



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:30 AM GMT

PDB ID : 3M3I  
Title : Hypothetical protein from Leishmania major  
Authors : Merritt, E.A.; Structural Genomics of Pathogenic Protozoa Consortium (SGPP)  
Deposited on : 2010-03-09  
Resolution : 2.35 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](mailto: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 (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

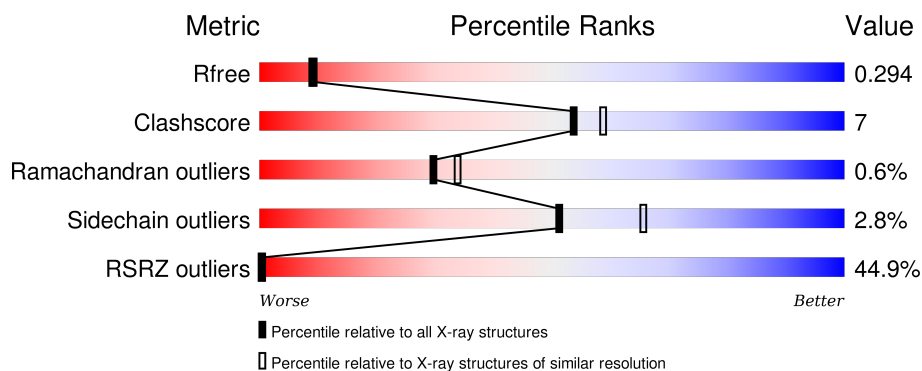
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.35 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	1352 (2.38-2.34)
Clashscore	102246	1456 (2.38-2.34)
Ramachandran outliers	100387	1435 (2.38-2.34)
Sidechain outliers	100360	1436 (2.38-2.34)
RSRZ outliers	91569	1358 (2.38-2.34)

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  $\geq 3$ , 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	225	<div> <div>30%</div> <div>68% 10% • 21%</div> </div>
1	B	225	<div> <div>39%</div> <div>64% 14% • 21%</div> </div>
1	C	225	<div> <div>29%</div> <div>64% 14% • 20%</div> </div>
1	D	225	<div> <div>26%</div> <div>67% 12% • 20%</div> </div>
1	E	225	<div> <div>28%</div> <div>66% 12% 21%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	225	<div><div>44%</div><div><div></div><div></div><div></div><div></div></div><div>69%</div><div>9%</div><div>22%</div></div>
1	G	225	<div><div>44%</div><div><div></div><div></div><div></div><div></div></div><div>63%</div><div>14%</div><div>22%</div></div>
1	H	225	<div><div>42%</div><div><div></div><div></div><div></div><div></div></div><div>68%</div><div>10%</div><div>21%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 11442 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 Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	177	Total	C	N	O	S	0	0	0
			1404	893	245	260	6			
1	B	177	Total	C	N	O	S	0	1	0
			1424	904	245	269	6			
1	C	179	Total	C	N	O	S	0	1	0
			1431	909	250	266	6			
1	D	179	Total	C	N	O	S	0	0	0
			1425	906	248	265	6			
1	E	177	Total	C	N	O	S	0	1	0
			1413	898	247	262	6			
1	F	176	Total	C	N	O	S	0	1	0
			1401	893	241	261	6			
1	G	176	Total	C	N	O	S	0	1	0
			1407	896	244	261	6			
1	H	177	Total	C	N	O	S	0	1	0
			1416	898	247	265	6			

There are 64 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
A	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
A	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
A	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
A	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
A	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
A	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
A	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
B	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
B	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
B	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
B	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
B	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
B	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
B	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
C	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
C	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
C	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
C	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
C	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
C	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
C	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
C	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
D	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
D	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
D	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
D	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
D	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
D	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
D	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
D	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
E	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
E	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
E	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
E	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
E	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
E	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
E	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
E	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
F	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
F	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
F	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
F	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
F	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
F	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
F	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
F	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
G	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
G	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
G	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
G	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
G	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
G	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
G	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	0	HIS	-	EXPRESSION TAG	UNP Q4FX13
H	-7	MET	-	EXPRESSION TAG	UNP Q4FX13
H	-6	ALA	-	EXPRESSION TAG	UNP Q4FX13
H	-5	HIS	-	EXPRESSION TAG	UNP Q4FX13
H	-4	HIS	-	EXPRESSION TAG	UNP Q4FX13
H	-3	HIS	-	EXPRESSION TAG	UNP Q4FX13
H	-2	HIS	-	EXPRESSION TAG	UNP Q4FX13
H	-1	HIS	-	EXPRESSION TAG	UNP Q4FX13
H	0	HIS	-	EXPRESSION TAG	UNP Q4FX13

