WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Dec 16, 2023 – 10:07 pm GMT

PDB ID	:	4CBL
Title	:	Pestivirus NS3 helicase
Authors	:	Tortorici, M.A.; Duquerroy, S.; Kwok, J.; Vonrhein, C.; Perez, J.; Lamp, B.; Bricogne, G.; Rumenapf, T.; Vachette, P.; Rey, F.A.
Deposited on	:	2013-10-14
Resolution	:	3.05 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:

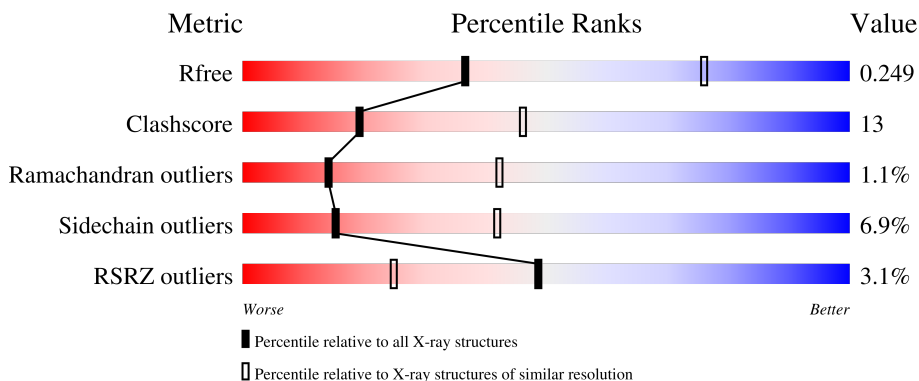|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 3.05 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	Percentile Ranks	Value
Rfree		0.249
Clashscore		13
Ramachandran outliers		1.1%
Sidechain outliers		6.9%
RSRZ outliers		3.1%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1754 (3.10-3.02)
Clashscore	141614	1864 (3.10-3.02)
Ramachandran outliers	138981	1794 (3.10-3.02)
Sidechain outliers	138945	1793 (3.10-3.02)
RSRZ outliers	127900	1713 (3.10-3.02)

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% 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	509	% 67% 24% • 5%
1	B	509	3% 68% 23% • 6%
1	C	509	2% 65% 25% • 8%
1	D	509	5% 64% 24% • 9%

WORLDWIDE PDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 14954 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
1	A	482	Total	C	N	O	S	0	0	0
			3793	2405	646	720	22			
1	B	479	Total	C	N	O	S	0	0	0
			3778	2391	642	724	21			
1	C	468	Total	C	N	O	S	0	0	0
			3715	2360	630	703	22			
1	D	462	Total	C	N	O	S	0	0	0
			3668	2330	621	696	21			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	183	MET	-	expression tag	UNP P19712
A	184	ALA	-	expression tag	UNP P19712
A	185	SER	-	expression tag	UNP P19712
A	186	HIS	-	expression tag	UNP P19712
A	187	HIS	-	expression tag	UNP P19712
A	188	HIS	-	expression tag	UNP P19712
A	189	HIS	-	expression tag	UNP P19712
A	190	HIS	-	expression tag	UNP P19712
A	191	HIS	-	expression tag	UNP P19712
A	192	HIS	-	expression tag	UNP P19712
A	193	MET	-	expression tag	UNP P19712
A	194	SER	-	expression tag	UNP P19712
A	195	GLY	-	expression tag	UNP P19712
A	196	ILE	-	expression tag	UNP P19712
A	197	GLN	-	expression tag	UNP P19712
A	198	THR	-	expression tag	UNP P19712
A	199	VAL	-	expression tag	UNP P19712
A	200	SER	-	expression tag	UNP P19712
A	201	LYS	-	expression tag	UNP P19712
A	202	SER	-	expression tag	UNP P19712
B	183	MET	-	expression tag	UNP P19712

