



wwPDB X-ray Structure Validation Summary Report ⓘ

May 24, 2020 – 04:51 am BST

PDB ID : 2YFI
Title : Crystal Structure of Biphenyl dioxygenase variant RR41 (BPDO-RR41)
Authors : Kumar, P.; Bolin, J.T.
Deposited on : 2011-04-06
Resolution : 2.15 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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.11

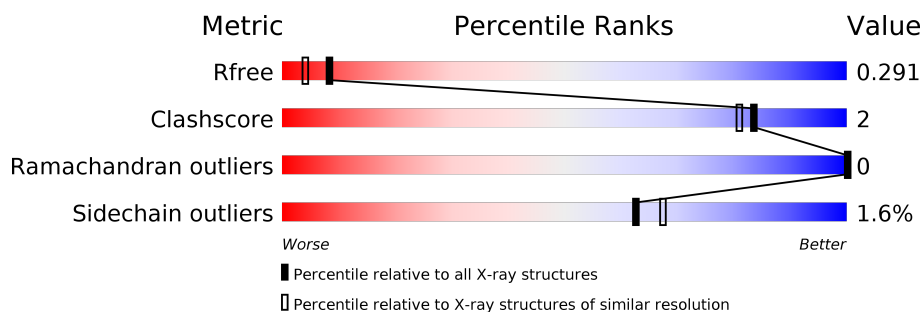
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION







The reported resolution of this entry is 2.15 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}	130704	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)

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 respectively. 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	459	 86% 7% 6%
1	C	459	 87% 6% • 6%
1	E	459	 88% 5% 6%
1	G	459	 87% 7% • 6%
1	I	459	 88% 6% 6%
1	K	459	 88% 6% 6%
2	B	188	 88% 7% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	188	<div><div></div><div>86%</div><div>10%<div><div></div><div></div><div></div></div></div></div>
2	F	188	<div><div></div><div>88%</div><div>7%<div><div></div><div></div><div></div></div></div></div>
2	H	188	<div><div></div><div>88%</div><div>8%<div><div></div><div></div><div></div></div></div></div>
2	J	188	<div><div></div><div>90%</div><div>6%<div><div></div><div></div><div></div></div></div></div>
2	L	188	<div><div></div><div>88%</div><div>7%<div><div></div><div></div><div></div></div></div></div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 30799 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 BIPHENYL DIOXYGENASE SUBUNIT ALPHA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	433	Total	C	N	O	S	0	0	0
			3430	2182	602	622	24			
1	C	433	Total	C	N	O	S	0	0	0
			3430	2182	602	622	24			
1	E	433	Total	C	N	O	S	0	0	0
			3430	2182	602	622	24			
1	G	433	Total	C	N	O	S	0	0	0
			3430	2182	602	622	24			
1	I	433	Total	C	N	O	S	0	0	0
			3430	2182	602	622	24			
1	K	433	Total	C	N	O	S	0	0	0
			3430	2182	602	622	24			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	335	ALA	THR	engineered mutation	UNP P37333
A	336	MET	PHE	engineered mutation	UNP P37333
A	338	GLN	ASN	engineered mutation	UNP P37333
A	341	VAL	ILE	engineered mutation	UNP P37333
A	409	PHE	LEU	engineered mutation	UNP P37333
C	335	ALA	THR	engineered mutation	UNP P37333
C	336	MET	PHE	engineered mutation	UNP P37333
C	338	GLN	ASN	engineered mutation	UNP P37333
C	341	VAL	ILE	engineered mutation	UNP P37333
C	409	PHE	LEU	engineered mutation	UNP P37333
E	335	ALA	THR	engineered mutation	UNP P37333
E	336	MET	PHE	engineered mutation	UNP P37333
E	338	GLN	ASN	engineered mutation	UNP P37333
E	341	VAL	ILE	engineered mutation	UNP P37333
E	409	PHE	LEU	engineered mutation	UNP P37333
G	335	ALA	THR	engineered mutation	UNP P37333
G	336	MET	PHE	engineered mutation	UNP P37333