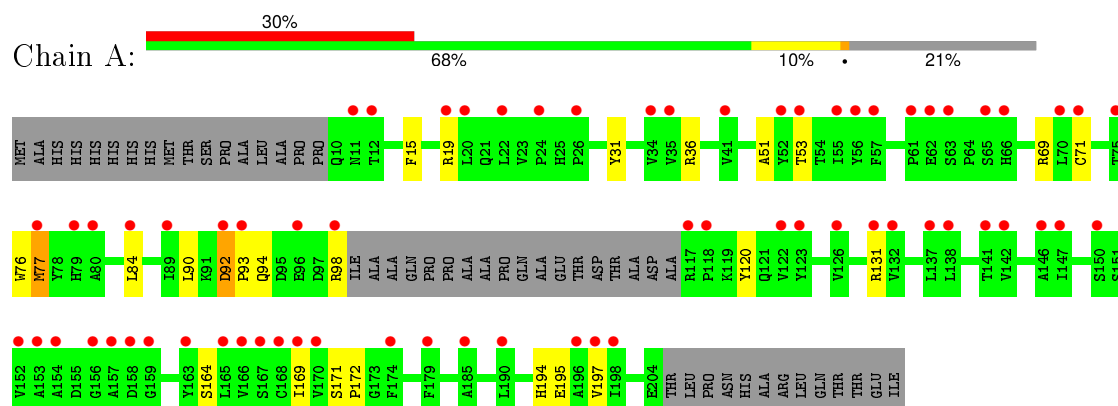
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	15	Total O 15 15	0	0
2	B	18	Total O 18 18	0	0
2	C	19	Total O 19 19	0	0
2	D	22	Total O 22 22	0	0
2	E	17	Total O 17 17	0	0
2	F	10	Total O 10 10	0	0
2	G	6	Total O 6 6	0	0
2	H	14	Total O 14 14	0	0

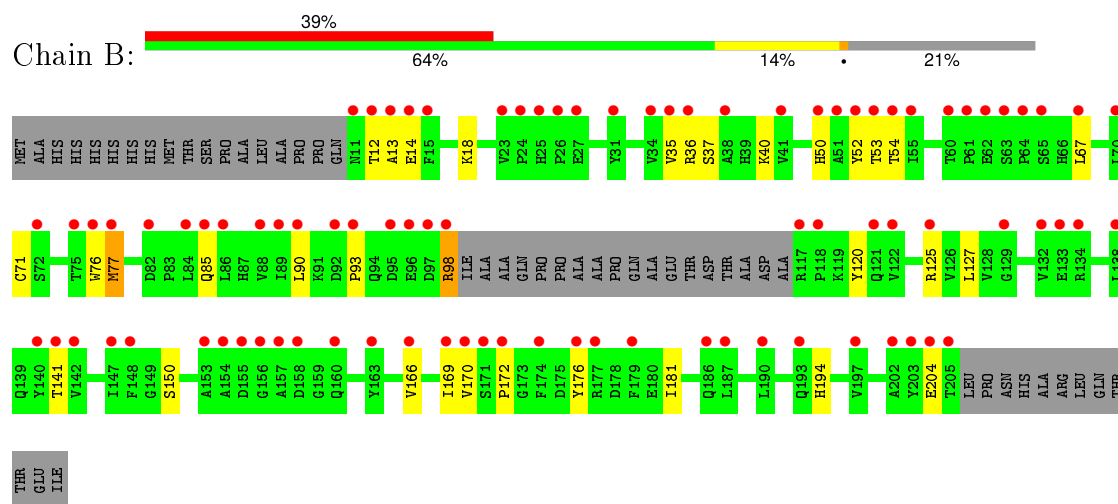
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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.

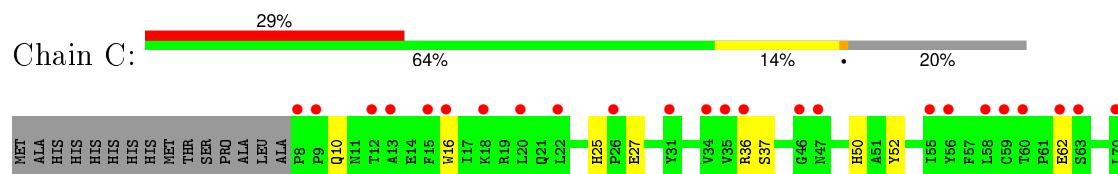
#### • Molecule 1: Putative uncharacterized protein



