



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 15, 2016 – 05:31 PM EST

PDB ID : 5LQ4  
Title : The Structure of ThcOx, the First Oxidase Protein from the Cyanobactin Pathways  
Authors : Bent, A.F.; Wagner, A.; Naismith, J.H.  
Deposited on : 2016-08-16  
Resolution : 2.65 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

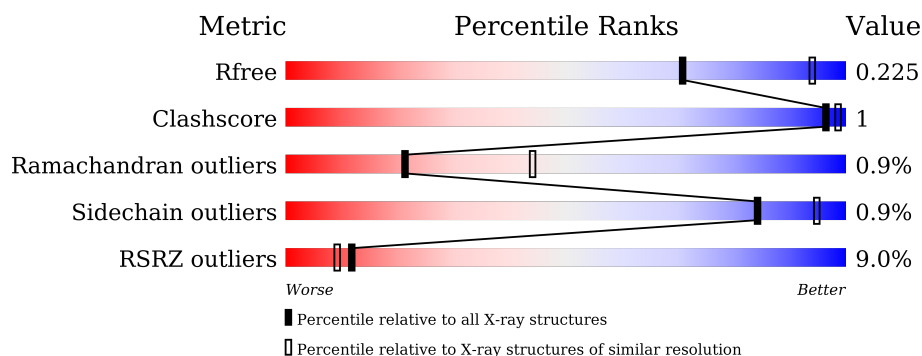
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.65 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	3152 (2.70-2.62)
Clashscore	102246	3524 (2.70-2.62)
Ramachandran outliers	100387	3469 (2.70-2.62)
Sidechain outliers	100360	3469 (2.70-2.62)
RSRZ outliers	91569	3161 (2.70-2.62)

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	467	<div> <div>6%</div> <div>94%</div> <div>...</div> </div>
2	B	473	<div> <div>11%</div> <div>93%</div> <div>..</div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 7355 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.

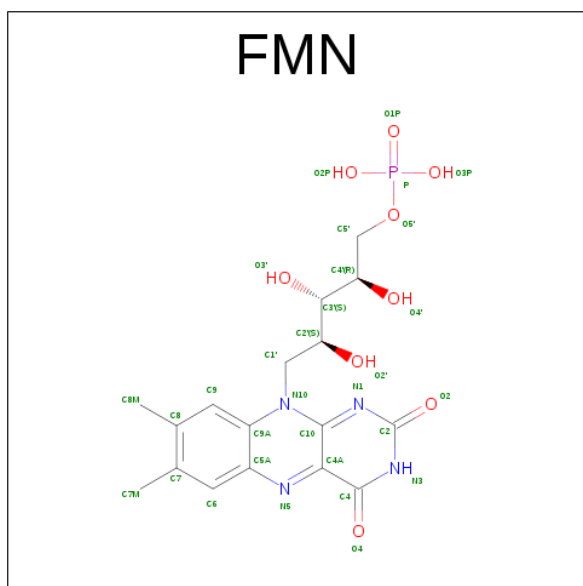
- Molecule 1 is a protein called CyaGox.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	457	Total	C	N	O	S	0	0	0
			3608	2325	594	680	9			

- Molecule 2 is a protein called CyaGox.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	465	Total	C	N	O	S	0	1	0
			3681	2369	608	694	10			

- Molecule 3 is FLAVIN MONONUCLEOTIDE (three-letter code: FMN) (formula: C<sub>17</sub>H<sub>21</sub>N<sub>4</sub>O<sub>9</sub>P).



