wwPDB X-ray Structure Validation Summary Report (i)

Aug 6, 2020 – 03:40 PM BST

PDB ID	:	6S3U
Title	:	Adhesin P140 from Mycoplasma Genitalium
Authors	:	Fita, I.; Aparicio, D.
Deposited on	:	2019-06-26
Resolution	:	3.24 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:

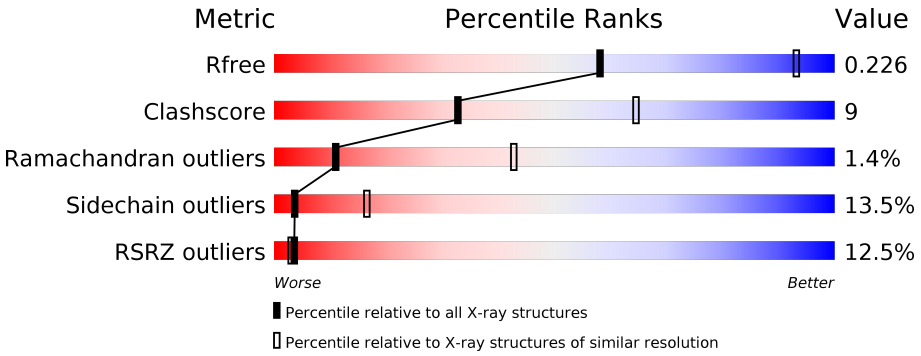
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 3.24 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	1619 (3.28-3.20)
Clashscore	141614	1755 (3.28-3.20)
Ramachandran outliers	138981	1728 (3.28-3.20)
Sidechain outliers	138945	1727 (3.28-3.20)
RSRZ outliers	127900	1567 (3.28-3.20)

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 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	1304	68% 26%
1	B	1304	68% 26%
1	C	1304	70% 25%
1	D	1304	69% 24%
1	E	1304	34% 69% 24%
1	F	1304	33% 69% 24%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 59767 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 Adhesin P1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1278	Total	C	N	O	S	0	0	0
			10016	6351	1692	1960	13			
1	B	1275	Total	C	N	O	S	0	0	0
			9999	6340	1689	1957	13			
1	C	1274	Total	C	N	O	S	0	0	0
			9993	6339	1688	1953	13			
1	D	1271	Total	C	N	O	S	0	0	0
			9976	6328	1685	1950	13			
1	E	1262	Total	C	N	O	S	0	0	0
			9886	6272	1666	1935	13			
1	F	1262	Total	C	N	O	S	0	0	0
			9897	6281	1668	1935	13			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1352	HIS	-	expression tag	UNP P20796
A	1353	HIS	-	expression tag	UNP P20796
A	1354	HIS	-	expression tag	UNP P20796
A	1355	HIS	-	expression tag	UNP P20796
A	1356	HIS	-	expression tag	UNP P20796
A	1357	HIS	-	expression tag	UNP P20796
B	1352	HIS	-	expression tag	UNP P20796
B	1353	HIS	-	expression tag	UNP P20796
B	1354	HIS	-	expression tag	UNP P20796
B	1355	HIS	-	expression tag	UNP P20796
B	1356	HIS	-	expression tag	UNP P20796
B	1357	HIS	-	expression tag	UNP P20796
C	1352	HIS	-	expression tag	UNP P20796
C	1353	HIS	-	expression tag	UNP P20796
C	1354	HIS	-	expression tag	UNP P20796
C	1355	HIS	-	expression tag	UNP P20796
C	1356	HIS	-	expression tag	UNP P20796

Continued on next page...

Continued from previous page...

