wwPDB X-ray Structure Validation Summary Report (i)

May 19, 2020 – 05:05 am BST

PDB ID	:	4RS4
Title	:	Crystal structure and mutational analysis of the endoribonuclease from human coronavirus 229E
Authors	:	Huo, T.; Liu, X.; Yang, C.; Rao, Z.
Deposited on	:	2014-11-06
Resolution	:	2.96 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 following versions of software and data (see references (1)) were used in the production of this report:

|		
MolProbity	:	4.02b-467
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 2.96 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.
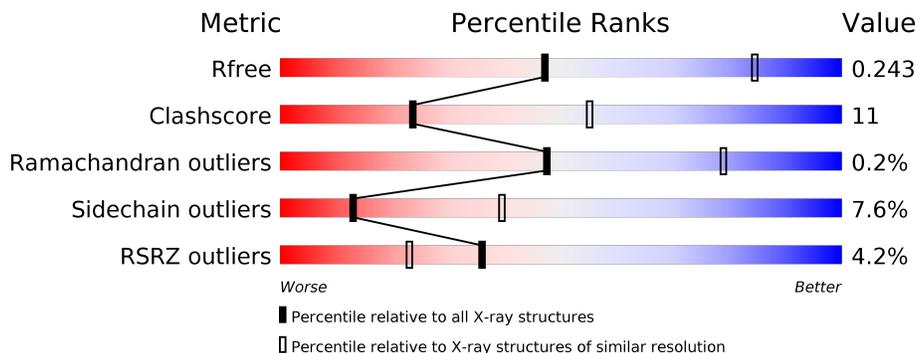
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3104 (3.00-2.92)
Clashscore	141614	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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 >=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	349	4% 73% 23% · ·
1	B	349	% 75% 20% · ·
1	C	349	9% 62% 31% · 5%
1	D	349	9% 62% 31% · ·
1	E	349	% 68% 28% · ·
1	F	349	% 73% 23% · ·

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 15989 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 Uridylate-specific endoribonuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	341	Total 2674	C 1724	N 432	O 504	S 14	0	0	0
1	B	345	Total 2695	C 1735	N 437	O 509	S 14	0	0	0
1	C	332	Total 2601	C 1679	N 418	O 490	S 14	0	0	0
1	D	337	Total 2636	C 1703	N 424	O 495	S 14	0	0	0
1	E	341	Total 2669	C 1721	N 430	O 504	S 14	0	0	0
1	F	346	Total 2714	C 1746	N 439	O 515	S 14	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	EXPRESSION TAG	UNP P0C6X1
A	17	SER	GLY	ENGINEERED MUTATION	UNP P0C6X1
A	142	ALA	THR	ENGINEERED MUTATION	UNP P0C6X1
A	219	MET	ILE	ENGINEERED MUTATION	UNP P0C6X1
A	252	SER	LEU	ENGINEERED MUTATION	UNP P0C6X1
B	0	SER	-	EXPRESSION TAG	UNP P0C6X1
B	17	SER	GLY	ENGINEERED MUTATION	UNP P0C6X1
B	142	ALA	THR	ENGINEERED MUTATION	UNP P0C6X1
B	219	MET	ILE	ENGINEERED MUTATION	UNP P0C6X1
B	252	SER	LEU	ENGINEERED MUTATION	UNP P0C6X1
C	0	SER	-	EXPRESSION TAG	UNP P0C6X1
C	17	SER	GLY	ENGINEERED MUTATION	UNP P0C6X1
C	142	ALA	THR	ENGINEERED MUTATION	UNP P0C6X1
C	219	MET	ILE	ENGINEERED MUTATION	UNP P0C6X1
C	252	SER	LEU	ENGINEERED MUTATION	UNP P0C6X1
D	0	SER	-	EXPRESSION TAG	UNP P0C6X1
D	17	SER	GLY	ENGINEERED MUTATION	UNP P0C6X1

Continued from previous page...

