



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 07:33 PM GMT

PDB ID : 4P8P  
Title : Crystal structure of M. tuberculosis DprE1 in complex with the non-covalent inhibitor QN127  
Authors : Neres, J.; Pojer, F.; Cole, S.T.  
Deposited on : 2014-03-31  
Resolution : 2.20 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](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

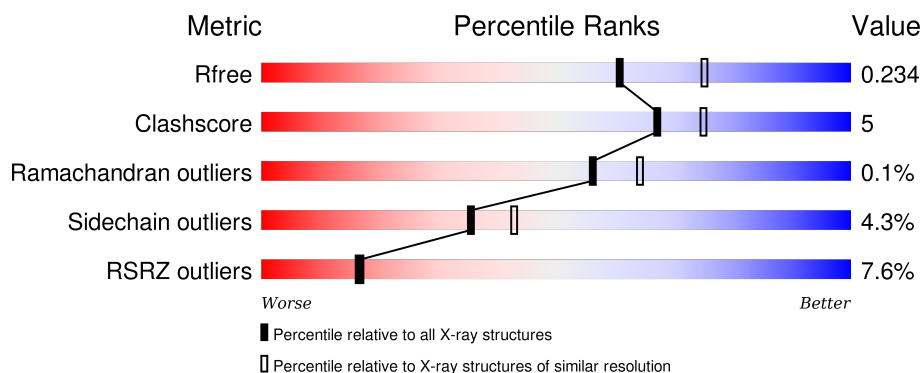
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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	480	
1	B	480	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7067 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 Probable decaprenylphosphoryl-beta-D-ribose oxidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	450	Total	C	N	O	S	0	0	0
			3461	2203	607	640	11			
1	B	420	Total	C	N	O	S	0	1	0
			3234	2052	575	597	10			

There are 38 discrepancies between the modelled and reference sequences:

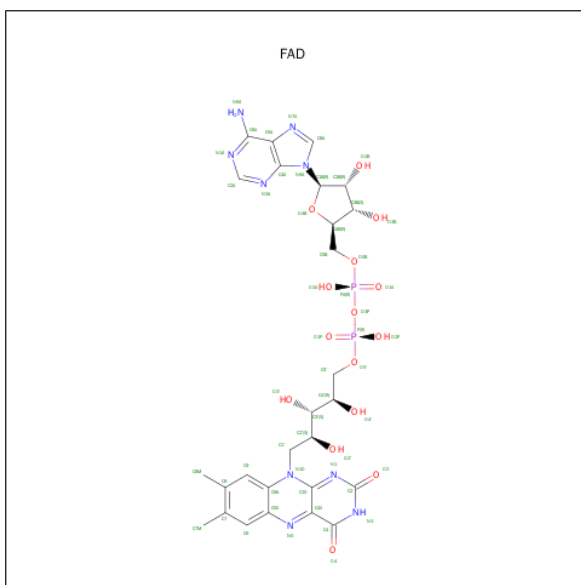
Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	GLY	-	expression tag	UNP P72056
A	-17	SER	-	expression tag	UNP P72056
A	-16	SER	-	expression tag	UNP P72056
A	-15	HIS	-	expression tag	UNP P72056
A	-14	HIS	-	expression tag	UNP P72056
A	-13	HIS	-	expression tag	UNP P72056
A	-12	HIS	-	expression tag	UNP P72056
A	-11	HIS	-	expression tag	UNP P72056
A	-10	HIS	-	expression tag	UNP P72056
A	-9	SER	-	expression tag	UNP P72056
A	-8	SER	-	expression tag	UNP P72056
A	-7	GLY	-	expression tag	UNP P72056
A	-6	LEU	-	expression tag	UNP P72056
A	-5	VAL	-	expression tag	UNP P72056
A	-4	PRO	-	expression tag	UNP P72056
A	-3	ARG	-	expression tag	UNP P72056
A	-2	GLY	-	expression tag	UNP P72056
A	-1	SER	-	expression tag	UNP P72056
A	0	HIS	-	expression tag	UNP P72056
B	-18	GLY	-	expression tag	UNP P72056
B	-17	SER	-	expression tag	UNP P72056
B	-16	SER	-	expression tag	UNP P72056
B	-15	HIS	-	expression tag	UNP P72056
B	-14	HIS	-	expression tag	UNP P72056
B	-13	HIS	-	expression tag	UNP P72056

*Continued on next page...*

Continued from previous page...

