wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 12:43 AM EDT

PDB ID	:	3SCJ
Title	:	Crystal structure of spike protein receptor-binding domain from a predicted SARS coronavirus civet strain complexed with human receptor ACE2
Authors	:	Wu, K.; Peng, G.; Wilken, M.; Geraghty, R.; Li, F.
Deposited on	:	2011-06-07
Resolution	:	3.00 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 (1)) were used in the production of this report:

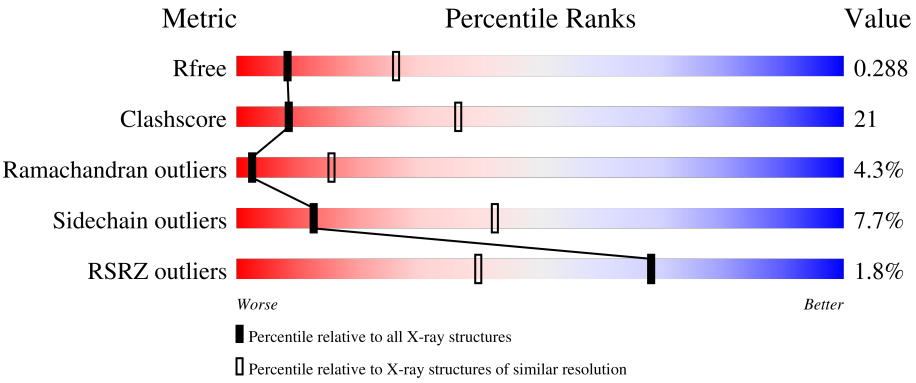
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	603	57% 36% 5%
1	B	603	2% 55% 38% 5%
2	E	186	4% 54% 32% 7% 6%
2	F	186	3% 58% 30% 6% 6%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 12544 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 Angiotensin-converting enzyme 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	597	Total	C	N	O	S	0	0	0
			4870	3115	806	920	29			
1	B	597	Total	C	N	O	S	0	0	0
			4870	3115	806	920	29			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	616	HIS	-	expression tag	UNP Q9BYF1
A	617	HIS	-	expression tag	UNP Q9BYF1
A	618	HIS	-	expression tag	UNP Q9BYF1
A	619	HIS	-	expression tag	UNP Q9BYF1
A	620	HIS	-	expression tag	UNP Q9BYF1
A	621	HIS	-	expression tag	UNP Q9BYF1
B	616	HIS	-	expression tag	UNP Q9BYF1
B	617	HIS	-	expression tag	UNP Q9BYF1
B	618	HIS	-	expression tag	UNP Q9BYF1
B	619	HIS	-	expression tag	UNP Q9BYF1
B	620	HIS	-	expression tag	UNP Q9BYF1
B	621	HIS	-	expression tag	UNP Q9BYF1

- Molecule 2 is a protein called Spike glycoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	174	Total	C	N	O	S	0	0	0
			1400	908	230	255	7			
2	F	174	Total	C	N	O	S	0	0	0
			1400	908	230	255	7			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	472	PRO	LEU	SEE REMARK 999	UNP P59594
E	479	ARG	ASN	conflict	UNP P59594
E	480	GLY	ASP	SEE REMARK 999	UNP P59594
E	503	HIS	-	expression tag	UNP P59594
E	504	HIS	-	expression tag	UNP P59594
E	505	HIS	-	expression tag	UNP P59594
E	506	HIS	-	expression tag	UNP P59594
E	507	HIS	-	expression tag	UNP P59594
E	508	HIS	-	expression tag	UNP P59594
F	472	PRO	LEU	SEE REMARK 999	UNP P59594
F	479	ARG	ASN	conflict	UNP P59594
F	480	GLY	ASP	SEE REMARK 999	UNP P59594
F	503	HIS	-	expression tag	UNP P59594
F	504	HIS	-	expression tag	UNP P59594
F	505	HIS	-	expression tag	UNP P59594
F	506	HIS	-	expression tag	UNP P59594
F	507	HIS	-	expression tag	UNP P59594
F	508	HIS	-	expression tag	UNP P59594

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Zn 1 1	0	0
3	B	1	Total Zn 1 1	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

