wwPDB X-ray Structure Validation Summary Report (i)

Dec 14, 2023 – 07:53 am GMT

PDB ID	:	2YFJ
Title	:	Crystal structure of Biphenyl dioxygenase variant RR41 with dibenzofuran
Authors	:	Kumar, P.; Sylvestre, M.; 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 (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

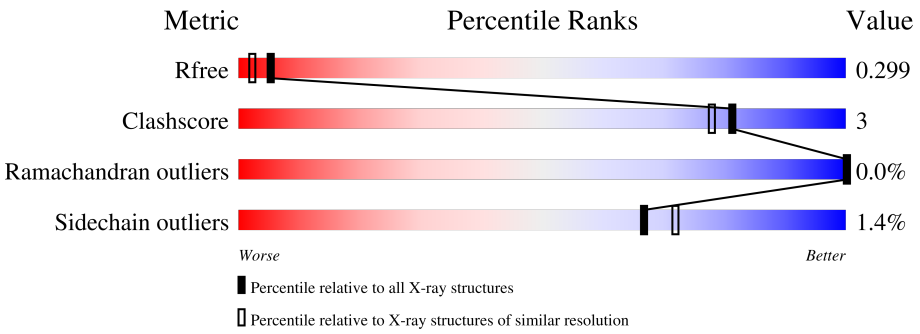
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.15 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 of 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 <=5%

Mol	Chain	Length	Quality of chain
1	A	459	88% 6% 6%
1	C	459	89% 5% 6%
1	E	459	87% 7% 6%
1	G	459	87% 7% 6%
1	I	459	87% 8% 6%
1	K	459	89% 5% 6%
2	B	188	86% 10% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	188	87% 9%
2	F	188	87% 8%
2	H	188	87% 9%
2	J	188	87% 9%
2	L	188	84% 12%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 30277 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
			Total	C	N	O	S			
1	A	433	3430	2182	602	622	24	0	0	0
1	C	433	3430	2182	602	622	24	0	0	0
1	E	433	3430	2182	602	622	24	0	0	0
1	G	433	3430	2182	602	622	24	0	0	0
1	I	433	3430	2182	602	622	24	0	0	0
1	K	433	3430	2182	602	622	24	0	0	0

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...

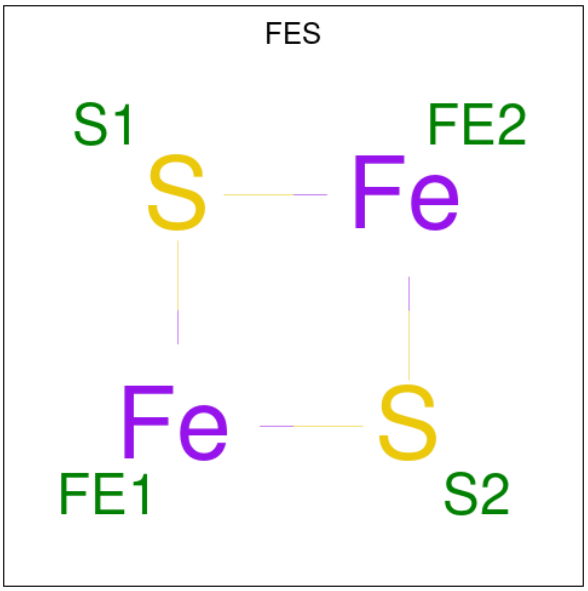
Continued from previous page...

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
			Total	C	N	O	S			
2	B	180	1496	948	265	279	4	0	0	0
2	D	180	1496	948	265	279	4	0	0	0
2	F	180	1496	948	265	279	4	0	0	0
2	H	180	1496	948	265	279	4	0	0	0
2	J	180	1496	948	265	279	4	0	0	0
2	L	180	1496	948	265	279	4	0	0	0

- Molecule 3 is FE2/S2 (INORGANIC) CLUSTER (three-letter code: FES) (formula: Fe_2S_2).

