wwPDB X-ray Structure Validation Summary Report (i)

Sep 14, 2020 – 01:49 AM BST

PDB ID	:	3SCK
Title	:	Crystal structure of spike protein receptor-binding domain from a predicted SARS coronavirus civet strain complexed with human-civet chimeric 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 following versions of software and data (see references (1)) were used in the production of this report:

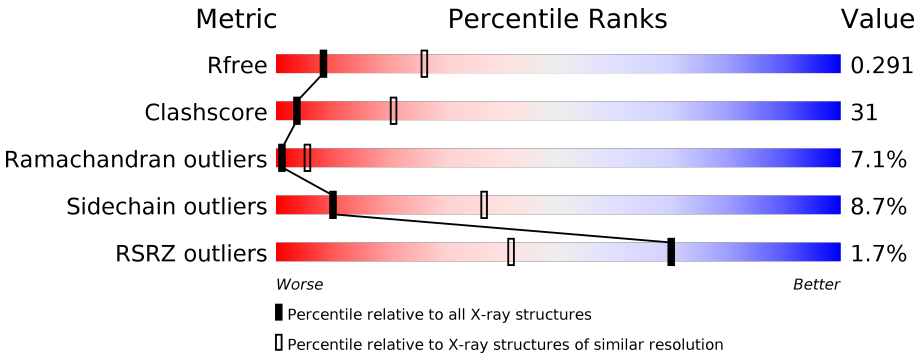
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.14.4.dev1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.4.dev1

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 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	603	45% 45% 9% ·
1	B	603	% 44% 46% 8% ·
2	E	185	5% 46% 37% 9% · 6%
2	F	185	3% 42% 41% 10% · 6%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 12518 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 chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	597	Total	C	N	O	S	0	0	0
			4863	3109	803	923	28			
1	B	597	Total	C	N	O	S	0	0	0
			4863	3109	803	923	28			

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	173	Total	C	N	O	S	0	0	0
			1394	905	229	254	6			
2	F	173	Total	C	N	O	S	0	0	0
			1394	905	229	254	6			

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	B	1	Total Zn 1 1	0	0
3	A	1	Total Zn 1 1	0	0

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

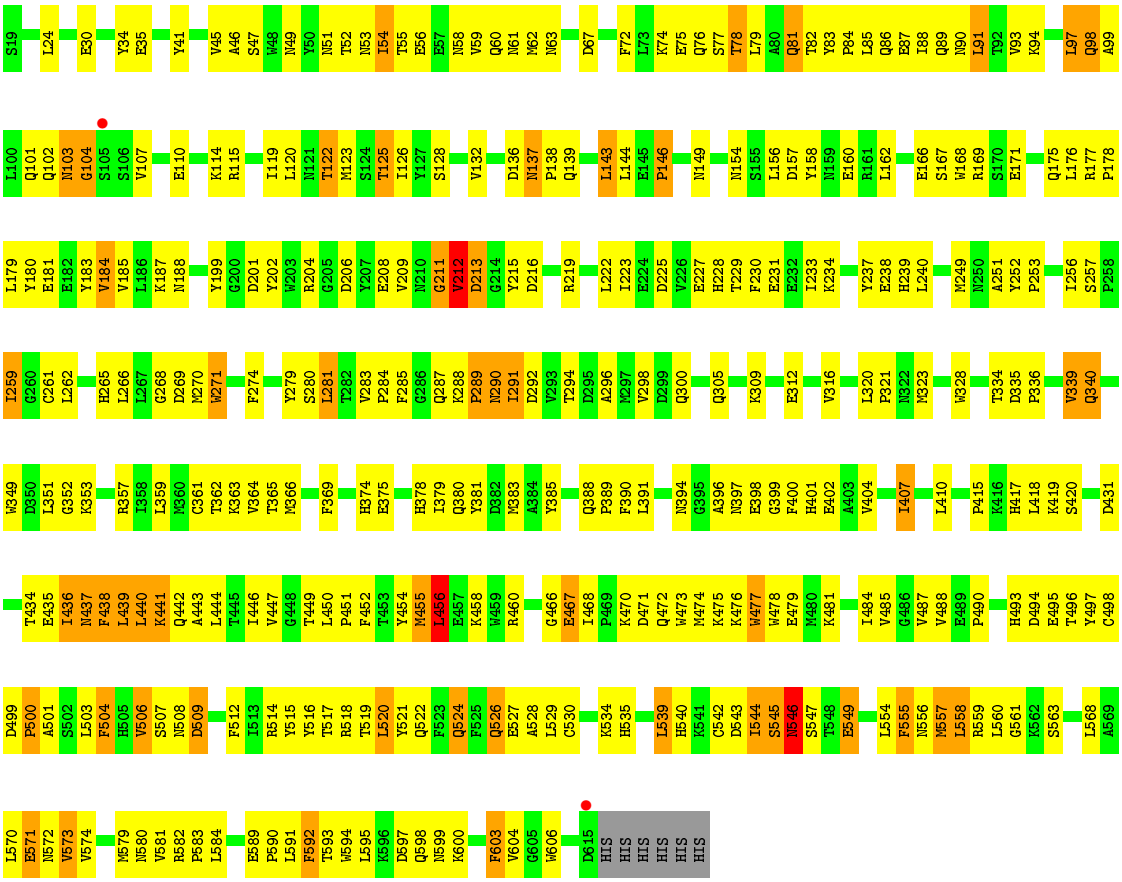
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total Cl 1 1	0	0
4	A	1	Total Cl 1 1	0	0

