



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 19, 2016 – 07:05 PM GMT

PDB ID : 4P5B
Title : Crystal structure of a UMP/dUMP methylase PolB from *Streptomyces cacaoi* bound with 5-Br dUMP
Authors : Li, Y.; Chen, W.; Li, J.; Xia, Z.; Deng, Z.; Zhou, J.
Deposited on : 2014-03-15
Resolution : 2.27 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 : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026982
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-20026982

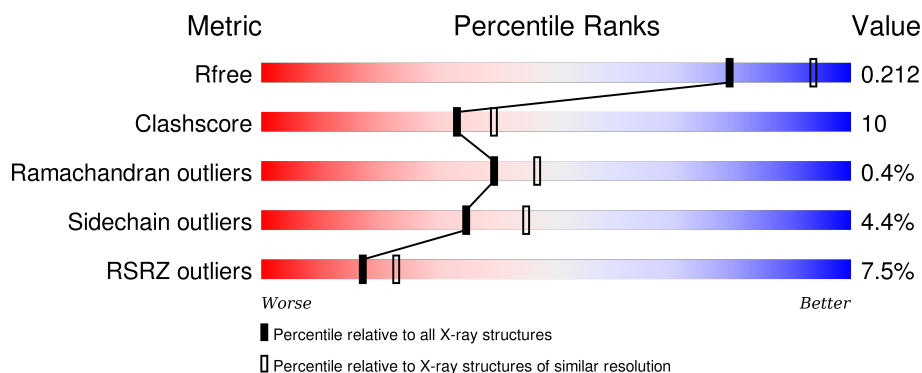
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.27 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	5193 (2.30-2.26)
Clashscore	102246	5929 (2.30-2.26)
Ramachandran outliers	100387	5851 (2.30-2.26)
Sidechain outliers	100360	5850 (2.30-2.26)
RSRZ outliers	91569	5204 (2.30-2.26)

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.

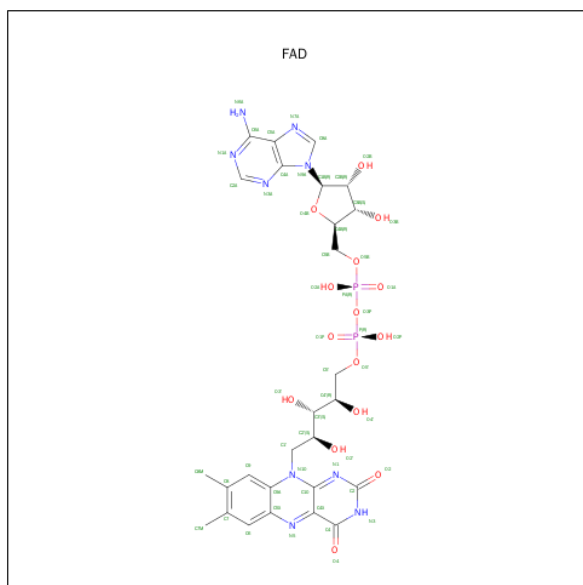
Mol	Chain	Length	Quality of chain
1	A	239	<div> <div>8%</div> <div>76%</div> <div>18%</div> <div>• •</div> </div>
1	B	239	<div> <div>9%</div> <div>76%</div> <div>18%</div> <div>5% •</div> </div>
1	C	239	<div> <div>8%</div> <div>75%</div> <div>21%</div> <div>• • •</div> </div>
1	D	239	<div> <div>5%</div> <div>79%</div> <div>19%</div> <div>•</div> </div>

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 Thymidylate synthase ThyX.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	229	Total 1813	C 1132	N 336	O 337	S 8	0	1	0
1	B	235	Total 1874	C 1169	N 346	O 351	S 8	0	1	0
1	C	234	Total 1854	C 1157	N 339	O 350	S 8	0	1	0
1	D	235	Total 1881	C 1175	N 346	O 352	S 8	0	1	0

- Molecule 2 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula: $\text{C}_{27}\text{H}_{33}\text{N}_9\text{O}_{15}\text{P}_2$).



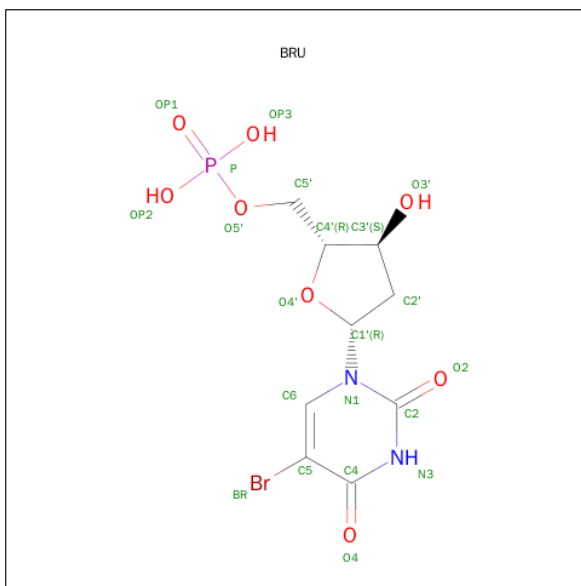
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total 53	C 27	N 9	O 15	P 2	0	0
2	B	1	Total 53	C 27	N 9	O 15	P 2	0	0

Continued on next page...