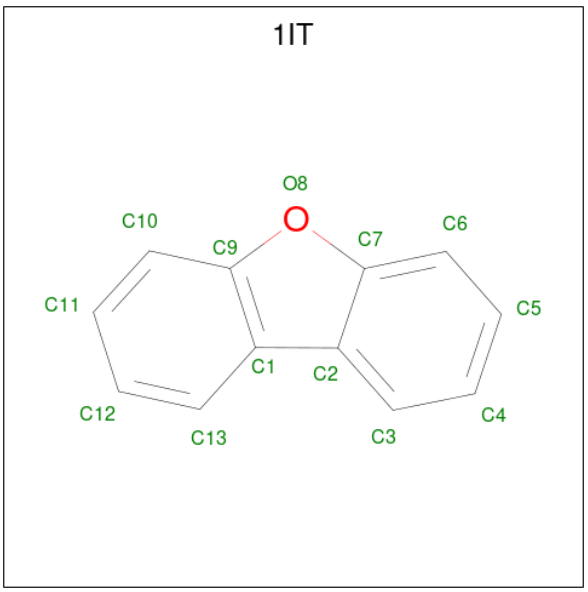


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	Fe	S		
3	A	1	4	2	2	0	0
3	C	1	4	2	2	0	0
3	E	1	4	2	2	0	0
3	G	1	4	2	2	0	0
3	I	1	4	2	2	0	0
3	K	1	4	2	2	0	0

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

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

- Molecule 5 is DIBENZOFURAN (three-letter code: 1IT) (formula: $C_{12}H_8O$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 13 12 1	0	0
5	C	1	Total C O 13 12 1	0	0
5	E	1	Total C O 13 12 1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	128	Total O 128 128	0	0
6	B	53	Total O 53 53	0	0
6	C	129	Total O 129 129	0	0
6	D	66	Total O 66 66	0	0
6	E	88	Total O 88 88	0	0
6	F	46	Total O 46 46	0	0
6	G	23	Total O 23 23	0	0
6	H	19	Total O 19 19	0	0

Continued on next page...

Continued from previous page...

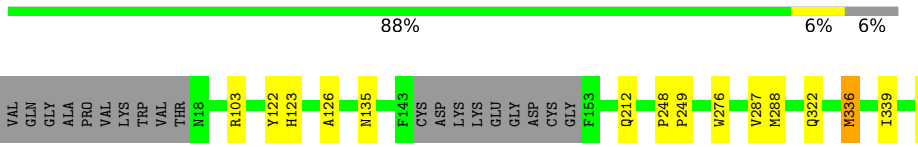
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	I	29	29	29	0	0
			Total	O		
6	J	31	31	31	0	0
			Total	O		
6	K	23	23	23	0	0
			Total	O		
6	L	17	17	17	0	0

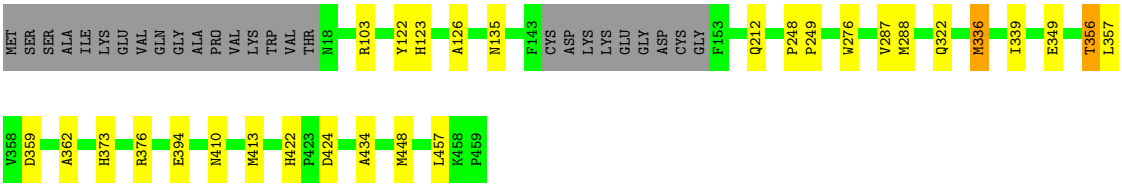
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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

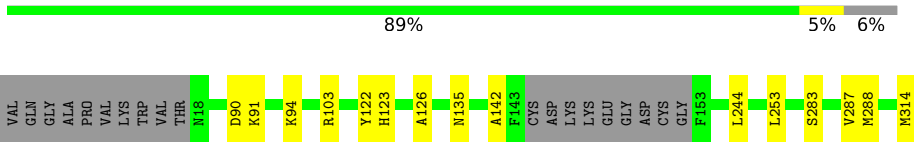
Chain A:

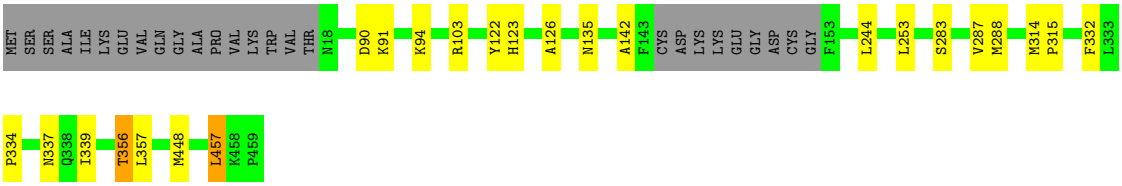




• Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA

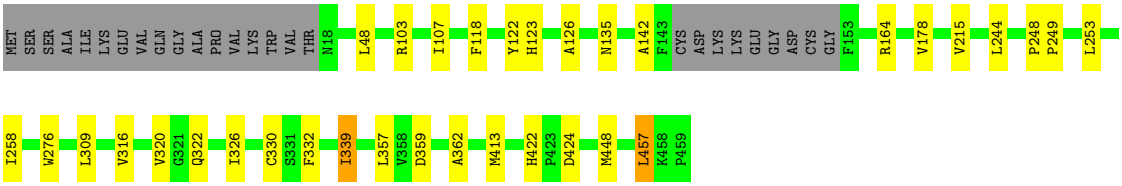
Chain C:





