wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:56 PM GMT

PDB ID	:	4B1B
Title	:	Crystal structure of Plasmodium falciparum oxidised Thioredoxin Reductase at 2.9 angstrom
Authors	:	Boumis, G.; Giardina, G.; Dimastrogiovanni, D.; Angelucci, F.; Saccoccia, F.; Brunori, M.; Bellelli, A.; Miele, A.E.
Deposited on	:	2012-07-09
Resolution	:	2.90 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

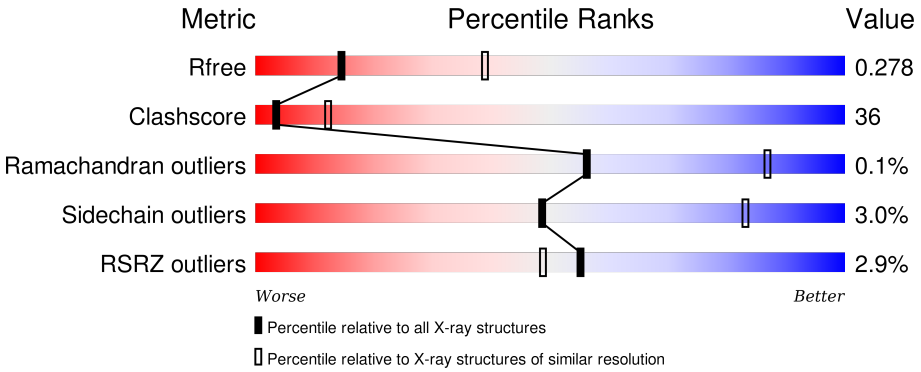
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1451 (2.90-2.90)
Clashscore	102246	1668 (2.90-2.90)
Ramachandran outliers	100387	1630 (2.90-2.90)
Sidechain outliers	100360	1632 (2.90-2.90)
RSRZ outliers	91569	1456 (2.90-2.90)

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 <=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	542	2% 48% 32% • 19%
1	B	542	3% 45% 33% • 19%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6887 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.

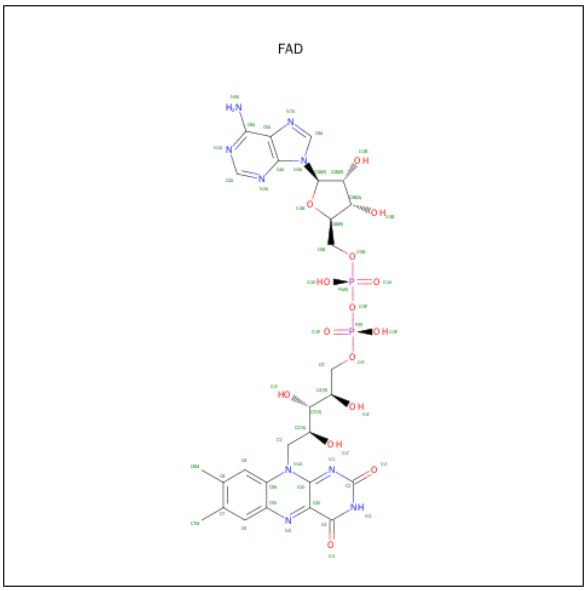
- Molecule 1 is a protein called THIOREDOXIN REDUCTASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	437	Total	C	N	O	S	0	1	0
			3389	2167	562	639	21			
1	B	437	Total	C	N	O	S	0	1	0
			3389	2167	562	639	21			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	EXPRESSION TAG	UNP Q25861
A	1	SER	-	EXPRESSION TAG	UNP Q25861
B	0	GLY	-	EXPRESSION TAG	UNP Q25861
B	1	SER	-	EXPRESSION TAG	UNP Q25861

- 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 water.