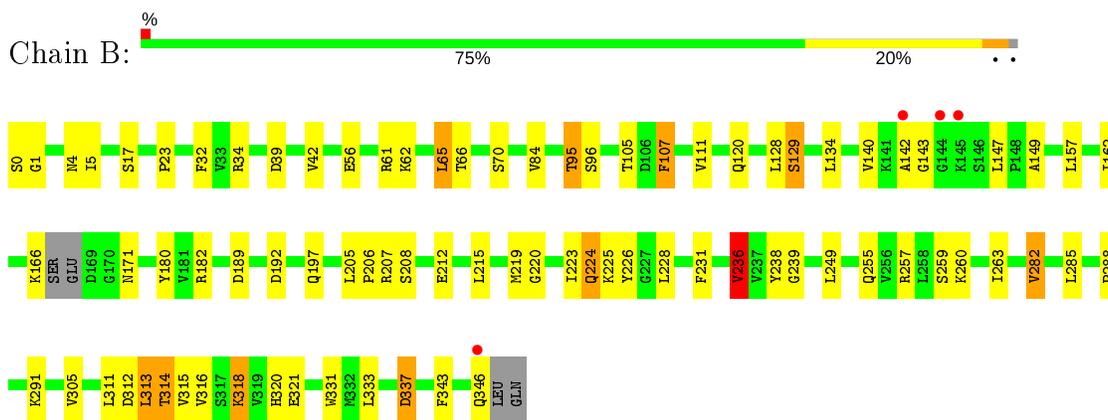| Chain	Residue	Modelled	Actual	Comment	Reference
D	142	ALA	THR	ENGINEERED MUTATION	UNP P0C6X1
D	219	MET	ILE	ENGINEERED MUTATION	UNP P0C6X1
D	252	SER	LEU	ENGINEERED MUTATION	UNP P0C6X1
E	0	SER	-	EXPRESSION TAG	UNP P0C6X1
E	17	SER	GLY	ENGINEERED MUTATION	UNP P0C6X1
E	142	ALA	THR	ENGINEERED MUTATION	UNP P0C6X1
E	219	MET	ILE	ENGINEERED MUTATION	UNP P0C6X1
E	252	SER	LEU	ENGINEERED MUTATION	UNP P0C6X1
F	0	SER	-	EXPRESSION TAG	UNP P0C6X1
F	17	SER	GLY	ENGINEERED MUTATION	UNP P0C6X1
F	142	ALA	THR	ENGINEERED MUTATION	UNP P0C6X1
F	219	MET	ILE	ENGINEERED MUTATION	UNP P0C6X1
F	252	SER	LEU	ENGINEERED MUTATION	UNP P0C6X1

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.
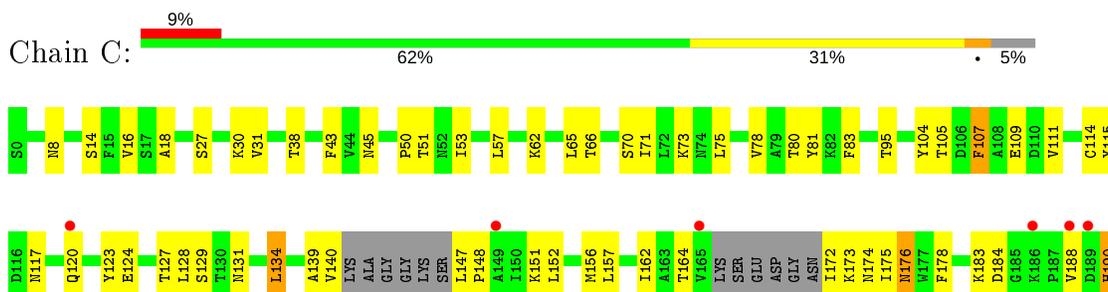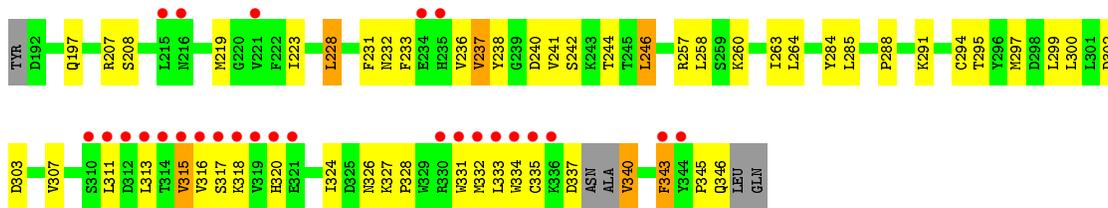
- Molecule 1: Uridylate-specific endoribonuclease

