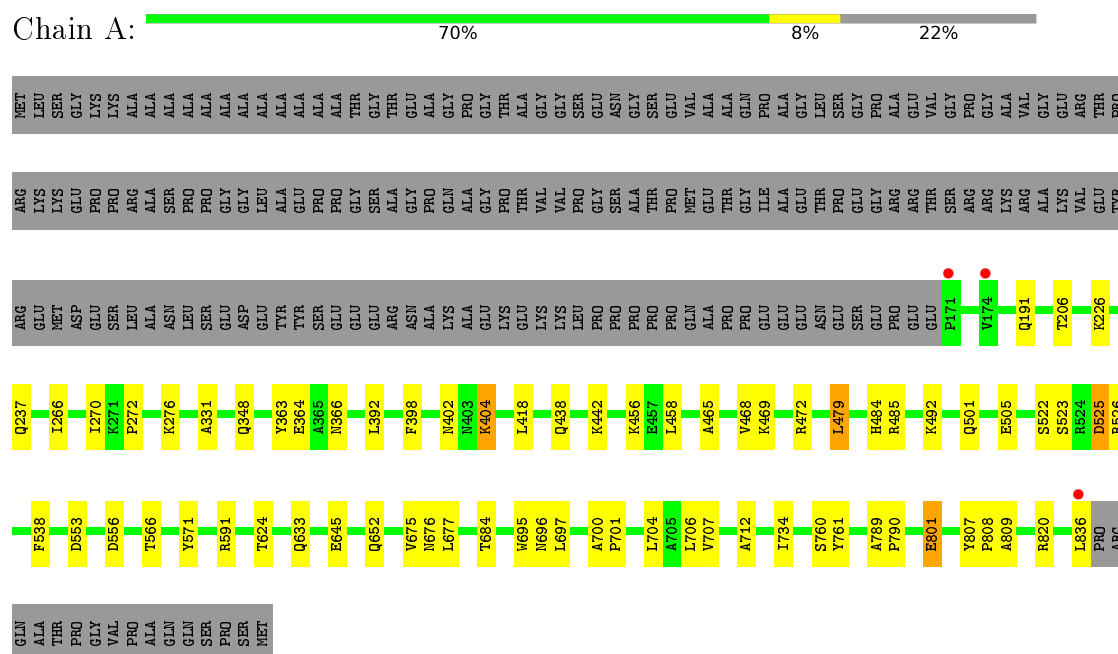


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

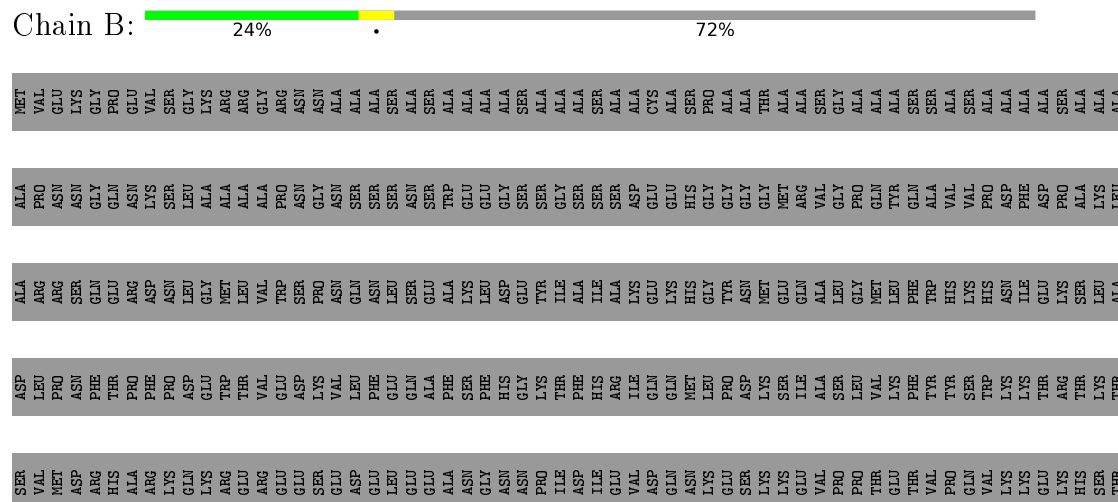
### 3 Residue-property plots

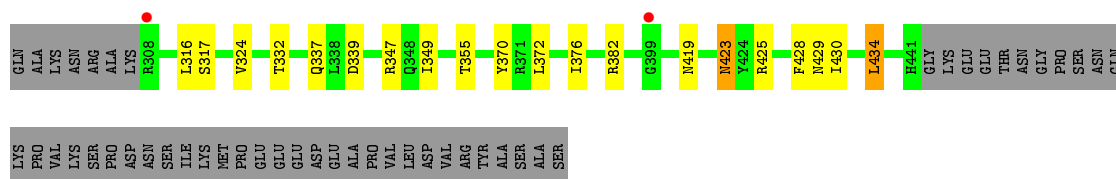
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Lysine-specific histone demethylase 1A



- Molecule 2: REST corepressor 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	120.37Å 180.39Å 235.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.31 48.92 – 3.31	Depositor EDS
% Data completeness (in resolution range)	98.9 (50.00-3.31) 99.1 (48.92-3.31)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.49 (at 3.33Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.199 , 0.223 0.207 , 0.228	Depositor DCC
$R_{free}$ test set	747 reflections (1.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	99.3	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 70.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	6356	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	108.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.08% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: FAD

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	0.49	0/5331	0.70	1/7231 (0.0%)
2	B	0.49	0/1102	0.69	0/1486
All	All	0.49	0/6433	0.70	1/8717 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	820	ARG	NE-CZ-NH1	5.43	123.02	120.30

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	5217	0	5259	28	0
2	B	1086	0	1098	9	0
3	A	53	0	31	2	0
All	All	6356	0	6388	33	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 3.