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	338	GLN	ASN	engineered mutation	UNP P37333
G	341	VAL	ILE	engineered mutation	UNP P37333
G	409	PHE	LEU	engineered mutation	UNP P37333
I	335	ALA	THR	engineered mutation	UNP P37333
I	336	MET	PHE	engineered mutation	UNP P37333
I	338	GLN	ASN	engineered mutation	UNP P37333
I	341	VAL	ILE	engineered mutation	UNP P37333
I	409	PHE	LEU	engineered mutation	UNP P37333
K	335	ALA	THR	engineered mutation	UNP P37333
K	336	MET	PHE	engineered mutation	UNP P37333
K	338	GLN	ASN	engineered mutation	UNP P37333
K	341	VAL	ILE	engineered mutation	UNP P37333
K	409	PHE	LEU	engineered mutation	UNP P37333

- Molecule 2 is a protein called BIPHENYL DIOXYGENASE SUBUNIT BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	180	Total	C	N	O	S	0	0	0
			1496	948	265	279	4			
2	D	180	Total	C	N	O	S	0	0	0
			1496	948	265	279	4			
2	F	180	Total	C	N	O	S	0	0	0
			1496	948	265	279	4			
2	H	180	Total	C	N	O	S	0	0	0
			1496	948	265	279	4			
2	J	180	Total	C	N	O	S	0	0	0
			1496	948	265	279	4			
2	L	180	Total	C	N	O	S	0	0	0
			1496	948	265	279	4			

- Molecule 3 is FE2/S2 (INORGANIC) CLUSTER (three-letter code: FES) (formula: Fe₂S₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	Fe	S	0	0
			4	2	2		
3	C	1	Total	Fe	S	0	0
			4	2	2		
3	E	1	Total	Fe	S	0	0
			4	2	2		
3	G	1	Total	Fe	S	0	0
			4	2	2		
3	I	1	Total	Fe	S	0	0
			4	2	2		
3	K	1	Total	Fe	S	0	0
			4	2	2		

- Molecule 4 is FE (II) ION (three-letter code: FE2) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	G	1	Total	Fe	0	0
			1	1		
4	K	1	Total	Fe	0	0
			1	1		
4	E	1	Total	Fe	0	0
			1	1		
4	I	1	Total	Fe	0	0
			1	1		
4	C	1	Total	Fe	0	0
			1	1		
4	A	1	Total	Fe	0	0
			1	1		

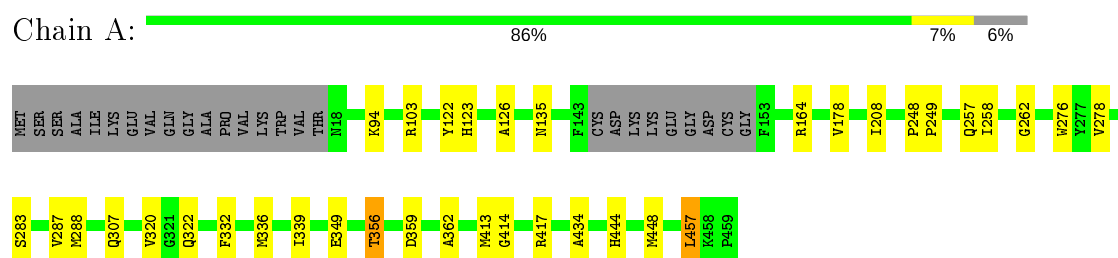
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	210	Total 210	O 210	0	0
5	B	94	Total 94	O 94	0	0
5	C	212	Total 212	O 212	0	0
5	D	99	Total 99	O 99	0	0
5	E	152	Total 152	O 152	0	0
5	F	94	Total 94	O 94	0	0
5	G	73	Total 73	O 73	0	0
5	H	79	Total 79	O 79	0	0
5	I	47	Total 47	O 47	0	0
5	J	46	Total 46	O 46	0	0
5	K	53	Total 53	O 53	0	0
5	L	54	Total 54	O 54	0	0

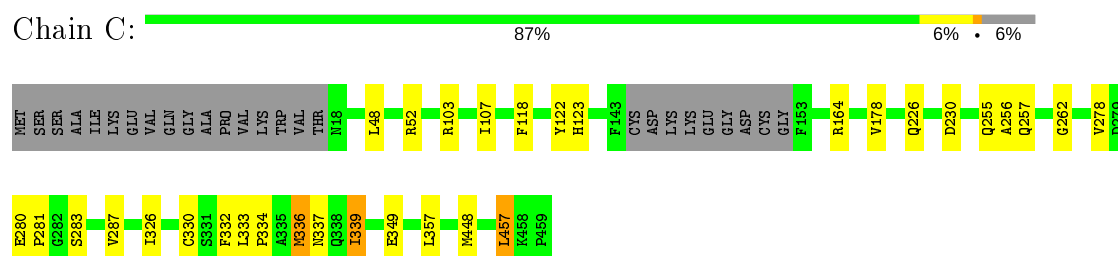
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.

