wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:56 PM GMT

PDB ID	:	4K03
Title	:	Crystal structure of Drosophila Cryprochrome
Authors	:	Berndt, A.; Wolf, E.
Deposited on	:	2013-04-03
Resolution	:	3.20 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:

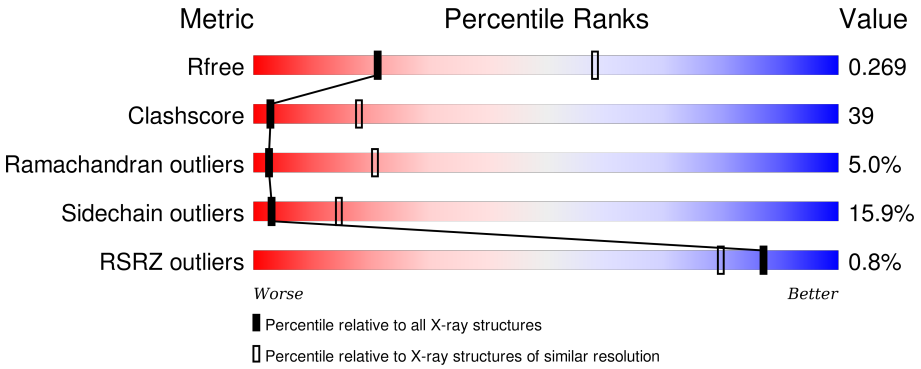
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 3.20 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	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	561	% 36% 46% 12% • 6%
1	B	561	% 37% 48% 11% • •

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 8746 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 Cryptochrome-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	529	Total	C	N	O	S	0	0	0
			4238	2707	749	758	24			
1	B	543	Total	C	N	O	S	0	0	0
			4350	2775	769	781	25			

There are 38 discrepancies between the modelled and reference sequences:

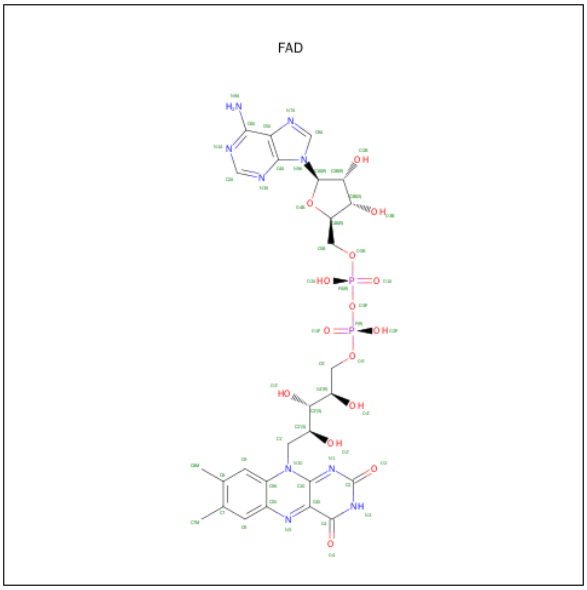
Chain	Residue	Modelled	Actual	Comment	Reference
A	-18	GLY	-	EXPRESSION TAG	UNP O77059
A	-17	ALA	-	EXPRESSION TAG	UNP O77059
A	-16	MET	-	EXPRESSION TAG	UNP O77059
A	-15	GLY	-	EXPRESSION TAG	UNP O77059
A	-14	SER	-	EXPRESSION TAG	UNP O77059
A	-13	GLY	-	EXPRESSION TAG	UNP O77059
A	-12	ILE	-	EXPRESSION TAG	UNP O77059
A	-11	GLN	-	EXPRESSION TAG	UNP O77059
A	-10	ARG	-	EXPRESSION TAG	UNP O77059
A	-9	PRO	-	EXPRESSION TAG	UNP O77059
A	-8	THR	-	EXPRESSION TAG	UNP O77059
A	-7	SER	-	EXPRESSION TAG	UNP O77059
A	-6	THR	-	EXPRESSION TAG	UNP O77059
A	-5	SER	-	EXPRESSION TAG	UNP O77059
A	-4	SER	-	EXPRESSION TAG	UNP O77059
A	-3	LEU	-	EXPRESSION TAG	UNP O77059
A	-2	VAL	-	EXPRESSION TAG	UNP O77059
A	-1	ALA	-	EXPRESSION TAG	UNP O77059
A	0	ALA	-	EXPRESSION TAG	UNP O77059
B	-18	GLY	-	EXPRESSION TAG	UNP O77059
B	-17	ALA	-	EXPRESSION TAG	UNP O77059
B	-16	MET	-	EXPRESSION TAG	UNP O77059
B	-15	GLY	-	EXPRESSION TAG	UNP O77059
B	-14	SER	-	EXPRESSION TAG	UNP O77059
B	-13	GLY	-	EXPRESSION TAG	UNP O77059

