



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 12:52 pm GMT

PDB ID : 2CJR
Title : CRYSTAL STRUCTURE OF OLIGOMERIZATION DOMAIN OF SARS
CORONAVIRUS NUCLEOCAPSID PROTEIN.
Authors : Chen, C.-Y.; Hsiao, C.-D.
Deposited on : 2006-04-06
Resolution : 2.50 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

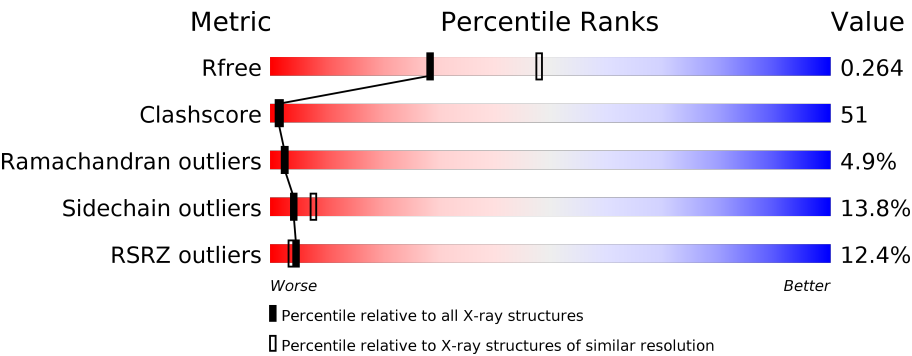
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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}	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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. 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 $\leq 5\%$. 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	128	<div><div>5%</div><div><div>49%</div><div>34%</div><div>7%</div><div>10%</div></div></div>
1	B	128	<div><div>7%</div><div><div>38%</div><div>40%</div><div>9%</div><div>12%</div></div></div>
1	C	128	<div><div>%</div><div><div>47%</div><div>34%</div><div>8%</div><div>12%</div></div></div>
1	D	128	<div><div>3%</div><div><div>43%</div><div>39%</div><div>8%</div><div>10%</div></div></div>
1	E	128	<div><div>20%</div><div><div>24%</div><div>46%</div><div>15%</div><div>14%</div></div></div>
1	F	128	<div><div>17%</div><div><div>32%</div><div>44%</div><div>11%</div><div>13%</div></div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	128	<div><div></div><div>15%</div><div>25%</div><div>48%</div><div>13%</div><div>15%</div></div>
1	H	128	<div><div></div><div>19%</div><div>23%</div><div>44%</div><div>18%</div><div>•</div><div>14%</div></div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 7973 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 NUCLEOCAPSID PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	115	Total	C	N	O	S	0	0	0
			912	577	165	168	2			
1	B	113	Total	C	N	O	S	0	0	0
			902	570	162	168	2			
1	C	113	Total	C	N	O	S	0	0	0
			898	568	162	166	2			
1	D	115	Total	C	N	O	S	0	0	0
			909	574	164	169	2			
1	E	110	Total	C	N	O	S	0	0	0
			880	558	159	161	2			
1	F	112	Total	C	N	O	S	0	0	1
			882	556	162	162	2			
1	G	109	Total	C	N	O	S	0	0	1
			867	547	159	159	2			
1	H	110	Total	C	N	O	S	0	0	0
			869	551	155	161	2			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	122	Total	O	0	0
			122	122		
2	B	130	Total	O	0	0
			130	130		
2	C	106	Total	O	0	0
			106	106		
2	D	113	Total	O	0	0
			113	113		
2	E	101	Total	O	0	0
			101	101		
2	F	95	Total	O	0	0
			95	95		

Continued on next page...

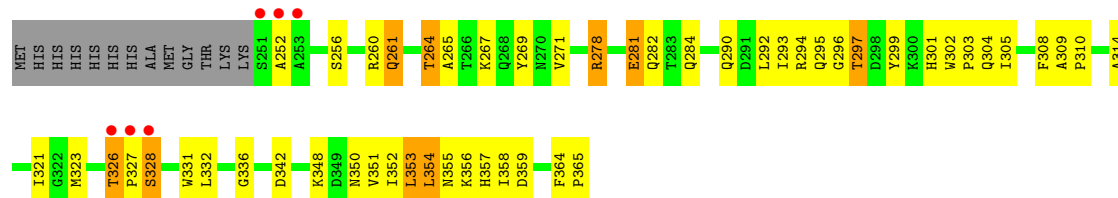
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	89	Total	O	0	0
			89	89		
2	H	98	Total	O	0	0
			98	98		

