wwPDB X-ray Structure Validation Summary Report (i)

Nov 1, 2021 – 11:43 PM EDT

PDB ID	:	1WNC
Title	:	Crystal structure of the SARS-CoV Spike protein fusion core
Authors	:	Xu, Y.; Lou, Z.; Liu, Y.; Pang, H.; Tien, P.; Gao, G.F.; Rao, Z.
Deposited on	:	2004-07-29
Resolution	:	2.80 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:

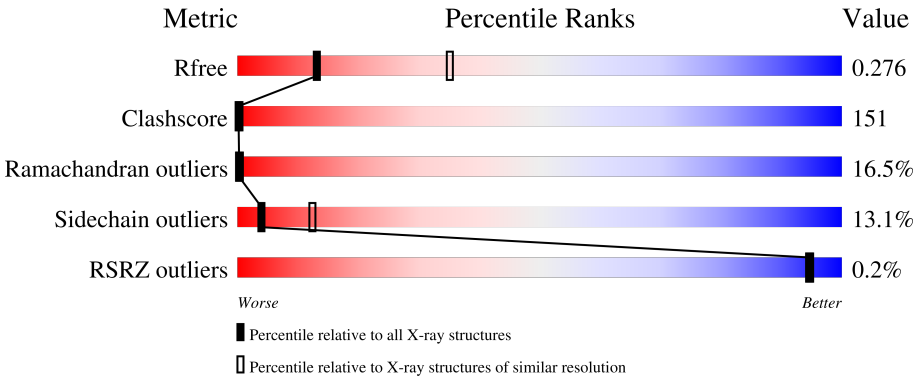
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	112	11% 42% 9% 38%
1	B	112	9% 52% 11% • 28%
1	C	112	• 50% 16% 30%
1	D	112	• 43% 21% • 31%
1	E	112	% • 50% 12% 34%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	112	43% 21% 33%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4131 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 E2 glycoprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	69	Total 533	C 326	N 97	O 110	0	0	0
1	B	81	Total 617	C 378	N 108	O 131	0	0	0
1	C	78	Total 604	C 369	N 107	O 128	0	0	0
1	D	77	Total 582	C 354	N 105	O 123	0	0	0
1	E	74	Total 568	C 348	N 103	O 117	0	0	0
1	F	75	Total 576	C 352	N 103	O 121	0	0	0

There are 132 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1122	LEU	-	linker	UNP P59594
A	1123	VAL	-	linker	UNP P59594
A	1124	PRO	-	linker	UNP P59594
A	1125	ARG	-	linker	UNP P59594
A	1126	GLY	-	linker	UNP P59594
A	1127	SER	-	linker	UNP P59594
A	1128	GLY	-	linker	UNP P59594
A	1129	GLY	-	linker	UNP P59594
A	1130	SER	-	linker	UNP P59594
A	1131	GLY	-	linker	UNP P59594
A	1132	GLY	-	linker	UNP P59594
A	1133	SER	-	linker	UNP P59594
A	1134	GLY	-	linker	UNP P59594
A	1135	GLY	-	linker	UNP P59594
A	1136	LEU	-	linker	UNP P59594
A	1137	GLU	-	linker	UNP P59594
A	1138	VAL	-	linker	UNP P59594

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1139	LEU	-	linker	UNP P59594
A	1140	PHE	-	linker	UNP P59594
A	1141	GLN	-	linker	UNP P59594
A	1142	GLY	-	linker	UNP P59594
A	1162	GLU	LYS	engineered mutation	UNP P59594
B	1122	LEU	-	linker	UNP P59594
B	1123	VAL	-	linker	UNP P59594
B	1124	PRO	-	linker	UNP P59594
B	1125	ARG	-	linker	UNP P59594
B	1126	GLY	-	linker	UNP P59594
B	1127	SER	-	linker	UNP P59594
B	1128	GLY	-	linker	UNP P59594
B	1129	GLY	-	linker	UNP P59594
B	1130	SER	-	linker	UNP P59594
B	1131	GLY	-	linker	UNP P59594
B	1132	GLY	-	linker	UNP P59594
B	1133	SER	-	linker	UNP P59594
B	1134	GLY	-	linker	UNP P59594
B	1135	GLY	-	linker	UNP P59594
B	1136	LEU	-	linker	UNP P59594
B	1137	GLU	-	linker	UNP P59594
B	1138	VAL	-	linker	UNP P59594
B	1139	LEU	-	linker	UNP P59594
B	1140	PHE	-	linker	UNP P59594
B	1141	GLN	-	linker	UNP P59594
B	1142	GLY	-	linker	UNP P59594
B	1162	GLU	LYS	engineered mutation	UNP P59594
C	1122	LEU	-	linker	UNP P59594
C	1123	VAL	-	linker	UNP P59594
C	1124	PRO	-	linker	UNP P59594
C	1125	ARG	-	linker	UNP P59594
C	1126	GLY	-	linker	UNP P59594
C	1127	SER	-	linker	UNP P59594
C	1128	GLY	-	linker	UNP P59594
C	1129	GLY	-	linker	UNP P59594
C	1130	SER	-	linker	UNP P59594
C	1131	GLY	-	linker	UNP P59594
C	1132	GLY	-	linker	UNP P59594
C	1133	SER	-	linker	UNP P59594
C	1134	GLY	-	linker	UNP P59594
C	1135	GLY	-	linker	UNP P59594
C	1136	LEU	-	linker	UNP P59594

