



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 11:37 pm GMT

PDB ID : 2GSW
Title : Crystal Structure of the Putative NADPH-dependent Azobenzene FMN-Reductase YhdA from *Bacillus subtilis*, Northeast Structural Genomics Target SR135
Authors : Forouhar, F.; Hussain, M.; Jayaraman, S.; Shen, J.; Cooper, B.; Cunningham, K.; Janjua, H.; Ma, L.-C.; Xiao, R.; Acton, T.B.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on : 2006-04-26
Resolution : 2.92 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
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
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)	:	recalc28949

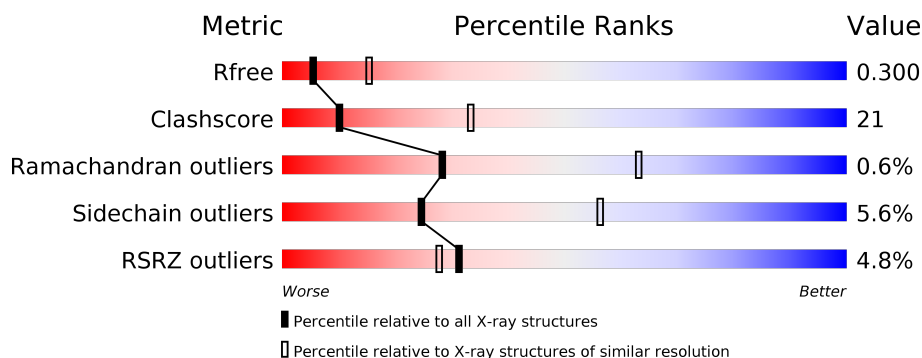
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.92 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	1813 (2.94-2.90)
Clashscore	112137	2045 (2.94-2.90)
Ramachandran outliers	110173	1997 (2.94-2.90)
Sidechain outliers	110143	1999 (2.94-2.90)
RSRZ outliers	101464	1825 (2.94-2.90)

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 $\leq 5\%$. 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	182	<div> <div>5%</div> <div> <div></div> <div>61%</div> <div>28%</div> <div>•</div> <div>8%</div> </div> </div>
1	B	182	<div> <div>5%</div> <div> <div></div> <div>62%</div> <div>27%</div> <div>•</div> <div>8%</div> </div> </div>
1	C	182	<div> <div>2%</div> <div> <div></div> <div>59%</div> <div>29%</div> <div>•</div> <div>8%</div> </div> </div>
1	D	182	<div> <div>5%</div> <div> <div></div> <div>61%</div> <div>27%</div> <div>•</div> <div>8%</div> </div> </div>

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 5307 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 yhdA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	168	Total	C	N	O	Se	0	0	0
			1285	819	220	241	5			
1	B	168	Total	C	N	O	Se	0	0	0
			1285	819	220	241	5			
1	C	168	Total	C	N	O	Se	0	0	0
			1285	819	220	241	5			
1	D	168	Total	C	N	O	Se	0	0	0
			1285	819	220	241	5			

There are 56 discrepancies between the modelled and reference sequences:

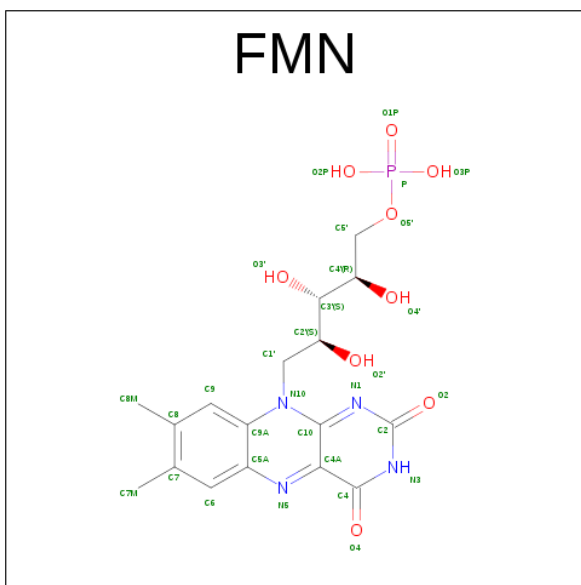
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	GB 2633257
A	3	MSE	MET	MODIFIED RESIDUE	GB 2633257
A	78	MSE	MET	MODIFIED RESIDUE	GB 2633257
A	118	MSE	MET	MODIFIED RESIDUE	GB 2633257
A	122	MSE	MET	MODIFIED RESIDUE	GB 2633257
A	165	MSE	MET	MODIFIED RESIDUE	GB 2633257
A	175	LEU	-	CLONING ARTIFACT	GB 2633257
A	176	GLU	-	CLONING ARTIFACT	GB 2633257
A	177	HIS	-	EXPRESSION TAG	GB 2633257
A	178	HIS	-	EXPRESSION TAG	GB 2633257
A	179	HIS	-	EXPRESSION TAG	GB 2633257
A	180	HIS	-	EXPRESSION TAG	GB 2633257
A	181	HIS	-	EXPRESSION TAG	GB 2633257
A	182	HIS	-	EXPRESSION TAG	GB 2633257
B	1	MSE	MET	MODIFIED RESIDUE	GB 2633257
B	3	MSE	MET	MODIFIED RESIDUE	GB 2633257
B	78	MSE	MET	MODIFIED RESIDUE	GB 2633257
B	118	MSE	MET	MODIFIED RESIDUE	GB 2633257
B	122	MSE	MET	MODIFIED RESIDUE	GB 2633257
B	165	MSE	MET	MODIFIED RESIDUE	GB 2633257
B	175	LEU	-	CLONING ARTIFACT	GB 2633257

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	176	GLU	-	CLONING ARTIFACT	GB 2633257
B	177	HIS	-	EXPRESSION TAG	GB 2633257
B	178	HIS	-	EXPRESSION TAG	GB 2633257
B	179	HIS	-	EXPRESSION TAG	GB 2633257
B	180	HIS	-	EXPRESSION TAG	GB 2633257
B	181	HIS	-	EXPRESSION TAG	GB 2633257
B	182	HIS	-	EXPRESSION TAG	GB 2633257
C	1	MSE	MET	MODIFIED RESIDUE	GB 2633257
C	3	MSE	MET	MODIFIED RESIDUE	GB 2633257
C	78	MSE	MET	MODIFIED RESIDUE	GB 2633257
C	118	MSE	MET	MODIFIED RESIDUE	GB 2633257
C	122	MSE	MET	MODIFIED RESIDUE	GB 2633257
C	165	MSE	MET	MODIFIED RESIDUE	GB 2633257
C	175	LEU	-	CLONING ARTIFACT	GB 2633257
C	176	GLU	-	CLONING ARTIFACT	GB 2633257
C	177	HIS	-	EXPRESSION TAG	GB 2633257
C	178	HIS	-	EXPRESSION TAG	GB 2633257
C	179	HIS	-	EXPRESSION TAG	GB 2633257
C	180	HIS	-	EXPRESSION TAG	GB 2633257
C	181	HIS	-	EXPRESSION TAG	GB 2633257
C	182	HIS	-	EXPRESSION TAG	GB 2633257
D	1	MSE	MET	MODIFIED RESIDUE	GB 2633257
D	3	MSE	MET	MODIFIED RESIDUE	GB 2633257
D	78	MSE	MET	MODIFIED RESIDUE	GB 2633257
D	118	MSE	MET	MODIFIED RESIDUE	GB 2633257
D	122	MSE	MET	MODIFIED RESIDUE	GB 2633257
D	165	MSE	MET	MODIFIED RESIDUE	GB 2633257
D	175	LEU	-	CLONING ARTIFACT	GB 2633257
D	176	GLU	-	CLONING ARTIFACT	GB 2633257
D	177	HIS	-	EXPRESSION TAG	GB 2633257
D	178	HIS	-	EXPRESSION TAG	GB 2633257
D	179	HIS	-	EXPRESSION TAG	GB 2633257
D	180	HIS	-	EXPRESSION TAG	GB 2633257
D	181	HIS	-	EXPRESSION TAG	GB 2633257
D	182	HIS	-	EXPRESSION TAG	GB 2633257