#### • Molecule 1: Putative uncharacterized protein



#### • Molecule 1: Putative uncharacterized protein

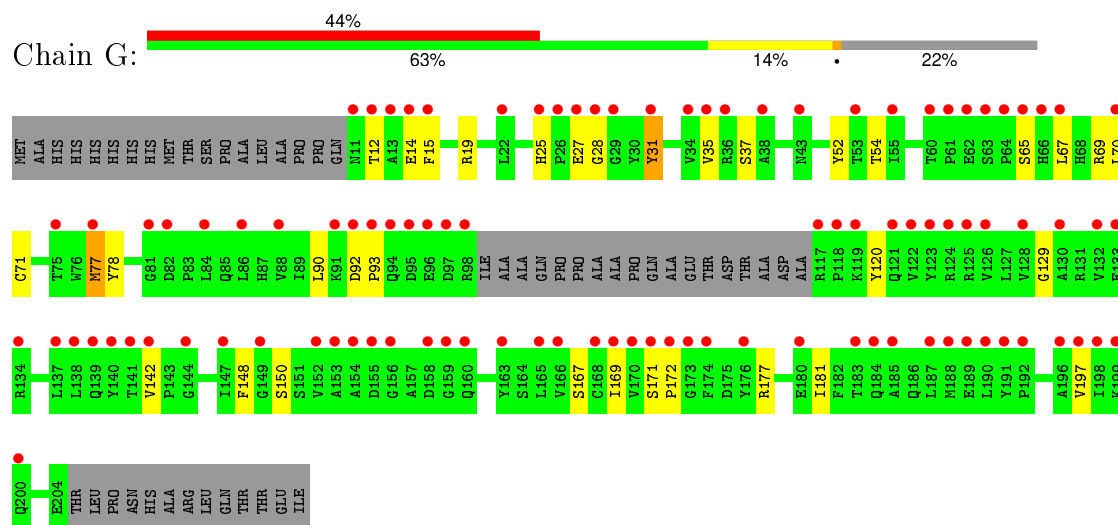






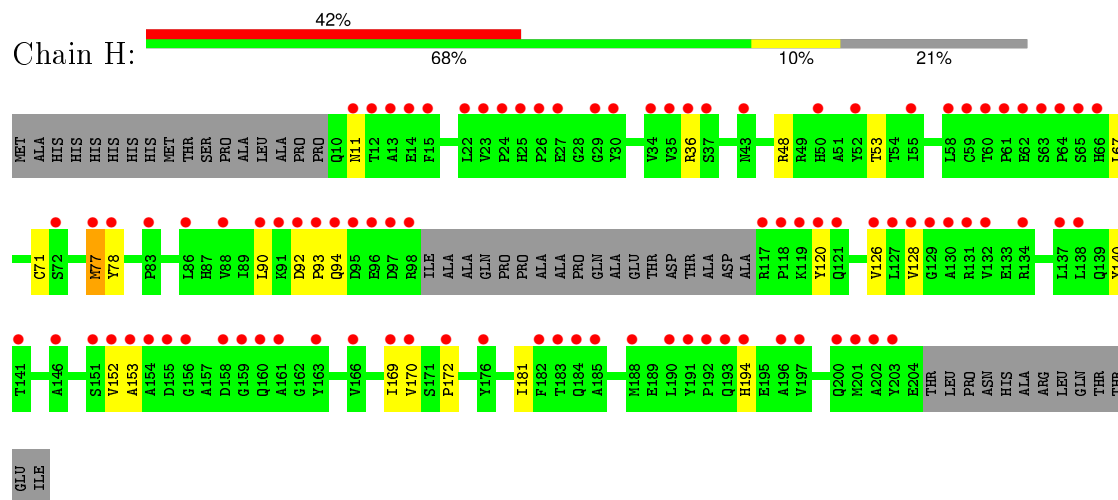
● Molecule 1: Putative uncharacterized protein

Chain G:



● Molecule 1: Putative uncharacterized protein

Chain H:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.98 Å   98.43 Å   100.76 Å 71.87°   80.19°   89.57°	Depositor
Resolution (Å)	46.72 – 2.35 46.72 – 2.35	Depositor EDS
% Data completeness (in resolution range)	92.8 (46.72-2.35) 82.9 (46.72-2.35)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.72 (at 2.34 Å)	Xtriage
Refinement program	REFMAC 5.5.0106	Depositor
R, $R_{free}$	0.239   ,   0.290 0.249   ,   0.294	Depositor DCC
$R_{free}$ test set	3677 reflections (5.26%)	DCC
Wilson B-factor (Å <sup>2</sup> )	35.1	Xtriage
Anisotropy	0.280	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 48.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 73637 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	11442	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 12.62% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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 [i](#)