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	1137	GLU	-	linker	UNP P59594
C	1138	VAL	-	linker	UNP P59594
C	1139	LEU	-	linker	UNP P59594
C	1140	PHE	-	linker	UNP P59594
C	1141	GLN	-	linker	UNP P59594
C	1142	GLY	-	linker	UNP P59594
C	1162	GLU	LYS	engineered mutation	UNP P59594
D	1122	LEU	-	linker	UNP P59594
D	1123	VAL	-	linker	UNP P59594
D	1124	PRO	-	linker	UNP P59594
D	1125	ARG	-	linker	UNP P59594
D	1126	GLY	-	linker	UNP P59594
D	1127	SER	-	linker	UNP P59594
D	1128	GLY	-	linker	UNP P59594
D	1129	GLY	-	linker	UNP P59594
D	1130	SER	-	linker	UNP P59594
D	1131	GLY	-	linker	UNP P59594
D	1132	GLY	-	linker	UNP P59594
D	1133	SER	-	linker	UNP P59594
D	1134	GLY	-	linker	UNP P59594
D	1135	GLY	-	linker	UNP P59594
D	1136	LEU	-	linker	UNP P59594
D	1137	GLU	-	linker	UNP P59594
D	1138	VAL	-	linker	UNP P59594
D	1139	LEU	-	linker	UNP P59594
D	1140	PHE	-	linker	UNP P59594
D	1141	GLN	-	linker	UNP P59594
D	1142	GLY	-	linker	UNP P59594
D	1162	GLU	LYS	engineered mutation	UNP P59594
E	1122	LEU	-	linker	UNP P59594
E	1123	VAL	-	linker	UNP P59594
E	1124	PRO	-	linker	UNP P59594
E	1125	ARG	-	linker	UNP P59594
E	1126	GLY	-	linker	UNP P59594
E	1127	SER	-	linker	UNP P59594
E	1128	GLY	-	linker	UNP P59594
E	1129	GLY	-	linker	UNP P59594
E	1130	SER	-	linker	UNP P59594
E	1131	GLY	-	linker	UNP P59594
E	1132	GLY	-	linker	UNP P59594
E	1133	SER	-	linker	UNP P59594
E	1134	GLY	-	linker	UNP P59594

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	1135	GLY	-	linker	UNP P59594
E	1136	LEU	-	linker	UNP P59594
E	1137	GLU	-	linker	UNP P59594
E	1138	VAL	-	linker	UNP P59594
E	1139	LEU	-	linker	UNP P59594
E	1140	PHE	-	linker	UNP P59594
E	1141	GLN	-	linker	UNP P59594
E	1142	GLY	-	linker	UNP P59594
E	1162	GLU	LYS	engineered mutation	UNP P59594
F	1122	LEU	-	linker	UNP P59594
F	1123	VAL	-	linker	UNP P59594
F	1124	PRO	-	linker	UNP P59594
F	1125	ARG	-	linker	UNP P59594
F	1126	GLY	-	linker	UNP P59594
F	1127	SER	-	linker	UNP P59594
F	1128	GLY	-	linker	UNP P59594
F	1129	GLY	-	linker	UNP P59594
F	1130	SER	-	linker	UNP P59594
F	1131	GLY	-	linker	UNP P59594
F	1132	GLY	-	linker	UNP P59594
F	1133	SER	-	linker	UNP P59594
F	1134	GLY	-	linker	UNP P59594
F	1135	GLY	-	linker	UNP P59594
F	1136	LEU	-	linker	UNP P59594
F	1137	GLU	-	linker	UNP P59594
F	1138	VAL	-	linker	UNP P59594
F	1139	LEU	-	linker	UNP P59594
F	1140	PHE	-	linker	UNP P59594
F	1141	GLN	-	linker	UNP P59594
F	1142	GLY	-	linker	UNP P59594
F	1162	GLU	LYS	engineered mutation	UNP P59594

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	100	Total O 100 100	0	0
2	B	124	Total O 124 124	0	0
2	C	108	Total O 108 108	0	0
2	D	99	Total O 99 99	0	0

Continued on next page...

