



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 14, 2020 – 07:03 pm BST

PDB ID : 3UPU  
Title : Crystal structure of the T4 Phage SF1B Helicase Dda  
Authors : He, X.; Yun, M.K.; Pemble IV, C.W.; Kreuzer, K.N.; Raney, K.D.; White, S.W.  
Deposited on : 2011-11-18  
Resolution : 3.30 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

<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

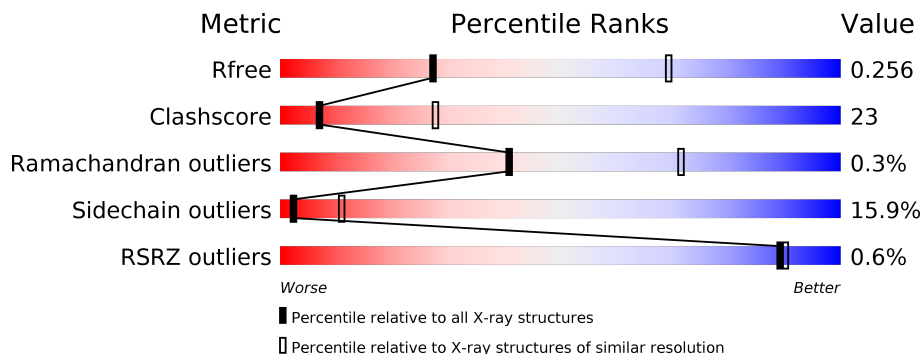
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.30 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 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  $\leq 5\%$ . 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	459	
1	B	459	
1	C	459	
2	D	8	
2	E	8	
2	F	8	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 11054 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 ATP-dependent DNA helicase dda.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	439	Total 3529	C 2280	N 582	O 655	S 5	Se 7	0	1	0
1	B	439	Total 3522	C 2275	N 580	O 655	S 5	Se 7	0	0	0
1	C	439	Total 3522	C 2275	N 580	O 655	S 5	Se 7	0	0	0

There are 75 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MSE	-	EXPRESSION TAG	UNP P32270
A	-18	GLY	-	EXPRESSION TAG	UNP P32270
A	-17	SER	-	EXPRESSION TAG	UNP P32270
A	-16	SER	-	EXPRESSION TAG	UNP P32270
A	-15	HIS	-	EXPRESSION TAG	UNP P32270
A	-14	HIS	-	EXPRESSION TAG	UNP P32270
A	-13	HIS	-	EXPRESSION TAG	UNP P32270
A	-12	HIS	-	EXPRESSION TAG	UNP P32270
A	-11	HIS	-	EXPRESSION TAG	UNP P32270
A	-10	HIS	-	EXPRESSION TAG	UNP P32270
A	-9	SER	-	EXPRESSION TAG	UNP P32270
A	-8	SER	-	EXPRESSION TAG	UNP P32270
A	-7	GLY	-	EXPRESSION TAG	UNP P32270
A	-6	LEU	-	EXPRESSION TAG	UNP P32270
A	-5	VAL	-	EXPRESSION TAG	UNP P32270
A	-4	PRO	-	EXPRESSION TAG	UNP P32270
A	-3	ARG	-	EXPRESSION TAG	UNP P32270
A	-2	GLY	-	EXPRESSION TAG	UNP P32270
A	-1	SER	-	EXPRESSION TAG	UNP P32270
A	0	HIS	-	EXPRESSION TAG	UNP P32270
A	38	ALA	LYS	ENGINEERED MUTATION	UNP P32270
A	54	GLU	GLY	SEE REMARK 999	UNP P32270
A	151	ASP	GLU	SEE REMARK 999	UNP P32270