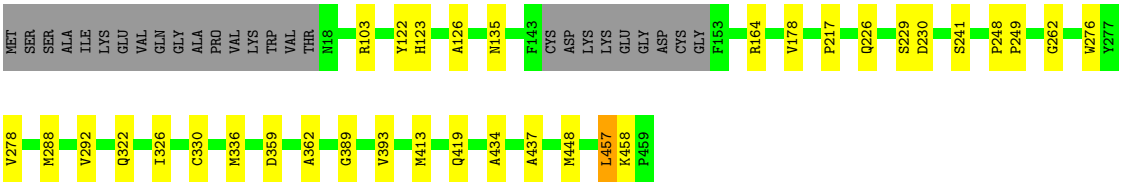
• Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA

Chain E:



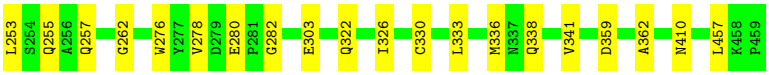
• Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA

Chain G:



• Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA

Chain I:

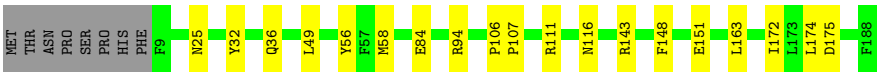


• Molecule 1: BIPHENYL DIOXYGENASE SUBUNIT ALPHA

Chain K:

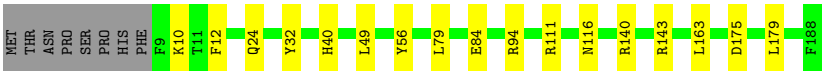
• Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

Chain B:



• Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

Chain D:



• Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

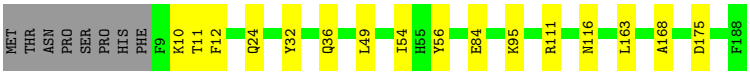
Chain F:

• Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

Chain H:

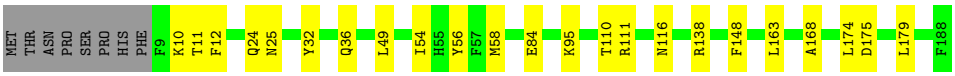
• Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

Chain J:



• Molecule 2: BIPHENYL DIOXYGENASE SUBUNIT BETA

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	86.98Å 278.12Å 92.96Å 90.00° 117.65° 90.00°	Depositor
Resolution (Å)	138.68 – 2.15 35.60 – 2.15	Depositor EDS
% Data completeness (in resolution range)	92.9 (138.68-2.15) 93.0 (35.60-2.15)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.13 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.204 , 0.233 0.281 , 0.299	Depositor DCC
R_{free} test set	9876 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	38.6	Xtriage
Anisotropy	0.149	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 36.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.020 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	30277	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% 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: 1IT, 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	0.33	0/3533	0.47	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	0.48	0/2068
2	D	0.34	0/1530	0.49	0/2068
2	F	0.33	0/1530	0.49	0/2068
2	H	0.31	0/1530	0.46	0/2068
2	J	0.32	0/1530	0.47	0/2068
2	L	0.31	0/1530	0.47	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	19	0
1	G	3430	0	3274	16	0
1	I	3430	0	3274	15	0
1	K	3430	0	3274	12	0
2	B	1496	0	1447	16	0
2	D	1496	0	1447	12	0
2	F	1496	0	1447	12	0
2	H	1496	0	1447	14	0
2	J	1496	0	1447	14	0
2	L	1496	0	1447	19	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	13	0	8	0	0
5	C	13	0	8	0	0
5	E	13	0	8	0	0
6	A	128	0	0	0	0
6	B	53	0	0	0	0
6	C	129	0	0	0	0
6	D	66	0	0	0	0
6	E	88	0	0	0	0
6	F	46	0	0	0	0
6	G	23	0	0	0	0
6	H	19	0	0	0	0
6	I	29	0	0	0	0
6	J	31	0	0	0	0
6	K	23	0	0	0	0
6	L	17	0	0	0	0
All	All	30277	0	28350	150	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 3.

