WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Nov 8, 2021 – 08:07 pm GMT

PDB ID	:	7OJS
Title	:	Complex structure 2 of the Bacillus subtilis CdaA c-di-AMP cyclase domain (CdaACD) and the phosphoglucomutase GlmM short variant (GlmMF369)
Authors	:	Pathania, M.; Grundling, A.G.; Freemont, P.
Deposited on	:	2021-05-17
Resolution	:	4.20 Å(reported)

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	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (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 ⓘ

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

The reported resolution of this entry is 4.20 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.
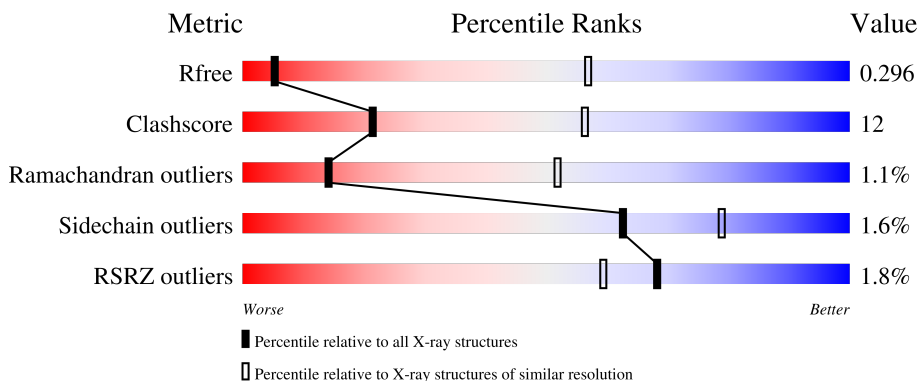
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)
RSRZ outliers	127900	1063 (4.70-3.70)

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	369	3% / 70% / 29%
1	B	369	5% / 72% / 27% .
1	C	369	2% / 72% / 28%
1	F	369	.% / 73% / 26% .
1	G	369	2% / 72% / 27% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	J	369	% 78% 22%
2	D	167	% 65% 21% · 13%
2	E	167	67% 18% · 13%
2	H	167	2% 57% 29% · 13%
2	I	167	2% 60% 27% · 13%
2	K	167	59% 25% · 13%
2	L	167	60% 25% · 13%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 23208 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 Phosphoglucosamine mutase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	368	Total	C	N	O	S	0	1	0
			2762	1731	464	553	14			
1	F	368	Total	C	N	O	S	0	1	0
			2762	1731	464	553	14			
1	A	368	Total	C	N	O	S	3	1	0
			2762	1731	464	553	14			
1	B	368	Total	C	N	O	S	0	1	0
			2762	1731	464	553	14			
1	G	368	Total	C	N	O	S	0	1	0
			2762	1731	464	553	14			
1	J	368	Total	C	N	O	S	3	1	0
			2762	1731	464	553	14			

- Molecule 2 is a protein called Cyclic di-AMP synthase CdaA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	146	Total	C	N	O	S	121	0	0
			1106	692	187	222	5			
2	I	146	Total	C	N	O	S	121	0	0
			1106	692	187	222	5			
2	D	146	Total	C	N	O	S	121	0	0
			1106	692	187	222	5			
2	E	146	Total	C	N	O	S	121	0	0
			1106	692	187	222	5			
2	K	146	Total	C	N	O	S	121	0	0
			1106	692	187	222	5			
2	L	146	Total	C	N	O	S	121	0	0
			1106	692	187	222	5			

3 Residue-property plots ⓘ

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.
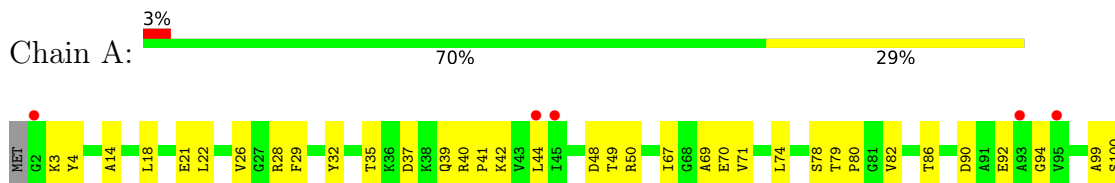
• Molecule 1: Phosphoglucosamine mutase

