



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 17, 2023 – 12:48 am GMT

PDB ID : 4CBM  
Title : Pestivirus NS3 helicase  
Authors : Tortorici, M.A.; Duquerroy, S.; Kwok, J.; Vonrhein, C.; Perez, J.; Lamp, B.; Bricogne, G.; Rummenapf, T.; Vachette, P.; Rey, F.A.  
Deposited on : 2013-10-14  
Resolution : 3.27 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
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

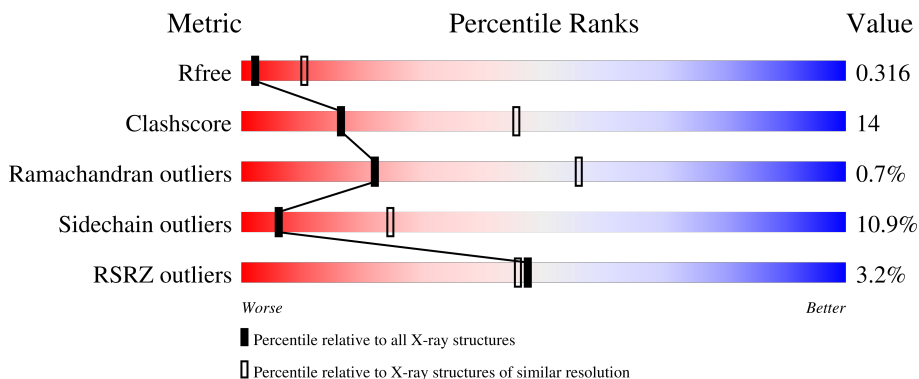
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.27 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	1177 (3.32-3.24)
Clashscore	141614	1044 (3.30-3.26)
Ramachandran outliers	138981	1026 (3.30-3.26)
Sidechain outliers	138945	1025 (3.30-3.26)
RSRZ outliers	127900	1141 (3.32-3.24)

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  $\geq 3$ , 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	516	 2% 61% 26% • 9%
1	B	516	 3% 58% 27% • 11%
1	C	516	 3% 61% 24% • 11%
1	D	516	 4% 60% 24% • 12%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 14616 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 SERINE PROTEASE NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	467	Total 3704	C 2349	N 632	O 702	S 21	0	0	0
1	B	458	Total 3641	C 2310	N 622	O 688	S 21	0	0	0
1	C	461	Total 3672	C 2331	N 624	O 695	S 22	0	0	0
1	D	453	Total 3598	C 2282	N 611	O 683	S 22	0	0	0

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	176	MET	-	expression tag	UNP P19712
A	177	ALA	-	expression tag	UNP P19712
A	178	SER	-	expression tag	UNP P19712
A	179	HIS	-	expression tag	UNP P19712
A	180	HIS	-	expression tag	UNP P19712
A	181	HIS	-	expression tag	UNP P19712
A	182	HIS	-	expression tag	UNP P19712
A	183	HIS	-	expression tag	UNP P19712
A	184	HIS	-	expression tag	UNP P19712
A	185	HIS	-	expression tag	UNP P19712
A	186	GLU	-	expression tag	UNP P19712
A	187	ASN	-	expression tag	UNP P19712
A	188	LEU	-	expression tag	UNP P19712
A	189	TYR	-	expression tag	UNP P19712
A	190	PHE	-	expression tag	UNP P19712
A	191	GLN	-	expression tag	UNP P19712
A	192	GLY	-	expression tag	UNP P19712
A	600	LYS	GLN	engineered mutation	UNP P19712
B	176	MET	-	expression tag	UNP P19712
B	177	ALA	-	expression tag	UNP P19712
B	178	SER	-	expression tag	UNP P19712