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	196	ILE	ASN	SEE REMARK 999	UNP P32270
A	357	GLY	ALA	SEE REMARK 999	UNP P32270
B	-19	MSE	-	EXPRESSION TAG	UNP P32270
B	-18	GLY	-	EXPRESSION TAG	UNP P32270
B	-17	SER	-	EXPRESSION TAG	UNP P32270
B	-16	SER	-	EXPRESSION TAG	UNP P32270
B	-15	HIS	-	EXPRESSION TAG	UNP P32270
B	-14	HIS	-	EXPRESSION TAG	UNP P32270
B	-13	HIS	-	EXPRESSION TAG	UNP P32270
B	-12	HIS	-	EXPRESSION TAG	UNP P32270
B	-11	HIS	-	EXPRESSION TAG	UNP P32270
B	-10	HIS	-	EXPRESSION TAG	UNP P32270
B	-9	SER	-	EXPRESSION TAG	UNP P32270
B	-8	SER	-	EXPRESSION TAG	UNP P32270
B	-7	GLY	-	EXPRESSION TAG	UNP P32270
B	-6	LEU	-	EXPRESSION TAG	UNP P32270
B	-5	VAL	-	EXPRESSION TAG	UNP P32270
B	-4	PRO	-	EXPRESSION TAG	UNP P32270
B	-3	ARG	-	EXPRESSION TAG	UNP P32270
B	-2	GLY	-	EXPRESSION TAG	UNP P32270
B	-1	SER	-	EXPRESSION TAG	UNP P32270
B	0	HIS	-	EXPRESSION TAG	UNP P32270
B	38	ALA	LYS	ENGINEERED MUTATION	UNP P32270
B	54	GLU	GLY	SEE REMARK 999	UNP P32270
B	151	ASP	GLU	SEE REMARK 999	UNP P32270
B	196	ILE	ASN	SEE REMARK 999	UNP P32270
B	357	GLY	ALA	SEE REMARK 999	UNP P32270
C	-19	MSE	-	EXPRESSION TAG	UNP P32270
C	-18	GLY	-	EXPRESSION TAG	UNP P32270
C	-17	SER	-	EXPRESSION TAG	UNP P32270
C	-16	SER	-	EXPRESSION TAG	UNP P32270
C	-15	HIS	-	EXPRESSION TAG	UNP P32270
C	-14	HIS	-	EXPRESSION TAG	UNP P32270
C	-13	HIS	-	EXPRESSION TAG	UNP P32270
C	-12	HIS	-	EXPRESSION TAG	UNP P32270
C	-11	HIS	-	EXPRESSION TAG	UNP P32270
C	-10	HIS	-	EXPRESSION TAG	UNP P32270
C	-9	SER	-	EXPRESSION TAG	UNP P32270
C	-8	SER	-	EXPRESSION TAG	UNP P32270
C	-7	GLY	-	EXPRESSION TAG	UNP P32270
C	-6	LEU	-	EXPRESSION TAG	UNP P32270
C	-5	VAL	-	EXPRESSION TAG	UNP P32270

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	-4	PRO	-	EXPRESSION TAG	UNP P32270
C	-3	ARG	-	EXPRESSION TAG	UNP P32270
C	-2	GLY	-	EXPRESSION TAG	UNP P32270
C	-1	SER	-	EXPRESSION TAG	UNP P32270
C	0	HIS	-	EXPRESSION TAG	UNP P32270
C	38	ALA	LYS	ENGINEERED MUTATION	UNP P32270
C	54	GLU	GLY	SEE REMARK 999	UNP P32270
C	151	ASP	GLU	SEE REMARK 999	UNP P32270
C	196	ILE	ASN	SEE REMARK 999	UNP P32270
C	357	GLY	ALA	SEE REMARK 999	UNP P32270

- Molecule 2 is a DNA chain called 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	8	Total	C	N	O	P	0	0	0
			157	80	16	54	7			
2	E	8	Total	C	N	O	P	0	0	0
			157	80	16	54	7			
2	F	8	Total	C	N	O	P	0	0	0
			157	80	16	54	7			