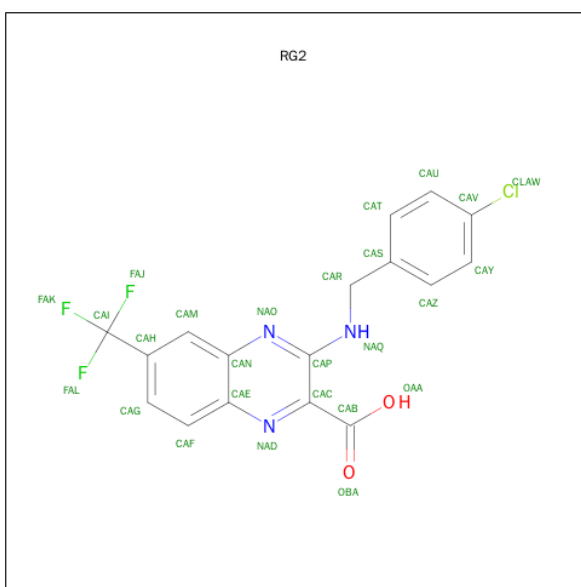
Chain	Residue	Modelled	Actual	Comment	Reference
B	-12	HIS	-	expression tag	UNP P72056
B	-11	HIS	-	expression tag	UNP P72056
B	-10	HIS	-	expression tag	UNP P72056
B	-9	SER	-	expression tag	UNP P72056
B	-8	SER	-	expression tag	UNP P72056
B	-7	GLY	-	expression tag	UNP P72056
B	-6	LEU	-	expression tag	UNP P72056
B	-5	VAL	-	expression tag	UNP P72056
B	-4	PRO	-	expression tag	UNP P72056
B	-3	ARG	-	expression tag	UNP P72056
B	-2	GLY	-	expression tag	UNP P72056
B	-1	SER	-	expression tag	UNP P72056
B	0	HIS	-	expression tag	UNP P72056

- Molecule 2 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula:  $C_{27}H_{33}N_9O_{15}P_2$ ).



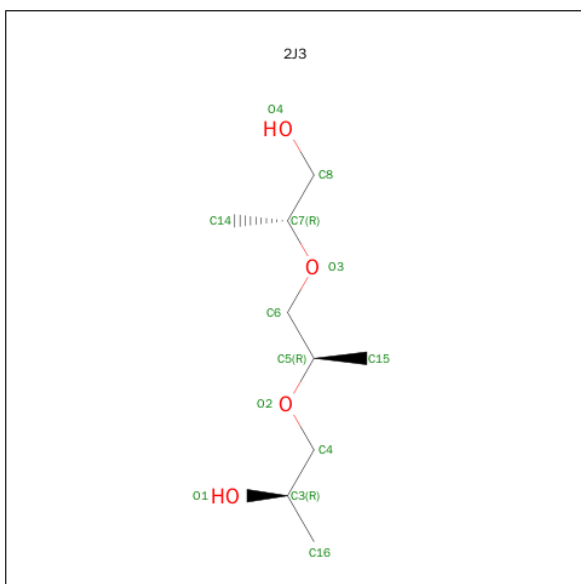
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	B	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

- Molecule 3 is 3-[(4-chlorobenzyl)amino]-6-(trifluoromethyl)quinoxaline-2-carboxylic acid (three-letter code: RG2) (formula:  $C_{17}H_{11}ClF_3N_3O_2$ ).



