W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 22, 2017 – 05:30 PM EST

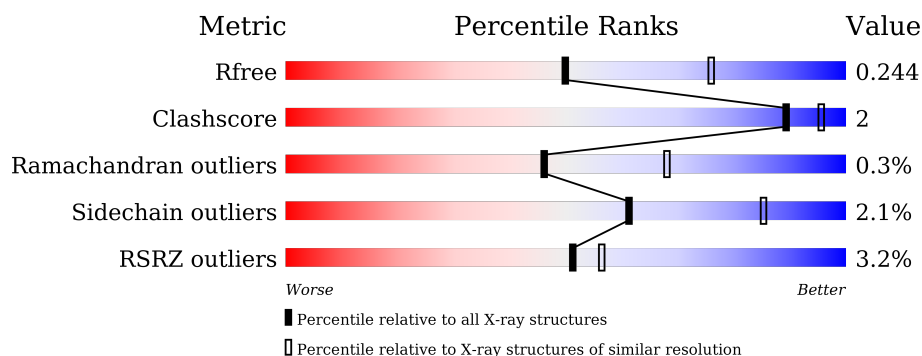|		
PDB ID	:	5KMS
Title	:	The structure of type II NADH dehydrogenase from Caldalkalibacillus thermarum complexed with NAD+ at 2.5 angstrom resolution.
Authors	:	Cook, G.M.; Aragao, D.; Nakatani, Y.
Deposited on	:	2016-06-27
Resolution	:	2.50 Å(reported)

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 2.50 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	Percentile Ranks	Value
Rfree		0.244
Clashscore		2
Ramachandran outliers		0.3%
Sidechain outliers		2.1%
RSRZ outliers		3.2%

Worse   *Better*

▮ Percentile relative to all X-ray structures

❘ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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	405	.% 90% 7% ·
1	B	405	.% 90% 7% ·
1	C	405	2% 92% 5% ·
1	D	405	8% 88% 9% ·

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 11969 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 FAD-dependent pyridine nucleotide-disulfide oxidoreductase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	394	Total 2897	C 1859	N 496	O 533	S 9	0	0	0
1	B	395	Total 2911	C 1865	N 497	O 540	S 9	0	0	0
1	C	394	Total 2885	C 1842	N 494	O 541	S 8	0	0	0
1	D	395	Total 2805	C 1801	N 477	O 519	S 8	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	400	HIS	-	expression tag	UNP F5L3B8
A	401	HIS	-	expression tag	UNP F5L3B8
A	402	HIS	-	expression tag	UNP F5L3B8
A	403	HIS	-	expression tag	UNP F5L3B8
A	404	HIS	-	expression tag	UNP F5L3B8
A	405	HIS	-	expression tag	UNP F5L3B8
B	400	HIS	-	expression tag	UNP F5L3B8
B	401	HIS	-	expression tag	UNP F5L3B8
B	402	HIS	-	expression tag	UNP F5L3B8
B	403	HIS	-	expression tag	UNP F5L3B8
B	404	HIS	-	expression tag	UNP F5L3B8
B	405	HIS	-	expression tag	UNP F5L3B8
C	400	HIS	-	expression tag	UNP F5L3B8
C	401	HIS	-	expression tag	UNP F5L3B8
C	402	HIS	-	expression tag	UNP F5L3B8
C	403	HIS	-	expression tag	UNP F5L3B8
C	404	HIS	-	expression tag	UNP F5L3B8
C	405	HIS	-	expression tag	UNP F5L3B8
D	400	HIS	-	expression tag	UNP F5L3B8
D	401	HIS	-	expression tag	UNP F5L3B8
D	402	HIS	-	expression tag	UNP F5L3B8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	403	HIS	-	expression tag	UNP F5L3B8
D	404	HIS	-	expression tag	UNP F5L3B8
D	405	HIS	-	expression tag	UNP F5L3B8