Continued from previous page...

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

- Molecule 3 is 5-BROMO-2'-DEOXYURIDINE-5'-MONOPHOSPHATE (three-letter code: BRU) (formula: $C_9H_{12}BrN_2O_8P$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
3	A	1	Total	Br	C	N	O	P	0	0
			21	1	9	2	8	1		
3	B	1	Total	Br	C	N	O	P	0	0
			21	1	9	2	8	1		
3	C	1	Total	Br	C	N	O	P	0	0
			21	1	9	2	8	1		
3	D	1	Total	Br	C	N	O	P	0	0
			21	1	9	2	8	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	S	0	0
			5	4	1		

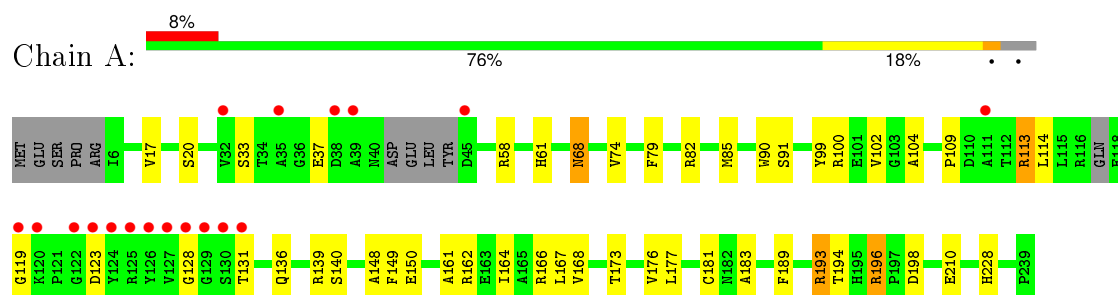
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	104	Total	O	0	0
			104	104		
5	B	97	Total	O	0	0
			97	97		
5	C	93	Total	O	0	0
			93	93		
5	D	108	Total	O	0	0
			108	108		

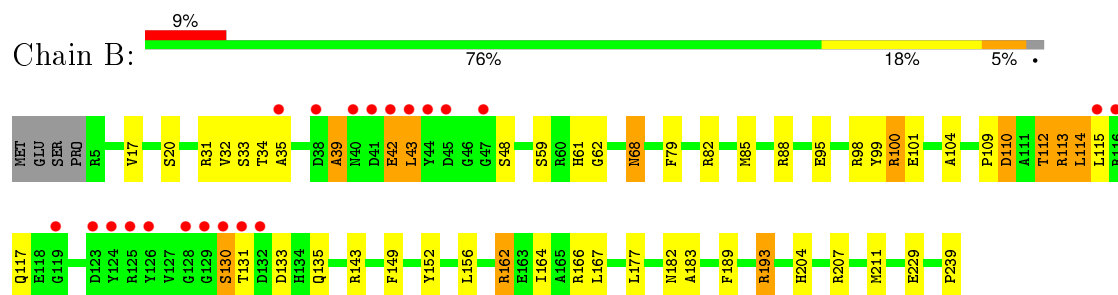
3 Residue-property plots

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.