- Molecule 2 is FLAVIN MONONUCLEOTIDE (three-letter code: FMN) (formula: C₁₇H₂₁N₄O₉P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			31	17	4	9	1		
2	B	1	Total	C	N	O	P	0	0
			31	17	4	9	1		
2	C	1	Total	C	N	O	P	0	0
			31	17	4	9	1		
2	D	1	Total	C	N	O	P	0	0
			31	17	4	9	1		

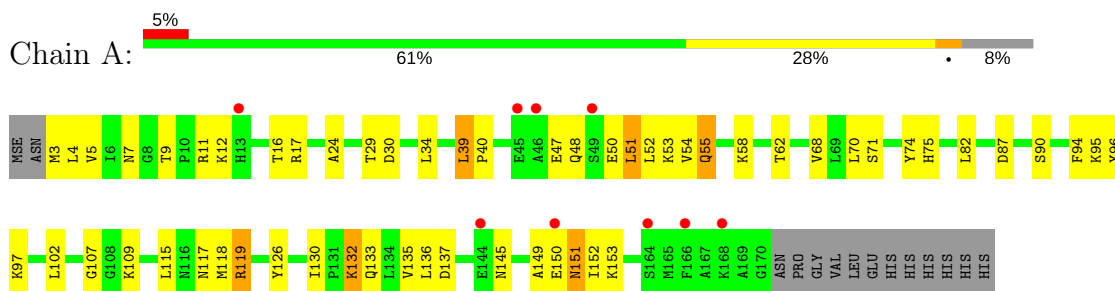
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	9	Total	O	0	0
			9	9		
3	B	5	Total	O	0	0
			5	5		
3	C	13	Total	O	0	0
			13	13		
3	D	16	Total	O	0	0
			16	16		

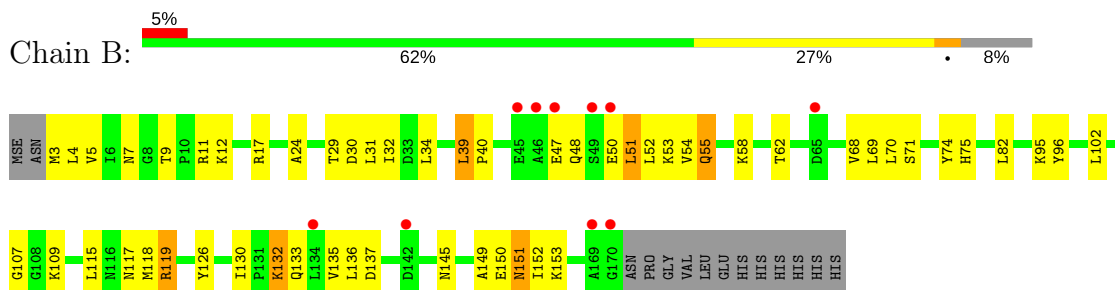
3 Residue-property plots [i](#)

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.

