wwPDB X-ray Structure Validation Summary Report (i)

May 20, 2024 – 03:05 AM JST

PDB ID	:	8JDI
Title	:	Crystal structure of Cas7-AcrIF25 complex
Authors	:	Yang, J.; Wang, J.; Wang, Y.
Deposited on	:	2023-05-14
Resolution	:	3.37 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 (i)) were used in the production of this report:

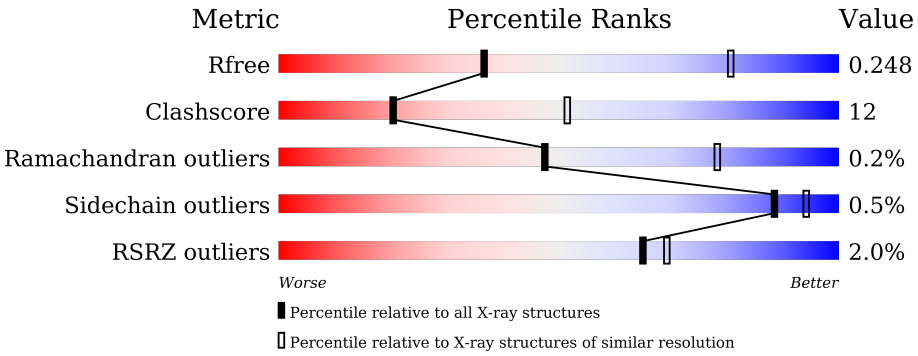
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 3.37 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.



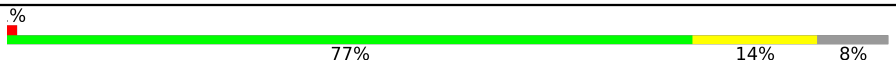
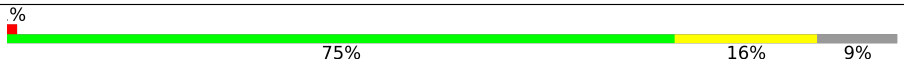
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1691 (3.46-3.30)
Clashscore	141614	1762 (3.46-3.30)
Ramachandran outliers	138981	1732 (3.46-3.30)
Sidechain outliers	138945	1731 (3.46-3.30)
RSRZ outliers	127900	1635 (3.46-3.30)

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	342	68% 21% 11%
1	B	342	2% 64% 20% 16%
1	C	342	2% 61% 29% • 10%
1	D	342	4% 67% 21% 12%
2	E	166	72% 19% 10%
2	F	166	4% 69% 21% • 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	166	
2	H	166	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 13487 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 CRISPR-associated protein Csy3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	304	Total	C	N	O	S	0	0	0
			2303	1452	417	432	2			
1	C	309	Total	C	N	O	S	0	0	0
			2296	1452	409	433	2			
1	B	288	Total	C	N	O	S	0	0	0
			2089	1320	375	393	1			
1	D	302	Total	C	N	O	S	0	0	0
			2255	1425	410	419	1			

- Molecule 2 is a protein called AcrIF25.

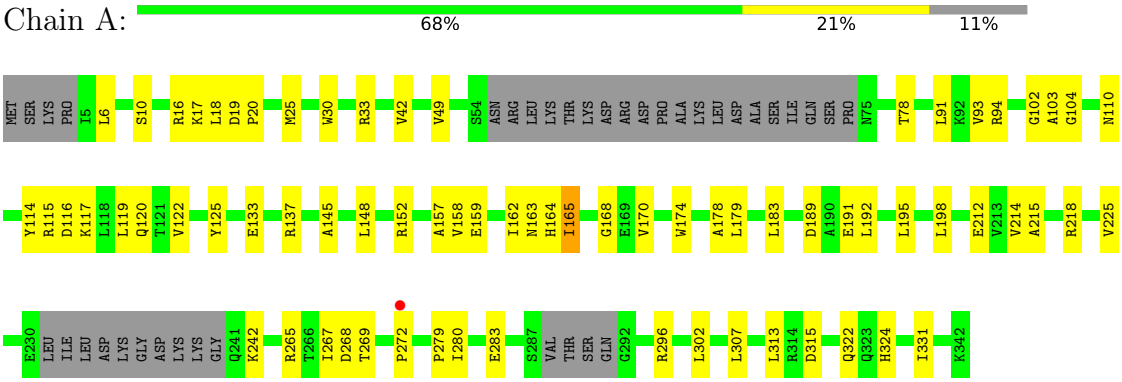
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	150	Total	C	N	O	S	0	0	0
			1118	702	198	213	5			
2	G	152	Total	C	N	O	S	0	1	0
			1163	728	207	223	5			
2	F	152	Total	C	N	O	S	0	1	0
			1126	703	200	218	5			
2	H	151	Total	C	N	O	S	0	0	0
			1137	714	200	219	4			

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.