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.

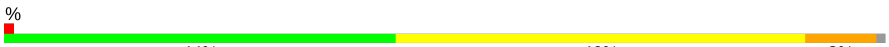
- Molecule 1: Angiotensin-converting enzyme 2 chimera

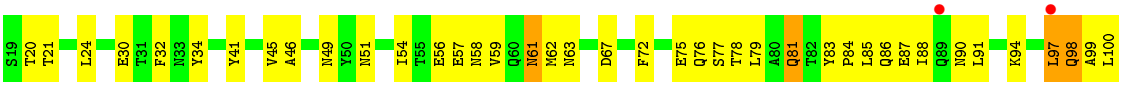
Chain A:

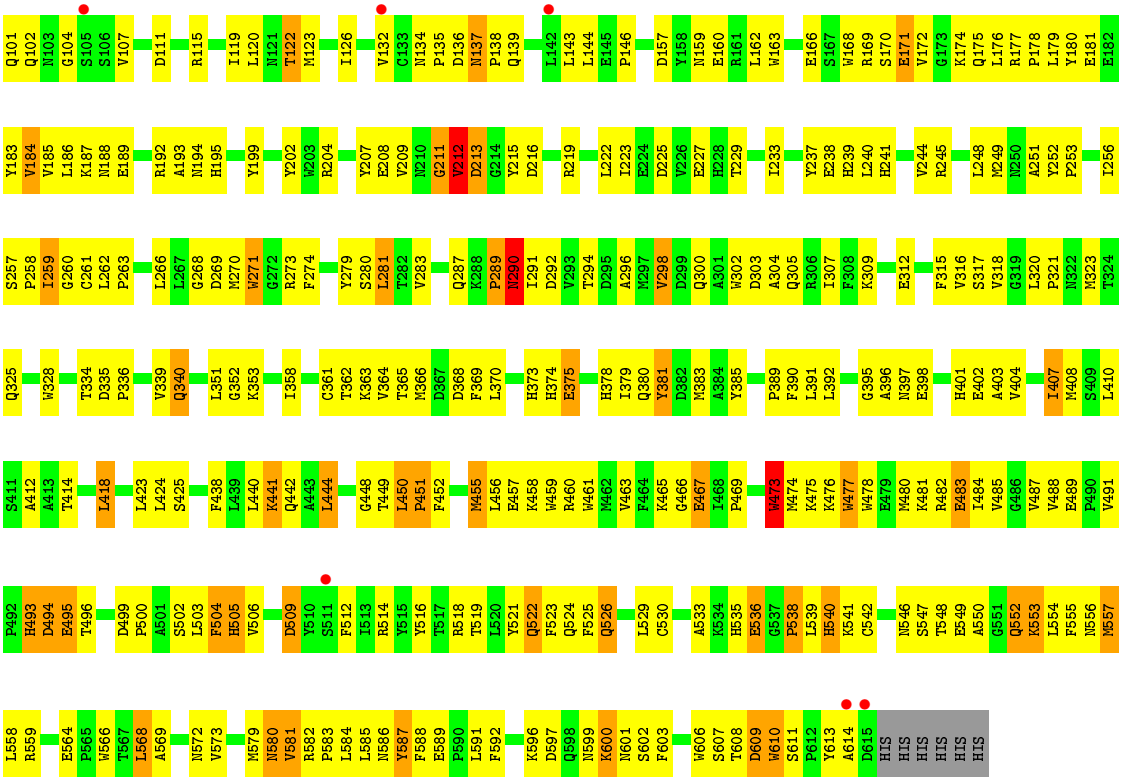


- Molecule 1: Angiotensin-converting enzyme 2 chimera

Chain B:

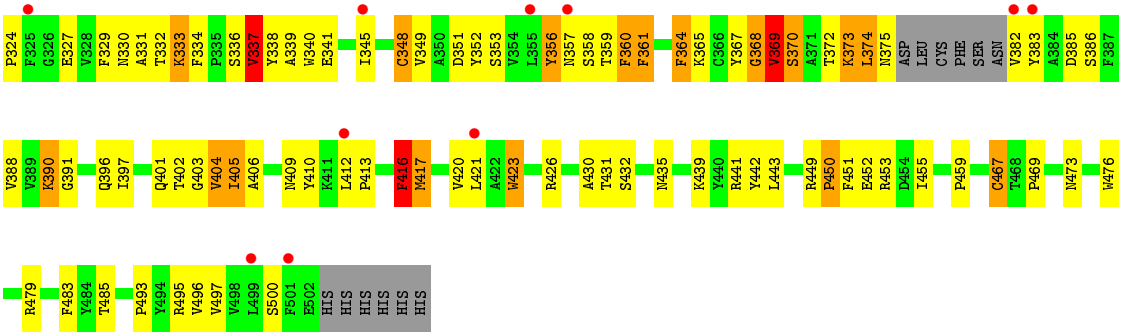