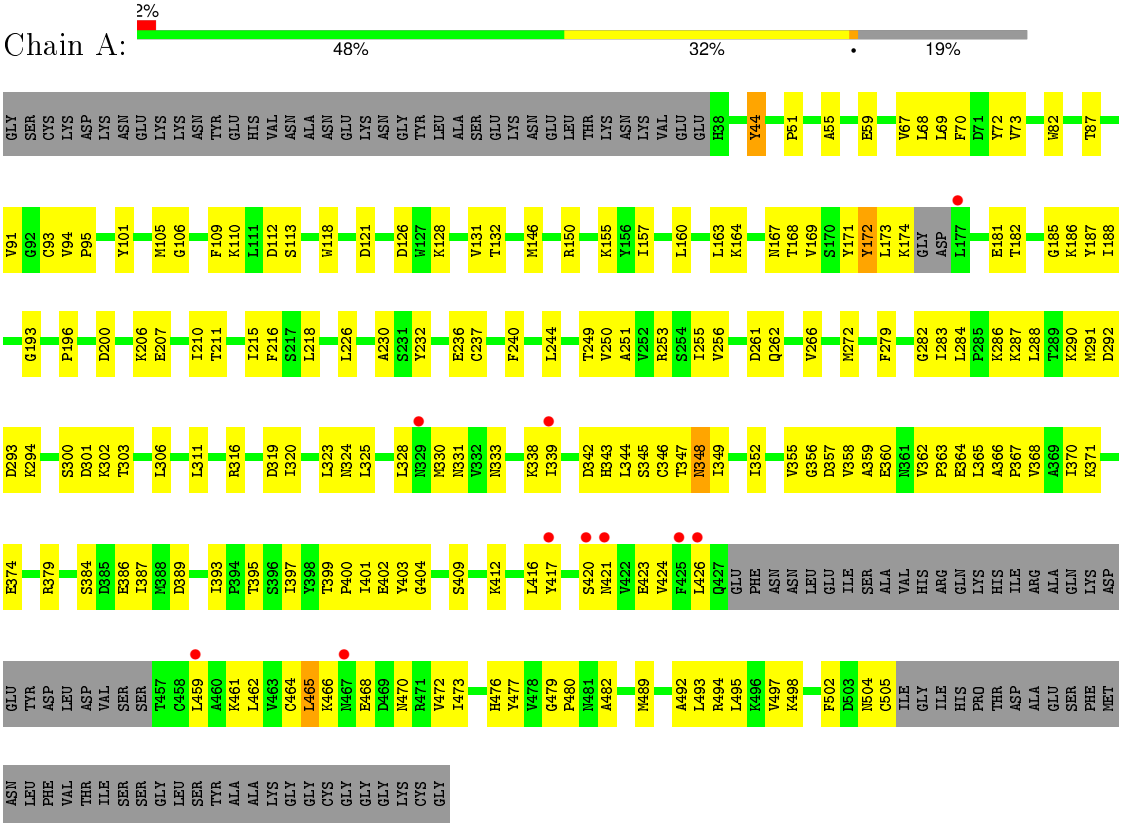
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	O	0	0
			2	2		
3	B	1	Total	O	0	0
			1	1		

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 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.