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	184	ALA	-	expression tag	UNP P19712
B	185	SER	-	expression tag	UNP P19712
B	186	HIS	-	expression tag	UNP P19712
B	187	HIS	-	expression tag	UNP P19712
B	188	HIS	-	expression tag	UNP P19712
B	189	HIS	-	expression tag	UNP P19712
B	190	HIS	-	expression tag	UNP P19712
B	191	HIS	-	expression tag	UNP P19712
B	192	HIS	-	expression tag	UNP P19712
B	193	MET	-	expression tag	UNP P19712
B	194	SER	-	expression tag	UNP P19712
B	195	GLY	-	expression tag	UNP P19712
B	196	ILE	-	expression tag	UNP P19712
B	197	GLN	-	expression tag	UNP P19712
B	198	THR	-	expression tag	UNP P19712
B	199	VAL	-	expression tag	UNP P19712
B	200	SER	-	expression tag	UNP P19712
B	201	LYS	-	expression tag	UNP P19712
B	202	SER	-	expression tag	UNP P19712
C	183	MET	-	expression tag	UNP P19712
C	184	ALA	-	expression tag	UNP P19712
C	185	SER	-	expression tag	UNP P19712
C	186	HIS	-	expression tag	UNP P19712
C	187	HIS	-	expression tag	UNP P19712
C	188	HIS	-	expression tag	UNP P19712
C	189	HIS	-	expression tag	UNP P19712
C	190	HIS	-	expression tag	UNP P19712
C	191	HIS	-	expression tag	UNP P19712
C	192	HIS	-	expression tag	UNP P19712
C	193	MET	-	expression tag	UNP P19712
C	194	SER	-	expression tag	UNP P19712
C	195	GLY	-	expression tag	UNP P19712
C	196	ILE	-	expression tag	UNP P19712
C	197	GLN	-	expression tag	UNP P19712
C	198	THR	-	expression tag	UNP P19712
C	199	VAL	-	expression tag	UNP P19712
C	200	SER	-	expression tag	UNP P19712
C	201	LYS	-	expression tag	UNP P19712
C	202	SER	-	expression tag	UNP P19712
D	183	MET	-	expression tag	UNP P19712
D	184	ALA	-	expression tag	UNP P19712
D	185	SER	-	expression tag	UNP P19712

Continued on next page...

Continued from previous page...

| Chain	Residue	Modelled	Actual	Comment	Reference
D	186	HIS	-	expression tag	UNP P19712
D	187	HIS	-	expression tag	UNP P19712
D	188	HIS	-	expression tag	UNP P19712
D	189	HIS	-	expression tag	UNP P19712
D	190	HIS	-	expression tag	UNP P19712
D	191	HIS	-	expression tag	UNP P19712
D	192	HIS	-	expression tag	UNP P19712
D	193	MET	-	expression tag	UNP P19712
D	194	SER	-	expression tag	UNP P19712
D	195	GLY	-	expression tag	UNP P19712
D	196	ILE	-	expression tag	UNP P19712
D	197	GLN	-	expression tag	UNP P19712
D	198	THR	-	expression tag	UNP P19712
D	199	VAL	-	expression tag	UNP P19712
D	200	SER	-	expression tag	UNP P19712
D	201	LYS	-	expression tag	UNP P19712
D	202	SER	-	expression tag	UNP P19712
A	588	TYR	ASN	engineered mutation	UNP P10725
B	588	TYR	ASN	engineered mutation	UNP P10725
C	588	TYR	ASN	engineered mutation	UNP P10725
D	588	TYR	ASN	engineered mutation	UNP P10725

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.