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	179	HIS	-	expression tag	UNP P19712
B	180	HIS	-	expression tag	UNP P19712
B	181	HIS	-	expression tag	UNP P19712
B	182	HIS	-	expression tag	UNP P19712
B	183	HIS	-	expression tag	UNP P19712
B	184	HIS	-	expression tag	UNP P19712
B	185	HIS	-	expression tag	UNP P19712
B	186	GLU	-	expression tag	UNP P19712
B	187	ASN	-	expression tag	UNP P19712
B	188	LEU	-	expression tag	UNP P19712
B	189	TYR	-	expression tag	UNP P19712
B	190	PHE	-	expression tag	UNP P19712
B	191	GLN	-	expression tag	UNP P19712
B	192	GLY	-	expression tag	UNP P19712
B	600	LYS	GLN	engineered mutation	UNP P19712
C	176	MET	-	expression tag	UNP P19712
C	177	ALA	-	expression tag	UNP P19712
C	178	SER	-	expression tag	UNP P19712
C	179	HIS	-	expression tag	UNP P19712
C	180	HIS	-	expression tag	UNP P19712
C	181	HIS	-	expression tag	UNP P19712
C	182	HIS	-	expression tag	UNP P19712
C	183	HIS	-	expression tag	UNP P19712
C	184	HIS	-	expression tag	UNP P19712
C	185	HIS	-	expression tag	UNP P19712
C	186	GLU	-	expression tag	UNP P19712
C	187	ASN	-	expression tag	UNP P19712
C	188	LEU	-	expression tag	UNP P19712
C	189	TYR	-	expression tag	UNP P19712
C	190	PHE	-	expression tag	UNP P19712
C	191	GLN	-	expression tag	UNP P19712
C	192	GLY	-	expression tag	UNP P19712
C	600	LYS	GLN	engineered mutation	UNP P19712
D	176	MET	-	expression tag	UNP P19712
D	177	ALA	-	expression tag	UNP P19712
D	178	SER	-	expression tag	UNP P19712
D	179	HIS	-	expression tag	UNP P19712
D	180	HIS	-	expression tag	UNP P19712
D	181	HIS	-	expression tag	UNP P19712
D	182	HIS	-	expression tag	UNP P19712
D	183	HIS	-	expression tag	UNP P19712
D	184	HIS	-	expression tag	UNP P19712

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	185	HIS	-	expression tag	UNP P19712
D	186	GLU	-	expression tag	UNP P19712
D	187	ASN	-	expression tag	UNP P19712
D	188	LEU	-	expression tag	UNP P19712
D	189	TYR	-	expression tag	UNP P19712
D	190	PHE	-	expression tag	UNP P19712
D	191	GLN	-	expression tag	UNP P19712
D	192	GLY	-	expression tag	UNP P19712
D	600	LYS	GLN	engineered mutation	UNP P19712

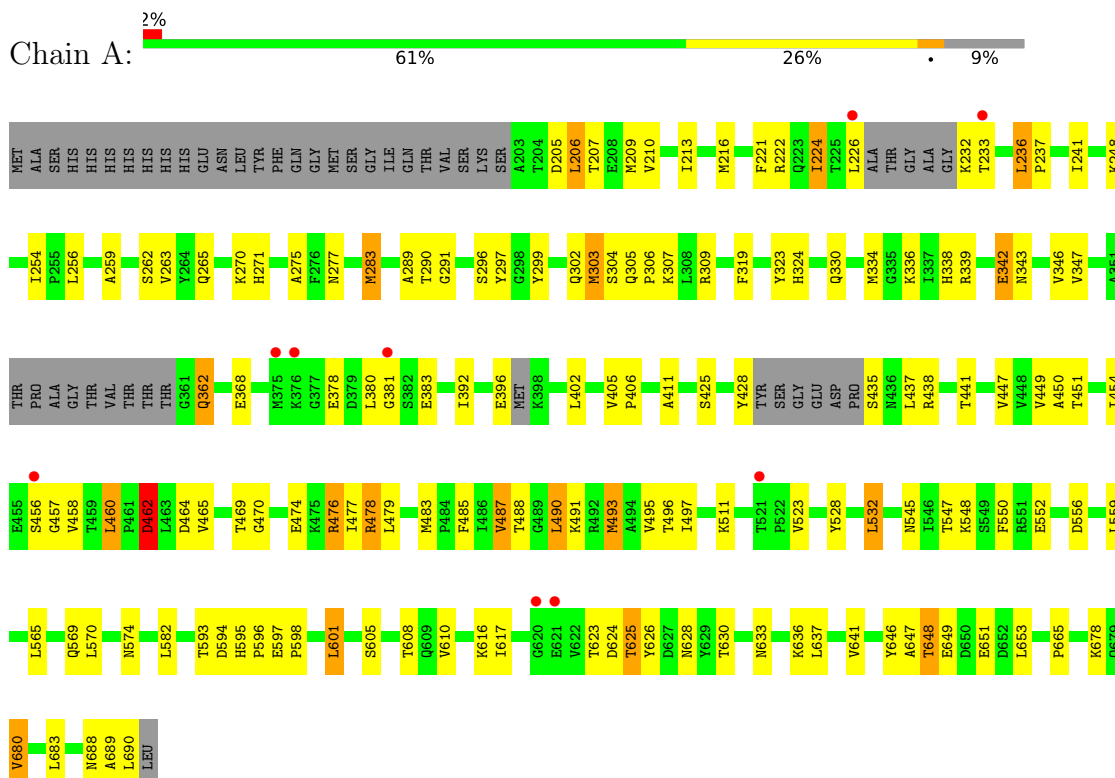
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O 1 1	0	0