All (33) 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:266:ILE:N	1:A:348:GLN:OE1	2.24	0.71
2:B:425:ARG:HA	2:B:430:ILE:HD13	1.74	0.69
1:A:525:ASP:OD1	1:A:525:ASP:N	2.35	0.60
1:A:760:SER:HB2	3:A:901:FAD:HM83	1.88	0.56
1:A:363:TYR:CE2	1:A:734:ILE:HG23	2.41	0.55
1:A:331:ALA:HA	3:A:901:FAD:N5	2.23	0.54
1:A:566:THR:HG21	1:A:697:LEU:HD13	1.92	0.51
2:B:428:PHE:HB2	2:B:430:ILE:CD1	2.41	0.50
1:A:695:TRP:CE3	1:A:697:LEU:HD11	2.47	0.49
1:A:761:TYR:CD1	1:A:809:ALA:HB1	2.48	0.49
1:A:484:HIS:CD2	2:B:372:LEU:HD13	2.48	0.49
2:B:430:ILE:HG22	2:B:434:LEU:HD11	1.94	0.48
1:A:700:ALA:HB1	1:A:701:PRO:HD2	1.96	0.47
1:A:363:TYR:CD2	1:A:734:ILE:HG23	2.51	0.46
1:A:418:LEU:CD1	2:B:324:VAL:HG21	2.46	0.46
1:A:442:LYS:HE3	2:B:355:THR:HG21	1.96	0.46
1:A:392:LEU:HD23	1:A:398:PHE:CD2	2.50	0.46
1:A:538:PHE:CE1	1:A:706:LEU:HD23	2.50	0.46
1:A:456:LYS:HA	2:B:370:TYR:CE2	2.51	0.45
1:A:801:GLU:CG	1:A:809:ALA:HA	2.47	0.45
2:B:428:PHE:HB2	2:B:430:ILE:HD11	1.99	0.44
2:B:423:ASN:OD1	2:B:423:ASN:N	2.44	0.43
1:A:707:VAL:HG12	1:A:712:ALA:HA	1.99	0.43
1:A:676:ASN:HB2	1:A:677:LEU:HD23	2.01	0.43
1:A:402:ASN:O	1:A:404:LYS:NZ	2.41	0.42
1:A:465:ALA:HB2	1:A:479:LEU:HD23	2.02	0.42
1:A:807:TYR:N	1:A:808:PRO:CD	2.84	0.41
1:A:538:PHE:CE1	1:A:706:LEU:CD2	3.03	0.41
1:A:553:ASP:O	1:A:556:ASP:HB2	2.20	0.41
1:A:789:ALA:HB1	1:A:790:PRO:HD2	2.02	0.40
1:A:270:ILE:O	1:A:272:PRO:HD3	2.21	0.40
1:A:468:VAL:O	1:A:472:ARG:NH1	2.55	0.40
1:A:501:GLN:O	1:A:505:GLU:HB2	2.21	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	664/852 (78%)	626 (94%)	36 (5%)	2 (0%)	46	81
2	B	132/482 (27%)	117 (89%)	15 (11%)	0	100	100
All	All	796/1334 (60%)	743 (93%)	51 (6%)	2 (0%)	46	81