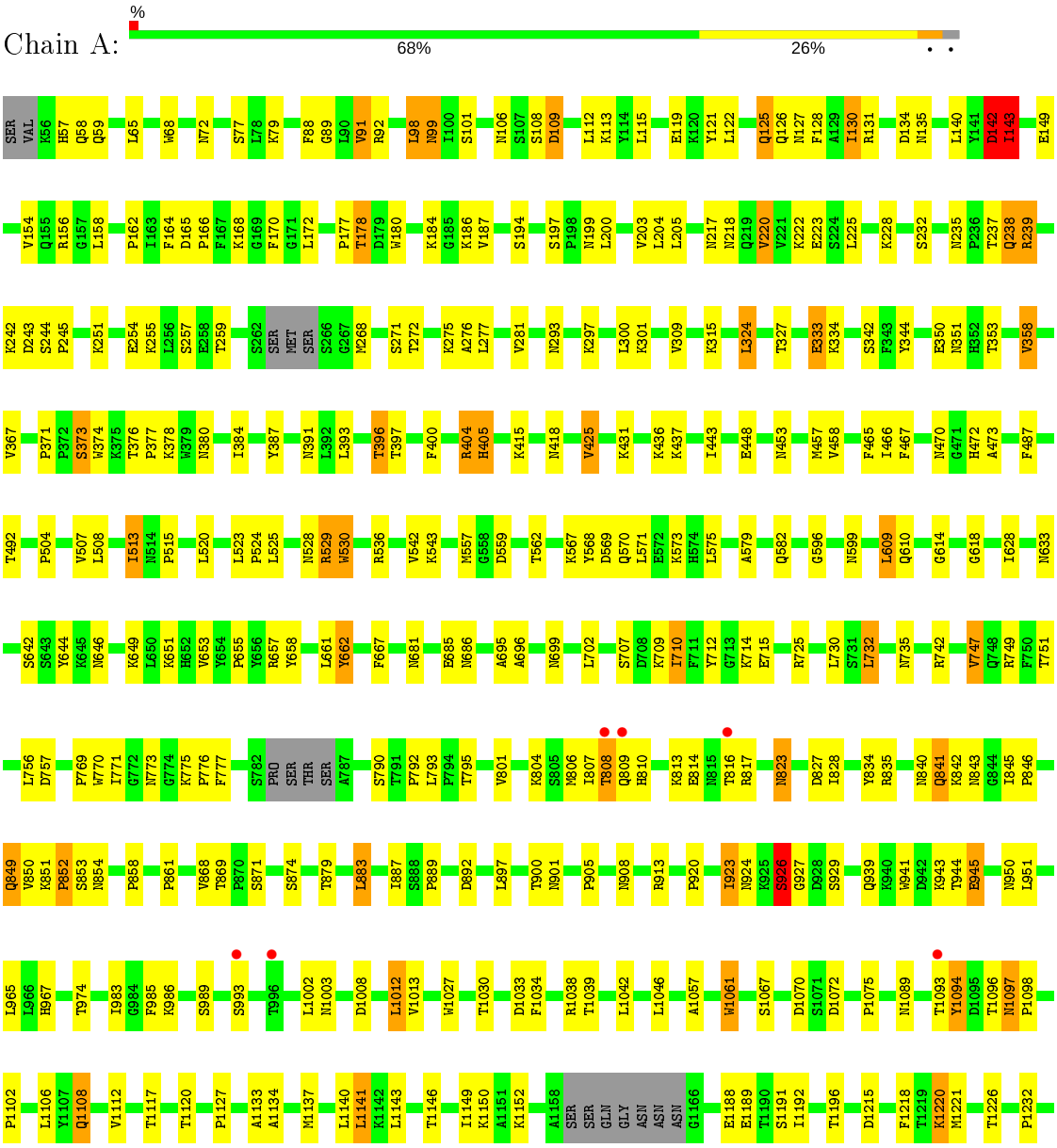
Chain	Residue	Modelled	Actual	Comment	Reference
C	1357	HIS	-	expression tag	UNP P20796
D	1352	HIS	-	expression tag	UNP P20796
D	1353	HIS	-	expression tag	UNP P20796
D	1354	HIS	-	expression tag	UNP P20796
D	1355	HIS	-	expression tag	UNP P20796
D	1356	HIS	-	expression tag	UNP P20796
D	1357	HIS	-	expression tag	UNP P20796
E	1352	HIS	-	expression tag	UNP P20796
E	1353	HIS	-	expression tag	UNP P20796
E	1354	HIS	-	expression tag	UNP P20796
E	1355	HIS	-	expression tag	UNP P20796
E	1356	HIS	-	expression tag	UNP P20796
E	1357	HIS	-	expression tag	UNP P20796
F	1352	HIS	-	expression tag	UNP P20796
F	1353	HIS	-	expression tag	UNP P20796
F	1354	HIS	-	expression tag	UNP P20796
F	1355	HIS	-	expression tag	UNP P20796
F	1356	HIS	-	expression tag	UNP P20796
F	1357	HIS	-	expression tag	UNP P20796

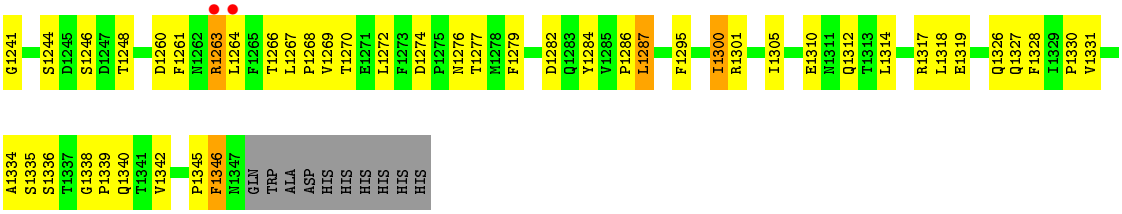
3 Residue-property plots (i)