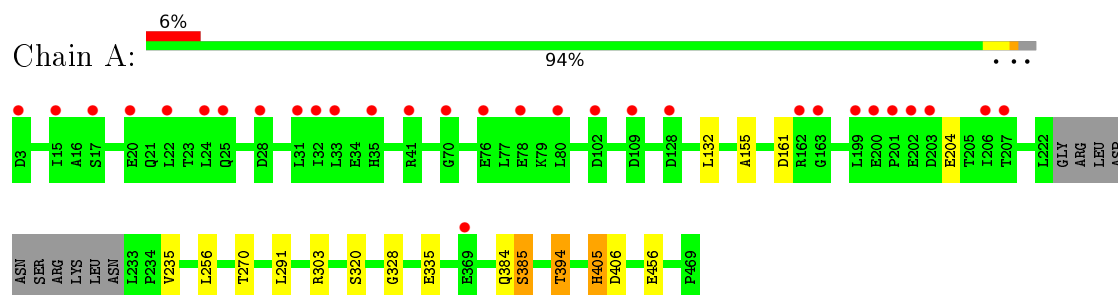
- Molecule 4 is water.

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

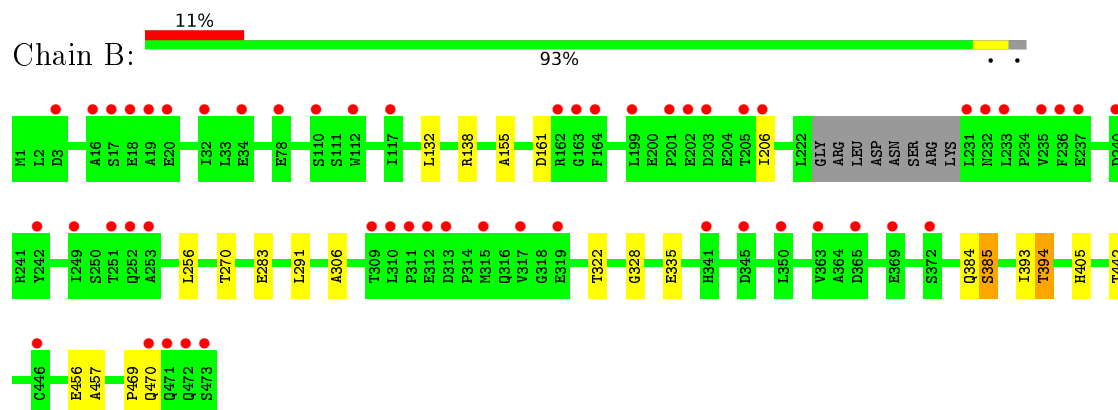
### 3 Residue-property plots

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

#### • Molecule 1: CyaGox



#### • Molecule 2: CyaGox



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	109.30 Å   109.30 Å   195.36 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	72.83 – 2.65 72.83 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.9 (72.83-2.65) 99.9 (72.83-2.65)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.59 (at 2.65 Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.197   ,   0.228 0.200   ,   0.225	Depositor DCC
$R_{free}$ test set	1752 reflections (5.25%)	DCC
Wilson B-factor (Å <sup>2</sup> )	62.6	Xtriage
Anisotropy	0.171	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 56.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7355	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% 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 [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	0.41	0/3693	0.65	1/5038 (0.0%)
2	B	0.41	0/3770	0.65	0/5140
All	All	0.41	0/7463	0.65	1/10178 (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	303	ARG	NE-CZ-NH2	-5.13	117.74	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	3608	0	3609	10	0
2	B	3681	0	3685	12	0
3	A	31	0	19	2	0
3	B	31	0	19	0	0
4	A	2	0	0	0	0
4	B	2	0	0	0	0
All	All	7355	0	7332	21	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 1.