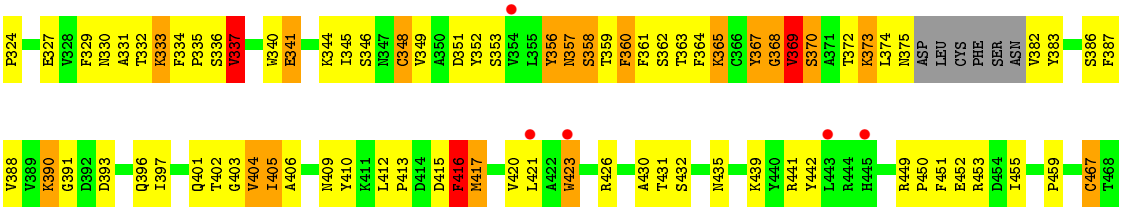
- Molecule 2: Spike glycoprotein

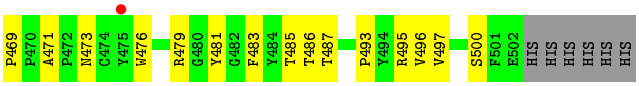
Chain E:



- Molecule 2: Spike glycoprotein

Chain F:




P469
P470
A471
P472
N473
C474
Y475
W476
R479
G480
Y481
G482
F483
Y484
T485
T486
T487
P493
Y494
R495
V496
V497
S500
F501
E502
HIS
HIS
HIS
HIS
HIS
HIS