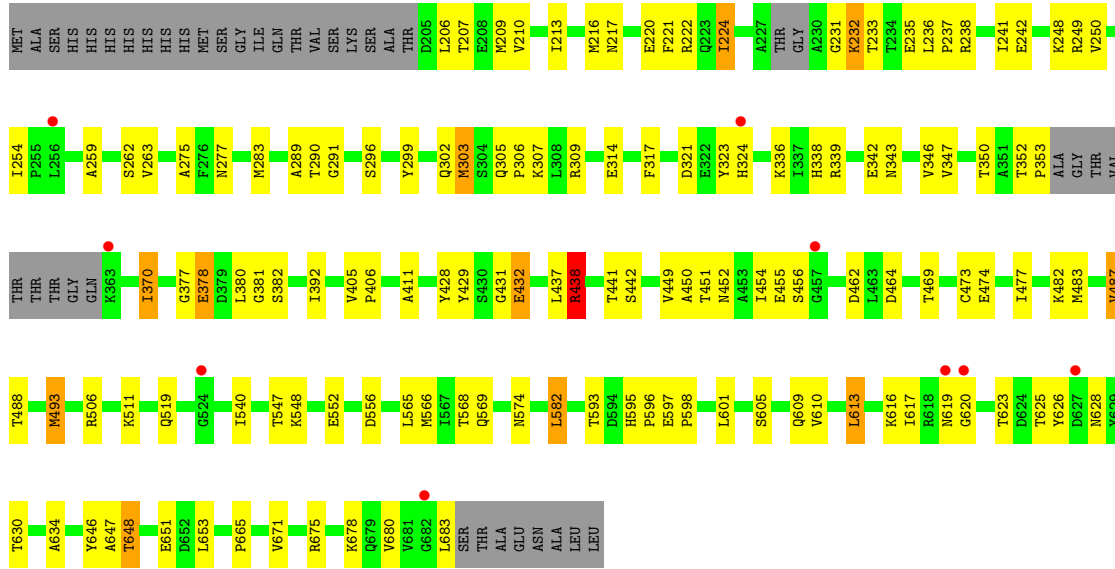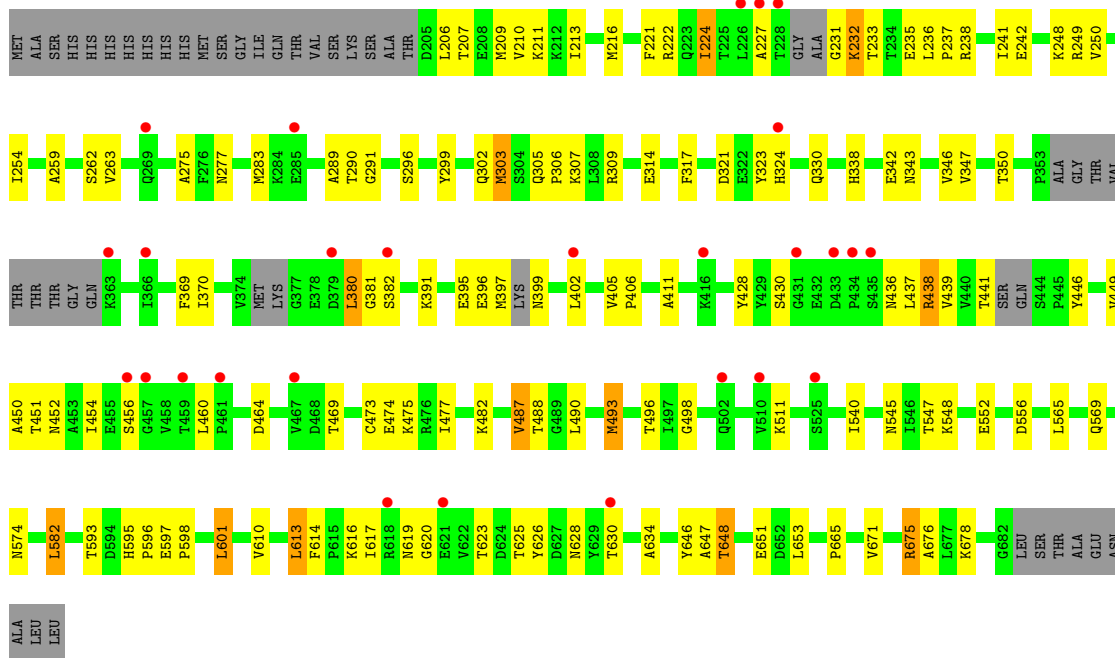• Molecule 1: SERINE PROTEASE NS3
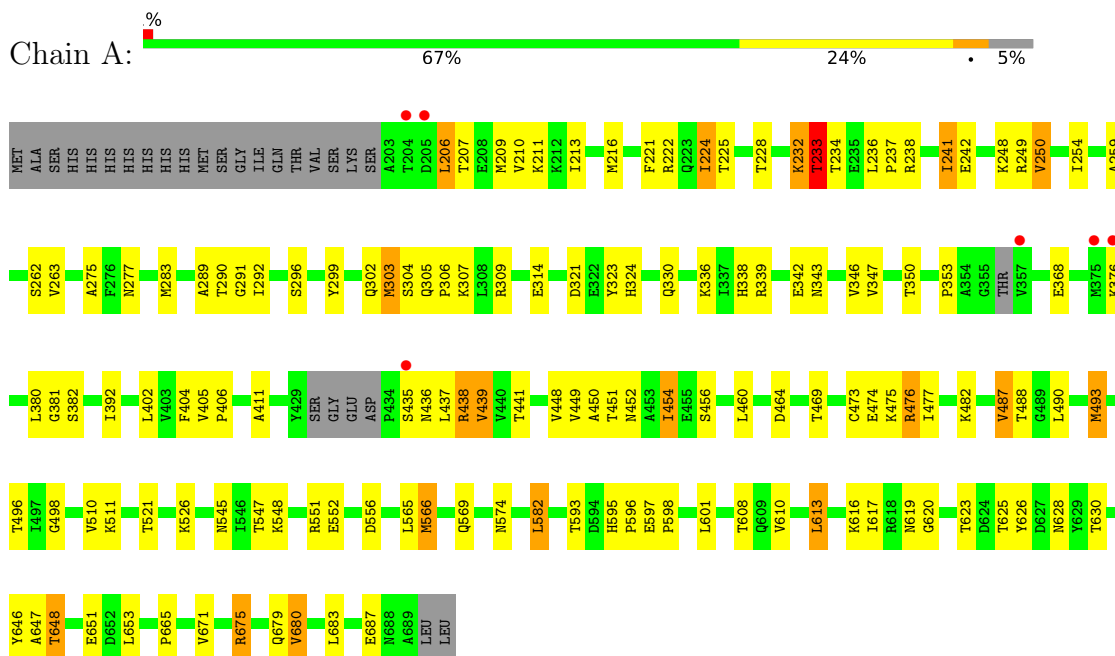
• Molecule 1: SERINE PROTEASE NS3