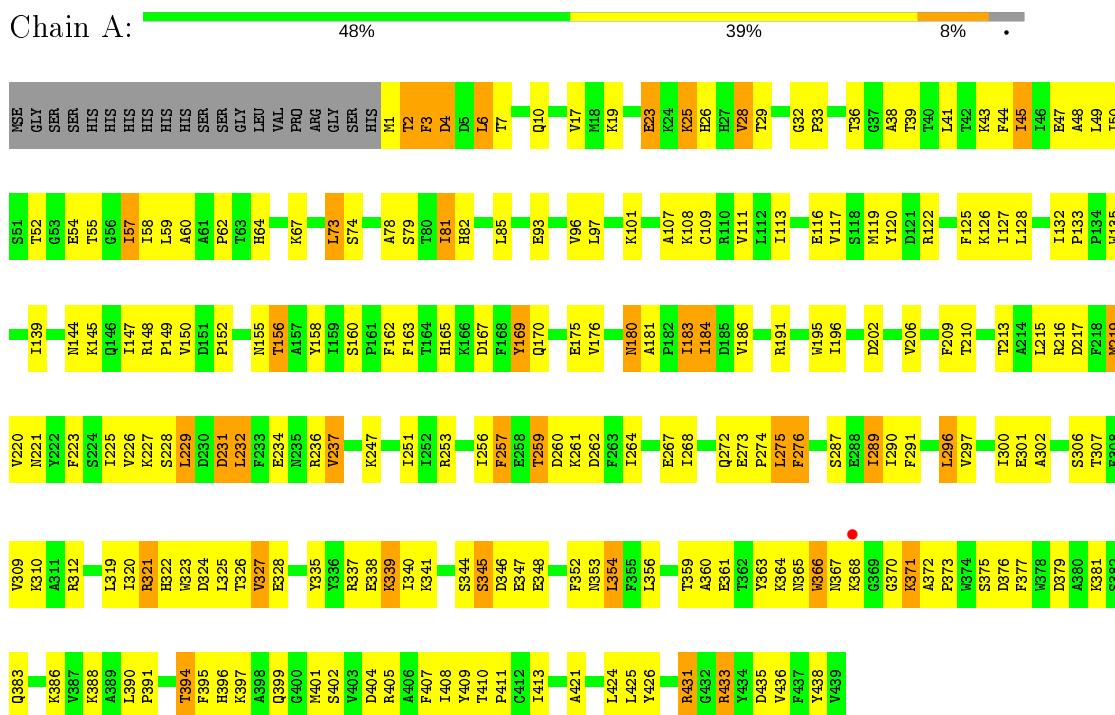
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	4	Total	O	0	0
			4	4		
3	B	2	Total	O	0	0
			2	2		
3	D	2	Total	O	0	0
			2	2		
3	E	1	Total	O	0	0
			1	1		
3	F	1	Total	O	0	0
			1	1		