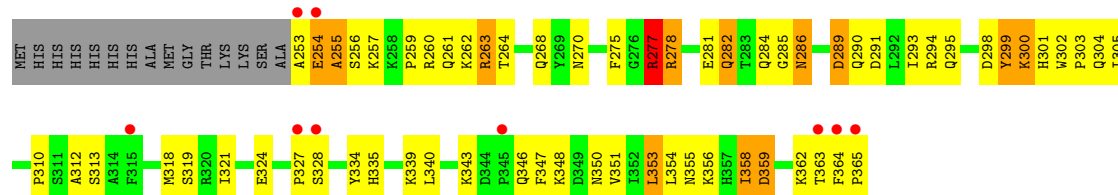
3 Residue-property plots [i](#)

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.

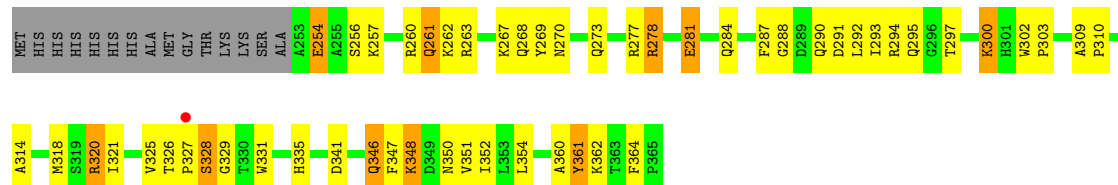
• Molecule 1: NUCLEOCAPSID PROTEIN



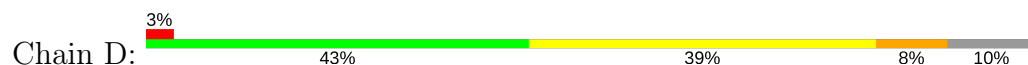
• Molecule 1: NUCLEOCAPSID PROTEIN



• Molecule 1: NUCLEOCAPSID PROTEIN

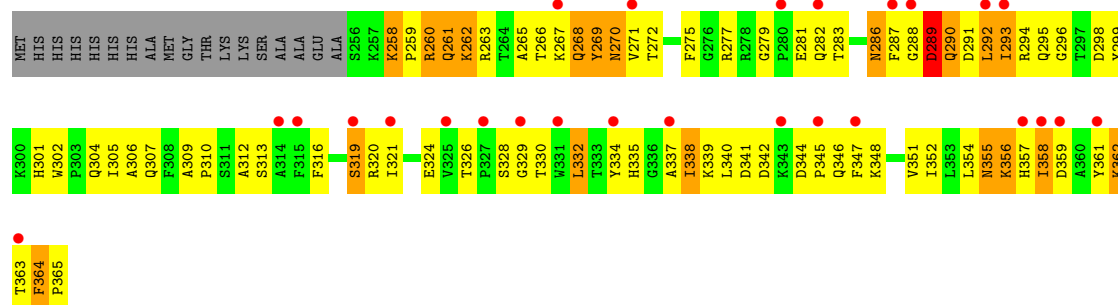


• Molecule 1: NUCLEOCAPSID PROTEIN

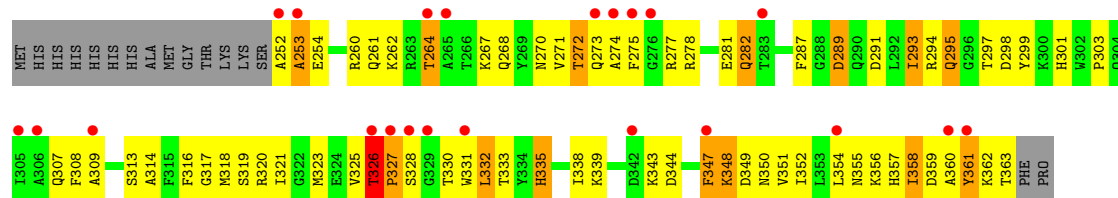




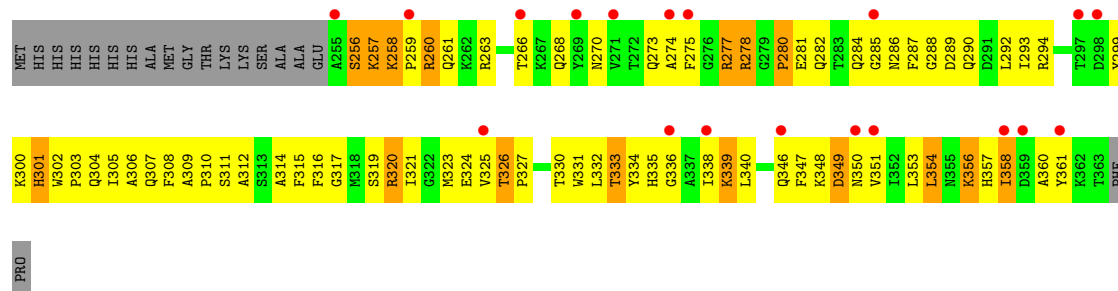
● Molecule 1: NUCLEOCAPSID PROTEIN



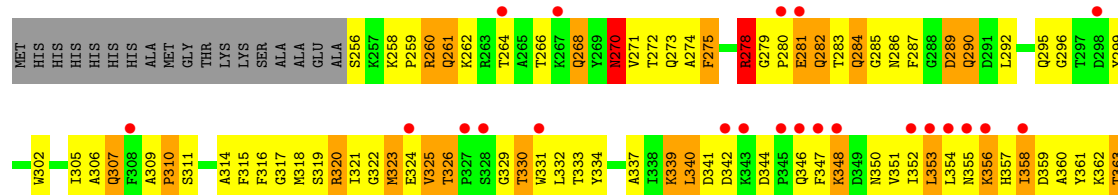
● Molecule 1: NUCLEOCAPSID PROTEIN



● Molecule 1: NUCLEOCAPSID PROTEIN



● Molecule 1: NUCLEOCAPSID PROTEIN





4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	159.42Å 84.20Å 105.18Å 90.00° 131.18° 90.00°	Depositor
Resolution (Å)	29.80 – 2.50 29.80 – 2.50	Depositor EDS
% Data completeness (in resolution range)	91.0 (29.80-2.50) 96.9 (29.80-2.50)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.12 (at 2.51Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.256 , 0.269 0.259 , 0.264	Depositor DCC
R_{free} test set	1754 reflections (5.24%)	DCC
Wilson B-factor (Å ²)	24.8	Xtriage
Anisotropy	0.630	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 75.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.023 for -h-2*1,-k,l	Xtriage
F_o, F_c correlation	0.90	EDS
Total number of atoms	7973	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.21 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 8.8831e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	1.22	1/936 (0.1%)	1.03	0/1263