The worst 5 of 150 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	1.59	0.84
1:C:339:ILE:HD11	1:C:357:LEU:HG	1.72	0.72
2:F:58:MET:HE1	2:F:174:LEU:HD22	1.76	0.68
1:I:262:GLY:HA2	1:I:278:VAL:HG23	1.75	0.67
1:A:123:HIS:HB2	3:A:900:FES:S2	2.35	0.66

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%)	415 (97%)	14 (3%)	0	100	100
1	C	429/459 (94%)	409 (95%)	20 (5%)	0	100	100
1	E	429/459 (94%)	416 (97%)	13 (3%)	0	100	100
1	G	429/459 (94%)	416 (97%)	13 (3%)	0	100	100
1	I	429/459 (94%)	408 (95%)	21 (5%)	0	100	100
1	K	429/459 (94%)	410 (96%)	18 (4%)	1 (0%)	47	46
2	B	178/188 (95%)	174 (98%)	4 (2%)	0	100	100
2	D	178/188 (95%)	173 (97%)	5 (3%)	0	100	100
2	F	178/188 (95%)	173 (97%)	5 (3%)	0	100	100
2	H	178/188 (95%)	174 (98%)	4 (2%)	0	100	100
2	J	178/188 (95%)	173 (97%)	5 (3%)	0	100	100
2	L	178/188 (95%)	170 (96%)	8 (4%)	0	100	100
All	All	3642/3882 (94%)	3511 (96%)	130 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	K	256	ALA

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	350/372 (94%)	344 (98%)	6 (2%)	60	65
1	C	350/372 (94%)	345 (99%)	5 (1%)	67	72
1	E	350/372 (94%)	343 (98%)	7 (2%)	55	59
1	G	350/372 (94%)	344 (98%)	6 (2%)	60	65
1	I	350/372 (94%)	343 (98%)	7 (2%)	55	59
1	K	350/372 (94%)	346 (99%)	4 (1%)	73	78
2	B	159/167 (95%)	158 (99%)	1 (1%)	86	90
2	D	159/167 (95%)	155 (98%)	4 (2%)	47	49
2	F	159/167 (95%)	156 (98%)	3 (2%)	57	61
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%)	158 (99%)	1 (1%)	86	90
All	All	3054/3234 (94%)	3010 (99%)	44 (1%)	67	72

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	336	MET
1	I	303	GLU
1	G	419	GLN
1	I	103	ARG
1	I	410	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 52 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	410	ASN
1	I	264	GLN
2	L	25	ASN
1	G	422	HIS
2	H	36	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 15 ligands modelled in this entry, 6 are monoatomic - leaving 9 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	$\#\|Z\| > 2$
5	1IT	A	1460	-	11,15,15	1.52	2 (18%)	12,21,21	1.83	4 (33%)
3	FES	C	900	1	0,4,4	-	-	-		
5	1IT	C	1451	-	11,15,15	1.50	2 (18%)	12,21,21	1.88	4 (33%)
3	FES	E	900	1	0,4,4	-	-	-		
5	1IT	E	1451	-	11,15,15	1.51	2 (18%)	12,21,21	1.83	4 (33%)
3	FES	A	900	1	0,4,4	-	-	-		
3	FES	G	900	1	0,4,4	-	-	-		
3	FES	I	900	1	0,4,4	-	-	-		
3	FES	K	900	1	0,4,4	-	-	-		

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
5	1IT	A	1460	-	-	-	0/3/3/3
3	FES	C	900	1	-	-	0/1/1/1
5	1IT	C	1451	-	-	-	0/3/3/3
3	FES	E	900	1	-	-	0/1/1/1
5	1IT	E	1451	-	-	-	0/3/3/3
3	FES	A	900	1	-	-	0/1/1/1
3	FES	G	900	1	-	-	0/1/1/1
3	FES	I	900	1	-	-	0/1/1/1
3	FES	K	900	1	-	-	0/1/1/1

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	1451	1IT	C2-C7	-2.28	1.38	1.43
5	A	1460	1IT	C1-C9	-2.26	1.38	1.43
5	C	1451	1IT	C2-C7	-2.24	1.38	1.43
5	E	1451	1IT	C1-C9	-2.22	1.38	1.43
5	C	1451	1IT	C1-C9	-2.21	1.38	1.43

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	C	1451	1IT	C3-C2-C7	3.61	122.66	120.38
5	C	1451	1IT	C13-C1-C9	3.50	122.59	120.38
5	E	1451	1IT	C13-C1-C9	3.46	122.56	120.38
5	A	1460	1IT	C3-C2-C7	3.45	122.56	120.38
5	A	1460	1IT	C13-C1-C9	3.44	122.56	120.38

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	C	900	FES	1	0
3	E	900	FES	1	0
3	A	900	FES	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	G	900	FES	1	0
3	I	900	FES	1	0
3	K	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 (i)

6.1 Protein, DNA and RNA chains (i)

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

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

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

6.3 Carbohydrates (i)

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

6.4 Ligands (i)

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

6.5 Other polymers (i)

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