### 5.1 Standard geometry [i](#)

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	0.51	0/1442	0.63	0/1959
1	B	0.48	0/1463	0.58	0/1989
1	C	0.53	0/1471	0.63	0/2000
1	D	0.53	0/1465	0.65	0/1991
1	E	0.53	1/1451 (0.1%)	0.64	0/1971
1	F	0.48	0/1440	0.59	0/1959
1	G	0.47	0/1446	0.58	0/1966
1	H	0.48	0/1454	0.60	0/1976
All	All	0.50	1/11632 (0.0%)	0.61	0/15811

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	168	CYS	CB-SG	-5.14	1.73	1.81

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	1404	0	1328	18	0
1	B	1424	0	1338	23	0
1	C	1431	0	1351	26	0
1	D	1425	0	1351	25	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	1413	0	1335	17	0
1	F	1401	0	1312	16	0
1	G	1407	0	1323	27	0
1	H	1416	0	1332	18	0
2	A	15	0	0	2	0
2	B	18	0	0	1	0
2	C	19	0	0	3	0
2	D	22	0	0	0	0
2	E	17	0	0	0	0
2	F	10	0	0	1	0
2	G	6	0	0	0	0
2	H	14	0	0	1	0
All	All	11442	0	10670	144	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 7.

The worst 5 of 144 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:169:ILE:CD1	1:H:169:ILE:HD12	1.85	1.06
1:A:169:ILE:HD12	1:H:169:ILE:HD12	1.06	1.06
1:E:195:GLU:OE1	1:F:19:ARG:NH1	1.95	0.99
1:A:169:ILE:HD12	1:H:169:ILE:CD1	1.95	0.96
1:A:195:GLU:OE1	1:G:19:ARG:NH1	2.04	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	173/225 (77%)	167 (96%)	5 (3%)	1 (1%)	30	34
1	B	174/225 (77%)	169 (97%)	4 (2%)	1 (1%)	30	34
1	C	176/225 (78%)	172 (98%)	3 (2%)	1 (1%)	30	34
1	D	175/225 (78%)	168 (96%)	6 (3%)	1 (1%)	30	34
1	E	174/225 (77%)	167 (96%)	6 (3%)	1 (1%)	30	34
1	F	173/225 (77%)	167 (96%)	5 (3%)	1 (1%)	30	34
1	G	173/225 (77%)	168 (97%)	4 (2%)	1 (1%)	30	34
1	H	174/225 (77%)	169 (97%)	4 (2%)	1 (1%)	30	34
All	All	1392/1800 (77%)	1347 (97%)	37 (3%)	8 (1%)	30	34

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	93	PRO
1	A	93	PRO
1	E	93	PRO
1	H	93	PRO
1	D	93	PRO

### 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	145/188 (77%)	140 (97%)	5 (3%)	44	57
1	B	149/188 (79%)	144 (97%)	5 (3%)	44	57
1	C	149/188 (79%)	146 (98%)	3 (2%)	63	77
1	D	149/188 (79%)	145 (97%)	4 (3%)	52	67
1	E	146/188 (78%)	140 (96%)	6 (4%)	37	48
1	F	144/188 (77%)	142 (99%)	2 (1%)	74	86
1	G	145/188 (77%)	139 (96%)	6 (4%)	37	48
1	H	147/188 (78%)	145 (99%)	2 (1%)	74	86

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1174/1504 (78%)	1141 (97%)	33 (3%)	51 65

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	92	ASP
1	E	31	TYR
1	G	197	VAL
1	D	98	ARG
1	D	126	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	21	GLN
1	G	121	GLN
1	G	25	HIS
1	D	85	GLN
1	G	68	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	177/225 (78%)	1.96	67 (37%) 0 0	57, 69, 99, 135	0
1	B	177/225 (78%)	2.26	88 (49%) 0 0	61, 73, 101, 162	0
1	C	179/225 (79%)	1.77	65 (36%) 0 0	57, 65, 83, 101	0
1	D	179/225 (79%)	1.77	58 (32%) 1 0	57, 65, 84, 113	0
1	E	177/225 (78%)	1.86	64 (36%) 0 0	55, 66, 84, 124	0
1	F	176/225 (78%)	2.46	100 (56%) 0 0	59, 74, 113, 205	0
1	G	176/225 (78%)	2.53	100 (56%) 0 0	59, 77, 121, 180	0
1	H	177/225 (78%)	2.35	94 (53%) 0 0	59, 74, 107, 142	0
All	All	1418/1800 (78%)	2.12	636 (44%) 0 0	55, 70, 103, 205	0

The worst 5 of 636 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	117	ARG	16.6
1	F	118	PRO	10.9
1	H	118	PRO	8.9
1	G	117	ARG	8.3
1	H	117	ARG	7.7

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.