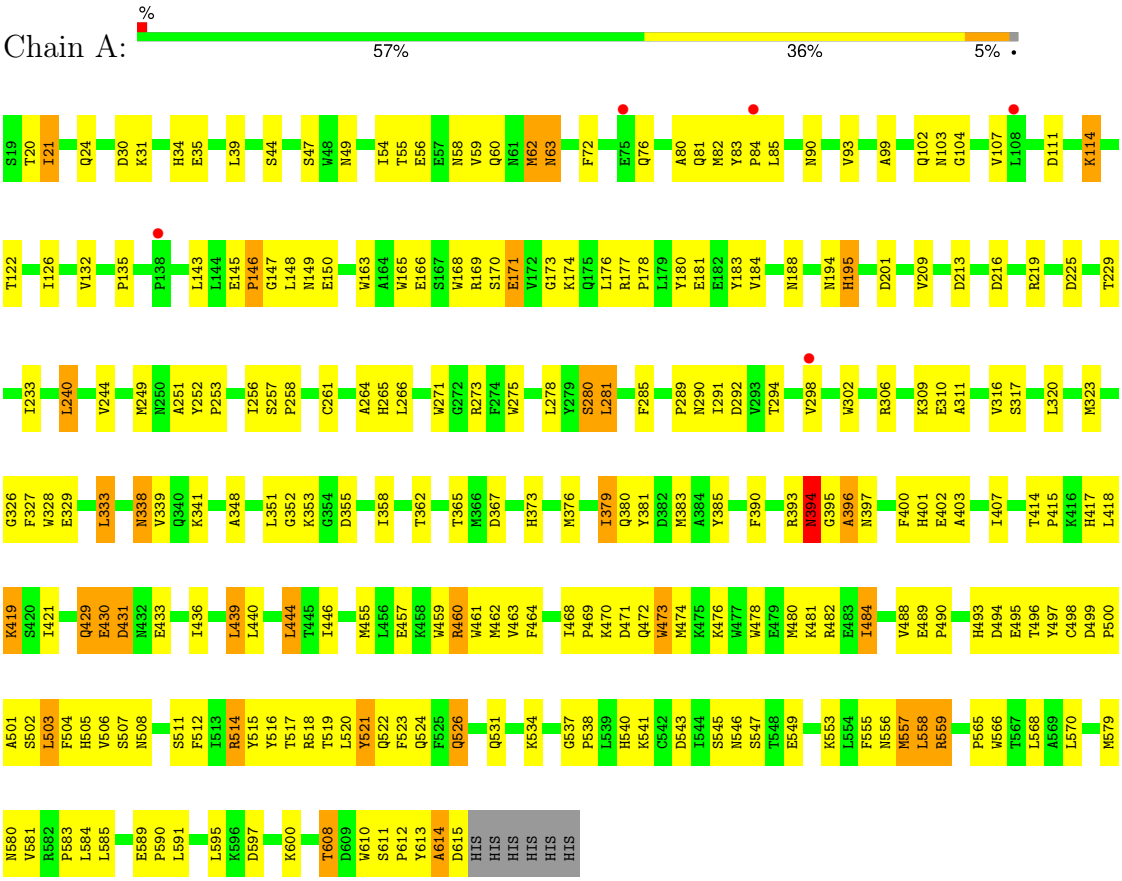
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Cl 1 1	0	0
4	B	1	Total Cl 1 1	0	0

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.

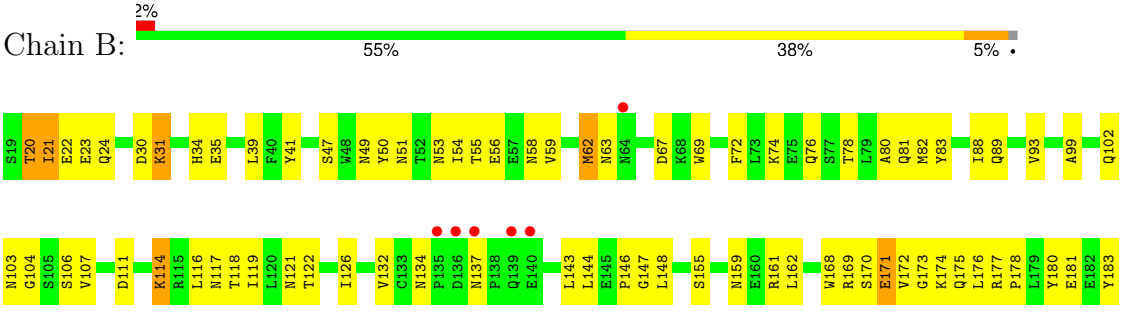
• Molecule 1: Angiotensin-converting enzyme 2