- Molecule 2 is FLAVIN-ADENINE DINUCLEOTIDE (three-letter code: FAD) (formula: $C_{27}H_{33}N_9O_{15}P_2$).
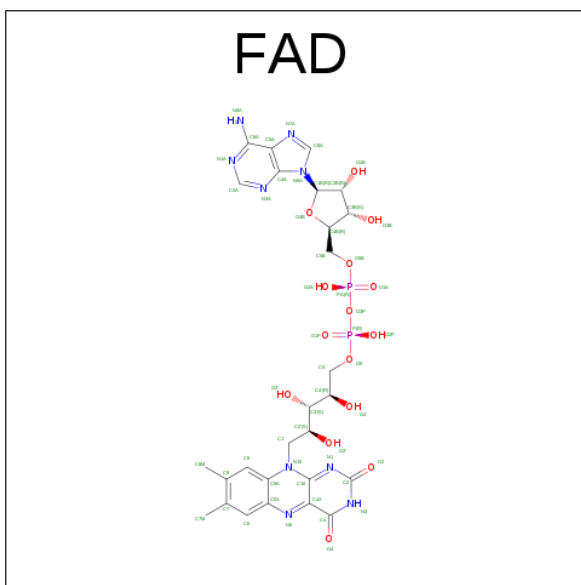
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	B	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	C	1	Total	C	N	O	P	0	0
			53	27	9	15	2		
2	D	1	Total	C	N	O	P	0	0
			53	27	9	15	2		

- Molecule 3 is NICOTINAMIDE-ADENINE-DINUCLEOTIDE (three-letter code: NAD) (formula: $C_{21}H_{27}N_7O_{14}P_2$).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total 44	C 21	N 7	O 14	P 2	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	60	Total 60	O 60	0	0
4	B	62	Total 62	O 62	0	0
4	C	62	Total 62	O 62	0	0
4	D	31	Total 31	O 31	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 (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.
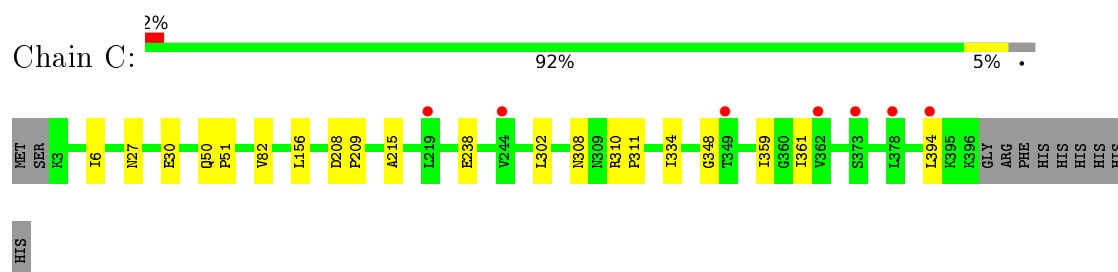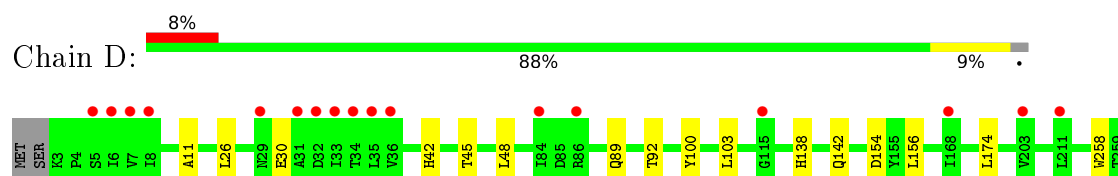
- Molecule 1: FAD-dependent pyridine nucleotide-disulfide oxidoreductase
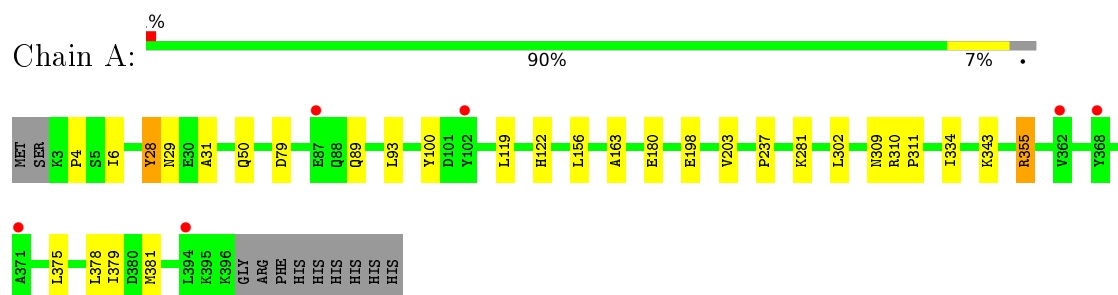
- Molecule 1: FAD-dependent pyridine nucleotide-disulfide oxidoreductase
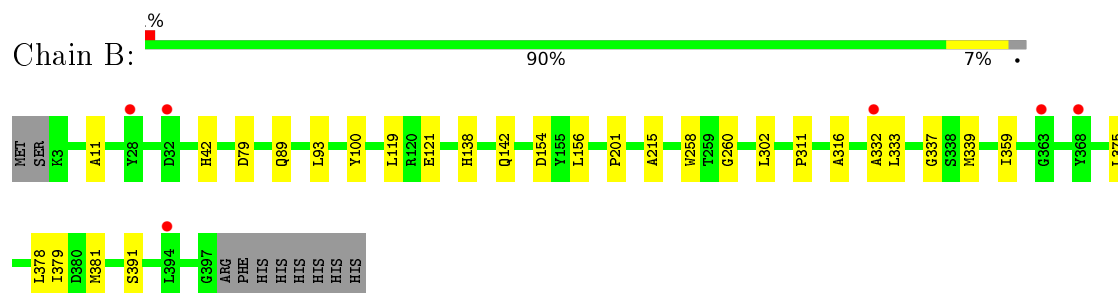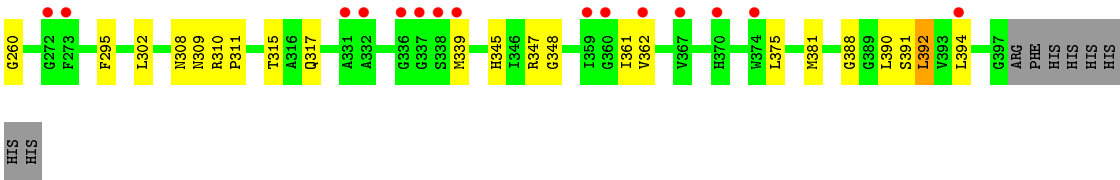