• Molecule 1: SERINE PROTEASE NS3

Chain C:

2%　　　65%　　　25%　　8%

• Molecule 1: SERINE PROTEASE NS3

Chain D:

5%　　　64%　　　24%　　9%
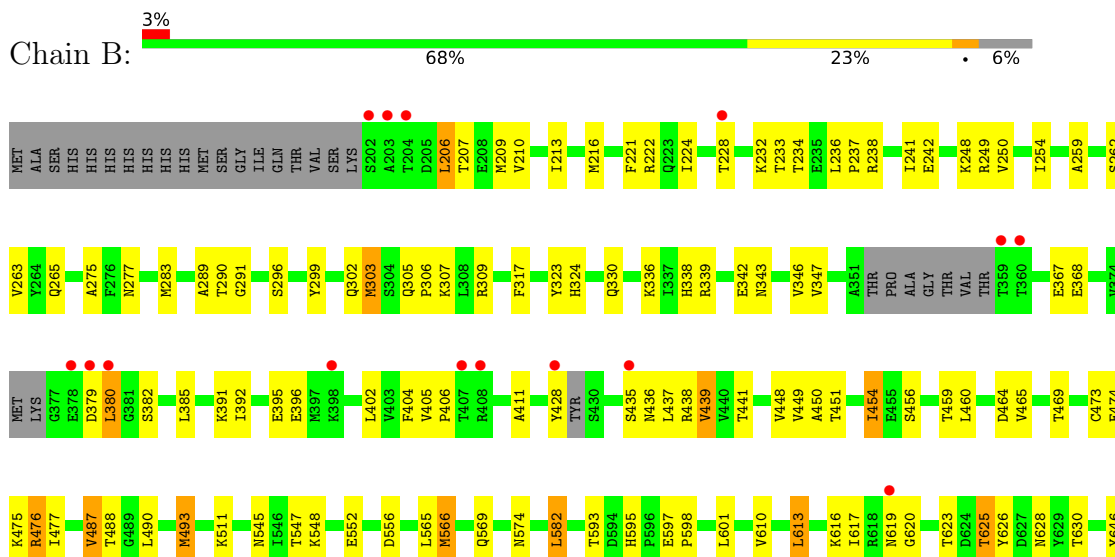
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	69.91Å 156.84Å 99.40Å 90.00° 97.21° 90.00°	Depositor
Resolution (Å)	20.00 – 3.05 19.99 – 3.05	Depositor EDS
% Data completeness (in resolution range)	94.9 (20.00-3.05) 94.9 (19.99-3.05)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.73 (at 3.04Å)	Xtriage
Refinement program	BUSTER 2.11.4	Depositor
R, R_{free}	0.194 , 0.226 0.218 , 0.249	Depositor DCC
R_{free} test set	1915 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	70.3	Xtriage
Anisotropy	0.539	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 73.5	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14954	wwPDB-VP
Average B, all atoms (Å^2)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.31% 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   ⓘ