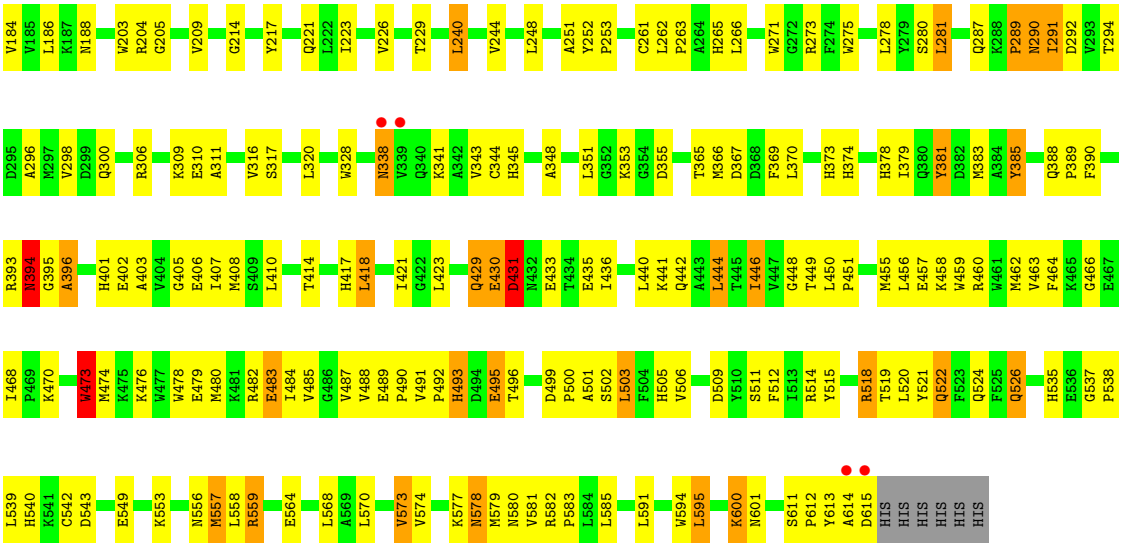
Chain A:



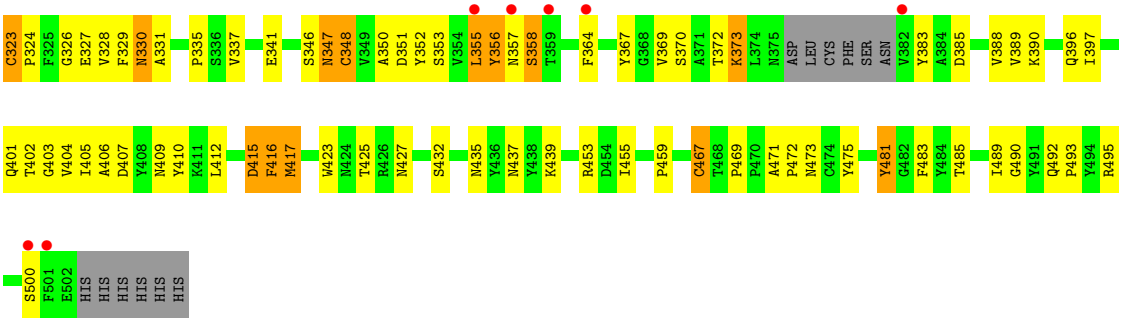
• Molecule 1: Angiotensin-converting enzyme 2

Chain B:

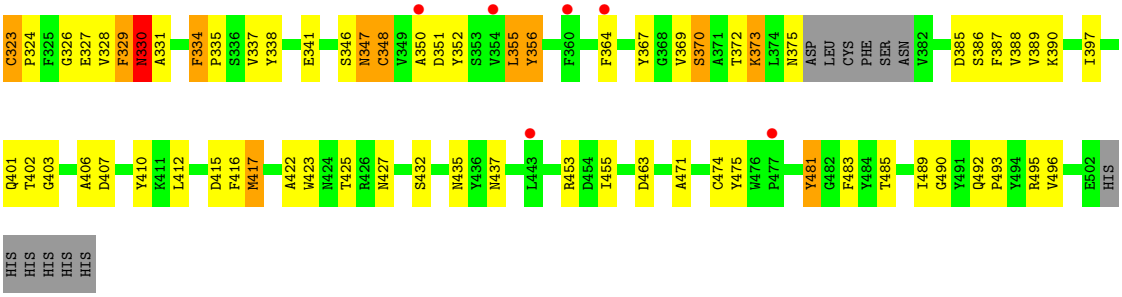




• Molecule 2: Spike glycoprotein



• Molecule 2: Spike glycoprotein



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	81.22Å 119.28Å 113.24Å 90.00° 92.19° 90.00°	Depositor
Resolution (Å)	47.27 – 3.00 47.27 – 3.00	Depositor EDS
% Data completeness (in resolution range)	87.3 (47.27-3.00) 87.2 (47.27-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.02 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.5.0109, CNS	Depositor
R, R_{free}	0.230 , 0.278 0.237 , 0.288	Depositor DCC
R_{free} test set	2153 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	77.7	Xtriage
Anisotropy	0.066	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 30.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.032 for h,-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	12544	wwPDB-VP
Average B, all atoms (Å^2)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, CL

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	0.50	0/5007	0.61	0/6803
1	B	0.51	0/5007	0.59	0/6803
2	E	0.54	0/1446	0.64	0/1970
2	F	0.56	0/1446	0.63	1/1970 (0.1%)
All	All	0.52	0/12906	0.61	1/17546 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	330	ASN	N-CA-C	-5.03	97.41	111.00

There are no chirality outliers.

There are no planarity outliers.

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	4870	0	4643	208	0
1	B	4870	0	4643	209	1
2	E	1400	0	1329	46	1
2	F	1400	0	1329	45	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	1	0	0	0	0
4	B	1	0	0	0	0
All	All	12544	0	11944	507	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 21.

The worst 5 of 507 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:261:CYS:HB2	1:A:488:VAL:CG2	1.87	1.03
1:A:402:GLU:HB3	1:A:518:ARG:HD2	1.42	1.01
1:B:261:CYS:HB2	1:B:488:VAL:HG13	1.42	1.00
1:B:74:LYS:HE2	1:B:106:SER:OG	1.64	0.95
1:A:132:VAL:HG12	1:A:171:GLU:HG3	1.47	0.94

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:B:287:GLN:NE2	2:E:469:PRO:O[2_556]	2.16	0.04

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	595/603 (99%)	504 (85%)	71 (12%)	20 (3%)	3 20
1	B	595/603 (99%)	498 (84%)	78 (13%)	19 (3%)	4 22
2	E	170/186 (91%)	131 (77%)	25 (15%)	14 (8%)	1 4
2	F	170/186 (91%)	130 (76%)	27 (16%)	13 (8%)	1 5

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1530/1578 (97%)	1263 (82%)	201 (13%)	66 (4%)	2 15

5 of 66 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	289	PRO
1	A	430	GLU
1	A	546	ASN
1	A	614	ALA
1	B	396	ALA

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	527/533 (99%)	487 (92%)	40 (8%)	13 43
1	B	527/533 (99%)	484 (92%)	43 (8%)	11 39
2	E	151/163 (93%)	140 (93%)	11 (7%)	14 44
2	F	151/163 (93%)	140 (93%)	11 (7%)	14 44
All	All	1356/1392 (97%)	1251 (92%)	105 (8%)	13 42

5 of 105 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	385	TYR
1	B	493	HIS
2	F	364	PHE
1	B	401	HIS
1	B	444	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	49	ASN
2	F	375	ASN
1	B	195	HIS
2	E	435	ASN
1	B	578	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)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 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	597/603 (99%)	-0.18	5 (0%)	86	65	42, 76, 124, 151	0
1	B	597/603 (99%)	-0.06	10 (1%)	70	41	47, 81, 127, 151	0
2	E	174/186 (93%)	0.13	7 (4%)	38	15	64, 89, 135, 144	0
2	F	174/186 (93%)	0.05	6 (3%)	45	19	67, 90, 134, 145	0
All	All	1542/1578 (97%)	-0.07	28 (1%)	68	40	42, 82, 132, 151	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	136	ASP	5.3
2	E	501	PHE	4.0
2	E	500	SER	3.6
2	E	357	ASN	3.3
1	B	615	ASP	3.3

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

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	CL	B	902	1/1	0.82	0.21	114,114,114,114	0
3	ZN	B	901	1/1	0.83	0.39	83,83,83,83	0
4	CL	A	902	1/1	0.89	0.50	96,96,96,96	0
3	ZN	A	901	1/1	0.94	0.34	82,82,82,82	0

6.5 Other polymers (i)

There are no such residues in this entry.