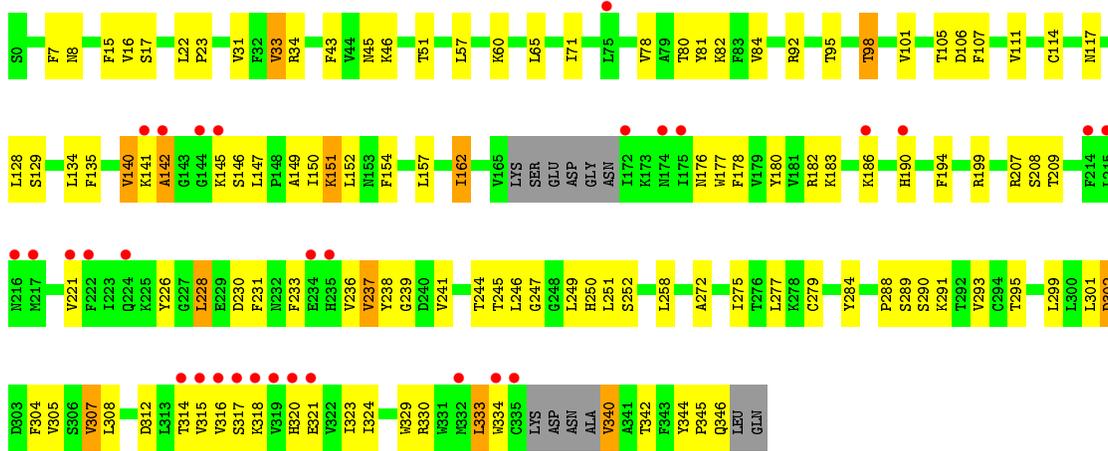- Molecule 1: Uridylate-specific endoribonuclease

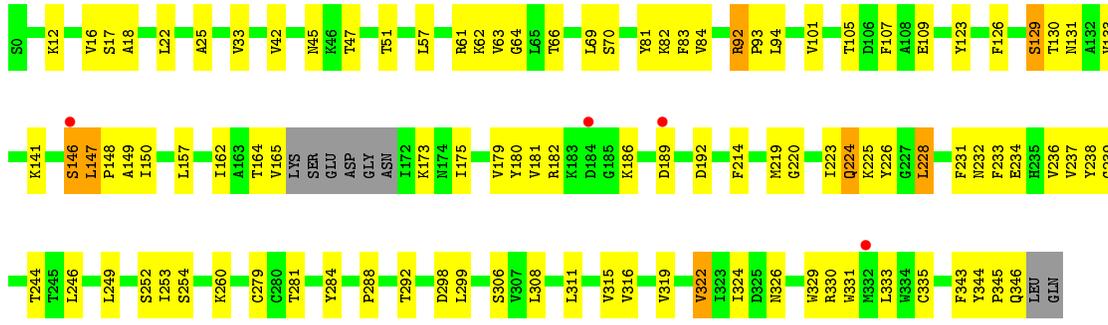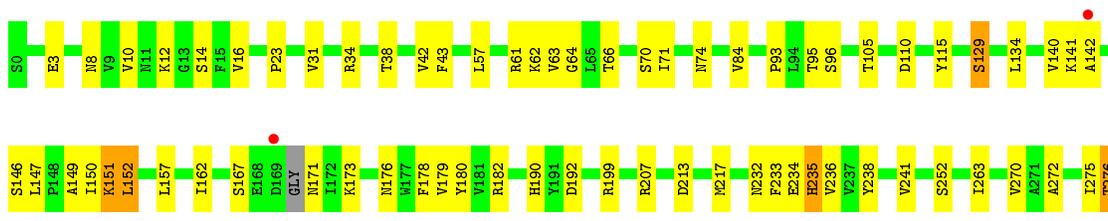- Molecule 1: Uridylate-specific endoribonuclease