Continued from previous page...

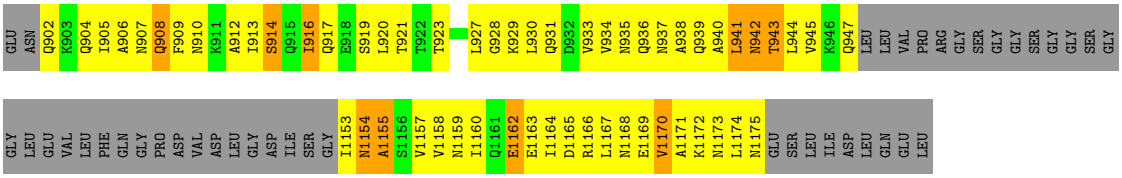
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	E	95	Total O 95 95	0	0
2	F	125	Total O 125 125	0	0

3 Residue-property plots (i)

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.

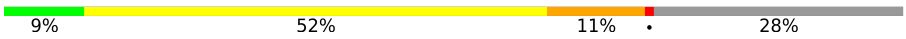
• Molecule 1: E2 glycoprotein

Chain A:



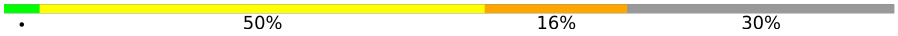
• Molecule 1: E2 glycoprotein

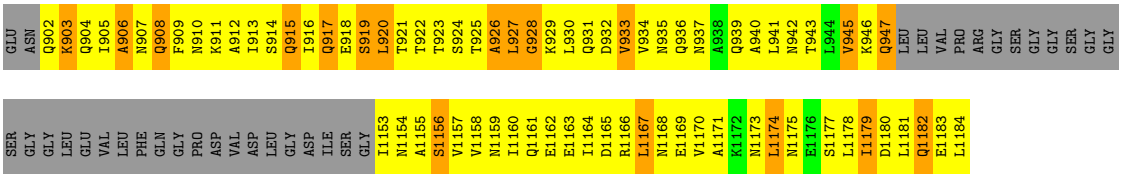
Chain B:



• Molecule 1: E2 glycoprotein

Chain C:



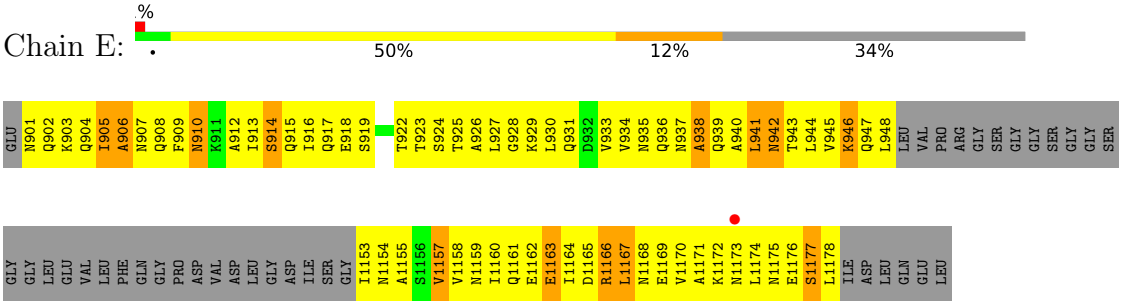


• Molecule 1: E2 glycoprotein

Chain D:

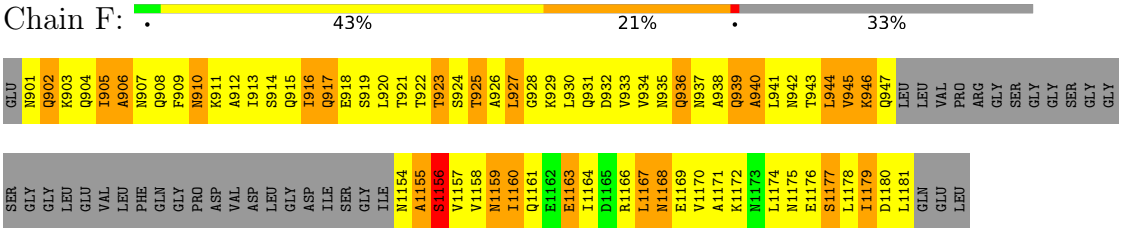
• Molecule 1: E2 glycoprotein

Chain E:



• Molecule 1: E2 glycoprotein

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	121.22Å 66.32Å 69.98Å 90.00° 107.35° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 47.19 – 2.80	Depositor EDS
% Data completeness (in resolution range)	(Not available) (50.00-2.80) 81.2 (47.19-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.22 (at 2.81Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.233 , 0.273 0.231 , 0.276	Depositor DCC
R_{free} test set	612 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	38.8	Xtriage
Anisotropy	0.583	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 95.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	4131	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.22% 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	0.57	0/532	0.80	0/718
1	B	0.83	1/616 (0.2%)	1.36	3/832 (0.4%)
1	C	0.63	0/603	0.89	1/813 (0.1%)
1	D	0.63	0/581	0.91	1/785 (0.1%)
1	E	0.58	0/567	0.80	0/766
1	F	0.66	0/575	0.85	1/777 (0.1%)
All	All	0.66	1/3474 (0.0%)	0.96	6/4691 (0.1%)

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

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	1168	ASN	C-N	-14.03	1.01	1.34

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	1168	ASN	O-C-N	-26.77	79.87	122.70
1	B	1168	ASN	CA-C-N	15.00	150.20	117.20
1	B	1168	ASN	C-N-CA	6.75	138.56	121.70
1	C	1182	GLN	N-CA-C	-6.55	93.33	111.00
1	F	1167	LEU	CA-CB-CG	5.93	128.94	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	1168	ASN	Mainchain

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	533	0	547	142	0
1	B	617	0	621	215	0
1	C	604	0	611	202	0
1	D	582	0	580	202	0
1	E	568	0	582	180	0
1	F	576	0	584	217	0
2	A	100	0	0	69	0
2	B	124	0	0	82	0
2	C	108	0	0	99	0
2	D	99	0	0	84	0
2	E	95	0	0	73	0
2	F	125	0	0	105	0
All	All	4131	0	3525	1058	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 151.

The worst 5 of 1058 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:E:941:LEU:HA	2:E:1:HOH:O	1.33	1.22
1:B:1154:ASN:HB2	2:F:81:HOH:O	1.50	1.11
1:C:1184:LEU:HA	2:C:482:HOH:O	1.50	1.11
1:A:1158:VAL:HB	1:C:933:VAL:HG21	1.18	1.10
1:E:1166:ARG:HA	1:E:1166:ARG:HH11	1.13	1.09

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	65/112 (58%)	44 (68%)	14 (22%)	7 (11%)	0 1
1	B	77/112 (69%)	47 (61%)	18 (23%)	12 (16%)	0 0
1	C	74/112 (66%)	45 (61%)	15 (20%)	14 (19%)	0 0
1	D	73/112 (65%)	40 (55%)	17 (23%)	16 (22%)	0 0
1	E	70/112 (62%)	47 (67%)	14 (20%)	9 (13%)	0 1
1	F	71/112 (63%)	43 (61%)	15 (21%)	13 (18%)	0 0
All	All	430/672 (64%)	266 (62%)	93 (22%)	71 (16%)	0 0

5 of 71 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	941	LEU
1	A	1154	ASN
1	A	1155	ALA
1	B	905	ILE
1	B	1180	ASP

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	61/94 (65%)	57 (93%)	4 (7%)	16 44
1	B	70/94 (74%)	66 (94%)	4 (6%)	20 50
1	C	69/94 (73%)	61 (88%)	8 (12%)	5 17

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	65/94 (69%)	49 (75%)	16 (25%)	0 2
1	E	65/94 (69%)	57 (88%)	8 (12%)	4 15
1	F	66/94 (70%)	54 (82%)	12 (18%)	1 5
All	All	396/564 (70%)	344 (87%)	52 (13%)	4 12

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	1167	LEU
1	E	941	LEU
1	F	1163	GLU
1	D	1169	GLU
1	E	910	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 45 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	931	GLN
1	F	908	GLN
1	E	935	ASN
1	E	1154	ASN
1	F	917	GLN

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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	1168:ASN	C	1169:GLU	N	1.01

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, 95^{th} 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	69/112 (61%)	-0.58	0 100 100	19, 41, 83, 95	0
1	B	81/112 (72%)	-0.68	0 100 100	13, 39, 63, 76	0
1	C	78/112 (69%)	-0.62	0 100 100	13, 38, 72, 80	0
1	D	77/112 (68%)	-0.53	0 100 100	17, 41, 61, 69	0
1	E	74/112 (66%)	-0.48	1 (1%) 75 70	13, 38, 79, 90	0
1	F	75/112 (66%)	-0.62	0 100 100	12, 34, 64, 82	0
All	All	454/672 (67%)	-0.59	1 (0%) 95 94	12, 39, 73, 95	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	1173	ASN	2.0

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 (i)

There are no such residues in this entry.