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	0.54	0/3866	0.71	2/5233 (0.0%)
1	B	0.51	0/3848	0.69	0/5208
1	C	0.53	1/3788 (0.0%)	0.74	5/5126 (0.1%)
1	D	0.48	0/3738	0.68	0/5057
All	All	0.52	1/15240 (0.0%)	0.71	7/20624 (0.0%)

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	B	0	1
1	C	0	1
All	All	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	483	MET	SD-CE	-7.27	1.37	1.77

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	232	LYS	N-CA-C	-6.66	93.01	111.00
1	C	432	GLU	CA-C-O	-6.41	106.63	120.10
1	C	675	ARG	CA-CB-CG	6.10	126.83	113.40
1	C	378	GLU	N-CA-C	5.71	126.43	111.00
1	A	437	LEU	CA-CB-CG	5.48	127.91	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	342	GLU	Sidechain
1	C	378	GLU	Mainchain

5.2   Too-close contacts   ⓘ

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	3793	0	3787	112	1
1	B	3778	0	3777	107	1
1	C	3715	0	3721	95	0
1	D	3668	0	3661	93	0
All	All	14954	0	14946	386	1

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 13.

The worst 5 of 386 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:323:TYR:CE1	1:B:324:HIS:CE1	2.04	1.43
1:C:324:HIS:HD2	1:C:350:THR:OG1	1.12	1.31
1:A:323:TYR:HE1	1:A:324:HIS:CE1	1.51	1.26
1:A:451:THR:O	1:A:454:ILE:HG13	1.36	1.26
1:B:451:THR:O	1:B:454:ILE:HG13	1.35	1.22

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:510:VAL:CG2	1:B:265:GLN:NE2[1_554]	2.10	0.10

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	476/509 (94%)	459 (96%)	12 (2%)	5 (1%)	14	42
1	B	471/509 (92%)	454 (96%)	14 (3%)	3 (1%)	25	55
1	C	462/509 (91%)	438 (95%)	17 (4%)	7 (2%)	10	35
1	D	450/509 (88%)	430 (96%)	14 (3%)	6 (1%)	12	38
All	All	1859/2036 (91%)	1781 (96%)	57 (3%)	21 (1%)	14	42

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	456	SER
1	B	456	SER
1	C	352	THR
1	C	432	GLU
1	C	456	SER

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	407/436 (93%)	373 (92%)	34 (8%)	11	34
1	B	410/436 (94%)	380 (93%)	30 (7%)	14	40
1	C	403/436 (92%)	379 (94%)	24 (6%)	19	46
1	D	398/436 (91%)	374 (94%)	24 (6%)	19	46
All	All	1618/1744 (93%)	1506 (93%)	112 (7%)	15	42

5 of 112 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	625	THR
1	D	675	ARG
1	C	392	ILE
1	D	648	THR
1	D	482	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 32 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	536	GLN
1	D	574	ASN
1	B	574	ASN
1	B	536	GLN
1	D	595	HIS

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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, 95th 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^2)	Q<0.9
1	A	482/509 (94%)	-0.14	6 (1%)	79	58	39, 71, 111, 147	0
1	B	479/509 (94%)	0.00	16 (3%)	46	23	41, 79, 123, 142	0
1	C	468/509 (91%)	-0.04	9 (1%)	66	43	41, 78, 117, 134	0
1	D	462/509 (90%)	0.22	27 (5%)	23	9	43, 89, 128, 151	0
All	All	1891/2036 (92%)	0.01	58 (3%)	49	25	39, 79, 121, 151	0

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	359	THR	4.7
1	D	363	LYS	3.9
1	A	204	THR	3.9
1	B	204	THR	3.8
1	A	205	ASP	3.7

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

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

6.3 Carbohydrates ⓘ

There are no monosaccharides 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.