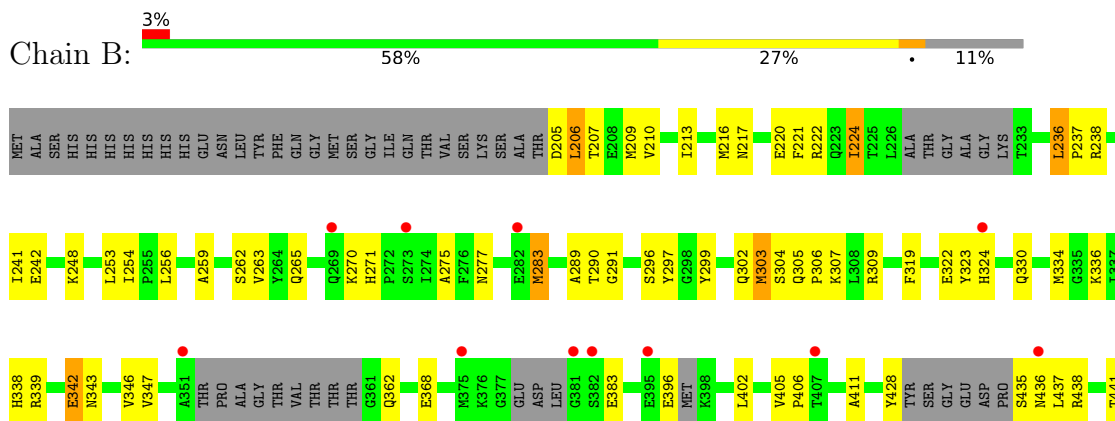
### 3 Residue-property plots

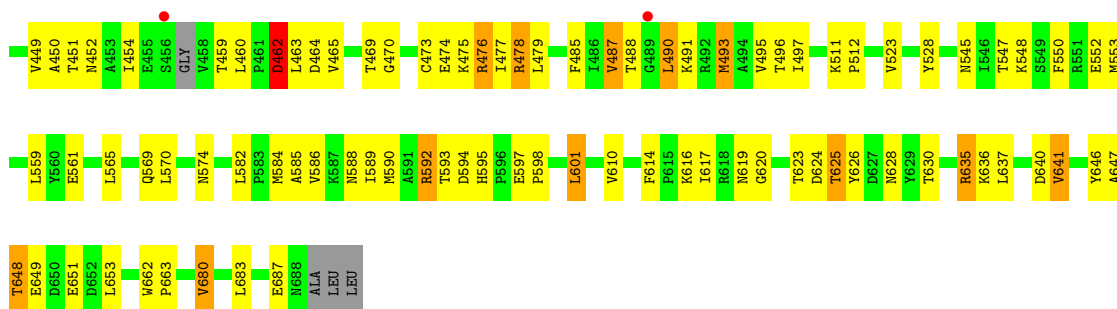
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: SERINE PROTEASE NS3