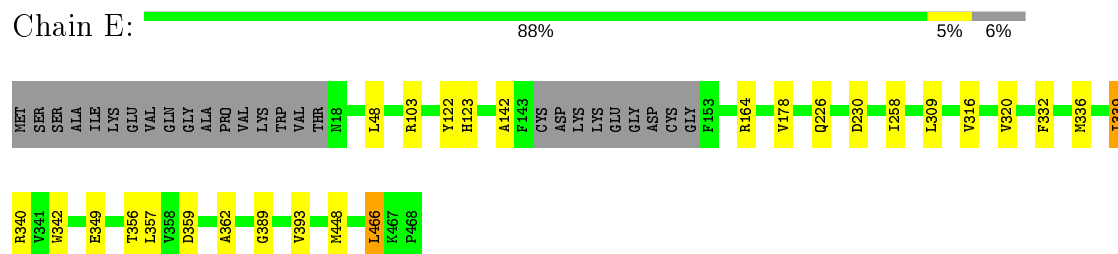
- Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA



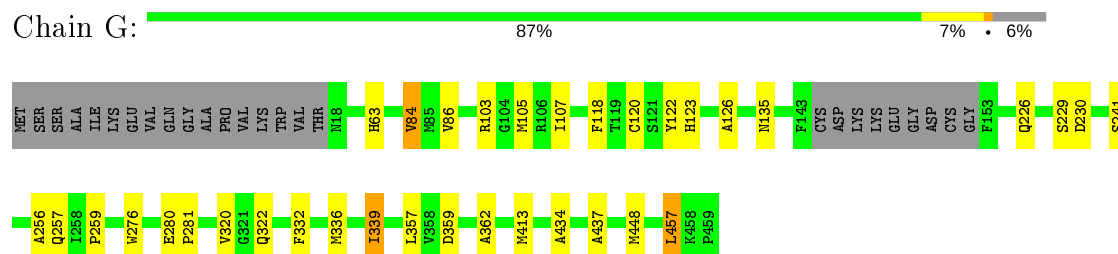
- Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA



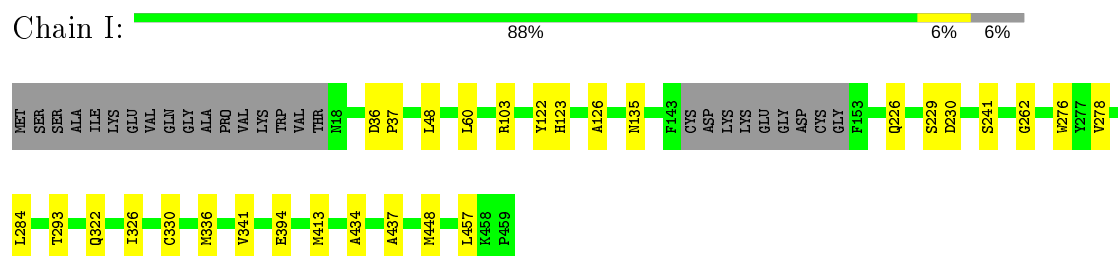
- Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA



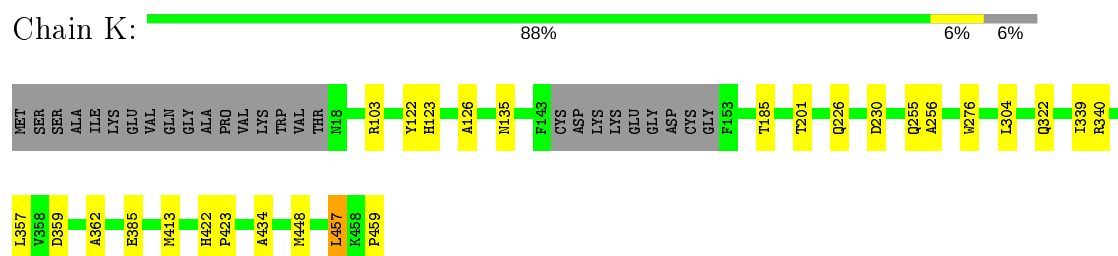
- Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA