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: Adhesin P1

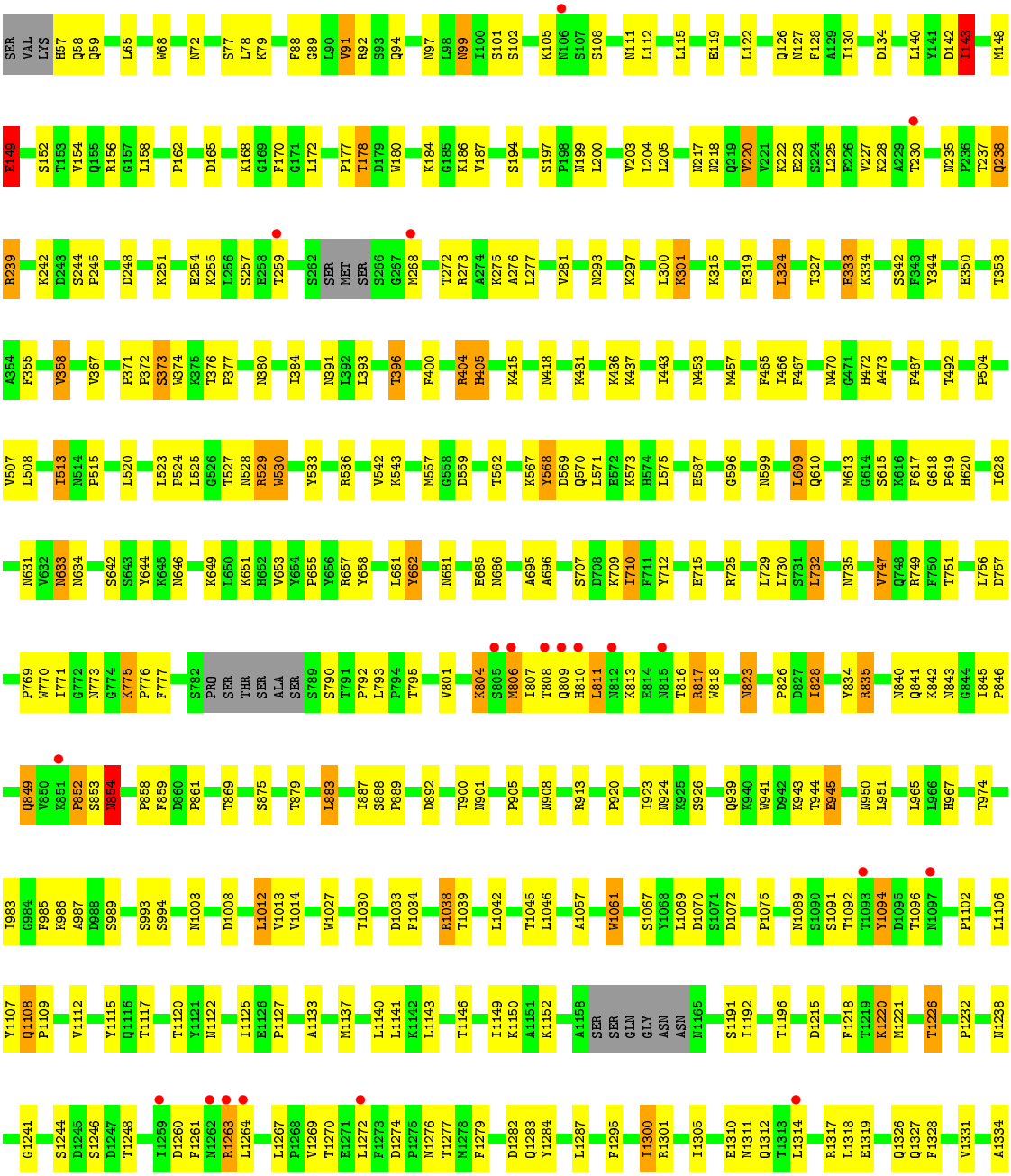
Chain A:

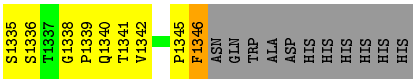




• Molecule 1: Adhesin P1

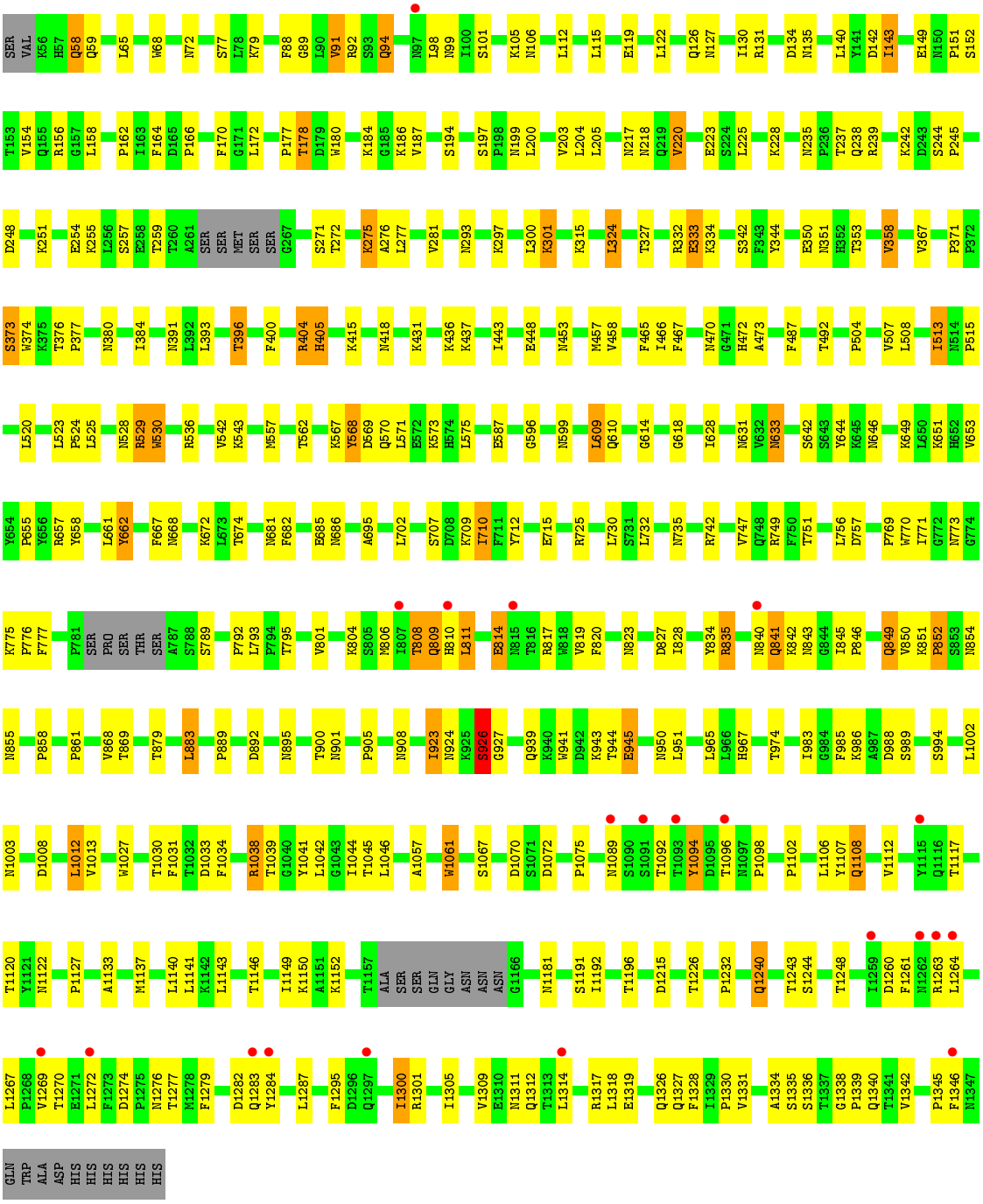
Chain B:





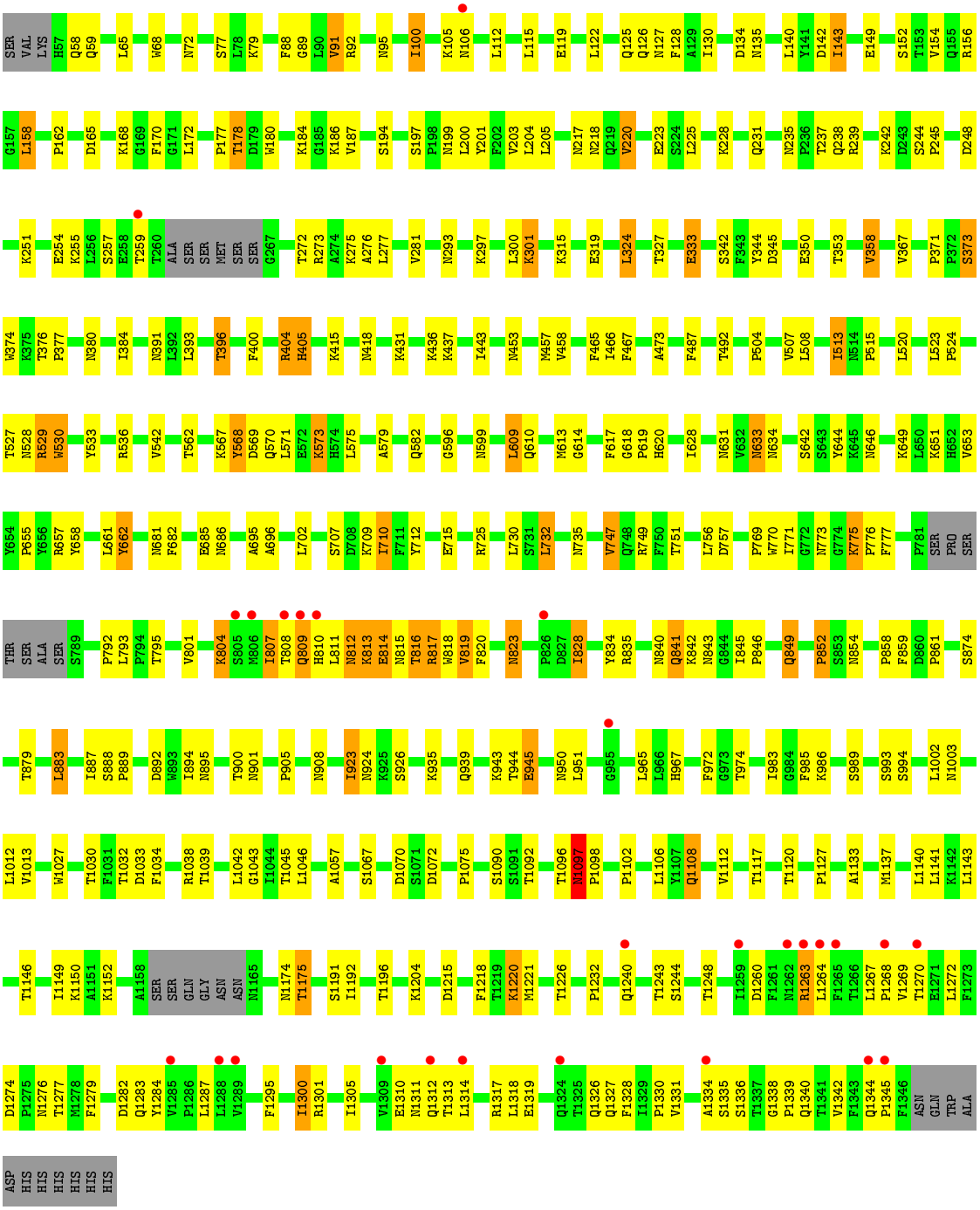
- Molecule 1: Adhesin P1

Chain C:



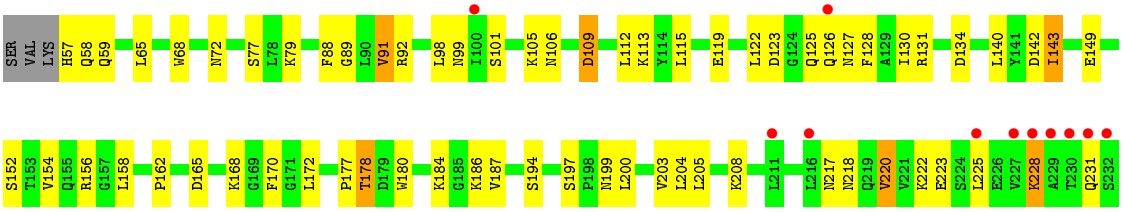
- Molecule 1: Adhesin P1

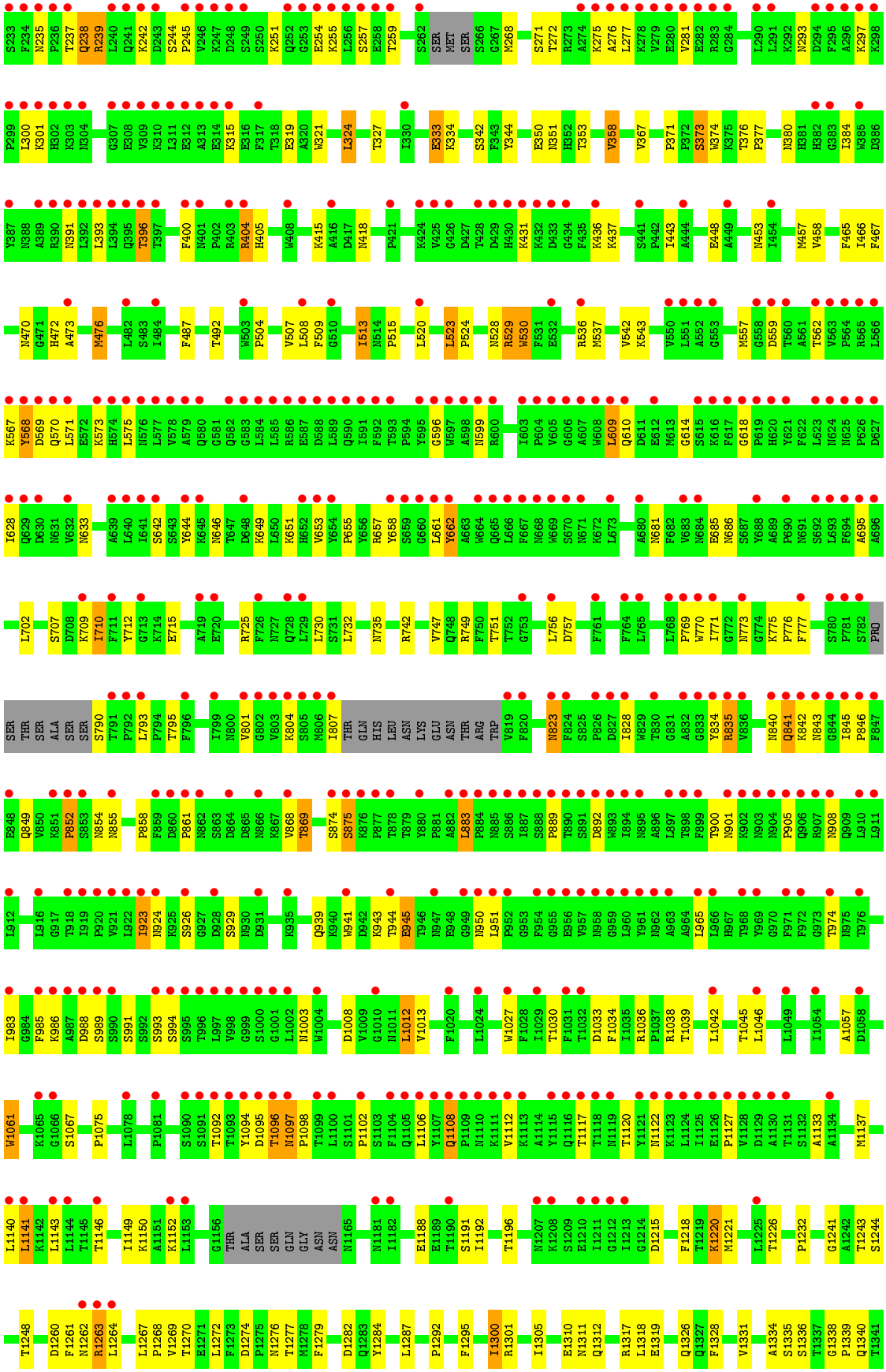
Chain D:



- Molecule 1: Adhesin P1

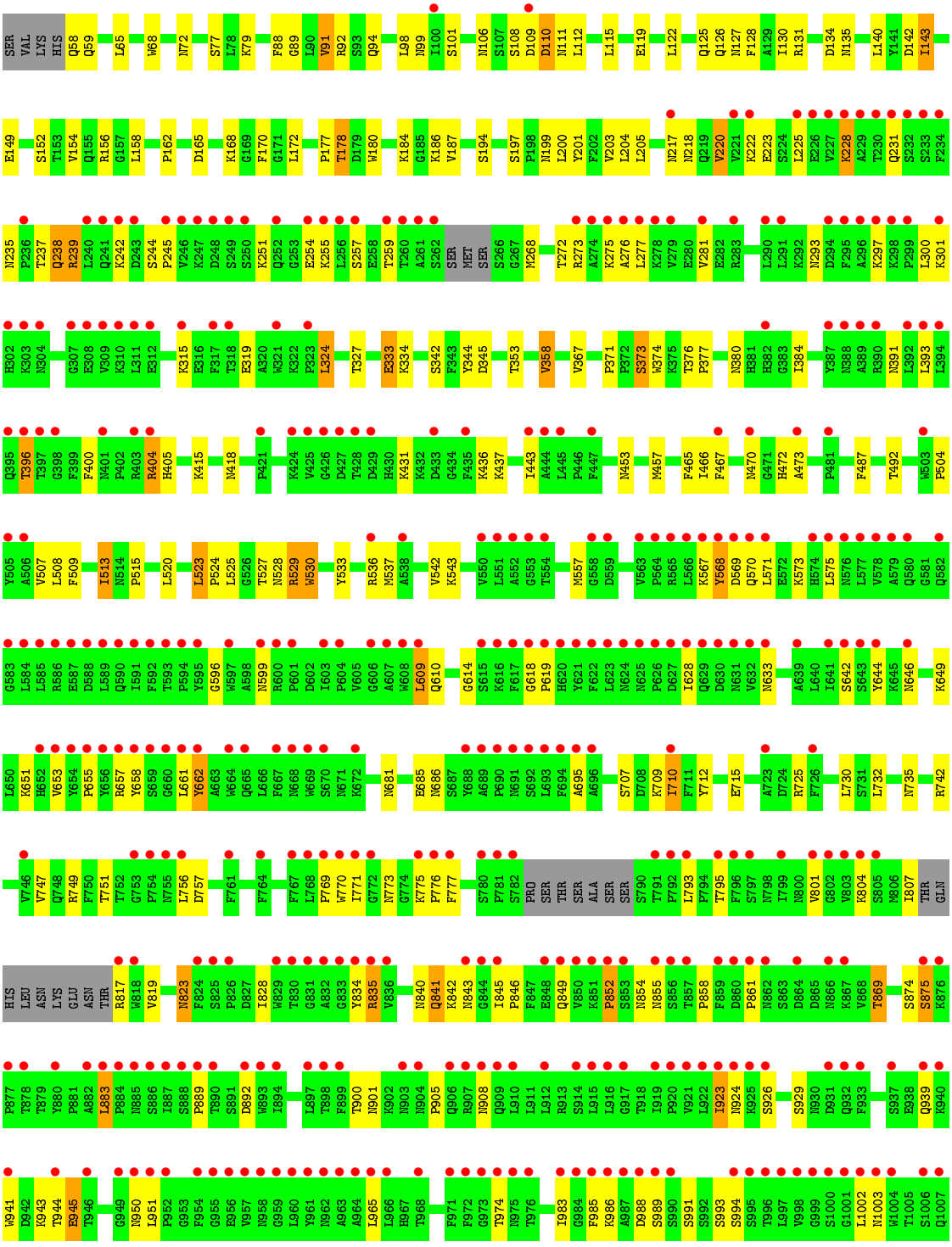
Chain E:

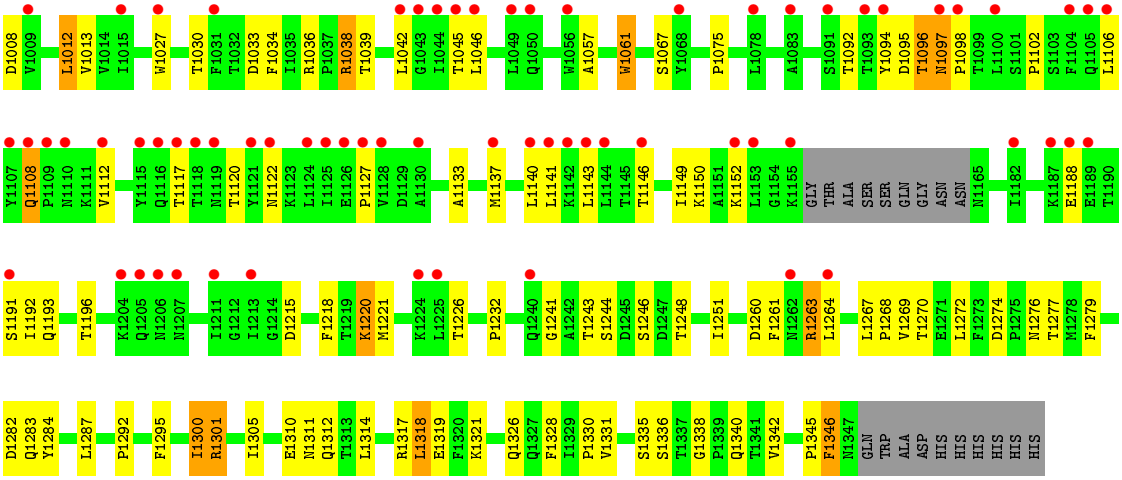




- Molecule 1: Adhesin P1

Chain F:




D1008 V1009 L1012 V1013 Y1014 I1015 W1027 T1030 F1031 T1032 D1033 F1034 I1035 R1036 P1037 R1038 T1039 L1042 G1043 I1044 T1045 L1046 L1049 Q1050 W1056 A1057 W1061 S1067 Y1068 P1075 L1078 A1083 S1091 T1092 T1093 Y1094 D1095 T1096 N1097 P1098 T1099 L1100 S1101 P1102 S1103 F1104 Q1105 L1106
Y1107 Q1108 P1109 N1110 K1111 V1112 Y1115 Q1116 T1117 T1118 N1119 T1120 Y1121 N1122 K1123 L1124 I1125 E1126 P1127 V1128 D1129 A1130 A1133 M1137 L1140 L1141 K1142 L1143 L1144 T1145 T1146 I1149 K1150 A1151 K1152 L1153 G1154 K1155 GLY THR ALA SER SER GLN GLY ASN ASN N1165 I1182 K1187 E1188 E1189 T1190
S1191 I1192 Q1193 T1196 K1204 Q1205 N1206 N1207 T1211 G1212 I1213 G1214 D1215 F1218 T1219 K1220 M1221 K1224 L1225 T1226 P1232 Q1240 G1241 A1242 T1243 S1244 D1245 S1246 D1247 T1248 I1251 D1260 F1261 N1262 R1263 L1264 L1267 P1268 V1269 T1270 E1271 L1272 F1273 D1274 P1275 N1276 T1277 M1278 F1279
D1282 Q1283 Y1284 L1287 P1292 F1295 I1300 R1301 I1305 E1310 N1311 Q1312 T1313 L1314 R1317 L1318 E1319 F1320 K1321 Q1326 Q1327 F1328 I1329 P1330 V1331 S1335 S1336 T1337 G1338 P1339 Q1340 T1341 V1342 P1345 F1346 N1347 GLN TRP ALA ASP HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	459.19Å 116.66Å 285.64Å 90.00° 124.20° 90.00°	Depositor
Resolution (Å)	37.36 – 3.24 236.25 – 3.24	Depositor EDS
% Data completeness (in resolution range)	61.7 (37.36-3.24) 61.8 (236.25-3.24)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.62 (at 3.26Å)	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
R, R_{free}	0.184 , 0.205 0.203 , 0.226	Depositor DCC
R_{free} test set	6155 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å^2)	59.0	Xtriage
Anisotropy	0.038	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 96.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	59767	wwPDB-VP
Average B, all atoms (Å^2)	91.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.81% 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	0.52	0/10269	0.78	6/13978 (0.0%)
1	B	0.49	0/10252	0.76	3/13956 (0.0%)
1	C	0.50	0/10246	0.76	1/13947 (0.0%)
1	D	0.50	0/10229	0.76	3/13925 (0.0%)
1	E	0.45	0/10135	0.73	2/13795 (0.0%)
1	F	0.45	0/10147	0.73	2/13812 (0.0%)
All	All	0.49	0/61278	0.75	17/83413 (0.0%)