TYR D192 Q197 R207 S208 L215 N216 M219 G220 V221 F222 I223 L228 F231 N232 F233 E234 H235 V236 V237 Y238 G239 D240 V241 S242 K243 T244 T245 L246 R257 L258 S259 K260 I263 L264 Y284 L285 P288 K291 C294 Y295 M297 D298 L299 L300 L301 D302

D303 V307 S310 L311 D312 L313 T314 V315 V316 S317 K318 V319 H320 E321 I324 D325 N326 K327 P328 W329 R330 W331 M332 L333 W334 C335 D336 D337 ASN ALA V340 F343 Y344 P345 Q346 LEU GLN

• Molecule 1: Uridylate-specific endoribonuclease

Chain D:

9%
62%   31%

S0 F7 N8 F15 V16 S17 L22 P23 V31 F32 V33 R34 F43 V44 N45 K46 T51 L57 K60 L65 I71 L75 V78 A79 T80 Y81 K82 F83 V84 R92 T95 T98 V101 T105 D106 F107 V111 C114 N117

L128 S129 L134 F135 V140 K141 A142 G143 G144 K145 S146 L147 P148 A149 I150 K151 L152 N153 F154 L157 I162 V165 LYS SER GLU ASP GLY ASN I172 K173 N174 I175 N176 W177 F178 V179 Y180 V181 R182 K183 K186 H190 F194 R199 R207 S208 T209 F214 L215

N216 M217 V221 F222 I223 Q224 K225 Y226 G227 L228 E229 D230 F231 N232 F233 E234 H235 V236 V237 Y238 G239 D240 V241 T244 T245 L246 G247 Q248 L249 H250 L251 S252 L258 A272 I275 T276 L277 K278 C279 Y284 P288 S289 S290 K291 T292 V293 C294 T295 L299 I300 L301 D302

D303 F304 V305 S306 V307 L308 D312 L313 T314 V315 V316 S317 K318 V319 H320 E321 V322 I323 I324 W329 R330 M331 M332 L333 W334 C335 LYS ASP ASN ALA V340 A341 T342 F343 Y344 P345 Q346 LEU GLN

• Molecule 1: Uridylate-specific endoribonuclease

Chain E:

%
68%   28%

S0 K12 V16 S17 A18 L22 A25 V33 V42 N45 K46 T47 T51 L57 R61 K62 V63 G64 L65 T66 L69 S70 Y81 K82 F83 V84 R92 P93 L94 V101 T105 D106 F107 A108 E109 Y123 F126 S129 T130 N131 A132 V133

K141 S146 L147 P148 A149 I150 L157 I162 A163 T164 V165 LYS SER GLU ASP GLY ASN I172 K173 N174 I175 V179 Y180 V181 R182 K183 D184 G185 K186 D189 D192 F214 M219 G220 I223 Q224 K225 Y226 G227 L228 F231 N232 F233 E234 H235 V236 V237 Y238 G239

T244 T245 L246 L249 S252 I253 S254 K260 C279 C280 T281 Y284 P288 T292 D298 L299 S306 V307 L308 L311 V315 V316 V319 V322 I323 I324 D325 N326 W329 R330 W331 M332 L333 W334 C335 F343 Y344 P345 Q346 LEU GLN

• Molecule 1: Uridylate-specific endoribonuclease

Chain F:

%
73%   23%

S0 E3 N8 V9 V10 N11 K12 G13 S14 F15 V16 P23 V31 R34 T38 V42 F43 L57 R61 K62 V63 G64 L65 T66 S70 I71 N74 V84 P93 L94 T95 S96 T105 D110 Y115 S129 L134 V140 K141 A142

S146 L147 P148 A149 I150 K151 L152 L157 I162 S167 E168 D169 GLY N171 I172 K173 N176 W177 F178 V179 V180 V181 R182 H190 Y191 V236 D192 R199 R207 D213 M217 N232 F233 E234 H235 V236 V237 Y238 V241 S252 I263 V270 A271 A272 I275 T276