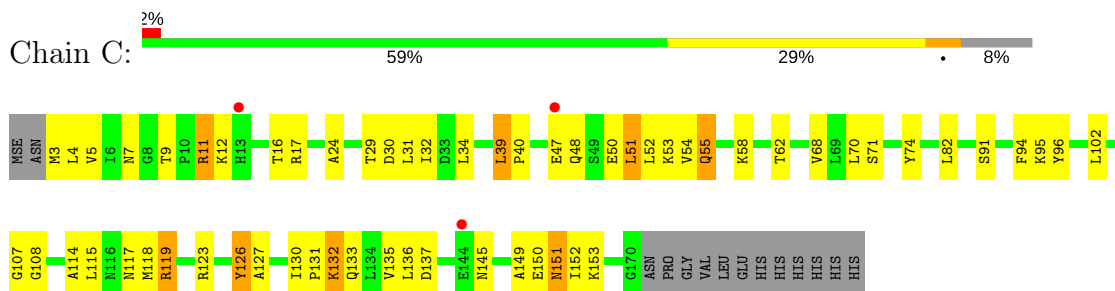
• Molecule 1: yhdA



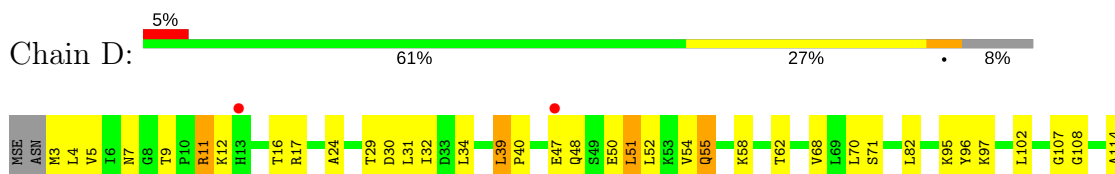
• Molecule 1: yhdA

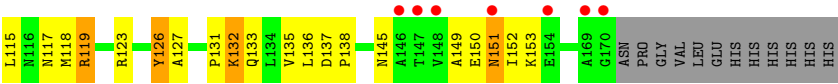


• Molecule 1: yhdA



• Molecule 1: yhdA





4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.44Å 99.30Å 69.44Å 90.00° 103.34° 90.00°	Depositor
Resolution (Å)	29.72 – 2.92 29.72 – 2.92	Depositor EDS
% Data completeness (in resolution range)	84.5 (29.72-2.92) 94.5 (29.72-2.92)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.43 (at 2.90Å)	Xtriage
Refinement program	XTALVIEW, CNS 1.1	Depositor
R, R_{free}	0.260 , 0.291 0.272 , 0.300	Depositor DCC
R_{free} test set	1354 reflections (10.46%)	DCC
Wilson B-factor (Å ²)	63.3	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 62.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	5307	wwPDB-VP
Average B, all atoms (Å ²)	59.0	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 26.59 % 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 2.5633e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.48	0/1300	0.77	5/1750 (0.3%)
1	B	0.45	0/1300	0.80	5/1750 (0.3%)
1	C	0.49	0/1300	0.79	5/1750 (0.3%)
1	D	0.45	0/1300	1.00	6/1750 (0.3%)
All	All	0.47	0/5200	0.85	21/7000 (0.3%)

There are no bond length outliers.

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	119	ARG	NE-CZ-NH1	-19.74	110.43	120.30
1	D	119	ARG	NE-CZ-NH2	18.41	129.50	120.30
1	B	11	ARG	NE-CZ-NH1	-13.95	113.33	120.30
1	D	11	ARG	NE-CZ-NH2	-13.42	113.59	120.30
1	A	11	ARG	NE-CZ-NH1	-13.05	113.77	120.30

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	1285	0	1326	54	0
1	B	1285	0	1326	54	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1285	0	1326	65	0
1	D	1285	0	1326	63	0
2	A	31	0	19	0	0
2	B	31	0	19	0	0
2	C	31	0	19	1	0
2	D	31	0	19	1	0
3	A	9	0	0	1	0
3	B	5	0	0	2	0
3	C	13	0	0	0	0
3	D	16	0	0	3	0
All	All	5307	0	5380	223	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 21.