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-12	ILE	-	EXPRESSION TAG	UNP O77059
B	-11	GLN	-	EXPRESSION TAG	UNP O77059
B	-10	ARG	-	EXPRESSION TAG	UNP O77059
B	-9	PRO	-	EXPRESSION TAG	UNP O77059
B	-8	THR	-	EXPRESSION TAG	UNP O77059
B	-7	SER	-	EXPRESSION TAG	UNP O77059
B	-6	THR	-	EXPRESSION TAG	UNP O77059
B	-5	SER	-	EXPRESSION TAG	UNP O77059
B	-4	SER	-	EXPRESSION TAG	UNP O77059
B	-3	LEU	-	EXPRESSION TAG	UNP O77059
B	-2	VAL	-	EXPRESSION TAG	UNP O77059
B	-1	ALA	-	EXPRESSION TAG	UNP O77059
B	0	ALA	-	EXPRESSION TAG	UNP O77059

- 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.

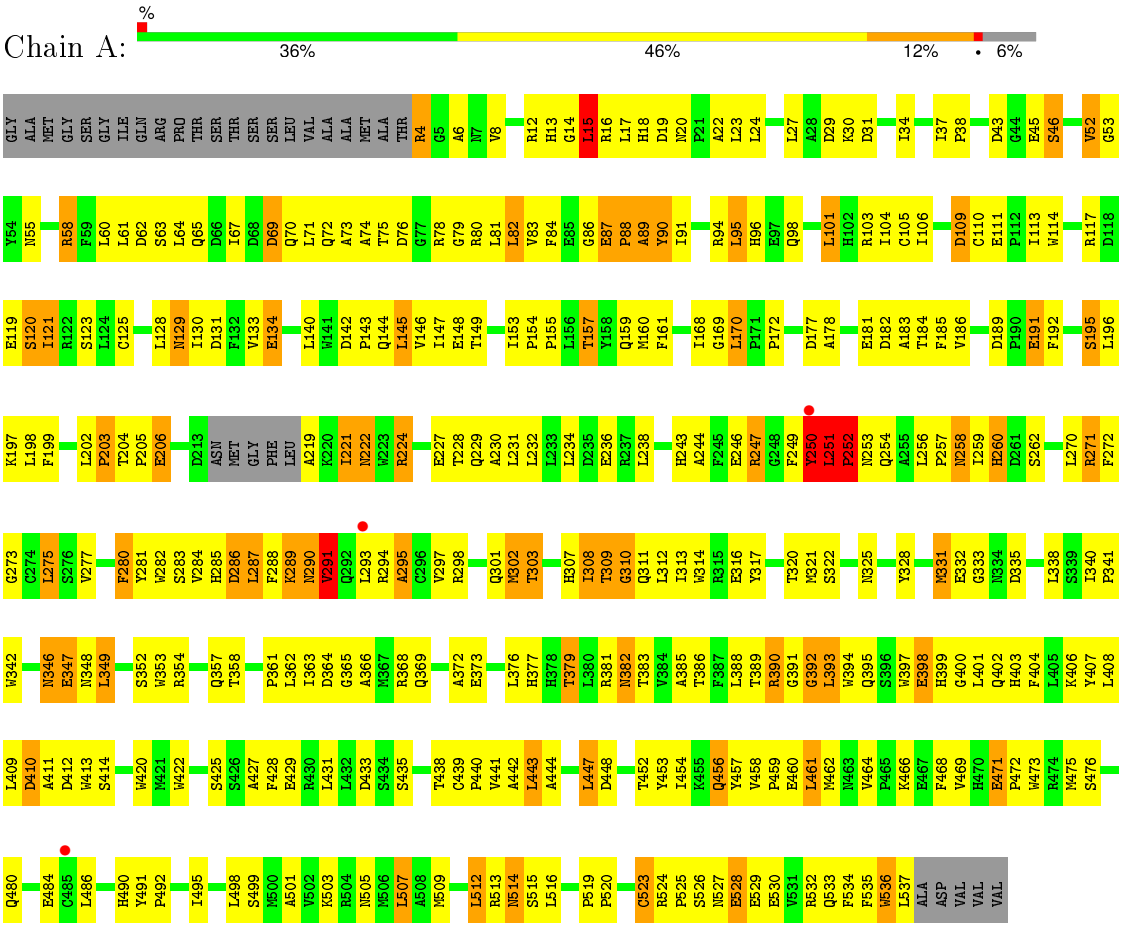
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	27	Total O 27 27	0	0
3	B	25	Total O 25 25	0	0