Chain C:

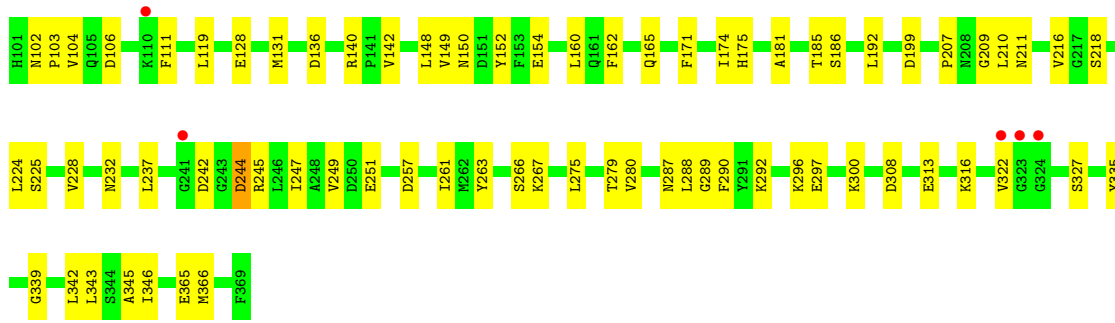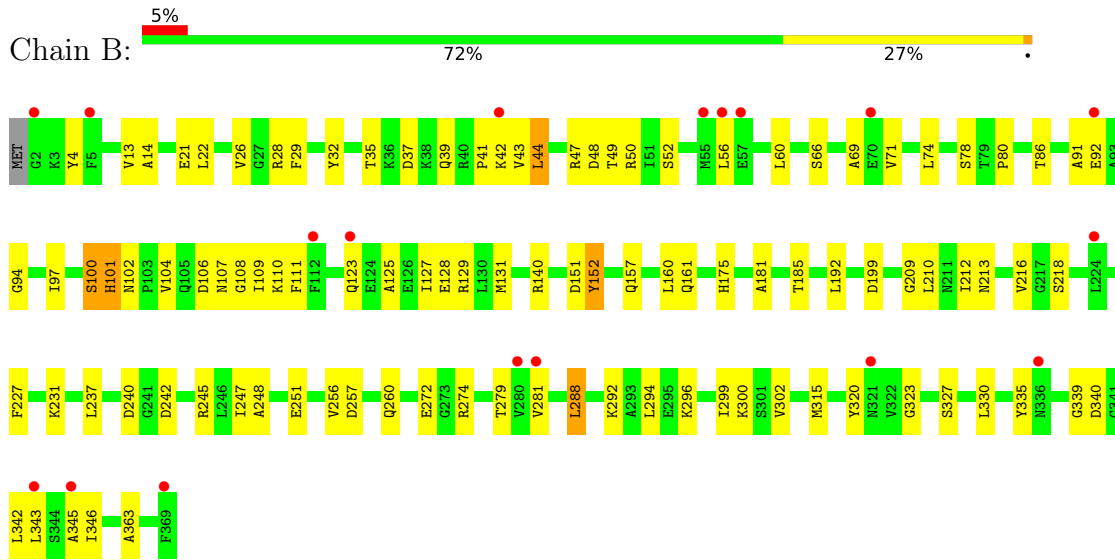• Molecule 1: Phosphoglucosamine mutase

Chain F:

• Molecule 1: Phosphoglucosamine mutase

Chain A:

• Molecule 1: Phosphoglucosamine mutase

Chain B:

• Molecule 1: Phosphoglucosamine mutase

Chain G:

• Molecule 1: Phosphoglucosamine mutase

Chain J:

K110 F111 M131 R140 L146 D151 E154 Q157 L160 L163 T166 A181 L192 L210 V216 G217 S218 S225 A226 F227 K231 L237 R245 L246 I247 A248 V249 D250 E251 K252 G253 V256 D257 I261 H268 R274 T279

T283 K292 K300 S301 D308 M315 N321 G324 E325 Q326 S327 L333 G339 D340 L343 S344 I346 M347 L348 T351 L359 F369

• Molecule 2: Cyclic di-AMP synthase CdaA

Chain H:

2%   57%   29%   ·   13%

E107 A108 K111 T112 A119 M123 R126 R127 A130 T133 I134 E135 T138 Y143 I144 P149 L150 N151 A152 K153 V154 E157 L158 F163 I164 P165 N166 T167 P168 L169 I175 M176 I181 Y187 L188 P189 K198 E199 L200 G201 T202 R203 H204 V208

G209 I210 S211 E212 V213 T214 T218 I219 I220 E223 V230 N233 G234 L245 M248 E252 PHE LYS LYS ASN THR ARG ASP THR SER ASN ARG TRP TYR TRP ARG GLY LYS LYS ASN GLY

• Molecule 2: Cyclic di-AMP synthase CdaA

Chain I:

2%   60%   27%   ·   13%

E107 Q110 K111 T112 I116 A119 I120 N121 Y122 R126 R127 L132 T133 M140 Y143 I144 A152 S156 E157 L158 L159 I160 I164 P165 N166 L169 A173 V174 I175 M176 A182 C186 Y187 L188 S191 L200 G201 T202 R203 H204 R205 A206 A207 V208

G209 I210 S211 E212 L217 T218 I219 E223 V230 A231 K232 L245 L249 E252 PHE LYS LYS ASN THR ARG ASP THR SER ASN ARG TRP TYR TRP ARG GLY LYS LYS ASN GLY

• Molecule 2: Cyclic di-AMP synthase CdaA

Chain D:

%   65%   21%   ·   13%

E107 Q110 K111 T112 I116 R126 R127 I128 G129 I134 I144 L150 N151 A152 L158 I164 P165 N166 L169 I175 M176 Y187 L188 E199 L200 G201 T202 R203 H204 A207 V208 G209 I210 S211 E212 V213 T214 L217 T218 E223 V230 H237

L240 E252 PHE LYS LYS ASN THR ARG ASP THR SER SER ASN ARG TRP TYR TRP ARG GLY LYS LYS ASN GLY

• Molecule 2: Cyclic di-AMP synthase CdaA

Chain E:

67%   18%   ·   13%

E107 Q110 K111 T112 M123 R127 I128 L132 A152 L158 I162 F163 I164 P165 L169 H170 D171 G172 M176 A185 C186 Y187 S197 L200 G201 T202 R203 H204 T214 D215 S216 L217 T218 I219 E223 E224 A231 E239 L249 E252 PHE LYS

LYS ASN THR ARG ASP THR SER ASN ARG TRP TYR TRP ARG GLY LYS LYS ASN GLY

• Molecule 2: Cyclic di-AMP synthase CdaA

Chain K:

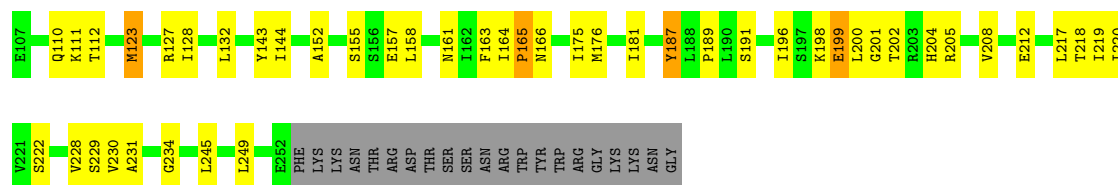59%   25%   •   13%
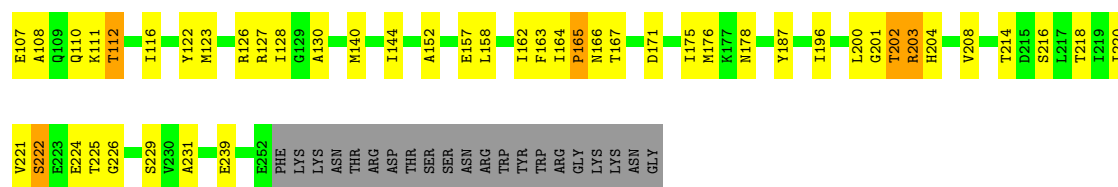
• Molecule 2: Cyclic di-AMP synthase CdaA

Chain L:

60%   25%   •   13%

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	62.80Å 228.54Å 153.19Å 90.00° 99.86° 90.00°	Depositor
Resolution (Å)	75.47 – 4.20 75.47 – 4.20	Depositor EDS
% Data completeness (in resolution range)	99.2 (75.47-4.20) 99.2 (75.47-4.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.29 (at 4.15Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, R_{free}	0.226 , 0.295 0.226 , 0.296	Depositor DCC
R_{free} test set	1497 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å^2)	103.1	Xtriage
Anisotropy	0.957	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< |L| > = 0.39, < L^2 > = 0.22$	Xtriage
Estimated twinning fraction	0.084 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	23208	wwPDB-VP
Average B, all atoms (Å^2)	116.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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	0.26	0/2806	0.46	0/3781
1	B	0.26	0/2806	0.46	0/3781
1	C	0.26	0/2806	0.47	0/3781
1	F	0.26	0/2806	0.47	0/3781
1	G	0.26	0/2806	0.48	0/3781
1	J	0.26	0/2806	0.46	0/3781
2	D	0.26	0/1118	0.53	1/1512 (0.1%)
2	E	0.28	0/1118	0.51	0/1512
2	H	0.27	0/1118	0.54	1/1512 (0.1%)
2	I	0.30	0/1118	0.58	1/1512 (0.1%)
2	K	0.31	0/1118	0.58	0/1512
2	L	0.27	0/1118	0.53	0/1512
All	All	0.27	0/23544	0.49	3/31758 (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	I	187	TYR	C-N-CA	8.79	143.67	121.70
2	H	200	LEU	CA-CB-CG	5.41	127.75	115.30
2	D	200	LEU	CA-CB-CG	5.33	127.57	115.30

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2762	0	2732	67	0
1	B	2762	0	2732	70	0
1	C	2762	0	2732	66	0
1	F	2762	0	2732	66	0
1	G	2762	0	2732	69	0
1	J	2762	0	2732	50	0
2	D	1106	0	1128	28	0
2	E	1106	0	1128	24	0
2	H	1106	0	1128	35	0
2	I	1106	0	1127	33	0
2	K	1106	0	1128	32	0
2	L	1106	0	1128	31	0
All	All	23208	0	23159	539	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 539 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:H:123:MET:HE1	2:H:163:PHE:HB2	1.55	0.87
1:J:42:LYS:H	1:J:92:GLU:HG2	1.41	0.84
1:J:41:PRO:HD2	1:J:69:ALA:HA	1.60	0.81
2:L:144:ILE:HG12	2:L:175:ILE:HG21	1.63	0.79
2:L:200:LEU:HD23	2:L:222:SER:OG	1.83	0.78

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	367/369 (100%)	342 (93%)	22 (6%)	3 (1%)	19	60