All (21) 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:335:GLU:HB2	1:A:394:THR:HG22	1.78	0.66
2:B:335:GLU:HB2	2:B:394:THR:HG22	1.79	0.64
1:A:384:GLN:O	1:A:385:SER:CB	2.57	0.53
2:B:384:GLN:O	2:B:385:SER:CB	2.58	0.51
1:A:155:ALA:HA	2:B:132:LEU:HD22	1.91	0.51
1:A:204:GLU:N	1:A:204:GLU:OE2	2.44	0.51
2:B:442:THR:CG2	2:B:457:ALA:HB1	2.41	0.50
3:A:901:FMN:O2'	3:A:901:FMN:O4'	2.25	0.49
2:B:138:ARG:NH1	2:B:206:ILE:HG21	2.29	0.47
2:B:394:THR:HG21	2:B:456:GLU:OE2	2.14	0.47
2:B:442:THR:HG21	2:B:457:ALA:HB1	1.97	0.47
1:A:394:THR:HG21	1:A:456:GLU:OE2	2.14	0.46
1:A:270:THR:HA	2:B:270:THR:HA	1.99	0.45
2:B:256:LEU:HD23	2:B:291:LEU:HD21	1.99	0.44
1:A:256:LEU:HD23	1:A:291:LEU:HD21	1.99	0.43
2:B:306:ALA:HB3	2:B:322:THR:HG22	2.02	0.42
2:B:393:ILE:N	2:B:393:ILE:HD12	2.36	0.41
1:A:132:LEU:HD22	2:B:155:ALA:HA	2.02	0.41
3:A:901:FMN:HO2'	3:A:901:FMN:HO4'	1.64	0.40
1:A:335:GLU:HB2	1:A:394:THR:CG2	2.49	0.40
1:A:405:HIS:CG	1:A:406:ASP:H	2.39	0.40

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	453/467 (97%)	429 (95%)	21 (5%)	3 (1%)	26 51

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	462/473 (98%)	436 (94%)	21 (4%)	5 (1%)	17	38
All	All	915/940 (97%)	865 (94%)	42 (5%)	8 (1%)	21	44

All (8) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	385	SER
1	A	405	HIS
2	B	385	SER
2	B	405	HIS
2	B	469	PRO
2	B	470	GLN
1	A	328	GLY
2	B	328	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	393/402 (98%)	389 (99%)	4 (1%)	82	94
2	B	402/408 (98%)	399 (99%)	3 (1%)	88	96
All	All	795/810 (98%)	788 (99%)	7 (1%)	84	95

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	161	ASP
1	A	235	VAL
1	A	320	SER
1	A	394	THR
2	B	161	ASP
2	B	283	GLU
2	B	394	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	64	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](#)

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	$\# Z  > 2$
3	FMN	A	901	-	32,33,33	1.59	6 (18%)	34,50,50	2.14	7 (20%)
3	FMN	B	901	-	32,33,33	1.62	6 (18%)	34,50,50	2.24	9 (26%)

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	FMN	A	901	-	-	0/18/18/18	0/3/3/3

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	FMN	B	901	-	-	0/18/18/18	0/3/3/3

All (12) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	901	FMN	C9A-N10	2.73	1.42	1.38
3	A	901	FMN	C9A-N10	2.78	1.42	1.38
3	B	901	FMN	C10-N10	2.81	1.42	1.39
3	A	901	FMN	C10-N10	2.83	1.42	1.39
3	B	901	FMN	C8-C7	3.19	1.49	1.41
3	A	901	FMN	C8-C7	3.20	1.49	1.41
3	A	901	FMN	C9A-C5A	3.40	1.49	1.42
3	B	901	FMN	C9A-C5A	3.43	1.49	1.42
3	A	901	FMN	C4-C4A	3.79	1.49	1.41
3	B	901	FMN	C4-C4A	3.93	1.49	1.41
3	A	901	FMN	C4A-C10	4.16	1.48	1.40
3	B	901	FMN	C4A-C10	4.29	1.48	1.40

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	901	FMN	C4-C4A-C10	-4.71	116.92	119.94
3	A	901	FMN	C4-C4A-C10	-4.58	117.01	119.94
3	B	901	FMN	C4A-C4-N3	-3.40	119.08	123.52
3	A	901	FMN	C4A-C4-N3	-3.38	119.10	123.52
3	B	901	FMN	N3-C2-N1	-3.02	122.61	127.69
3	A	901	FMN	N3-C2-N1	-2.86	122.88	127.69
3	B	901	FMN	O4'-C4'-C5'	-2.59	104.46	110.09
3	B	901	FMN	O5'-P-O1P	-2.32	101.24	107.08
3	B	901	FMN	C1'-N10-C9A	2.13	121.30	118.83
3	A	901	FMN	C1'-N10-C9A	2.89	122.19	118.83
3	A	901	FMN	C4-C4A-N5	2.90	122.23	118.70
3	B	901	FMN	C4-C4A-N5	3.39	122.83	118.70
3	A	901	FMN	C4A-N5-C5A	4.26	121.74	116.72
3	B	901	FMN	C4A-N5-C5A	4.60	122.15	116.72
3	A	901	FMN	C4-N3-C2	7.71	121.59	115.16
3	B	901	FMN	C4-N3-C2	7.91	121.76	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	FMN	2	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