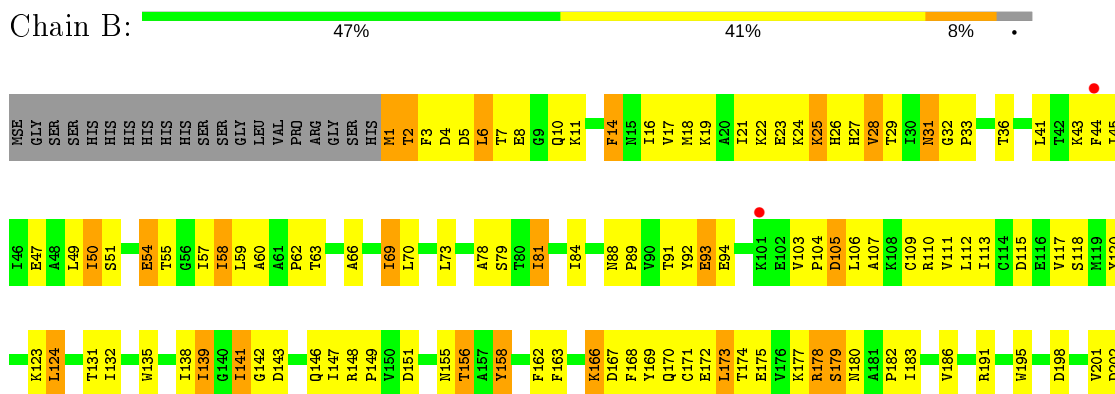
### 3 Residue-property plots

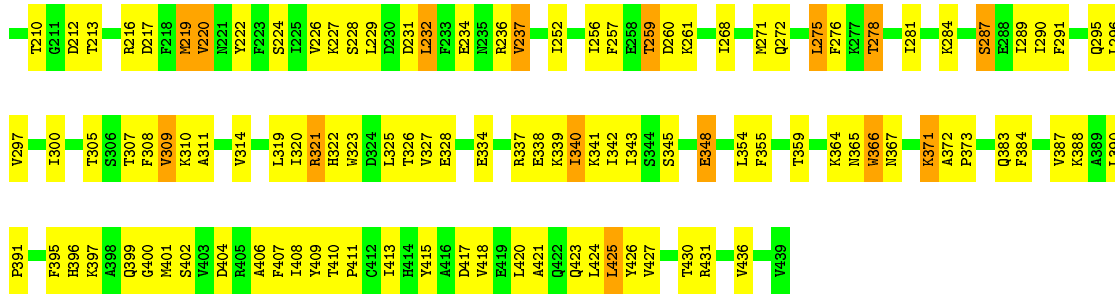
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are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

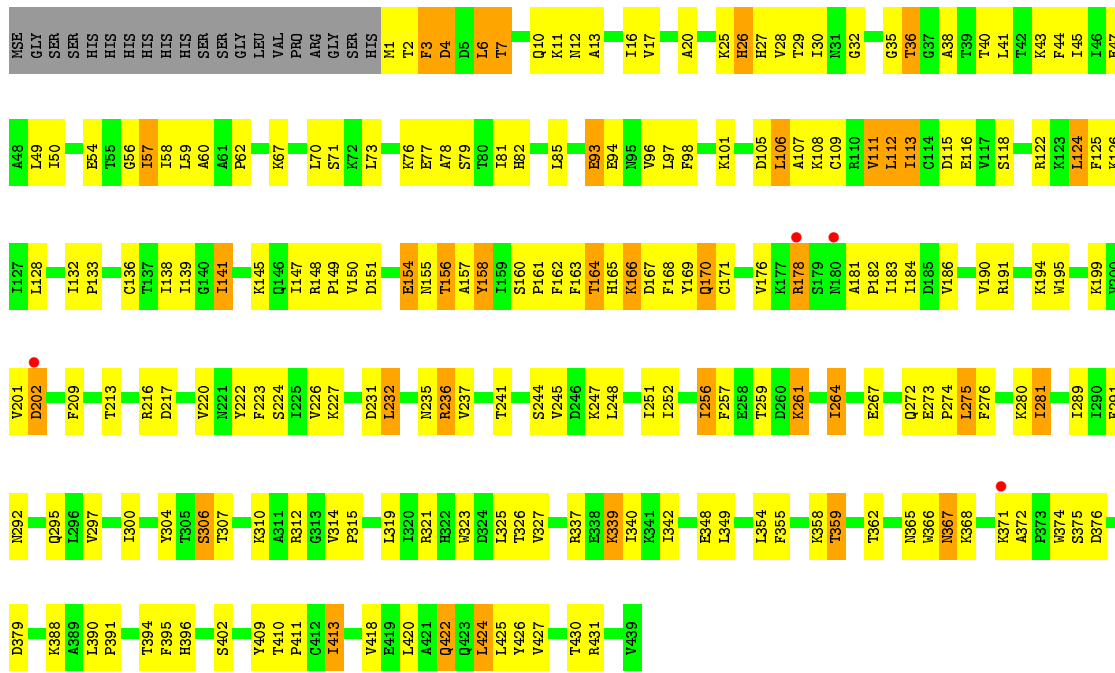


- Molecule 1: ATP-dependent DNA helicase dda





• Molecule 1: ATP-dependent DNA helicase dda



• Molecule 2: 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*T)-3'



• Molecule 2: 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*T)-3'



• Molecule 2: 5'-D(\*TP\*TP\*TP\*TP\*TP\*TP\*TP\*T)-3'