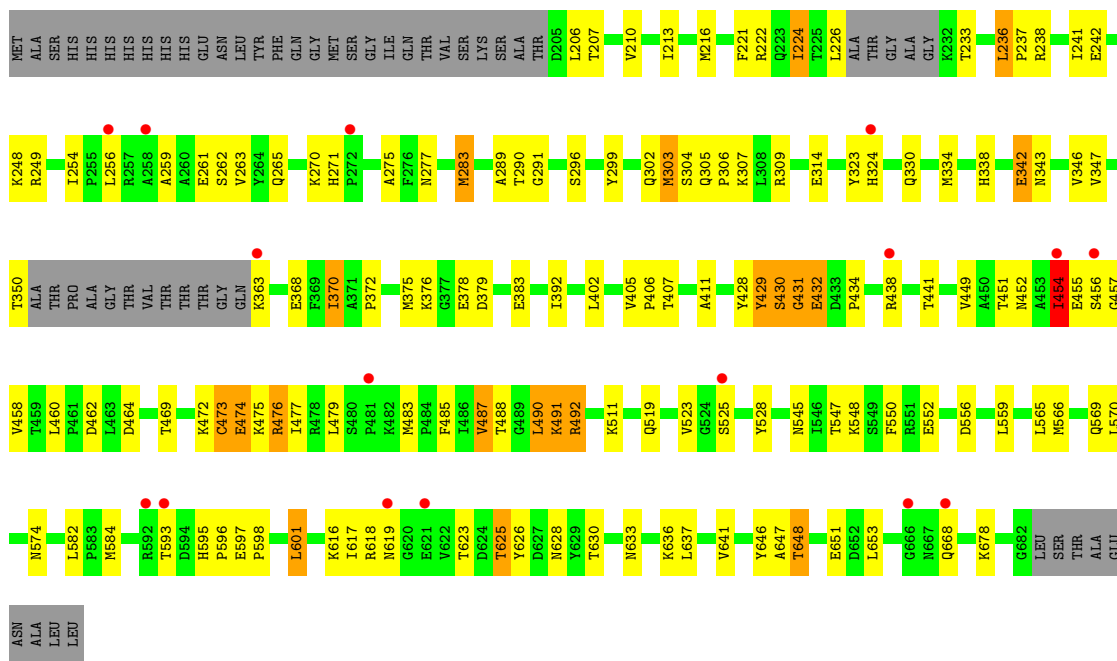


- Molecule 1: SERINE PROTEASE NS3

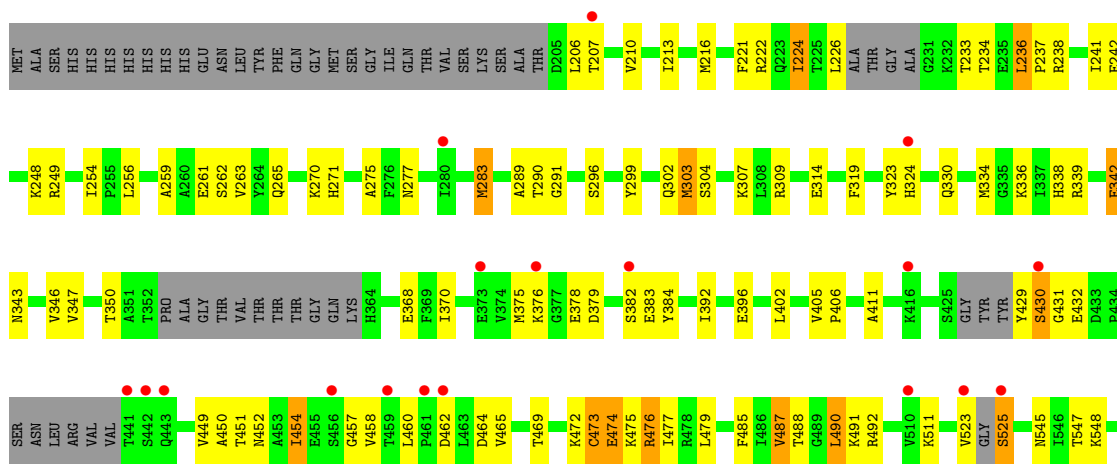


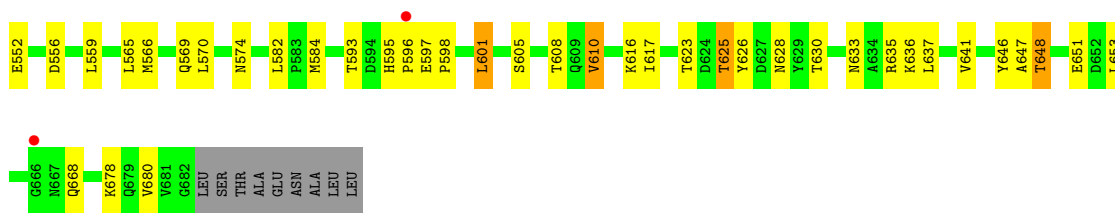


• Molecule 1: SERINE PROTEASE NS3



• Molecule 1: SERINE PROTEASE NS3







## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.10Å 159.59Å 99.23Å 90.00° 97.24° 90.00°	Depositor
Resolution (Å)	20.00 – 3.27 19.98 – 3.27	Depositor EDS
% Data completeness (in resolution range)	94.0 (20.00-3.27) 94.0 (19.98-3.27)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.28 (at 3.29Å)	Xtrriage
Refinement program	BUSTER 2.11.4	Depositor
R, $R_{free}$	0.233 , 0.283 0.264 , 0.316	Depositor DCC
$R_{free}$ test set	1585 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.2	Xtrriage
Anisotropy	1.199	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 59.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	14616	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	64.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% 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 [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	0/3771	0.71	3/5098 (0.1%)
1	B	0.53	0/3706	0.71	4/5007 (0.1%)
1	C	0.53	0/3743	0.73	3/5062 (0.1%)
1	D	0.49	0/3664	0.68	1/4951 (0.0%)
All	All	0.52	0/14884	0.71	11/20118 (0.1%)

There are no bond length outliers.

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	497	ILE	N-CA-C	-8.15	89.00	111.00
1	B	497	ILE	N-CA-C	-8.03	89.33	111.00
1	C	454	ILE	C-N-CA	8.03	141.76	121.70
1	C	431	GLY	C-N-CA	7.21	139.74	121.70
1	A	462	ASP	CA-CB-CG	6.46	127.62	113.40

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	3704	0	3724	119	0
1	B	3641	0	3657	120	0
1	C	3672	0	3685	102	0
1	D	3598	0	3602	98	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	1	0	0	2	0
All	All	14616	0	14668	419	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 14.