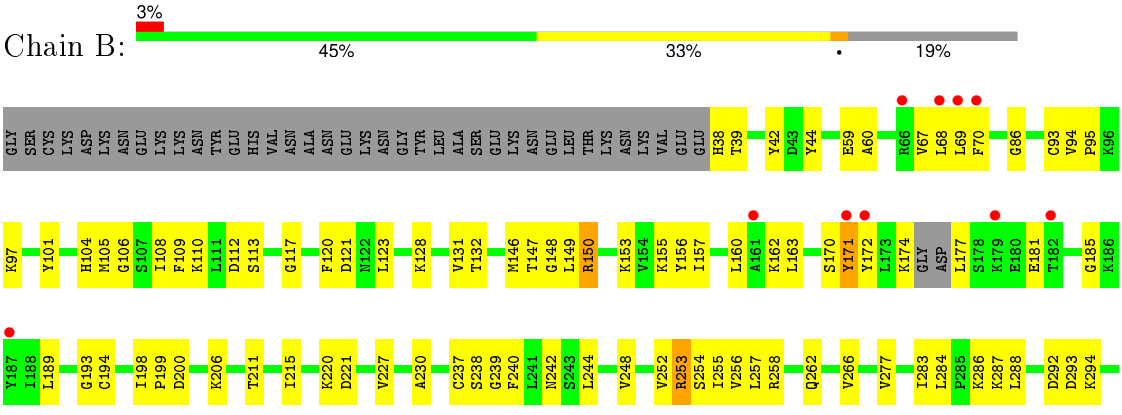
- Molecule 1: THIOREDOXIN REDUCTASE

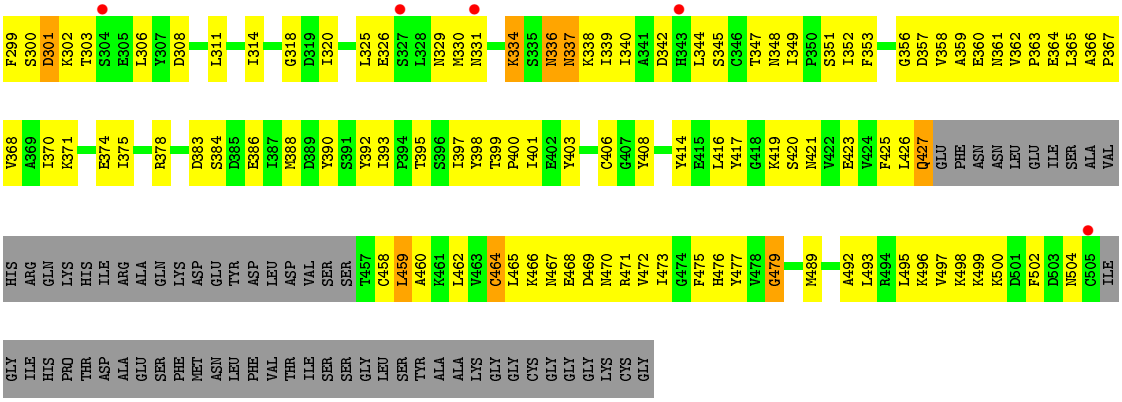
Chain A:



- Molecule 1: THIOREDOXIN REDUCTASE

Chain B:




F299 S300 D301 K302 T303 S304 E305 L306 Y307 D308 L311 I314 G318 D319 I320 L325 E326 S327 L328 N329 M330 N331 K334 S335 N336 N337 K338 I339 I340 A341 D342 H343 L344 S345 C346 T347 N348 I349 P350 S351 I352 F353 G356 D357 V358 A359 E360 N361 V362 P363 E364 L365 A366 P367
V368 A369 I370 K371 E374 I375 R378 D383 S384 D385 E386 I387 M388 D389 Y390 S391 Y392 I393 P394 T395 S396 I397 Y398 T399 P400 I401 E402 Y403 C406 G407 Y408 Y414 E415 L416 Y417 G418 K419 S420 N421 V422 E423 V424 F425 L426 Q427 GLU PHE ASN ASN LEU GLU ILE SER ALA VAL
HIS ARG GLN LYS HIS ILE ARG ALA GLN LYS ASP GLU TYR ASP LEU ASP VAL SER SER T457 C458 L459 A460 K461 L462 V463 C464 L465 K466 N467 E468 D469 N470 R471 V472 I473 G474 F475 H476 Y477 V478 G479 M489 A492 L493 R494 L495 K496 V497 K498 K499 K500 D501 F502 D503 N504 C505 ILE
GLY ILE HIS PRO THR ASP ALA GLU SER PHE MET ASN LEU PHE VAL THR ILE SER SER GLY LEU SER TYR ALA ALA LYS GLY GLY CYS GLY GLY GLY LYS CYS GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	63.19Å 109.24Å 182.41Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.91 – 2.90 46.91 – 2.90	Depositor EDS
% Data completeness (in resolution range)	96.7 (46.91-2.90) 96.7 (46.91-2.90)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.13 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.265 , 0.288 0.259 , 0.278	Depositor DCC
R_{free} test set	1394 reflections (5.27%)	DCC
Wilson B-factor (Å^2)	38.8	Xtriage
Anisotropy	0.481	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 29.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	1 of 27907 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	6887	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.35	0/3455	0.36	0/4659
1	B	0.38	0/3455	0.41	2/4659 (0.0%)
All	All	0.36	0/6910	0.39	2/9318 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	479	GLY	C-N-CD	5.80	140.57	128.40
1	B	458	CYS	O-C-N	5.26	131.12	122.70