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	801	GLU
1	A	364	GLU

### 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	566/699 (81%)	538 (95%)	28 (5%)	31	69
2	B	118/395 (30%)	105 (89%)	13 (11%)	8	31
All	All	684/1094 (62%)	643 (94%)	41 (6%)	24	62

All (41) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	191	GLN
1	A	206	THR
1	A	226	LYS
1	A	237	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	276	LYS
1	A	366	ASN
1	A	404	LYS
1	A	438	GLN
1	A	458	LEU
1	A	469	LYS
1	A	479	LEU
1	A	485	ARG
1	A	492	LYS
1	A	522	SER
1	A	523	SER
1	A	525	ASP
1	A	526	ARG
1	A	571	TYR
1	A	591	ARG
1	A	624	THR
1	A	633	GLN
1	A	645	GLU
1	A	652	GLN
1	A	675	VAL
1	A	684	THR
1	A	696	ASN
1	A	704	LEU
1	A	836	LEU
2	B	316	LEU
2	B	317	SER
2	B	332	THR
2	B	337	GLN
2	B	339	ASP
2	B	347	ARG
2	B	349	ILE
2	B	376	ILE
2	B	382	ARG
2	B	419	ASN
2	B	423	ASN
2	B	429	ASN
2	B	434	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	237	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	402	ASN
1	A	422	HIS
1	A	438	GLN
1	A	484	HIS
1	A	633	GLN
2	B	337	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](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
3	FAD	A	901	-	52,58,58	1.34	6 (11%)	52,89,89	2.57	15 (28%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	FAD	A	901	-	-	0/30/50/50	0/6/6/6

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	901	FAD	C2A-N3A	2.05	1.35	1.32
3	A	901	FAD	C8-C7	2.96	1.48	1.41
3	A	901	FAD	C9A-N10	3.29	1.43	1.38
3	A	901	FAD	C5A-C4A	3.33	1.48	1.40
3	A	901	FAD	C4X-C10	3.45	1.47	1.40
3	A	901	FAD	C9A-C5X	3.80	1.50	1.42

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	901	FAD	N3A-C2A-N1A	-8.27	122.37	128.87
3	A	901	FAD	N3-C2-N1	-5.08	119.14	127.69
3	A	901	FAD	C4-C4X-C10	-5.04	116.72	119.94
3	A	901	FAD	C4X-C10-N10	-3.77	117.78	120.52
3	A	901	FAD	C1B-N9A-C4A	-3.48	122.92	126.81
3	A	901	FAD	C4B-O4B-C1B	-2.62	106.86	109.64
3	A	901	FAD	C4X-C4-N3	-2.43	120.34	123.52
3	A	901	FAD	C4X-N5-C5X	2.22	119.33	116.72
3	A	901	FAD	O2A-PA-O1A	2.39	125.01	112.56
3	A	901	FAD	O4'-C4'-C3'	2.73	115.98	108.96
3	A	901	FAD	C6-C5X-C9A	2.87	122.28	119.11
3	A	901	FAD	N6A-C6A-N1A	2.93	123.44	118.52
3	A	901	FAD	C5X-C9A-N10	3.41	120.13	117.58
3	A	901	FAD	C1'-N10-C9A	3.98	123.44	118.83
3	A	901	FAD	C4-N3-C2	9.19	122.82	115.16

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	901	FAD	2	0

## 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, 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	666/852 (78%)	0.01	3 (0%) 91 92	61, 101, 140, 171	0
2	B	134/482 (27%)	0.12	2 (1%) 76 76	93, 128, 156, 177	0
All	All	800/1334 (59%)	0.03	5 (0%) 90 90	61, 108, 145, 177	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	399	GLY	2.9
1	A	174	VAL	2.5
1	A	171	PRO	2.4
2	B	308	ARG	2.1
1	A	836	LEU	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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	FAD	A	901	53/53	0.96	0.24	-0.04	64,75,88,100	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.