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 – 2.99	Depositor EDS
% Data completeness (in resolution range)	78.7 (47.27-3.00) 76.8 (47.27-2.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.16 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.5.0109, CNS	Depositor
R, R_{free}	0.239 , 0.285 0.245 , 0.291	Depositor DCC
R_{free} test set	2146 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	84.7	Xtriage
Anisotropy	0.175	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 48.3	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	12518	wwPDB-VP
Average B, all atoms (Å^2)	105.0	wwPDB-VP

Xtriage'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.*

[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	0.57	0/4999	0.67	2/6796 (0.0%)
1	B	0.55	0/4999	0.64	0/6796
2	E	0.60	0/1440	0.65	0/1961
2	F	0.57	0/1440	0.62	0/1961
All	All	0.56	0/12878	0.65	2/17514 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	290	ASN	N-CA-C	5.38	125.52	111.00
1	A	291	ILE	N-CA-C	-5.22	96.91	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	4863	0	4636	297	0
1	B	4863	0	4636	319	0
2	E	1394	0	1327	73	0
2	F	1394	0	1327	85	0
3	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	1	0	0	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
All	All	12518	0	11926	765	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 31.

The worst 5 of 765 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)
2:E:324:PRO:HG3	2:E:348:CYS:SG	1.87	1.12
1:A:456:LEU:HD22	1:A:477:TRP:HH2	1.16	1.03
2:F:324:PRO:HG3	2:F:348:CYS:SG	2.01	1.00
1:B:582:ARG:HB3	1:B:583:PRO:HD3	1.46	0.97
1:A:410:LEU:HD23	1:A:526:GLN:HG3	1.46	0.97

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	595/603 (99%)	452 (76%)	105 (18%)	38 (6%)	1 7
1	B	595/603 (99%)	444 (75%)	112 (19%)	39 (7%)	1 6
2	E	169/185 (91%)	125 (74%)	29 (17%)	15 (9%)	1 3
2	F	169/185 (91%)	119 (70%)	33 (20%)	17 (10%)	0 2
All	All	1528/1576 (97%)	1140 (75%)	279 (18%)	109 (7%)	1 5

5 of 109 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	54	ILE
1	A	137	ASN
1	A	212	VAL
1	A	289	PRO
1	A	340	GLN

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%)	483 (92%)	44 (8%)	11 39
1	B	527/533 (99%)	483 (92%)	44 (8%)	11 39
2	E	150/162 (93%)	136 (91%)	14 (9%)	9 33
2	F	150/162 (93%)	134 (89%)	16 (11%)	6 26
All	All	1354/1390 (97%)	1236 (91%)	118 (9%)	10 37

5 of 118 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	240	LEU
1	B	418	LEU
2	F	364	PHE
1	B	259	ILE
1	B	298	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 65 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	64	ASN
1	B	139	GLN
2	E	409	ASN
1	B	81	GLN
1	B	98	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

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

5.5 Carbohydrates

There are no monosaccharides in this entry.

5.6 Ligand geometry

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

There are no such residues in this entry.

5.8 Polymer linkage issues

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%)	-0.13	2 (0%) 94 84	56, 98, 143, 159	0
1	B	597/603 (99%)	-0.02	8 (1%) 77 51	55, 99, 144, 158	0
2	E	173/185 (93%)	0.19	10 (5%) 23 7	79, 116, 149, 157	0
2	F	173/185 (93%)	0.16	6 (3%) 44 18	81, 117, 148, 157	0
All	All	1540/1576 (97%)	-0.02	26 (1%) 70 41	55, 103, 146, 159	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	615	ASP	10.2
1	B	105	SER	4.9
2	E	421	LEU	4.0
2	E	501	PHE	4.0
1	A	105	SER	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	A	902	1/1	0.89	0.61	138,138,138,138	0
4	CL	B	902	1/1	0.95	0.28	108,108,108,108	0
3	ZN	B	901	1/1	0.96	0.46	98,98,98,98	0
3	ZN	A	901	1/1	0.96	0.46	102,102,102,102	0

6.5 Other polymers (i)

There are no such residues in this entry.