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.

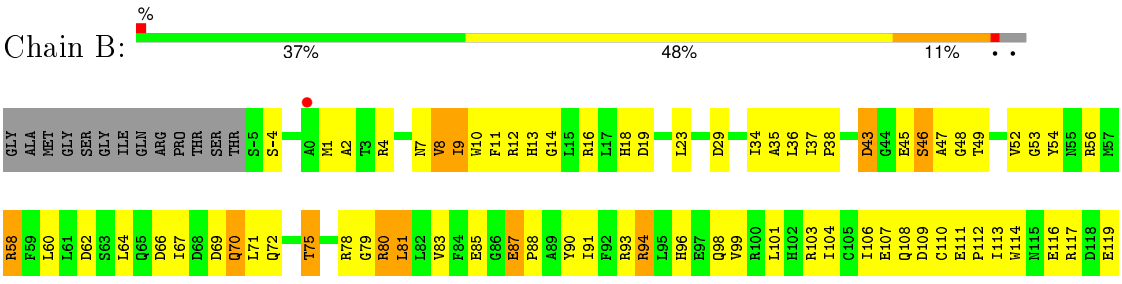
- Molecule 1: Cryptochrome-1

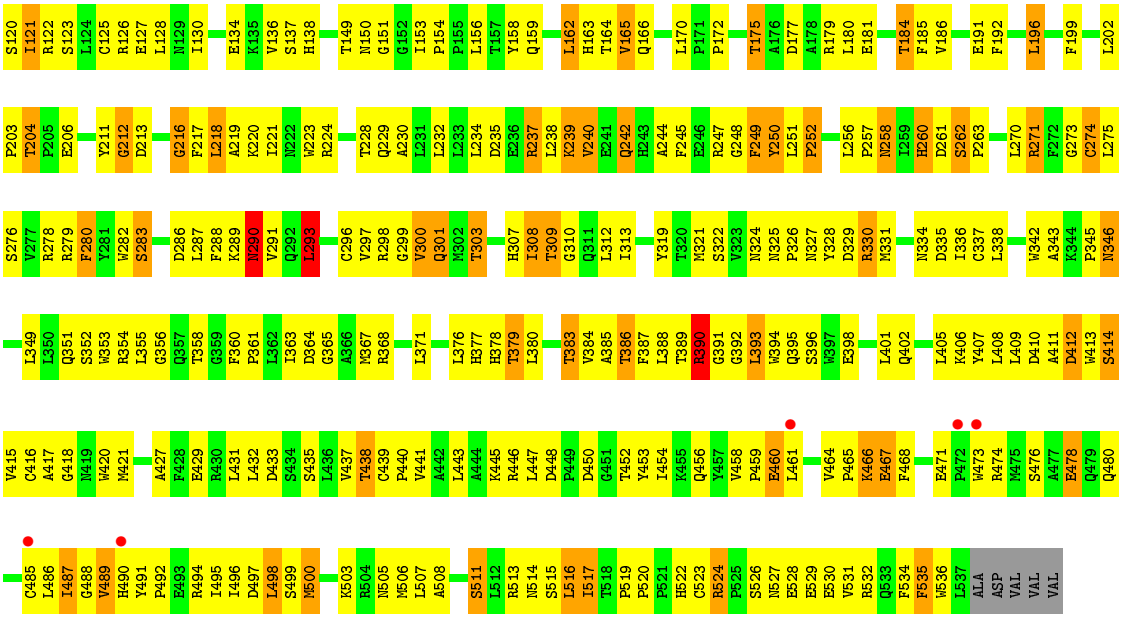
Chain A:



- Molecule 1: Cryptochrome-1

Chain B:




S120 I121 R122 S123 L124 C125 R126 E127 L128 N129 I130 E134 K135 V136 S137 H138 T149 N150 G151 G152 I153 P154 P155 L156 T157 Y158 Q159 L162 H163 T164 V165 Q166 L170 P171 P172 T175 A176 D177 A178 R179 L180 E181 T184 F185 V186 E191 F192 L196 F199 L202
P203 T204 P205 E206 Y211 G212 D213 G216 F217 L218 A219 K220 I221 N222 W223 R224 T228 Q229 A230 L231 L232 L233 L234 D235 E236 R237 L238 K239 V240 E241 Q242 H243 A244 F245 E246 R247 G248 F249 Y250 L251 P252 L256 P257 N258 I259 H260 D261 S262 P263 L270 R271 F272 G273 C274 L275
S276 V277 R278 R279 F280 Y281 W282 S283 D286 L287 F288 K289 N290 V291 Q292 L293 C296 V297 R298 G299 V300 Q301 M302 T303 H307 I308 T309 G310 Q311 L312 I313 Y319 T320 M321 S322 V323 N324 N325 P326 N327 Y328 D329 R330 M331 N334 D335 I336 C337 L338 W342 A343 K344 P345 N346
L349 I350 Q351 S352 W353 R354 L355 G356 Q357 T358 G359 F360 P361 I362 I363 D364 G365 A366 M367 R368 L371 L376 H377 H378 T379 L380 T383 V384 A385 T386 F387 I388 T389 R390 G391 G392 I393 W394 Q395 S396 W397 E398 L401 Q402 L405 K406 Y407 I408 L409 D410 A411 D412 W413 S414
V415 C416 A417 G418 N419 W420 M421 A427 F428 E429 R430 L431 L432 D433 S434 S435 L436 V437 T438 C439 P440 V441 A442 L443 A444 K445 R446 L447 D448 P449 D450 G451 T452 Y453 I454 K455 Q456 Y457 V458 P459 E460 L461 V464 P465 K466 E467 F468 E471 P472 W473 R474 M475 S476 A477 E478 Q479 Q480
C485 I486 I487 G488 W489 H490 Y491 P492 E493 R494 I495 I496 D497 I498 S499 N500 K503 R504 N505 M506 L507 A508 S511 L512 R513 N514 S515 L516 I517 T518 P519 P520 P521 H522 C523 R524 P525 S526 N527 E528 E529 E530 V531 R532 Q533 F534 F535 W536 L537 ALA ASP VAL VAL VAL

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	71.12Å 121.81Å 79.72Å 90.00° 114.78° 90.00°	Depositor
Resolution (Å)	40.49 – 3.20 46.60 – 3.20	Depositor EDS
% Data completeness (in resolution range)	99.7 (40.49-3.20) 99.7 (46.60-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 3.19Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.4_486)	Depositor
R, R_{free}	0.188 , 0.269 0.188 , 0.269	Depositor DCC
R_{free} test set	1041 reflections (5.39%)	DCC
Wilson B-factor (Å^2)	58.7	Xtriage
Anisotropy	0.178	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 85.4	EDS
Estimated twinning fraction	0.035 for h,-k,-h-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Outliers	0 of 20371 reflections	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	8746	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.70% 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	0.52	0/4355	0.71	3/5929 (0.1%)
1	B	0.50	0/4469	0.72	1/6082 (0.0%)
All	All	0.51	0/8824	0.72	4/12011 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	251	LEU	CA-CB-CG	8.68	135.26	115.30
1	A	393	LEU	N-CA-C	6.16	127.62	111.00
1	A	291	VAL	N-CA-C	-5.91	95.03	111.00
1	B	232	LEU	CA-CB-CG	5.01	126.83	115.30