- Molecule 1: FAD-dependent pyridine nucleotide-disulfide oxidoreductase

- Molecule 1: FAD-dependent pyridine nucleotide-disulfide oxidoreductase

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	72.88Å 114.01Å 130.64Å 90.00° 91.77° 90.00°	Depositor
Resolution (Å)	47.89 – 2.50 47.89 – 2.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.89-2.50) 100.0 (47.89-2.50)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.91 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.214 , 0.244 0.216 , 0.244	Depositor DCC
R_{free} test set	3723 reflections (5.04%)	DCC
Wilson B-factor (Å^2)	51.3	Xtriage
Anisotropy	0.118	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 50.0	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.026 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	11969	wwPDB-VP
Average B, all atoms (Å^2)	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.43% 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, 0.333 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: FAD, NAD

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	0.22	0/2955	0.42	0/4029
1	B	0.22	0/2970	0.42	0/4050
1	C	0.22	0/2939	0.42	0/4006
1	D	0.22	0/2860	0.42	0/3911
All	All	0.22	0/11724	0.42	0/15996

There are no bond length outliers.

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	2897	0	2837	14	0
1	B	2911	0	2833	14	0
1	C	2885	0	2810	10	0
1	D	2805	0	2669	18	0
2	A	53	0	31	0	0
2	B	53	0	31	1	0
2	C	53	0	31	0	0
2	D	53	0	31	1	0
3	A	44	0	24	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	60	0	0	0	0
4	B	62	0	0	0	0
4	C	62	0	0	0	0
4	D	31	0	0	0	0
All	All	11969	0	11297	55	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 2.