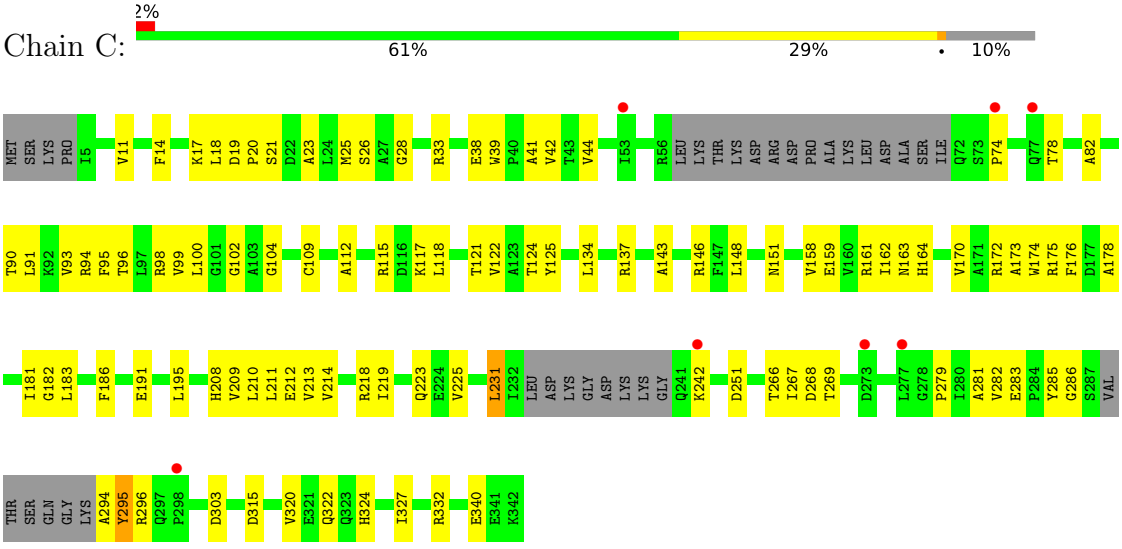
• Molecule 1: CRISPR-associated protein Csy3

Chain A:



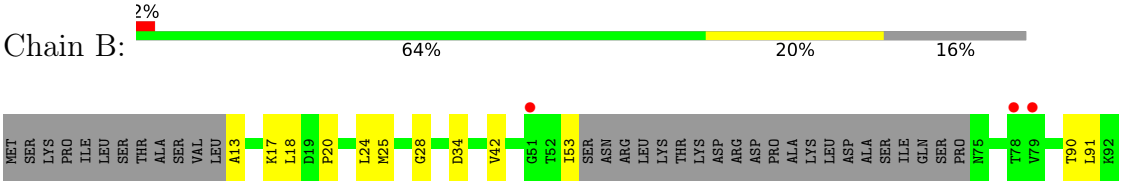
• Molecule 1: CRISPR-associated protein Csy3

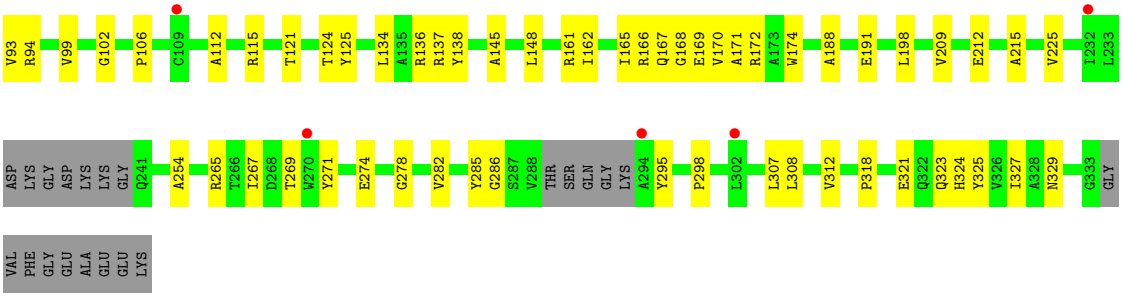
Chain C:



• Molecule 1: CRISPR-associated protein Csy3

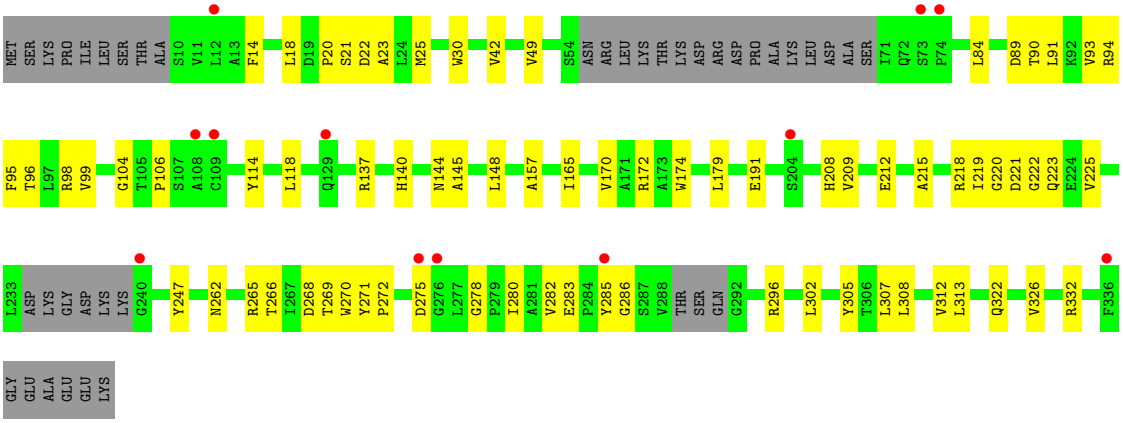
Chain B:





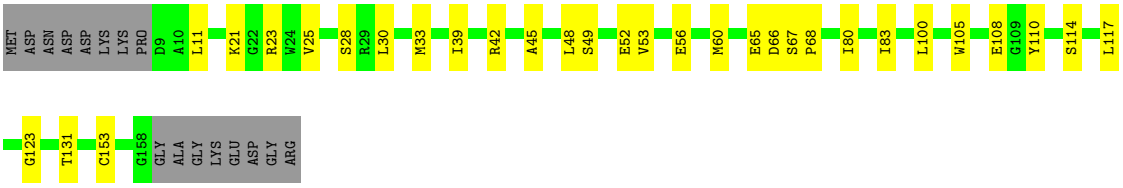
• Molecule 1: CRISPR-associated protein Csy3

Chain D:



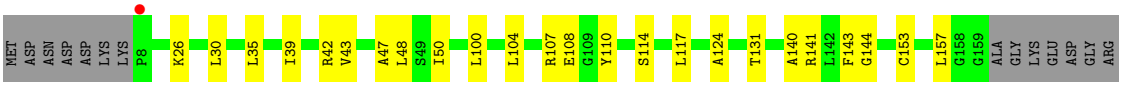
• Molecule 2: AcrIF25

Chain E:



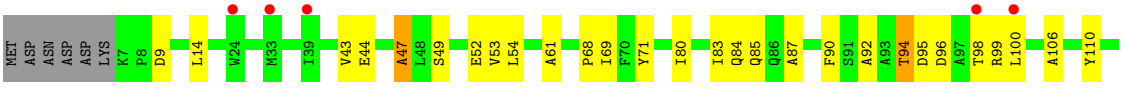
• Molecule 2: AcrIF25

Chain G:



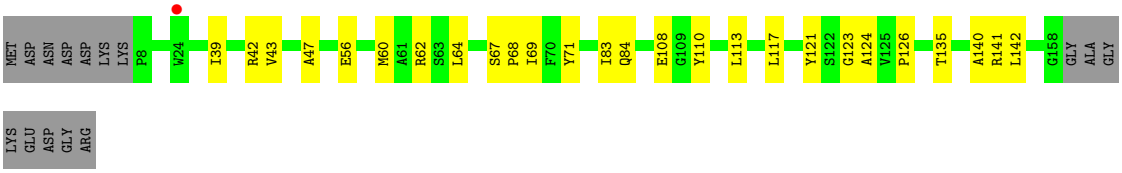
• Molecule 2: AcrIF25

Chain F:



- Molecule 2: AcrIF25

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	108.67Å 119.01Å 215.99Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.43 – 3.37 49.43 – 3.37	Depositor EDS
% Data completeness (in resolution range)	89.3 (49.43-3.37) 89.3 (49.43-3.37)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.64 (at 3.40Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.218 , 0.247 0.218 , 0.248	Depositor DCC
R_{free} test set	1752 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	59.2	Xtriage
Anisotropy	0.074	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 37.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	13487	wwPDB-VP
Average B, all atoms (Å^2)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.23% 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	0.42	0/2345	0.55	0/3187
1	B	0.40	0/2128	0.53	0/2910
1	C	0.36	0/2338	0.58	1/3188 (0.0%)
1	D	0.40	0/2298	0.59	0/3131
2	E	0.48	0/1138	0.52	0/1548
2	F	0.45	0/1150	0.61	2/1566 (0.1%)
2	G	0.26	0/1187	0.48	0/1608
2	H	0.34	0/1158	0.53	0/1573
All	All	0.39	0/13742	0.55	3/18711 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	92	ALA	N-CA-C	-6.00	94.79	111.00
1	C	74	PRO	N-CA-CB	5.81	110.27	103.30
2	F	92	ALA	N-CA-CB	5.20	117.38	110.10

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	2303	0	2231	50	0
1	B	2089	0	1930	54	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2296	0	2178	79	0
1	D	2255	0	2146	56	0
2	E	1118	0	1075	28	0
2	F	1126	0	1058	35	0
2	G	1163	0	1143	28	0
2	H	1137	0	1102	22	0
All	All	13487	0	12863	326	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 12.

The worst 5 of 326 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:C:231:LEU:HD12	1:C:231:LEU:H	1.24	1.00
1:C:268:ASP:OD2	1:C:296:ARG:NH1	1.95	1.00
2:E:67:SER:HB2	2:E:68:PRO:CD	1.97	0.94
2:G:104:LEU:O	2:G:108:GLU:HG3	1.66	0.94
1:D:137:ARG:NH1	1:D:270:TRP:CZ2	2.40	0.88

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	296/342 (86%)	278 (94%)	17 (6%)	1 (0%)	41 73
1	B	280/342 (82%)	267 (95%)	13 (5%)	0	100 100
1	C	301/342 (88%)	279 (93%)	21 (7%)	1 (0%)	41 73
1	D	294/342 (86%)	276 (94%)	17 (6%)	1 (0%)	41 73
2	E	148/166 (89%)	145 (98%)	3 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	F	151/166 (91%)	146 (97%)	4 (3%)	1 (1%)	22 56
2	G	151/166 (91%)	141 (93%)	10 (7%)	0	100 100
2	H	149/166 (90%)	143 (96%)	6 (4%)	0	100 100
All	All	1770/2032 (87%)	1675 (95%)	91 (5%)	4 (0%)	47 78

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	340	GLU
1	D	286	GLY
1	A	168	GLY
2	F	47	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	224/273 (82%)	223 (100%)	1 (0%)	91 95
1	B	187/273 (68%)	187 (100%)	0	100 100
1	C	217/273 (80%)	213 (98%)	4 (2%)	59 79
1	D	212/273 (78%)	212 (100%)	0	100 100
2	E	107/128 (84%)	107 (100%)	0	100 100
2	F	106/128 (83%)	105 (99%)	1 (1%)	78 89
2	G	116/128 (91%)	116 (100%)	0	100 100
2	H	111/128 (87%)	111 (100%)	0	100 100
All	All	1280/1604 (80%)	1274 (100%)	6 (0%)	88 94

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	251	ASP
1	C	295	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	94	THR
1	C	195	LEU
1	A	165	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	323	GLN
2	F	111	GLN
2	H	111	GLN
1	D	229	GLN
1	B	322	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)

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	304/342 (88%)	-0.05	1 (0%) 94 96	17, 39, 68, 79	0
1	B	288/342 (84%)	0.31	8 (2%) 53 57	28, 58, 100, 112	0
1	C	309/342 (90%)	0.10	7 (2%) 60 64	32, 57, 88, 103	0
1	D	302/342 (88%)	0.17	12 (3%) 38 41	18, 49, 93, 107	0
2	E	150/166 (90%)	-0.07	0 100 100	27, 50, 79, 95	0
2	F	152/166 (91%)	0.28	6 (3%) 39 43	42, 70, 105, 118	0
2	G	152/166 (91%)	-0.06	1 (0%) 87 91	27, 47, 76, 91	0
2	H	151/166 (90%)	-0.00	1 (0%) 87 91	20, 42, 87, 96	0
All	All	1808/2032 (88%)	0.10	36 (1%) 65 69	17, 52, 91, 118	0

The worst 5 of 36 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	109	CYS	4.7
1	C	74	PRO	4.2
1	B	294	ALA	4.0
1	C	273	ASP	3.7
1	B	79	VAL	3.4

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

There are no such residues in this entry.