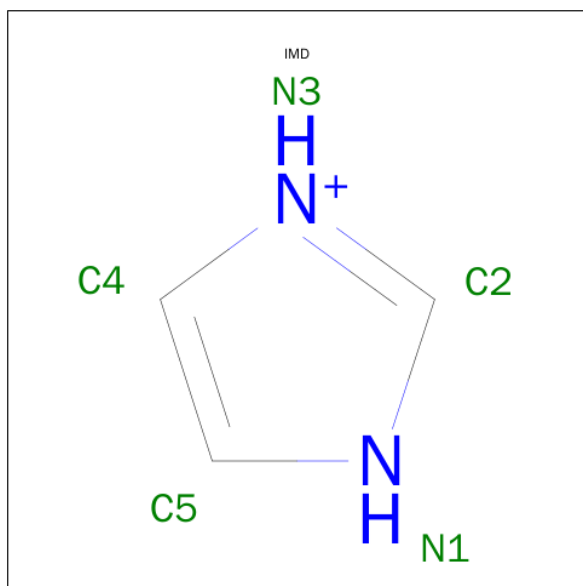
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	C	Cl	F	N	O	0	0
			26	17	1	3	3	2		
3	B	1	Total	C	Cl	F	N	O	0	0
			26	17	1	3	3	2		

- Molecule 4 is (2R)-2-([(2R)-2-([(2R)-2-hydroxypropyl]oxy)propyl]oxy)propan-1-ol (three-letter code: 2J3) (formula: C<sub>9</sub>H<sub>20</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			13	9	4		

- Molecule 5 is IMIDAZOLE (three-letter code: IMD) (formula:  $C_3H_5N_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	N	0	0
			5	3	2		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	112	Total	O	0	0
			112	112		
6	B	84	Total	O	0	0
			84	84		

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 ( $\text{RSRZ} > 2$ ). 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.

- Chain A:
- 
- Sequence logo for Chain A. The y-axis represents information content in bits (0.00 to 0.10). The x-axis lists amino acids. A color scale at the top indicates conservation levels: 7% (red), 84% (green), 9% (yellow), and 6% (grey).

- Chain B:
- 
- 7% 73% 14% 13%
- GLY SER SER SER HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET LEU SER VAL GLY ALA THR T8 T9 A10 T11 R12 W16 G17 R18 T19 A20 P21 S22 V23 A24 N25 V26 T29 F33 M34 I35 V36 K37 A40 R41 V42 A43 E44 S45 G46 C47
- G48 R49 G50 A51 Y60 M63 G68 G69 L70 V71 T75 P76 H81 S82 T83 D84 Y90 D91 I92 D93 D99 K103 W111 P116 G117 T118 R119 Q120 V121 H132 A154 D167 G182 V205 T206 E211 L215 W230 F231 L240
- A258 K266 F267 P270 G291 I292 G293 L310 Y314 H315 P316 L332 A154 D167 G182 V205 T206 E211 L215 W230 F231 L240
- F347 K348 K349 I350 Q355 F362 M364 A375 S378 D394 D404 R405 R413 K418 R421 T422 L447 R448 F449 F450 M454 L461

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	77.62Å 83.96Å 80.87Å 90.00° 103.25° 90.00°	Depositor
Resolution (Å)	49.20 – 2.20 49.17 – 2.20	Depositor EDS
% Data completeness (in resolution range)	98.3 (49.20-2.20) 98.4 (49.17-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.66 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.195 , 0.234 0.194 , 0.234	Depositor DCC
$R_{free}$ test set	2553 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	37.8	Xtriage
Anisotropy	0.018	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 33.1	EDS
Estimated twinning fraction	0.024 for l,-k,h	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 50451 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7067	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.43% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: RG2, 2J3, IMD, 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	0.51	5/3544 (0.1%)	0.60	0/4815
1	B	0.50	3/3311 (0.1%)	0.57	0/4492
All	All	0.51	8/6855 (0.1%)	0.58	0/9307

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	323	TRP	CD2-CE2	5.19	1.47	1.41
1	A	172	TRP	CD2-CE2	5.12	1.47	1.41
1	A	230	TRP	CD2-CE2	5.12	1.47	1.41
1	B	16	TRP	CD2-CE2	5.12	1.47	1.41
1	A	437	TRP	CD2-CE2	5.10	1.47	1.41

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	3461	0	3430	26	0
1	B	3234	0	3218	41	0
2	A	53	0	31	2	0
2	B	53	0	31	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	26	0	10	0	0
3	B	26	0	10	1	0
4	A	13	0	20	1	0
5	A	5	0	5	2	0
6	A	112	0	0	1	0
6	B	84	0	0	3	0
All	All	7067	0	6755	66	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 5.