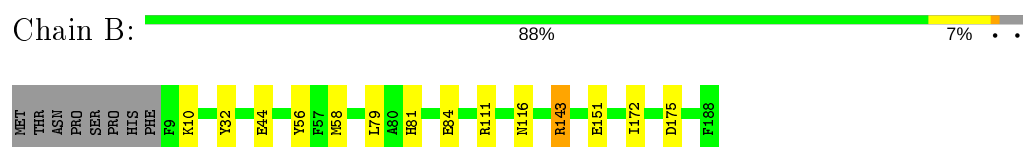
- Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA



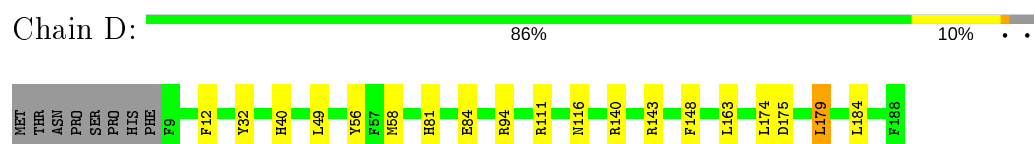
- Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA



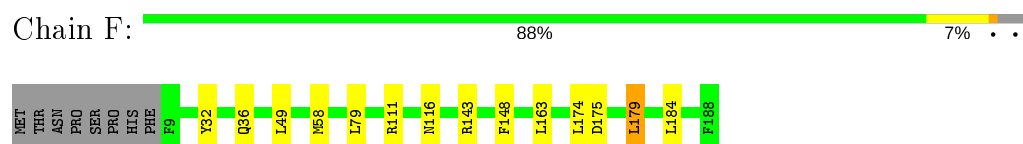
- Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA



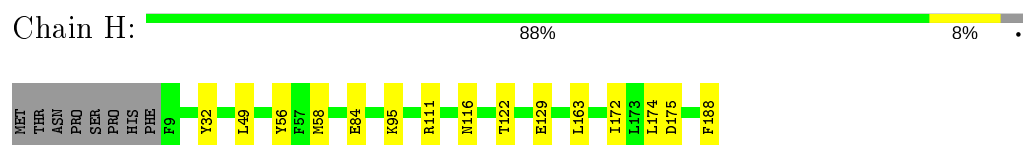
- Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA



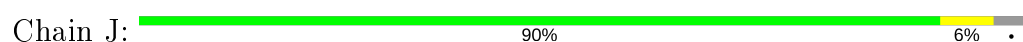
- Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

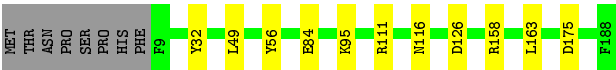


- Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

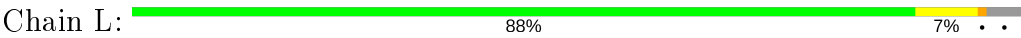


- Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA





● Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	86.97Å 277.81Å 92.93Å 90.00° 117.61° 90.00°	Depositor
Resolution (Å)	138.68 – 2.15 39.69 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.0 (138.68-2.15) 99.0 (39.69-2.15)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.17 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.200 , 0.232 0.271 , 0.291	Depositor DCC
R_{free} test set	10445 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	41.7	Xtriage
Anisotropy	0.102	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 43.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.017 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	30799	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.84% 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, 0.333 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: FE2, FES

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	0.34	0/3533	0.48	0/4796
1	C	0.34	0/3533	0.48	0/4796
1	E	0.32	0/3533	0.47	0/4796
1	G	0.31	0/3533	0.45	0/4796
1	I	0.31	0/3533	0.45	0/4796
1	K	0.31	0/3533	0.45	0/4796
2	B	0.33	0/1530	0.49	0/2068
2	D	0.35	0/1530	0.49	0/2068
2	F	0.33	0/1530	0.49	0/2068
2	H	0.33	0/1530	0.47	0/2068
2	J	0.31	0/1530	0.47	0/2068
2	L	0.32	0/1530	0.48	0/2068
All	All	0.32	0/30378	0.47	0/41184