The worst 5 of 419 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:450:ALA:CB	1:A:454:ILE:HG21	1.46	1.44
1:A:450:ALA:HB1	1:A:454:ILE:CG2	1.47	1.43
1:B:450:ALA:CB	1:B:454:ILE:HG21	1.48	1.42
1:B:450:ALA:HB1	1:B:454:ILE:CG2	1.49	1.41
1:C:324:HIS:HD2	1:C:350:THR:OG1	1.10	1.33

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	457/516 (89%)	437 (96%)	17 (4%)	3 (1%)	22 56
1	B	444/516 (86%)	432 (97%)	12 (3%)	0	100 100
1	C	455/516 (88%)	434 (95%)	15 (3%)	6 (1%)	12 42
1	D	441/516 (86%)	422 (96%)	15 (3%)	4 (1%)	17 50
All	All	1797/2064 (87%)	1725 (96%)	59 (3%)	13 (1%)	22 56

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	458	VAL

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	432	GLU
1	C	454	ILE
1	C	455	GLU
1	C	458	VAL

### 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	403/442 (91%)	361 (90%)	42 (10%)	7	26
1	B	397/442 (90%)	359 (90%)	38 (10%)	8	29
1	C	401/442 (91%)	354 (88%)	47 (12%)	5	22
1	D	392/442 (89%)	346 (88%)	46 (12%)	5	22
All	All	1593/1768 (90%)	1420 (89%)	173 (11%)	6	24

5 of 173 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	582	LEU
1	D	382	SER
1	C	628	ASN
1	D	236	LEU
1	D	460	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	324	HIS
1	C	536	GLN
1	D	536	GLN
1	C	343	ASN
1	C	574	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 monosaccharides 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	467/516 (90%)	0.16	9 (1%) 66 64	23, 58, 96, 119	0
1	B	458/516 (88%)	0.22	13 (2%) 53 51	15, 60, 89, 116	0
1	C	461/516 (89%)	0.23	16 (3%) 44 42	25, 63, 96, 126	0
1	D	453/516 (87%)	0.40	20 (4%) 34 33	25, 67, 106, 135	0
All	All	1839/2064 (89%)	0.25	58 (3%) 47 46	15, 62, 98, 135	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	454	ILE	4.0
1	D	441	THR	4.0
1	C	363	LYS	3.9
1	D	442	SER	3.8
1	D	461	PRO	3.7

### 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 monosaccharides in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.