### 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	457/467 (97%)	0.52	30 (6%) 22 19	40, 70, 115, 141	0
2	B	465/473 (98%)	0.77	53 (11%) 7 5	41, 72, 125, 160	0
All	All	922/940 (98%)	0.65	83 (9%) 12 9	40, 71, 119, 160	0

All (83) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	201	PRO	9.5
2	B	345	ASP	9.3
2	B	162	ARG	6.5
2	B	19	ALA	6.0
1	A	20	GLU	6.0
2	B	312	GLU	5.9
2	B	232	ASN	5.0
2	B	231	LEU	5.0
2	B	203	ASP	4.8
2	B	201	PRO	4.8
1	A	35	HIS	4.5
2	B	470	GLN	4.4
1	A	202	GLU	4.3
1	A	199	LEU	4.2
2	B	249	ILE	4.1
2	B	34	GLU	3.9
1	A	17	SER	3.8
2	B	310	LEU	3.8
2	B	313	ASP	3.7
2	B	251	THR	3.7
1	A	24	LEU	3.6
1	A	22	LEU	3.5
1	A	33	LEU	3.5
2	B	471	GLN	3.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	233	LEU	3.4
1	A	31	LEU	3.3
2	B	365	ASP	3.3
1	A	203	ASP	3.2
1	A	78	GLU	3.2
2	B	17	SER	3.2
1	A	28	ASP	3.1
2	B	473	SER	3.1
2	B	163	GLY	3.1
2	B	117	ILE	2.9
2	B	446	CYS	2.9
2	B	199	LEU	2.9
1	A	15	ILE	2.9
2	B	18	GLU	2.8
2	B	311	PRO	2.8
2	B	16	ALA	2.8
1	A	200	GLU	2.7
1	A	206	ILE	2.7
2	B	319	GLU	2.7
2	B	252	GLN	2.7
2	B	202	GLU	2.7
2	B	363	VAL	2.7
2	B	253	ALA	2.6
2	B	237	GLU	2.6
2	B	240	ASP	2.6
2	B	164	PHE	2.5
1	A	207	THR	2.5
1	A	3	ASP	2.5
2	B	32	ILE	2.4
2	B	235	VAL	2.4
2	B	206	ILE	2.4
1	A	32	ILE	2.3
2	B	242	TYR	2.3
2	B	309	THR	2.3
1	A	163	GLY	2.3
1	A	80	LEU	2.3
2	B	317	VAL	2.3
2	B	110	SER	2.3
1	A	41	ARG	2.3
2	B	350	LEU	2.2
1	A	369	GLU	2.2
2	B	369	GLU	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	25	GLN	2.2
1	A	102	ASP	2.2
1	A	76	GLU	2.2
2	B	315	MET	2.2
2	B	20	GLU	2.1
2	B	472	GLN	2.1
1	A	109	ASP	2.1
2	B	205	THR	2.1
2	B	236	PHE	2.1
1	A	70	GLY	2.1
2	B	112	TRP	2.1
2	B	372	SER	2.1
2	B	78	GLU	2.1
2	B	341	HIS	2.0
1	A	128	ASP	2.0
2	B	3	ASP	2.0
1	A	162	ARG	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(Å <sup>2</sup> )	Q<0.9
3	FMN	B	901	31/31	0.95	0.20	0.64	38,44,52,57	0
3	FMN	A	901	31/31	0.96	0.15	-0.83	42,48,60,68	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.