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	4238	0	4076	343	0
1	B	4350	0	4190	342	0
2	A	53	0	31	3	0
2	B	53	0	31	1	0
3	A	27	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	25	0	0	1	0
All	All	8746	0	8328	670	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 39.

The worst 5 of 670 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:297:VAL:HG13	1:A:298:ARG:H	0.97	1.09
1:A:297:VAL:HG13	1:A:298:ARG:N	1.79	0.98
1:A:192:PHE:CZ	1:A:196:LEU:HD22	2.04	0.93
1:A:297:VAL:CG1	1:A:298:ARG:H	1.81	0.92
1:A:17:LEU:HD12	1:A:70:GLN:OE1	1.70	0.91

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	525/561 (94%)	409 (78%)	90 (17%)	26 (5%)	3 21
1	B	541/561 (96%)	444 (82%)	70 (13%)	27 (5%)	3 21
All	All	1066/1122 (95%)	853 (80%)	160 (15%)	53 (5%)	3 21

5 of 53 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	15	LEU
1	A	221	ILE
1	A	252	PRO

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	254	GLN
1	A	291	VAL

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	447/488 (92%)	377 (84%)	70 (16%)	3	15
1	B	458/488 (94%)	384 (84%)	74 (16%)	3	14
All	All	905/976 (93%)	761 (84%)	144 (16%)	3	14

5 of 144 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	512	LEU
1	B	83	VAL
1	B	460	GLU
1	A	523	CYS
1	B	46	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	505	ASN
1	A	533	GLN
1	B	301	GLN
1	A	403	HIS
1	B	96	HIS

5.3.3 RNA

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	$\#\|Z\| > 2$
2	FAD	A	601	-	48,58,58	1.08	5 (10%)	54,89,89	2.63	10 (18%)
2	FAD	B	601	-	48,58,58	1.06	5 (10%)	54,89,89	2.62	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	601	-	-	0/30/50/50	0/6/6/6
2	FAD	B	601	-	-	0/30/50/50	0/6/6/6

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	FAD	C4X-N5	-2.42	1.29	1.33
2	A	601	FAD	C10-N10	-2.24	1.36	1.39
2	A	601	FAD	C10-N1	-2.17	1.32	1.35
2	B	601	FAD	C10-N10	-2.11	1.36	1.39
2	B	601	FAD	C10-N1	-2.02	1.32	1.35

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	FAD	C4X-C4-N3	-6.85	114.22	123.59
2	A	601	FAD	C4X-C4-N3	-6.57	114.60	123.59
2	B	601	FAD	C4X-C10-N10	-6.36	116.77	120.52
2	A	601	FAD	C4X-C10-N10	-6.25	116.84	120.52
2	B	601	FAD	C4-C4X-C10	-3.88	117.46	119.94

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	FAD	3	0
2	B	601	FAD	1	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	529/561 (94%)	-0.29	3 (0%)	90	84	23, 52, 87, 135	0
1	B	543/561 (96%)	-0.19	6 (1%)	82	72	17, 54, 92, 161	0
All	All	1072/1122 (95%)	-0.24	9 (0%)	87	80	17, 54, 90, 161	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	485	CYS	3.0
1	B	0	ALA	2.9
1	B	490	HIS	2.7
1	B	461	LEU	2.7
1	B	473	TRP	2.3

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	A	601	53/53	0.96	0.19	0.09	37,37,60,60	0
2	FAD	B	601	53/53	0.96	0.17	-0.28	39,39,63,63	0

6.5 Other polymers (i)

There are no such residues in this entry.