The worst 5 of 223 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:3:MSE:HE2	1:A:29:THR:HG21	1.46	0.93
1:B:3:MSE:HE2	1:B:29:THR:HG21	1.49	0.92
1:C:3:MSE:HE2	1:C:29:THR:HG21	1.51	0.90
1:D:3:MSE:HE2	1:D:29:THR:HG21	1.53	0.90
1:A:48:GLN:HA	1:A:51:LEU:HD13	1.57	0.85

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	166/182 (91%)	143 (86%)	22 (13%)	1 (1%)	28	63
1	B	166/182 (91%)	142 (86%)	23 (14%)	1 (1%)	28	63

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	166/182 (91%)	142 (86%)	23 (14%)	1 (1%)	28	63
1	D	166/182 (91%)	140 (84%)	25 (15%)	1 (1%)	28	63
All	All	664/728 (91%)	567 (85%)	93 (14%)	4 (1%)	28	63

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	51	LEU
1	B	51	LEU
1	C	51	LEU
1	D	51	LEU

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	138/145 (95%)	130 (94%)	8 (6%)	23	55
1	B	138/145 (95%)	130 (94%)	8 (6%)	23	55
1	C	138/145 (95%)	130 (94%)	8 (6%)	23	55
1	D	138/145 (95%)	131 (95%)	7 (5%)	28	61
All	All	552/580 (95%)	521 (94%)	31 (6%)	25	56

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	145	ASN
1	C	53	LYS
1	D	132	LYS
1	B	151	ASN
1	C	55	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	151	ASN
1	C	13	HIS
1	D	55	GLN
1	B	84	ASN
1	B	133	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](#)

4 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	FMN	A	201	-	31,33,33	3.46	12 (38%)	38,50,50	3.20	11 (28%)
2	FMN	B	202	-	31,33,33	3.59	12 (38%)	38,50,50	3.13	10 (26%)
2	FMN	C	203	-	31,33,33	3.58	13 (41%)	38,50,50	3.09	10 (26%)
2	FMN	D	204	-	31,33,33	3.63	13 (41%)	38,50,50	3.06	11 (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
2	FMN	A	201	-	-	0/16/18/18	0/3/3/3
2	FMN	B	202	-	-	0/16/18/18	0/3/3/3
2	FMN	C	203	-	-	0/16/18/18	0/3/3/3
2	FMN	D	204	-	-	0/16/18/18	0/3/3/3

The worst 5 of 50 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	202	FMN	C2'-C3'	-2.96	1.47	1.53
2	C	203	FMN	C2'-C3'	-2.94	1.47	1.53
2	D	204	FMN	C2'-C3'	-2.65	1.48	1.53
2	A	201	FMN	C2'-C3'	-2.39	1.48	1.53
2	D	204	FMN	C6-C5A	2.44	1.45	1.41

The worst 5 of 42 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	202	FMN	C4A-C4-N3	-5.23	116.04	123.48
2	D	204	FMN	C4A-C4-N3	-5.10	116.23	123.48
2	A	201	FMN	C4A-C4-N3	-4.80	116.65	123.48
2	C	203	FMN	C4A-C4-N3	-4.79	116.66	123.48
2	D	204	FMN	C7M-C7-C6	-4.71	108.52	120.34

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	203	FMN	1	0
2	D	204	FMN	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, 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	163/182 (89%)	0.34	9 (5%) 26 21	25, 63, 77, 84	0
1	B	163/182 (89%)	0.41	10 (6%) 22 18	27, 64, 78, 84	0
1	C	163/182 (89%)	0.32	3 (1%) 69 67	24, 62, 77, 83	0
1	D	163/182 (89%)	0.38	9 (5%) 26 21	26, 63, 77, 83	0
All	All	652/728 (89%)	0.36	31 (4%) 31 28	24, 63, 77, 84	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	169	ALA	5.2
1	A	46	ALA	4.7
1	B	46	ALA	4.0
1	B	170	GLY	3.5
1	D	147	THR	3.1

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, 95th 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(\AA^2)	Q<0.9
2	FMN	A	201	31/31	0.78	0.30	1.50	78,84,88,89	0
2	FMN	C	203	31/31	0.77	0.29	1.47	76,89,91,92	0
2	FMN	B	202	31/31	0.84	0.26	1.06	87,96,97,97	0
2	FMN	D	204	31/31	0.83	0.26	0.90	88,91,93,94	0

6.5 Other polymers [i](#)

There are no such residues in this entry.