●

T600
T601
T602
T603
T604
T605
T606
T607



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	225.54Å 107.25Å 85.47Å 90.00° 94.13° 90.00°	Depositor
Resolution (Å)	32.27 – 3.30 36.04 – 3.30	Depositor EDS
% Data completeness (in resolution range)	91.1 (32.27-3.30) 90.8 (36.04-3.30)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.81 (at 3.32Å)	Xtrriage
Refinement program	PHENIX 1.7_650	Depositor
R, $R_{free}$	0.204 , 0.258 0.197 , 0.256	Depositor DCC
$R_{free}$ test set	1466 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	92.4	Xtrriage
Anisotropy	0.443	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 64.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11054	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	110.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.65% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality

### 5.1 Standard geometry

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	0.28	0/3607	0.51	0/4869
1	B	0.26	0/3596	0.47	0/4854
1	C	0.25	0/3596	0.47	0/4854
2	D	0.51	0/172	1.48	0/264
2	E	0.50	0/172	1.42	0/264
2	F	0.52	0/172	1.39	0/264
All	All	0.28	0/11315	0.57	0/15369

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	4
1	C	0	1
All	All	0	5

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	259	THR	Peptide
1	A	344	SER	Peptide
1	A	366	TRP	Peptide
1	A	4	ASP	Peptide
1	C	108	LYS	Peptide

## 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	3529	0	3549	175	0
1	B	3522	0	3542	163	0
1	C	3522	0	3542	156	0
2	D	157	0	98	8	0
2	E	157	0	98	5	0
2	F	157	0	98	9	0
3	A	4	0	0	0	0
3	B	2	0	0	0	0
3	D	2	0	0	0	0
3	E	1	0	0	0	0
3	F	1	0	0	0	0
All	All	11054	0	10927	504	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 23.

The worst 5 of 504 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:B:6:LEU:HD12	1:B:11:LYS:HG2	1.31	1.07
1:B:276:PHE:HE1	1:B:289:ILE:HG12	1.18	1.04
1:B:321:ARG:HG3	1:B:321:ARG:HH11	1.22	1.00
1:A:232:LEU:HD22	1:A:232:LEU:H	1.26	0.95
1:B:276:PHE:CE1	1:B:289:ILE:HG12	2.05	0.92

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	438/459 (95%)	408 (93%)	29 (7%)	1 (0%)	47	77
1	B	437/459 (95%)	397 (91%)	39 (9%)	1 (0%)	47	77
1	C	437/459 (95%)	402 (92%)	33 (8%)	2 (0%)	29	61
All	All	1312/1377 (95%)	1207 (92%)	101 (8%)	4 (0%)	41	71

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	93	GLU
1	C	93	GLU
1	A	93	GLU
1	C	256	ILE

### 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	386/394 (98%)	330 (86%)	56 (14%)	3	14
1	B	385/394 (98%)	316 (82%)	69 (18%)	2	8
1	C	385/394 (98%)	326 (85%)	59 (15%)	2	12
All	All	1156/1182 (98%)	972 (84%)	184 (16%)	2	11

5 of 184 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	171	CYS
1	B	257	PHE
1	C	337	ARG
1	B	173	LEU
1	B	210	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	192	ASN
1	B	292	ASN
1	C	165	HIS
1	B	95	ASN
1	B	155	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](#)

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	432/459 (94%)	-0.09	1 (0%) 95 96	62, 98, 136, 169	0
1	B	432/459 (94%)	-0.04	2 (0%) 91 91	64, 106, 159, 184	0
1	C	432/459 (94%)	0.03	4 (0%) 84 84	81, 115, 154, 180	0
2	D	8/8 (100%)	0.05	0 100 100	86, 91, 111, 160	0
2	E	8/8 (100%)	0.11	0 100 100	95, 104, 149, 197	0
2	F	8/8 (100%)	0.64	1 (12%) 3 3	106, 112, 161, 217	0
All	All	1320/1401 (94%)	-0.03	8 (0%) 89 90	62, 107, 155, 217	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	600	DT	5.1
1	C	180	ASN	3.5
1	C	371	LYS	3.0
1	C	202	ASP	2.6
1	C	178	ARG	2.4

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.