1	B	1.11	1/926 (0.1%)	1.13	4/1250 (0.3%)
1	C	1.11	0/922	1.02	1/1245 (0.1%)
1	D	1.24	1/933 (0.1%)	1.15	2/1260 (0.2%)
1	E	1.17	1/904 (0.1%)	1.12	1/1220 (0.1%)
1	F	1.06	0/904	1.03	1/1220 (0.1%)
1	G	1.11	2/889 (0.2%)	1.10	2/1199 (0.2%)
1	H	1.15	1/893 (0.1%)	1.13	3/1208 (0.2%)
All	All	1.15	7/7307 (0.1%)	1.09	14/9865 (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	F	0	1
1	G	0	1
1	H	0	1
All	All	0	3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	314	ALA	CA-CB	6.38	1.65	1.52
1	D	364	PHE	CE1-CZ	5.60	1.48	1.37
1	H	275	PHE	CE2-CZ	5.57	1.48	1.37
1	B	312	ALA	CA-CB	-5.38	1.41	1.52
1	E	269	TYR	CD2-CE2	5.27	1.47	1.39

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	278	ARG	NE-CZ-NH2	-11.05	114.77	120.30
1	B	277	ARG	NE-CZ-NH1	-9.84	115.38	120.30
1	F	289	ASP	CB-CG-OD1	7.21	124.78	118.30
1	D	277	ARG	NE-CZ-NH1	6.63	123.61	120.30
1	E	332	LEU	CA-CB-CG	6.12	129.37	115.30

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	303	PRO	Peptide
1	G	326	THR	Peptide
1	H	310	PRO	Peptide

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	912	0	884	54	0
1	B	902	0	870	80	0
1	C	898	0	866	79	0
1	D	909	0	876	73	0
1	E	880	0	852	134	1
1	F	882	0	859	111	0
1	G	867	0	847	119	0
1	H	869	0	828	152	1
2	A	122	0	0	19	0
2	B	130	0	0	27	1
2	C	106	0	0	14	0
2	D	113	0	0	21	1
2	E	101	0	0	38	0
2	F	95	0	0	15	0
2	G	89	0	0	28	0
2	H	98	0	0	35	0
All	All	7973	0	6882	710	2

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 51.