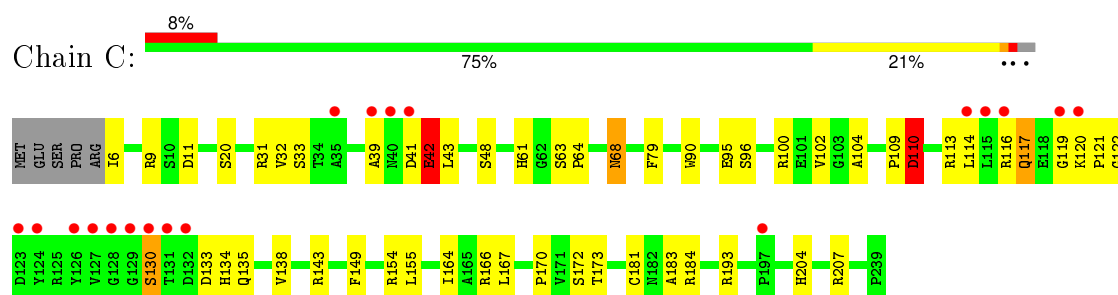
• Molecule 1: Thymidylate synthase ThyX



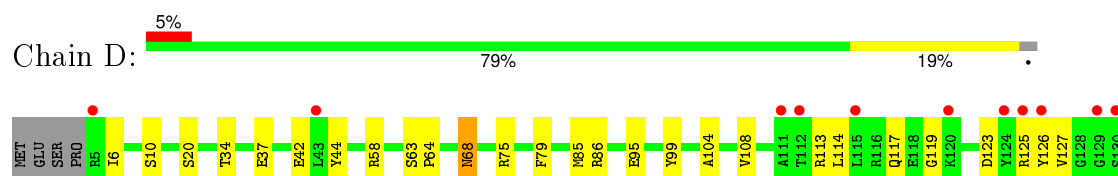
• Molecule 1: Thymidylate synthase ThyX



• Molecule 1: Thymidylate synthase ThyX



• Molecule 1: Thymidylate synthase ThyX





4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	170.79Å 59.50Å 118.98Å 90.00° 120.31° 90.00°	Depositor
Resolution (Å)	50.00 – 2.27 34.24 – 2.27	Depositor EDS
% Data completeness (in resolution range)	97.6 (50.00-2.27) 97.6 (34.24-2.27)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.00 (at 2.27Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.164 , 0.216 0.163 , 0.212	Depositor DCC
R_{free} test set	2332 reflections (5.26%)	DCC
Wilson B-factor (Å ²)	26.9	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 54.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 46653 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	8125	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.41% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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 ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: BRU, SO4, 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	1.04	1/1851 (0.1%)	0.96	4/2505 (0.2%)
1	B	1.02	0/1915	0.91	2/2594 (0.1%)
1	C	1.05	0/1895	0.91	5/2569 (0.2%)
1	D	1.01	1/1923 (0.1%)	0.91	0/2605
All	All	1.03	2/7584 (0.0%)	0.92	11/10273 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	229	GLU	CG-CD	5.35	1.59	1.51
1	A	91	SER	CB-OG	5.00	1.48	1.42

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	196	ARG	NE-CZ-NH1	-9.68	115.46	120.30
1	A	193	ARG	NE-CZ-NH2	-6.57	117.01	120.30
1	A	196	ARG	NE-CZ-NH2	6.31	123.45	120.30
1	C	207	ARG	NE-CZ-NH1	5.86	123.23	120.30
1	C	154	ARG	NE-CZ-NH2	-5.84	117.38	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	1813	0	1761	40	0
1	B	1874	0	1818	57	0
1	C	1854	0	1785	38	0
1	D	1881	0	1825	38	0
2	A	53	0	31	3	0
2	B	53	0	31	5	0
2	C	53	0	31	2	0
2	D	53	0	31	3	0
3	A	21	0	10	1	0
3	B	21	0	10	4	0
3	C	21	0	10	2	0
3	D	21	0	10	1	0
4	A	5	0	0	0	0
5	A	104	0	0	6	0
5	B	97	0	0	6	0
5	C	93	0	0	3	1
5	D	108	0	0	3	0
All	All	8125	0	7353	155	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 10.

The worst 5 of 155 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)
2:B:302:FAD:O1A	5:B:463:HOH:O	1.70	1.09
2:A:301:FAD:O1A	5:A:475:HOH:O	1.72	1.06
2:D:302:FAD:O1A	5:D:449:HOH:O	1.77	1.01
1:D:68:ASN:HD21	1:D:183:ALA:H	1.07	1.00
1:D:10:SER:HB3	1:D:108:VAL:O	1.61	0.99

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)
5:C:401:HOH:O	5:C:409:HOH:O[2_556]	2.04	0.16

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	224/239 (94%)	220 (98%)	4 (2%)	0	100	100
1	B	234/239 (98%)	225 (96%)	7 (3%)	2 (1%)	21	22
1	C	233/239 (98%)	222 (95%)	9 (4%)	2 (1%)	21	22
1	D	234/239 (98%)	228 (97%)	6 (3%)	0	100	100
All	All	925/956 (97%)	895 (97%)	26 (3%)	4 (0%)	39	47

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	42	GLU
1	C	117	GLN
1	B	39	ALA
1	C	42	GLU

5.3.2 Protein sidechains [i](#)

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	189/201 (94%)	182 (96%)	7 (4%)	41	54
1	B	196/201 (98%)	184 (94%)	12 (6%)	23	29
1	C	193/201 (96%)	182 (94%)	11 (6%)	25	32
1	D	197/201 (98%)	193 (98%)	4 (2%)	63	78
All	All	775/804 (96%)	741 (96%)	34 (4%)	35	45