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	367/369 (100%)	347 (95%)	17 (5%)	3 (1%)	19	60
1	C	367/369 (100%)	339 (92%)	25 (7%)	3 (1%)	19	60
1	F	367/369 (100%)	343 (94%)	22 (6%)	2 (0%)	29	68
1	G	367/369 (100%)	340 (93%)	24 (6%)	3 (1%)	19	60
1	J	367/369 (100%)	344 (94%)	21 (6%)	2 (0%)	29	68
2	D	144/167 (86%)	128 (89%)	14 (10%)	2 (1%)	11	47
2	E	144/167 (86%)	122 (85%)	18 (12%)	4 (3%)	5	33
2	H	144/167 (86%)	128 (89%)	14 (10%)	2 (1%)	11	47
2	I	144/167 (86%)	129 (90%)	13 (9%)	2 (1%)	11	47
2	K	144/167 (86%)	126 (88%)	13 (9%)	5 (4%)	3	29
2	L	144/167 (86%)	127 (88%)	14 (10%)	3 (2%)	7	39
All	All	3066/3216 (95%)	2815 (92%)	217 (7%)	34 (1%)	14	52

5 of 34 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	100	SER
1	A	100	SER
1	B	100	SER
2	K	165	PRO
2	L	202	THR

5.3.2　Protein sidechains ⓘ

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	294/294 (100%)	290 (99%)	4 (1%)	67	80
1	B	294/294 (100%)	291 (99%)	3 (1%)	76	86
1	C	294/294 (100%)	291 (99%)	3 (1%)	76	86
1	F	294/294 (100%)	292 (99%)	2 (1%)	84	90
1	G	294/294 (100%)	289 (98%)	5 (2%)	60	78

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	J	294/294 (100%)	292 (99%)	2 (1%)	84	90
2	D	120/139 (86%)	116 (97%)	4 (3%)	38	61
2	E	120/139 (86%)	117 (98%)	3 (2%)	47	68
2	H	120/139 (86%)	118 (98%)	2 (2%)	60	78
2	I	120/139 (86%)	117 (98%)	3 (2%)	47	68
2	K	120/139 (86%)	115 (96%)	5 (4%)	30	55
2	L	120/139 (86%)	117 (98%)	3 (2%)	47	68
All	All	2484/2598 (96%)	2445 (98%)	39 (2%)	62	79

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	366	MET
2	K	222	SER
1	J	44	LEU
2	K	178	ASN
2	L	123	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	213	ASN
1	F	213	ASN
1	G	213	ASN
2	K	204	HIS

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 ⓘ

There are no ligands in this entry.

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, 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	368/369 (99%)	0.18	10 (2%)	54	44	77, 119, 150, 176	0
1	B	368/369 (99%)	0.50	18 (4%)	29	25	97, 135, 174, 192	0
1	C	368/369 (99%)	0.26	7 (1%)	66	58	77, 113, 149, 169	0
1	F	368/369 (99%)	0.16	5 (1%)	75	65	70, 102, 134, 158	0
1	G	368/369 (99%)	0.21	7 (1%)	66	58	83, 129, 156, 176	0
1	J	368/369 (99%)	0.29	2 (0%)	91	86	71, 118, 145, 164	0
2	D	146/167 (87%)	0.01	1 (0%)	87	82	76, 100, 140, 173	27 (18%)
2	E	146/167 (87%)	-0.06	0	100	100	71, 104, 119, 150	27 (18%)
2	H	146/167 (87%)	0.13	3 (2%)	63	54	72, 104, 137, 174	27 (18%)
2	I	146/167 (87%)	0.20	3 (2%)	63	54	77, 107, 129, 145	27 (18%)
2	K	146/167 (87%)	-0.01	0	100	100	68, 94, 115, 137	27 (18%)
2	L	146/167 (87%)	0.01	0	100	100	62, 96, 132, 155	27 (18%)
All	All	3084/3216 (95%)	0.20	56 (1%)	68	59	62, 113, 153, 192	162 (5%)

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	2	GLY	4.4
1	A	324	GLY	4.2
2	H	133	THR	3.9
1	G	2	GLY	3.5
1	J	324	GLY	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 ⓘ

There are no ligands in this entry.

6.5   Other polymers ⓘ

There are no such residues in this entry.