The worst 5 of 55 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:D:89:GLN:NE2	1:D:100:TYR:O	2.22	0.71
1:A:6:ILE:HD11	1:A:334:ILE:HD11	1.77	0.66
1:C:302:LEU:HD11	1:C:311:PRO:HB3	1.79	0.63
1:C:6:ILE:HD11	1:C:334:ILE:HD11	1.85	0.59
1:D:317:GLN:OE1	1:D:347:ARG:N	2.37	0.58

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	392/405 (97%)	383 (98%)	6 (2%)	3 (1%)	24	41
1	B	393/405 (97%)	385 (98%)	8 (2%)	0	100	100
1	C	392/405 (97%)	385 (98%)	7 (2%)	0	100	100
1	D	393/405 (97%)	385 (98%)	7 (2%)	1 (0%)	46	68
All	All	1570/1620 (97%)	1538 (98%)	28 (2%)	4 (0%)	46	68

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	28	TYR
1	A	29	ASN
1	D	309	ASN
1	A	309	ASN

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	286/329 (87%)	279 (98%)	7 (2%)	57	82
1	B	287/329 (87%)	282 (98%)	5 (2%)	68	89
1	C	285/329 (87%)	282 (99%)	3 (1%)	80	94
1	D	261/329 (79%)	252 (97%)	9 (3%)	44	72
All	All	1119/1316 (85%)	1095 (98%)	24 (2%)	61	85

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	381	MET
1	C	238	GLU
1	D	391	SER
1	B	391	SER
1	C	156	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	89	GLN
1	A	196	ASN
1	C	39	ASN
1	D	89	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)

5 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	601	-	52,58,58	4.07	19 (36%)	52,89,89	2.69	11 (21%)
3	NAD	A	602	-	42,48,48	3.52	14 (33%)	46,73,73	1.66	5 (10%)
2	FAD	B	601	-	52,58,58	4.05	19 (36%)	52,89,89	2.74	11 (21%)
2	FAD	C	601	-	52,58,58	4.06	19 (36%)	52,89,89	2.76	12 (23%)
2	FAD	D	601	-	52,58,58	4.07	19 (36%)	52,89,89	2.70	11 (21%)

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	601	-	-	0/30/50/50	0/6/6/6
3	NAD	A	602	-	-	0/22/62/62	0/5/5/5
2	FAD	B	601	-	-	0/30/50/50	0/6/6/6
2	FAD	C	601	-	-	0/30/50/50	0/6/6/6

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	FAD	D	601	-	-	0/30/50/50	0/6/6/6

The worst 5 of 90 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	601	FAD	C2B-C1B	-15.48	1.29	1.53
2	C	601	FAD	C2B-C1B	-15.45	1.29	1.53
2	B	601	FAD	C2B-C1B	-15.43	1.29	1.53
2	A	601	FAD	C2B-C1B	-15.22	1.29	1.53
3	A	602	NAD	C2D-C3D	-10.84	1.24	1.53

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	601	FAD	N3A-C2A-N1A	-11.51	119.83	128.87
2	B	601	FAD	N3A-C2A-N1A	-11.33	119.97	128.87
2	D	601	FAD	N3A-C2A-N1A	-11.25	120.03	128.87
2	A	601	FAD	N3A-C2A-N1A	-11.24	120.05	128.87
3	A	602	NAD	N3A-C2A-N1A	-8.79	121.97	128.87

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	602	NAD	1	0
2	B	601	FAD	1	0
2	D	601	FAD	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	394/405 (97%)	0.12	6 (1%)	76	79	31, 57, 83, 98	0
1	B	395/405 (97%)	0.09	6 (1%)	76	79	29, 54, 80, 102	0
1	C	394/405 (97%)	0.04	7 (1%)	71	75	31, 54, 80, 112	0
1	D	395/405 (97%)	0.47	32 (8%)	15	16	37, 71, 98, 108	0
All	All	1578/1620 (97%)	0.18	51 (3%)	51	56	29, 59, 89, 112	0

The worst 5 of 51 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	363	GLY	6.3
1	D	331	ALA	6.1
1	D	332	ALA	4.9
1	D	336	GLY	4.8
1	D	272	GLY	4.6

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(Å^2)	Q<0.9
2	FAD	B	601	53/53	0.97	0.18	0.73	25,40,50,55	0
3	NAD	A	602	44/44	0.93	0.16	0.25	50,64,72,79	0
2	FAD	A	601	53/53	0.97	0.16	0.15	33,45,53,57	0
2	FAD	C	601	53/53	0.98	0.15	-0.25	31,41,55,64	0
2	FAD	D	601	53/53	0.95	0.15	-0.38	45,55,69,71	0

6.5 Other polymers (i)

There are no such residues in this entry.