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	3430	0	3274	20	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	3430	0	3274	17	0
1	E	3430	0	3274	14	0
1	G	3430	0	3274	18	0
1	I	3430	0	3274	13	0
1	K	3430	0	3274	13	0
2	B	1496	0	1447	12	0
2	D	1496	0	1447	14	0
2	F	1496	0	1447	11	0
2	H	1496	0	1447	12	0
2	J	1496	0	1447	8	0
2	L	1496	0	1447	9	0
3	A	4	0	0	1	0
3	C	4	0	0	1	0
3	E	4	0	0	1	0
3	G	4	0	0	1	0
3	I	4	0	0	1	0
3	K	4	0	0	1	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
4	E	1	0	0	0	0
4	G	1	0	0	0	0
4	I	1	0	0	0	0
4	K	1	0	0	0	0
5	A	210	0	0	1	0
5	B	94	0	0	0	0
5	C	212	0	0	1	0
5	D	99	0	0	0	0
5	E	152	0	0	0	0
5	F	94	0	0	0	0
5	G	73	0	0	2	0
5	H	79	0	0	0	0
5	I	47	0	0	0	0
5	J	46	0	0	0	0
5	K	53	0	0	0	0
5	L	54	0	0	0	0
All	All	30799	0	28326	137	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 137 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:413:MET:HG3	1:E:142:ALA:HB1	1.69	0.74
2:F:58:MET:HE1	2:F:174:LEU:HD22	1.77	0.67
1:C:339:ILE:HD11	1:C:357:LEU:HG	1.78	0.66
1:G:86:VAL:HB	5:G:2017:HOH:O	1.97	0.64
1:K:123:HIS:HB2	3:K:900:FES:S2	2.38	0.63

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	429/459 (94%)	417 (97%)	12 (3%)	0	100	100
1	C	429/459 (94%)	411 (96%)	18 (4%)	0	100	100
1	E	429/459 (94%)	414 (96%)	15 (4%)	0	100	100
1	G	429/459 (94%)	411 (96%)	18 (4%)	0	100	100
1	I	429/459 (94%)	412 (96%)	17 (4%)	0	100	100
1	K	429/459 (94%)	409 (95%)	20 (5%)	0	100	100
2	B	178/188 (95%)	174 (98%)	4 (2%)	0	100	100
2	D	178/188 (95%)	172 (97%)	6 (3%)	0	100	100
2	F	178/188 (95%)	173 (97%)	5 (3%)	0	100	100
2	H	178/188 (95%)	172 (97%)	6 (3%)	0	100	100
2	J	178/188 (95%)	173 (97%)	5 (3%)	0	100	100
2	L	178/188 (95%)	173 (97%)	5 (3%)	0	100	100
All	All	3642/3882 (94%)	3511 (96%)	131 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	350/372 (94%)	342 (98%)	8 (2%)	50	53
1	C	350/372 (94%)	343 (98%)	7 (2%)	55	59
1	E	350/372 (94%)	342 (98%)	8 (2%)	50	53
1	G	350/372 (94%)	343 (98%)	7 (2%)	55	59
1	I	350/372 (94%)	345 (99%)	5 (1%)	67	72
1	K	350/372 (94%)	347 (99%)	3 (1%)	78	83
2	B	159/167 (95%)	156 (98%)	3 (2%)	57	61
2	D	159/167 (95%)	156 (98%)	3 (2%)	57	61
2	F	159/167 (95%)	158 (99%)	1 (1%)	86	90
2	H	159/167 (95%)	159 (100%)	0	100	100
2	J	159/167 (95%)	159 (100%)	0	100	100
2	L	159/167 (95%)	155 (98%)	4 (2%)	47	49
All	All	3054/3234 (94%)	3005 (98%)	49 (2%)	62	67

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	103	ARG
1	E	339	ILE
2	L	10	LYS
1	E	258	ILE
1	E	466	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	36	GLN
1	K	226	GLN
1	E	422	HIS
1	C	391	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	D	131	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](#)

Of 12 ligands modelled in this entry, 6 are monoatomic - leaving 6 for Mogul analysis.

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	FES	E	900	1	0,4,4	0.00	-	-		
3	FES	C	900	1	0,4,4	0.00	-	-		
3	FES	G	900	1	0,4,4	0.00	-	-		
3	FES	K	900	1	0,4,4	0.00	-	-		
3	FES	A	900	1	0,4,4	0.00	-	-		
3	FES	I	900	1	0,4,4	0.00	-	-		

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	FES	E	900	1	-	-	0/1/1/1
3	FES	C	900	1	-	-	0/1/1/1
3	FES	G	900	1	-	-	0/1/1/1
3	FES	K	900	1	-	-	0/1/1/1
3	FES	A	900	1	-	-	0/1/1/1
3	FES	I	900	1	-	-	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	E	900	FES	1	0
3	C	900	FES	1	0
3	G	900	FES	1	0
3	K	900	FES	1	0
3	A	900	FES	1	0
3	I	900	FES	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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.