W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Oct 6, 2024 – 06:24 PM EDT

PDB ID	:	3E6M
Title	:	The crystal structure of a MarR family transcriptional regulator from Silicibacter pomeroyi DSS.
Authors	:	Tan, K.; Volkart, L.; Freeman, L.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2008-08-15
Resolution	:	2.20 Å(reported)

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

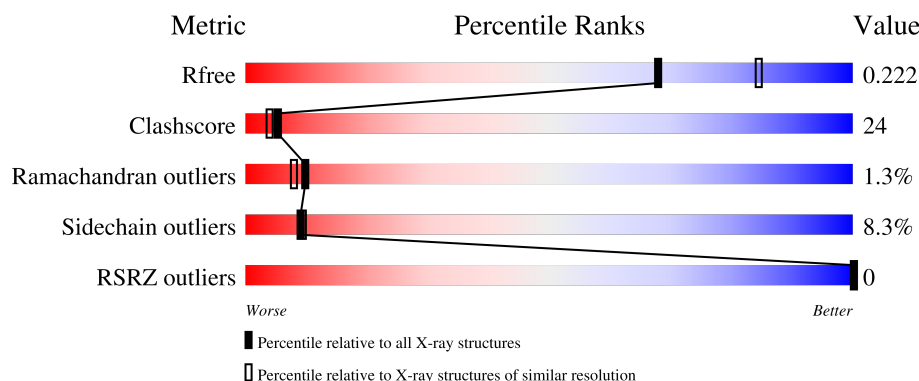|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 2.20 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	Percentile Ranks	Value
Rfree		0.222
Clashscore		24
Ramachandran outliers		1.3%
Sidechain outliers		8.3%
RSRZ outliers		0

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	5791 (2.20-2.20)
Clashscore	180529	6634 (2.20-2.20)
Ramachandran outliers	177936	6560 (2.20-2.20)
Sidechain outliers	177891	6561 (2.20-2.20)
RSRZ outliers	164620	5791 (2.20-2.20)

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	161	52% 34% 5% 9%
1	B	161	52% 34% 6% 9%
1	C	161	50% 34% 7% 9%
1	D	161	52% 38% • 7%
1	E	161	52% 35% 5% 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	161	55% 30% 5% 9%
1	G	161	47% 41% 8%
1	H	161	61% 29% 9%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 9532 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 MarR family Transcriptional regulator.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	146	Total	C	N	O	S	Se	0	0	0
			1115	708	192	213	1	1			
1	B	147	Total	C	N	O	S	Se	0	0	0
			1132	717	192	221	1	1			
1	C	146	Total	C	N	O	S	Se	0	0	0
			1119	709	192	216	1	1			
1	D	149	Total	C	N	O	S	Se	0	0	0
			1142	723	194	223	1	1			
1	E	147	Total	C	N	O	S	Se	0	0	0
			1135	721	193	219	1	1			
1	F	146	Total	C	N	O	S	Se	0	0	0
			1128	715	192	219	1	1			
1	G	148	Total	C	N	O	S	Se	0	0	0
			1144	726	195	221	1	1			
1	H	147	Total	C	N	O	S	Se	0	0	0
			1132	717	192	221	1	1			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q5LLB9
A	-1	ASN	-	expression tag	UNP Q5LLB9
A	0	ALA	-	expression tag	UNP Q5LLB9
B	-2	SER	-	expression tag	UNP Q5LLB9
B	-1	ASN	-	expression tag	UNP Q5LLB9
B	0	ALA	-	expression tag	UNP Q5LLB9
C	-2	SER	-	expression tag	UNP Q5LLB9
C	-1	ASN	-	expression tag	UNP Q5LLB9
C	0	ALA	-	expression tag	UNP Q5LLB9
D	-2	SER	-	expression tag	UNP Q5LLB9
D	-1	ASN	-	expression tag	UNP Q5LLB9
D	0	ALA	-	expression tag	UNP Q5LLB9
E	-2	SER	-	expression tag	UNP Q5LLB9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	-1	ASN	-	expression tag	UNP Q5LLB9
E	0	ALA	-	expression tag	UNP Q5LLB9
F	-2	SER	-	expression tag	UNP Q5LLB9
F	-1	ASN	-	expression tag	UNP Q5LLB9
F	0	ALA	-	expression tag	UNP Q5LLB9
G	-2	SER	-	expression tag	UNP Q5LLB9
G	-1	ASN	-	expression tag	UNP Q5LLB9
G	0	ALA	-	expression tag	UNP Q5LLB9
H	-2	SER	-	expression tag	UNP Q5LLB9
H	-1	ASN	-	expression tag	UNP Q5LLB9
H	0	ALA	-	expression tag	UNP Q5LLB9

- Molecule 2 is water.