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	3389	0	3427	246	1
1	B	3389	0	3427	279	1
2	A	53	0	31	4	0
2	B	53	0	31	2	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0
All	All	6887	0	6916	497	1

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 36.

The worst 5 of 497 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:489:MET:HE1	1:B:502:PHE:CZ	1.68	1.27
1:B:489:MET:CE	1:B:502:PHE:HZ	1.49	1.26
1:A:489:MET:CE	1:A:502:PHE:CZ	2.23	1.20
1:B:162:LYS:CE	1:B:172:TYR:HE1	1.56	1.17
1:B:492:ALA:HB1	1:B:497:VAL:HG21	1.22	1.16

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:343:HIS:CG	1:B:38:HIS:CD2[3_544]	2.13	0.07

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	432/542 (80%)	399 (92%)	33 (8%)	0	100	100
1	B	432/542 (80%)	402 (93%)	29 (7%)	1 (0%)	52	84
All	All	864/1084 (80%)	801 (93%)	62 (7%)	1 (0%)	56	87

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	147	THR

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	371/457 (81%)	361 (97%)	10 (3%)	52 84
1	B	371/457 (81%)	359 (97%)	12 (3%)	46 81
All	All	742/914 (81%)	720 (97%)	22 (3%)	48 83

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	505	CYS
1	B	171	TYR
1	B	459	LEU
1	B	93	CYS
1	B	150	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	487	GLN
1	B	158	ASN
1	B	427	GLN
1	A	476	HIS
1	B	348	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 (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

2 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	1506	-	48,58,58	1.19	6 (12%)	54,89,89	2.09	6 (11%)
2	FAD	B	1506	-	48,58,58	1.20	5 (10%)	54,89,89	2.08	7 (12%)

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	1506	-	-	0/30/50/50	0/6/6/6
2	FAD	B	1506	-	-	0/30/50/50	0/6/6/6

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1506	FAD	C5X-N5	2.01	1.38	1.35
2	A	1506	FAD	C5X-N5	2.02	1.38	1.35
2	A	1506	FAD	C1'-N10	2.03	1.50	1.48
2	B	1506	FAD	C2A-N1A	2.41	1.38	1.33
2	A	1506	FAD	C2A-N1A	2.42	1.38	1.33

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1506	FAD	N3A-C2A-N1A	-11.54	120.06	128.89
2	B	1506	FAD	N3A-C2A-N1A	-11.35	120.20	128.89

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1506	FAD	P-O3P-PA	-4.21	120.90	132.73
2	A	1506	FAD	P-O3P-PA	-4.11	121.19	132.73
2	B	1506	FAD	C4X-C4-N3	-2.75	119.83	123.59

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1506	FAD	4	0
2	B	1506	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

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^{th} 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	437/542 (80%)	0.14	10 (2%)	64	59	17, 27, 66, 77	2 (0%)
1	B	437/542 (80%)	0.22	15 (3%)	49	41	17, 34, 64, 86	6 (1%)
All	All	874/1084 (80%)	0.18	25 (2%)	55	49	17, 30, 65, 86	8 (0%)

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	177	LEU	4.8
1	A	426	LEU	4.2
1	B	172	TYR	4.2
1	A	459	LEU	3.8
1	A	467	ASN	3.7

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

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^{th} 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(Å^2)	Q<0.9
2	FAD	B	1506	53/53	0.95	0.18	-0.35	26,30,37,39	0
2	FAD	A	1506	53/53	0.96	0.17	-0.41	18,19,20,20	0

6.5 Other polymers (i)

There are no such residues in this entry.