The worst 5 of 710 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:D:305:ILE:HD13	1:D:354:LEU:HD23	1.21	1.20
1:C:335:HIS:HE1	1:D:282:GLN:NE2	1.39	1.20
2:F:2016:HOH:O	1:G:277:ARG:HG2	1.44	1.17
1:H:305:ILE:HG12	1:H:353:LEU:HD21	1.20	1.13
1:C:335:HIS:CE1	1:D:282:GLN:HE22	1.64	1.13

All (2) 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:E:359:ASP:O	1:H:352:ILE:CG2[4_455]	1.93	0.27
2:B:2069:HOH:O	2:D:2097:HOH:O[4_455]	2.06	0.14

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	113/128 (88%)	98 (87%)	12 (11%)	3 (3%)	6	9
1	B	111/128 (87%)	99 (89%)	9 (8%)	3 (3%)	6	9
1	C	111/128 (87%)	105 (95%)	4 (4%)	2 (2%)	10	17
1	D	113/128 (88%)	104 (92%)	5 (4%)	4 (4%)	4	5
1	E	108/128 (84%)	82 (76%)	17 (16%)	9 (8%)	1	1
1	F	110/128 (86%)	84 (76%)	17 (16%)	9 (8%)	1	1
1	G	107/128 (84%)	83 (78%)	18 (17%)	6 (6%)	2	2
1	H	108/128 (84%)	87 (81%)	14 (13%)	7 (6%)	1	1
All	All	881/1024 (86%)	742 (84%)	96 (11%)	43 (5%)	2	2

5 of 43 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	342	ASP
1	E	260	ARG
1	E	355	ASN
1	E	356	LYS
1	E	358	ILE

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	93/106 (88%)	84 (90%)	9 (10%)	9	18
1	B	93/106 (88%)	81 (87%)	12 (13%)	5	9
1	C	92/106 (87%)	84 (91%)	8 (9%)	12	23
1	D	93/106 (88%)	82 (88%)	11 (12%)	6	11
1	E	91/106 (86%)	75 (82%)	16 (18%)	2	3
1	F	90/106 (85%)	78 (87%)	12 (13%)	4	8
1	G	90/106 (85%)	77 (86%)	13 (14%)	4	7
1	H	89/106 (84%)	69 (78%)	20 (22%)	1	1
All	All	731/848 (86%)	630 (86%)	101 (14%)	4	8

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	272	THR
1	F	260	ARG
1	H	323	MET
1	E	289	ASP
1	E	319	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 40 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	286	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	307	GLN
1	H	284	GLN
1	D	295	GLN
1	D	357	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 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	115/128 (89%)	0.17	6 (5%) 28 29	3, 16, 43, 55	0
1	B	113/128 (88%)	0.16	9 (7%) 13 13	3, 20, 38, 67	0
1	C	113/128 (88%)	0.13	1 (0%) 84 85	2, 22, 44, 46	0
1	D	115/128 (89%)	0.17	4 (3%) 44 47	3, 15, 41, 52	0
1	E	110/128 (85%)	1.30	26 (23%) 1 0	18, 38, 58, 64	0
1	F	112/128 (87%)	1.24	22 (19%) 1 1	23, 43, 57, 59	0
1	G	109/128 (85%)	1.17	19 (17%) 2 1	15, 41, 54, 58	0
1	H	110/128 (85%)	1.16	24 (21%) 1 1	20, 38, 54, 65	0
All	All	897/1024 (87%)	0.68	111 (12%) 4 4	2, 31, 53, 67	0

The worst 5 of 111 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	329	GLY	7.7
1	D	255	ALA	6.5
1	F	253	ALA	6.0
1	F	326	THR	6.0
1	G	269	TYR	6.0

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.