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	A	0	1
1	C	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	476	MET	CB-CG-SD	7.17	133.92	112.40
1	E	861	PRO	N-CA-C	-6.71	94.65	112.10
1	F	861	PRO	N-CA-C	-6.65	94.81	112.10
1	A	109	ASP	C-N-CA	6.20	137.21	121.70
1	A	806	MET	C-N-CA	6.17	137.12	121.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	926	SER	Mainchain
1	C	926	SER	Mainchain

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	10016	0	9726	182	0
1	B	9999	0	9707	183	0
1	C	9993	0	9706	182	0
1	D	9976	0	9687	205	0
1	E	9886	0	9597	156	0
1	F	9897	0	9610	163	0
All	All	59767	0	58033	1029	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 9.

The worst 5 of 1029 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:100:ILE:CD1	1:D:100:ILE:CG1	1.76	1.57
1:D:812:ASN:ND2	1:D:815:ASN:HD22	1.28	1.31
1:F:1279:PHE:CE2	1:F:1301:ARG:HD2	1.78	1.18
1:A:1266:THR:HG21	1:C:1264:LEU:HD11	1.19	1.15
1:A:1264:LEU:HG	1:C:1283:GLN:HB2	1.30	1.09

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	1270/1304 (97%)	1128 (89%)	122 (10%)	20 (2%)	9 40
1	B	1267/1304 (97%)	1136 (90%)	114 (9%)	17 (1%)	12 44
1	C	1266/1304 (97%)	1128 (89%)	123 (10%)	15 (1%)	13 46
1	D	1263/1304 (97%)	1127 (89%)	117 (9%)	19 (2%)	10 41
1	E	1252/1304 (96%)	1118 (89%)	117 (9%)	17 (1%)	11 43
1	F	1252/1304 (96%)	1114 (89%)	119 (10%)	19 (2%)	10 41
All	All	7570/7824 (97%)	6751 (89%)	712 (9%)	107 (1%)	11 43

5 of 107 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	108	SER
1	A	1094	TYR
1	A	1097	ASN
1	A	1241	GLY
1	B	811	LEU

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	1121/1146 (98%)	964 (86%)	157 (14%)	3 15
1	B	1120/1146 (98%)	965 (86%)	155 (14%)	3 15
1	C	1118/1146 (98%)	973 (87%)	145 (13%)	4 18
1	D	1117/1146 (98%)	969 (87%)	148 (13%)	4 17
1	E	1107/1146 (97%)	958 (86%)	149 (14%)	4 16
1	F	1108/1146 (97%)	958 (86%)	150 (14%)	4 16
All	All	6691/6876 (97%)	5787 (86%)	904 (14%)	4 16

5 of 904 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	883	LEU
1	D	415	LYS
1	F	747	VAL
1	C	1012	LEU
1	D	100	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 174 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	815	ASN
1	D	391	ASN
1	F	733	ASN
1	C	854	ASN
1	D	71	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 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, 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	1278/1304 (98%)	0.15	8 (0%) 89 85	13, 49, 106, 165	0
1	B	1275/1304 (97%)	0.28	20 (1%) 72 63	24, 60, 127, 238	0
1	C	1274/1304 (97%)	0.24	21 (1%) 72 63	11, 50, 128, 234	0
1	D	1271/1304 (97%)	0.32	27 (2%) 63 53	18, 58, 132, 228	0
1	E	1262/1304 (96%)	1.91	445 (35%) 0 0	25, 156, 238, 252	0
1	F	1262/1304 (96%)	1.75	431 (34%) 0 0	21, 149, 229, 244	0
All	All	7622/7824 (97%)	0.77	952 (12%) 3 3	11, 68, 224, 252	0

The worst 5 of 952 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	920	PRO	20.2
1	F	836	VAL	14.6
1	F	920	PRO	14.3
1	E	1125	ILE	14.1
1	F	961	TYR	13.9

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 (i)

There are no such residues in this entry.