The worst 5 of 66 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:103:LYS:HE3	1:B:267:PHE:CE2	1.73	1.22
1:B:103:LYS:HE3	1:B:267:PHE:CZ	2.03	0.94
1:A:406:ARG:HH11	1:A:406:ARG:HG2	1.41	0.85
1:A:119:ARG:HB2	1:A:119:ARG:HH11	1.41	0.84
1:B:103:LYS:CE	1:B:267:PHE:CZ	2.72	0.72

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

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	446/480 (93%)	439 (98%)	6 (1%)	1 (0%)	52	59
1	B	415/480 (86%)	407 (98%)	8 (2%)	0	100	100
All	All	861/960 (90%)	846 (98%)	14 (2%)	1 (0%)	56	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	321	GLY

### 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	360/384 (94%)	346 (96%)	14 (4%)	39	48
1	B	337/384 (88%)	321 (95%)	16 (5%)	32	39
All	All	697/768 (91%)	667 (96%)	30 (4%)	35	43

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	441	ARG
1	B	49	ARG
1	B	413	ARG
1	B	41	ARG
1	B	92	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	364	ASN
1	B	364	ASN
1	B	81	HIS
1	A	132	HIS
1	B	132	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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

6 ligands are 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$
2	FAD	A	501	-	48,58,58	1.32	6 (12%)	54,89,89	2.23	13 (24%)
3	RG2	A	502	-	24,28,28	1.33	3 (12%)	36,41,41	1.21	2 (5%)
4	2J3	A	503	-	12,12,12	0.72	0	11,14,14	0.57	0
5	IMD	A	504	-	3,5,5	0.33	0	4,5,5	0.73	0
2	FAD	B	501	-	48,58,58	1.40	6 (12%)	54,89,89	2.27	11 (20%)
3	RG2	B	502	-	24,28,28	1.34	3 (12%)	36,41,41	1.08	2 (5%)

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
2	FAD	A	501	-	-	0/30/50/50	0/6/6/6
3	RG2	A	502	-	-	0/11/15/15	0/3/3/3
4	2J3	A	503	-	-	0/12/12/12	0/0/0/0
5	IMD	A	504	-	-	0/0/0/0	0/1/1/1
2	FAD	B	501	-	-	0/30/50/50	0/6/6/6
3	RG2	B	502	-	-	0/11/15/15	0/3/3/3

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	501	FAD	C9A-N10	2.05	1.41	1.38
3	A	502	RG2	CAC-NAD	2.38	1.35	1.32
3	A	502	RG2	CAV-CLAW	2.46	1.80	1.74
2	A	501	FAD	C9A-N10	2.49	1.42	1.38
3	B	502	RG2	CAV-CLAW	2.54	1.80	1.74

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	501	FAD	N3A-C2A-N1A	-8.00	122.77	128.89
2	A	501	FAD	N3A-C2A-N1A	-7.92	122.83	128.89
2	A	501	FAD	C4-C4X-C10	-6.80	115.59	119.94
2	B	501	FAD	C4-C4X-C10	-6.47	115.80	119.94
3	A	502	RG2	CAC-CAP-NAO	-3.56	119.09	122.02

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	501	FAD	2	0
4	A	503	2J3	1	0
5	A	504	IMD	2	0
3	B	502	RG2	1	0

## 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, 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	450/480 (93%)	0.23	34 (7%) 17 16	22, 36, 88, 117	0
1	B	420/480 (87%)	0.28	32 (7%) 17 16	26, 43, 77, 95	0
All	All	870/960 (90%)	0.25	66 (7%) 17 16	22, 40, 81, 117	0

The worst 5 of 66 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	273	LEU	6.6
1	A	323	TRP	6.3
1	A	321	GLY	5.7
1	B	8	THR	5.7
1	A	281	ASN	5.3

### 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	RG2	B	502	26/26	0.91	0.17	0.49	49,52,66,70	0
2	FAD	A	501	53/53	0.97	0.15	0.46	26,29,30,32	0
4	2J3	A	503	13/13	0.82	0.20	0.33	71,74,77,77	0
2	FAD	B	501	53/53	0.98	0.13	-0.44	28,34,37,38	0
3	RG2	A	502	26/26	0.95	0.14	-0.51	35,38,63,76	0
5	IMD	A	504	5/5	0.96	0.14	-	39,39,41,41	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.