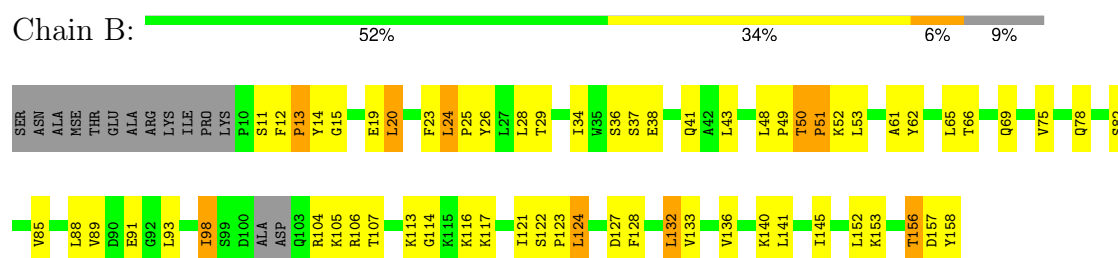| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	65	Total 65	O 65	0	0
2	B	57	Total 57	O 57	0	0
2	C	59	Total 59	O 59	0	0
2	D	50	Total 50	O 50	0	0
2	E	69	Total 69	O 69	0	0
2	F	58	Total 58	O 58	0	0
2	G	62	Total 62	O 62	0	0
2	H	65	Total 65	O 65	0	0

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.

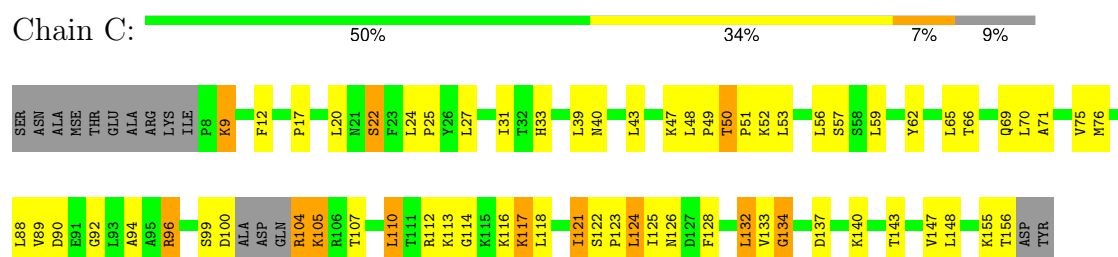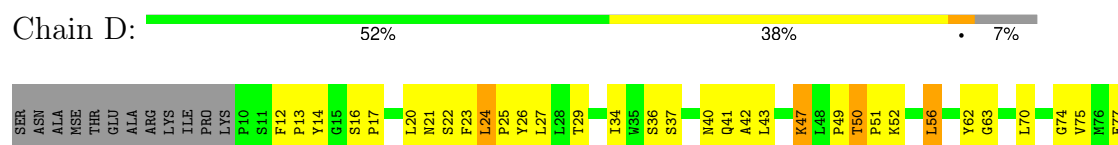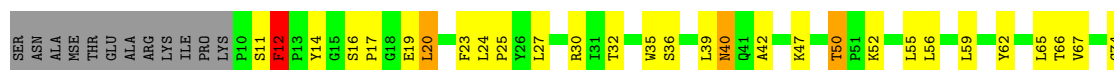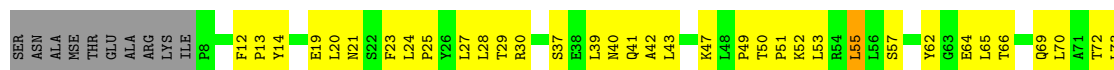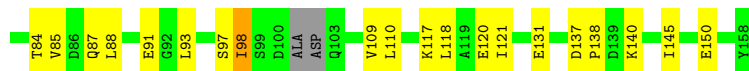• Molecule 1: MarR family Transcriptional regulator

Chain A:

• Molecule 1: MarR family Transcriptional regulator

Chain B:

• Molecule 1: MarR family Transcriptional regulator

Chain C:

• Molecule 1: MarR family Transcriptional regulator

Chain D:

Q78 S82 R83 T84 V85 D86 Q87 L88 L93 R96 S99 Q103 R104 K105 R106 T107 V108 V109 R112 K116 K117 L118 A119 S122 P123 L124 I125 N126 D127 V133 G134 N135 V136 D137 K140 L141 I145 L148 T156 D157 Y158

• Molecule 1: MarR family Transcriptional regulator

Chain E: 52% 35% 5% 9%

SER ASN ALA MSE THR GLU ALA ARG LYS ILE P8 K9 P13 L20 N21 S22 F23 L24 P25 Y26 L27 L28 T29 I34 W35 S36 S37 E38 L39 N40 Q41 A42 K47 T50 P51 K52 L55 L56 S57 A61 Y62 G63 E64 L65 Q69 L70 A71 T72 L73 G74 V75 M76

E77 Q78 S79 T80 T81 T84 V85 L88 V89 D90 R96 S97 I98 S99 ASP ALA ASP GLN K105 R104 V109 L110 T111 R112 K113 G114 K115 L118 I121 S122 P123 L124 I125 V133 G134 N135 K140 L141 Q142 E146 V147 L152 Y158

• Molecule 1: MarR family Transcriptional regulator

Chain F: 55% 30% 5% • 9%

SER ASN ALA MSE THR GLU ALA ARG LYS ILE PRO LYS P10 S11 F12 P13 Y14 G15 S16 P17 G18 E19 L20 F23 L24 P25 Y26 L27 R30 I31 T32 W35 S36 L39 N40 Q41 A42 K47 T50 P51 K52 L55 L56 L59 Y62 L65 T66 V67 G74

V75 M76 V85 D86 Q87 L88 E91 G92 L93 I98 S99 ASP ALA ASP Q103 R104 K105 R106 T107 L110 T111 K115 L118 S122 L132 V133 G134 N135 V136 D137 P138 D139 K140 C144 L148 K155 T156 D157 Y158

• Molecule 1: MarR family Transcriptional regulator

Chain G: 47% 41% • • 8%

SER ASN ALA MSE THR GLU ALA ARG LYS ILE P8 F12 P13 Y14 E19 L20 N21 S22 F23 L24 P25 Y26 L27 L28 T29 R30 S37 E38 L39 N40 Q41 A42 L43 K47 L48 P49 T50 P51 K52 L53 R54 L55 L56 S57 Y62 G63 E64 L65 T66 Q69 L70 A71 T72 L73

G74 V75 M76 E77 Q78 S79 T80 T84 V85 V89 L93 R96 S97 I98 S99 ASP ALA ASP Q103 R106 T107 V108 V109 L110 T111 G114 K115 K117 L118 A119 E120 I121 S122 P123 L124 I125 N126 D127 E131 V136 L141 Q142 I145 K155 T156 D157 Y158

• Molecule 1: MarR family Transcriptional regulator

Chain H: 61% 29% • 9%

SER ASN ALA MSE THR GLU ALA ARG LYS ILE PRO LYS P10 Y14 P17 L20 N21 S22 F23 L24 P25 Y26 L27 L28 T29 R30 I31 T32 W35 S36 S37 E38 L39 N40 T50 P51 K52 L55 L56 S60 G63 V67 L73 M76 E77 Q78 S79 T80 T81

T84 V85 D86 Q87 L88 E91 G92 L93 S97 I98 S99 D100 ALA ASP Q103 V109 L110 K117 L118 A119 E120 I121 E131 D137 P138 D139 K140 I145 E150 Y158

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	75.43Å 93.20Å 101.56Å 90.00° 90.02° 90.00°	Depositor
Resolution (Å)	39.60 – 2.20 39.60 – 2.20	Depositor EDS
% Data completeness (in resolution range)	96.1 (39.60-2.20) 97.8 (39.60-2.20)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.00 (at 2.20Å)	Xtriage
Refinement program	PHENIX (phenix.refine)	Depositor
R, R_{free}	0.185 , 0.230 0.185 , 0.222	Depositor DCC
R_{free} test set	3517 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	39.2	Xtriage
Anisotropy	0.756	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 50.5	EDS
L-test for twinning[2]	$< |L| > = 0.52, < L^2 > = 0.35$	Xtriage
Estimated twinning fraction	0.479 for h,-k,-l	Xtriage
Reported twinning fraction	0.486 for h,-k,-l	Depositor
Outliers	1 of 70328 reflections (0.001%)	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	9532	wwPDB-VP
Average B, all atoms (Å^2)	61.0	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 54.99 % 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 3.3753e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	0.39	0/1131	0.59	0/1529
1	B	0.37	0/1148	0.65	0/1551
1	C	0.42	0/1135	0.61	0/1534
1	D	0.39	0/1159	0.60	0/1568
1	E	0.39	0/1152	0.60	0/1556
1	F	0.39	0/1144	0.64	0/1545
1	G	0.39	0/1161	0.58	0/1568
1	H	0.39	0/1148	0.56	0/1551
All	All	0.39	0/9178	0.60	0/12402

There are no bond length outliers.