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	114	LEU
1	C	6	ILE
1	D	113	ARG
1	B	135	GLN
1	A	166	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	195	HIS
1	C	68	ASN
1	D	68	ASN
1	B	117	GLN
1	D	93	ASN

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](#)

9 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	301	-	52,58,58	1.77	12 (23%)	52,89,89	2.41	24 (46%)
3	BRU	A	302	-	17,22,22	1.69	3 (17%)	22,33,33	2.71	9 (40%)
4	SO4	A	303	-	4,4,4	0.27	0	6,6,6	0.54	0
3	BRU	B	301	-	17,22,22	1.33	2 (11%)	22,33,33	2.46	8 (36%)
2	FAD	B	302	-	52,58,58	1.79	10 (19%)	52,89,89	2.37	18 (34%)
2	FAD	C	301	-	52,58,58	2.03	12 (23%)	52,89,89	2.19	13 (25%)
3	BRU	C	302	-	17,22,22	1.56	2 (11%)	22,33,33	2.70	6 (27%)
3	BRU	D	301	-	17,22,22	1.73	5 (29%)	22,33,33	2.75	6 (27%)
2	FAD	D	302	-	52,58,58	2.09	11 (21%)	52,89,89	2.20	13 (25%)

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	301	-	-	0/30/50/50	0/6/6/6
3	BRU	A	302	-	-	0/6/22/22	0/2/2/2
4	SO4	A	303	-	-	0/0/0/0	0/0/0/0
3	BRU	B	301	-	-	0/6/22/22	0/2/2/2
2	FAD	B	302	-	-	0/30/50/50	0/6/6/6
2	FAD	C	301	-	-	0/30/50/50	0/6/6/6
3	BRU	C	302	-	-	0/6/22/22	0/2/2/2
3	BRU	D	301	-	-	0/6/22/22	0/2/2/2
2	FAD	D	302	-	-	0/30/50/50	0/6/6/6

The worst 5 of 57 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	301	BRU	P-OP3	-2.82	1.45	1.54
2	A	301	FAD	C1'-N10	-2.24	1.46	1.48
3	A	302	BRU	P-OP2	-2.13	1.47	1.54
3	D	301	BRU	O4'-C1'	-2.13	1.37	1.42
2	A	301	FAD	C6-C5X	-2.07	1.38	1.41

The worst 5 of 97 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	301	FAD	N3A-C2A-N1A	-8.24	122.39	128.87
2	B	302	FAD	N3A-C2A-N1A	-7.88	122.68	128.87

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	301	BRU	O5'-P-OP1	-6.29	91.27	107.08
3	A	302	BRU	C5-C4-N3	-6.13	117.46	124.00
2	B	302	FAD	O5B-PA-O1A	-6.12	84.15	109.21

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

8 monomers are involved in 21 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	FAD	3	0
3	A	302	BRU	1	0
3	B	301	BRU	4	0
2	B	302	FAD	5	0
2	C	301	FAD	2	0
3	C	302	BRU	2	0
3	D	301	BRU	1	0
2	D	302	FAD	3	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	229/239 (95%)	0.06	18 (7%) 15 21	12, 24, 74, 95	0
1	B	235/239 (98%)	-0.04	21 (8%) 12 16	13, 26, 67, 77	0
1	C	234/239 (97%)	-0.08	19 (8%) 15 20	13, 26, 72, 83	0
1	D	235/239 (98%)	-0.21	12 (5%) 32 39	15, 26, 62, 76	0
All	All	933/956 (97%)	-0.07	70 (7%) 17 23	12, 25, 69, 95	0

The worst 5 of 70 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	129	GLY	10.8
1	A	124	TYR	9.6
1	A	128	GLY	8.4
1	B	43	LEU	7.8
1	A	127	VAL	6.9

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	FAD	D	302	53/53	0.89	0.17	1.23	32,49,54,56	0
2	FAD	B	302	53/53	0.91	0.16	0.76	24,43,50,56	0
2	FAD	C	301	53/53	0.89	0.16	0.60	29,42,51,52	0
3	BRU	C	302	21/21	0.96	0.14	0.13	19,25,31,39	0
3	BRU	A	302	21/21	0.97	0.13	-0.10	13,18,23,30	0
4	SO4	A	303	5/5	0.97	0.10	-0.34	54,55,56,58	0
2	FAD	A	301	53/53	0.94	0.11	-0.45	17,34,40,43	0
3	BRU	B	301	21/21	0.96	0.11	-0.56	12,17,21,29	0
3	BRU	D	301	21/21	0.98	0.10	-0.81	16,22,25,36	0

6.5 Other polymers [i](#)

There are no such residues in this entry.