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	127.93Å 143.17Å 174.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.96 38.25 – 2.96	Depositor EDS
% Data completeness (in resolution range)	90.4 (30.00-2.96) 89.3 (38.25-2.96)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.59 (at 2.95Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.208 , 0.255 0.195 , 0.243	Depositor DCC
R_{free} test set	3387 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	56.9	Xtriage
Anisotropy	0.595	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 45.5	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	15989	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.46% 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)

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	0.31	0/2732	0.49	1/3708 (0.0%)
1	B	0.30	0/2753	0.49	0/3737
1	C	0.28	0/2655	0.46	0/3604
1	D	0.28	0/2692	0.47	0/3653
1	E	0.30	0/2726	0.47	0/3700
1	F	0.29	0/2772	0.47	0/3761
All	All	0.29	0/16330	0.47	1/22163 (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	D	0	1
1	F	0	1
All	All	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	109	GLU	N-CA-C	-6.04	94.69	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	142	ALA	Peptide
1	F	142	ALA	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	2674	0	2657	58	0
1	B	2695	0	2668	58	0
1	C	2601	0	2568	68	0
1	D	2636	0	2621	74	0
1	E	2669	0	2652	65	0
1	F	2714	0	2691	49	0
All	All	15989	0	15857	358	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 11.

The worst 5 of 358 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:315:VAL:HG11	1:B:318:LYS:HE2	1.42	1.01
1:F:157:LEU:HB2	1:F:162:ILE:HD13	1.58	0.86
1:A:288:PRO:HG3	1:D:162:ILE:HD11	1.57	0.85
1:D:334:TRP:HZ3	1:D:344:TYR:HE2	1.24	0.84
1:A:164:THR:HG23	1:A:174:ASN:HA	1.61	0.82

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	337/349 (97%)	325 (96%)	12 (4%)	0	100 100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	341/349 (98%)	328 (96%)	12 (4%)	1 (0%)	41	73
1	C	322/349 (92%)	307 (95%)	14 (4%)	1 (0%)	41	73
1	D	331/349 (95%)	311 (94%)	18 (5%)	2 (1%)	25	60
1	E	337/349 (97%)	321 (95%)	16 (5%)	0	100	100
1	F	342/349 (98%)	332 (97%)	10 (3%)	0	100	100
All	All	2010/2094 (96%)	1924 (96%)	82 (4%)	4 (0%)	47	79

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	236	VAL
1	D	237	VAL
1	D	190	HIS
1	C	246	LEU

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	298/306 (97%)	281 (94%)	17 (6%)	20	52
1	B	299/306 (98%)	277 (93%)	22 (7%)	13	41
1	C	290/306 (95%)	263 (91%)	27 (9%)	9	30
1	D	293/306 (96%)	266 (91%)	27 (9%)	9	30
1	E	297/306 (97%)	275 (93%)	22 (7%)	13	41
1	F	303/306 (99%)	283 (93%)	20 (7%)	16	46
All	All	1780/1836 (97%)	1645 (92%)	135 (8%)	13	39

5 of 135 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	315	VAL
1	D	162	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	235	HIS
1	C	324	ILE
1	D	84	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	326	ASN
1	D	117	ASN
1	F	171	ASN
1	C	232	ASN
1	E	232	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, 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	341/349 (97%)	0.11	14 (4%)	37	24	32, 68, 115, 153	0
1	B	345/349 (98%)	-0.04	4 (1%)	79	63	33, 56, 100, 137	0
1	C	332/349 (95%)	0.44	32 (9%)	8	4	45, 80, 125, 144	0
1	D	337/349 (96%)	0.49	30 (8%)	9	5	37, 85, 145, 171	0
1	E	341/349 (97%)	0.02	4 (1%)	79	63	39, 62, 99, 136	0
1	F	346/349 (99%)	-0.02	2 (0%)	89	78	35, 52, 93, 143	0
All	All	2042/2094 (97%)	0.17	86 (4%)	36	23	32, 67, 118, 171	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	144	GLY	6.9
1	D	317	SER	6.7
1	D	316	VAL	6.7
1	D	314	THR	5.4
1	D	319	VAL	4.8

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 (i)

There are no such residues in this entry.