There are no bond angle outliers.

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	1115	0	1157	64	0
1	B	1132	0	1166	65	0
1	C	1119	0	1162	70	0
1	D	1142	0	1174	71	0
1	E	1135	0	1180	56	0
1	F	1128	0	1168	62	0
1	G	1144	0	1188	74	0
1	H	1132	0	1166	44	0

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	65	0	0	6	0
2	B	57	0	0	5	0
2	C	59	0	0	6	0
2	D	50	0	0	5	0
2	E	69	0	0	2	0
2	F	58	0	0	3	0
2	G	62	0	0	0	0
2	H	65	0	0	3	0
All	All	9532	0	9361	436	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 24.

The worst 5 of 436 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:66:THR:HG21	1:C:105:LYS:HE3	1.45	0.97
1:F:40:ASN:HD21	1:F:50:THR:H	1.12	0.95
1:G:156:THR:C	1:G:158:TYR:H	1.70	0.94
1:C:56:LEU:HB3	1:C:118:LEU:HD11	1.52	0.91
1:A:90:ASP:HB3	2:A:176:HOH:O	1.70	0.90

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	142/161 (88%)	131 (92%)	10 (7%)	1 (1%)	19	19
1	B	143/161 (89%)	124 (87%)	17 (12%)	2 (1%)	9	7
1	C	142/161 (88%)	126 (89%)	13 (9%)	3 (2%)	5	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	147/161 (91%)	122 (83%)	23 (16%)	2 (1%)	9	7
1	E	143/161 (89%)	134 (94%)	8 (6%)	1 (1%)	19	19
1	F	142/161 (88%)	134 (94%)	4 (3%)	4 (3%)	4	2
1	G	144/161 (89%)	132 (92%)	10 (7%)	2 (1%)	9	7
1	H	143/161 (89%)	134 (94%)	9 (6%)	0	100	100
All	All	1146/1288 (89%)	1037 (90%)	94 (8%)	15 (1%)	10	8

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	75	VAL
1	C	134	GLY
1	F	11	SER
1	F	134	GLY
1	G	157	ASP

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	125/138 (91%)	111 (89%)	14 (11%)	5	4
1	B	128/138 (93%)	119 (93%)	9 (7%)	12	14
1	C	127/138 (92%)	112 (88%)	15 (12%)	4	4
1	D	128/138 (93%)	121 (94%)	7 (6%)	18	22
1	E	129/138 (94%)	117 (91%)	12 (9%)	7	7
1	F	128/138 (93%)	119 (93%)	9 (7%)	12	14
1	G	130/138 (94%)	120 (92%)	10 (8%)	10	11
1	H	128/138 (93%)	119 (93%)	9 (7%)	12	14
All	All	1023/1104 (93%)	938 (92%)	85 (8%)	9	9

5 of 85 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	12	PHE
1	G	98	ILE
1	F	40	ASN
1	F	156	THR
1	G	157	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	129	HIS
1	H	40	ASN
1	H	78	GLN
1	G	129	HIS
1	D	21	ASN

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 oligosaccharides 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(Å²)	Q<0.9
1	A	145/161 (90%)	-1.31	0	100	100	42, 56, 90, 112	0
1	B	146/161 (90%)	-1.37	0	100	100	38, 62, 102, 129	0
1	C	145/161 (90%)	-1.33	0	100	100	27, 56, 89, 111	0
1	D	148/161 (91%)	-1.37	0	100	100	27, 59, 98, 120	0
1	E	146/161 (90%)	-1.35	0	100	100	37, 55, 90, 142	0
1	F	145/161 (90%)	-1.34	0	100	100	40, 60, 106, 154	0
1	G	147/161 (91%)	-1.38	0	100	100	34, 54, 98, 138	0
1	H	146/161 (90%)	-1.33	0	100	100	41, 58, 96, 146	0
All	All	1168/1288 (90%)	-1.35	0	100	100	27, 58, 98, 154	0

There are no RSRZ outliers to report.

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.