wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 05:17 AM EDT

PDB ID	:	6GEF
Title	:	X-ray structure of the Yersinia pseudotuberculosis ATPase DotB
Authors	:	Prevost, M.S.; Waksman, G.
Deposited on	:	2018-04-26
Resolution	:	2.75 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:

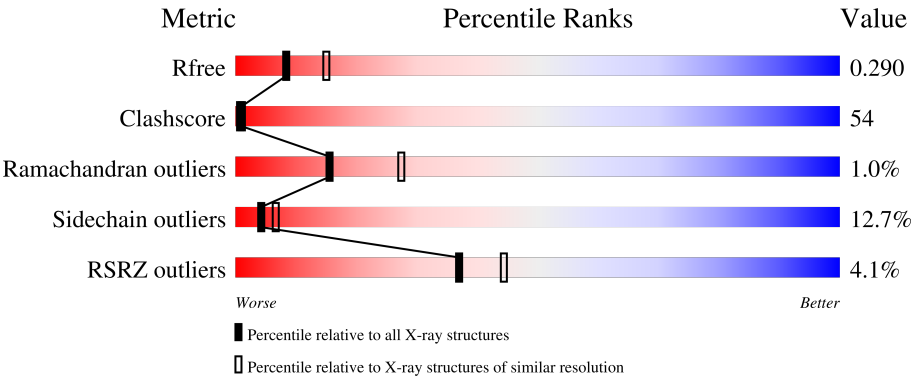
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.75 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	1235 (2.78-2.74)
Clashscore	141614	1277 (2.78-2.74)
Ramachandran outliers	138981	1257 (2.78-2.74)
Sidechain outliers	138945	1257 (2.78-2.74)
RSRZ outliers	127900	1207 (2.78-2.74)

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	402	5% 42% 46% 9% ·
1	B	402	2% 38% 51% 6% ·
1	C	402	4% 40% 46% 10% ·
1	D	402	8% 40% 47% 9% ·
1	E	402	2% 43% 44% 8% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	402	

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 18021 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 Type IV secretion system protein DotB.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	388	Total	C	N	O	S	0	0	0
			3027	1910	526	578	13			
1	B	384	Total	C	N	O	S	0	0	0
			2986	1884	516	573	13			
1	C	385	Total	C	N	O	S	0	0	0
			3009	1898	524	574	13			
1	D	385	Total	C	N	O	S	0	0	0
			2982	1883	516	570	13			
1	E	385	Total	C	N	O	S	0	0	0
			2995	1889	522	571	13			
1	F	386	Total	C	N	O	S	0	0	0
			3022	1907	526	576	13			

There are 96 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	MET	-	initiating methionine	UNP A0A0U1QTI9
A	-12	ALA	-	expression tag	UNP A0A0U1QTI9
A	-11	SER	-	expression tag	UNP A0A0U1QTI9
A	-10	TRP	-	expression tag	UNP A0A0U1QTI9
A	-9	SER	-	expression tag	UNP A0A0U1QTI9
A	-8	HIS	-	expression tag	UNP A0A0U1QTI9
A	-7	PRO	-	expression tag	UNP A0A0U1QTI9
A	-6	GLN	-	expression tag	UNP A0A0U1QTI9
A	-5	PHE	-	expression tag	UNP A0A0U1QTI9
A	-4	GLU	-	expression tag	UNP A0A0U1QTI9
A	-3	LYS	-	expression tag	UNP A0A0U1QTI9
A	-2	ILE	-	expression tag	UNP A0A0U1QTI9
A	-1	GLU	-	expression tag	UNP A0A0U1QTI9
A	0	GLY	-	expression tag	UNP A0A0U1QTI9
A	1	ARG	-	expression tag	UNP A0A0U1QTI9
A	388	SER	-	expression tag	UNP A0A0U1QTI9
B	-13	MET	-	initiating methionine	UNP A0A0U1QTI9

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-12	ALA	-	expression tag	UNP A0A0U1QTI9
B	-11	SER	-	expression tag	UNP A0A0U1QTI9
B	-10	TRP	-	expression tag	UNP A0A0U1QTI9
B	-9	SER	-	expression tag	UNP A0A0U1QTI9
B	-8	HIS	-	expression tag	UNP A0A0U1QTI9
B	-7	PRO	-	expression tag	UNP A0A0U1QTI9
B	-6	GLN	-	expression tag	UNP A0A0U1QTI9
B	-5	PHE	-	expression tag	UNP A0A0U1QTI9
B	-4	GLU	-	expression tag	UNP A0A0U1QTI9
B	-3	LYS	-	expression tag	UNP A0A0U1QTI9
B	-2	ILE	-	expression tag	UNP A0A0U1QTI9
B	-1	GLU	-	expression tag	UNP A0A0U1QTI9
B	0	GLY	-	expression tag	UNP A0A0U1QTI9
B	1	ARG	-	expression tag	UNP A0A0U1QTI9
B	388	SER	-	expression tag	UNP A0A0U1QTI9
C	-13	MET	-	initiating methionine	UNP A0A0U1QTI9
C	-12	ALA	-	expression tag	UNP A0A0U1QTI9
C	-11	SER	-	expression tag	UNP A0A0U1QTI9
C	-10	TRP	-	expression tag	UNP A0A0U1QTI9
C	-9	SER	-	expression tag	UNP A0A0U1QTI9
C	-8	HIS	-	expression tag	UNP A0A0U1QTI9
C	-7	PRO	-	expression tag	UNP A0A0U1QTI9
C	-6	GLN	-	expression tag	UNP A0A0U1QTI9
C	-5	PHE	-	expression tag	UNP A0A0U1QTI9
C	-4	GLU	-	expression tag	UNP A0A0U1QTI9
C	-3	LYS	-	expression tag	UNP A0A0U1QTI9
C	-2	ILE	-	expression tag	UNP A0A0U1QTI9
C	-1	GLU	-	expression tag	UNP A0A0U1QTI9
C	0	GLY	-	expression tag	UNP A0A0U1QTI9
C	1	ARG	-	expression tag	UNP A0A0U1QTI9
C	388	SER	-	expression tag	UNP A0A0U1QTI9
D	-13	MET	-	initiating methionine	UNP A0A0U1QTI9
D	-12	ALA	-	expression tag	UNP A0A0U1QTI9
D	-11	SER	-	expression tag	UNP A0A0U1QTI9
D	-10	TRP	-	expression tag	UNP A0A0U1QTI9
D	-9	SER	-	expression tag	UNP A0A0U1QTI9
D	-8	HIS	-	expression tag	UNP A0A0U1QTI9
D	-7	PRO	-	expression tag	UNP A0A0U1QTI9
D	-6	GLN	-	expression tag	UNP A0A0U1QTI9
D	-5	PHE	-	expression tag	UNP A0A0U1QTI9
D	-4	GLU	-	expression tag	UNP A0A0U1QTI9
D	-3	LYS	-	expression tag	UNP A0A0U1QTI9

Continued on next page...

Continued from previous page...

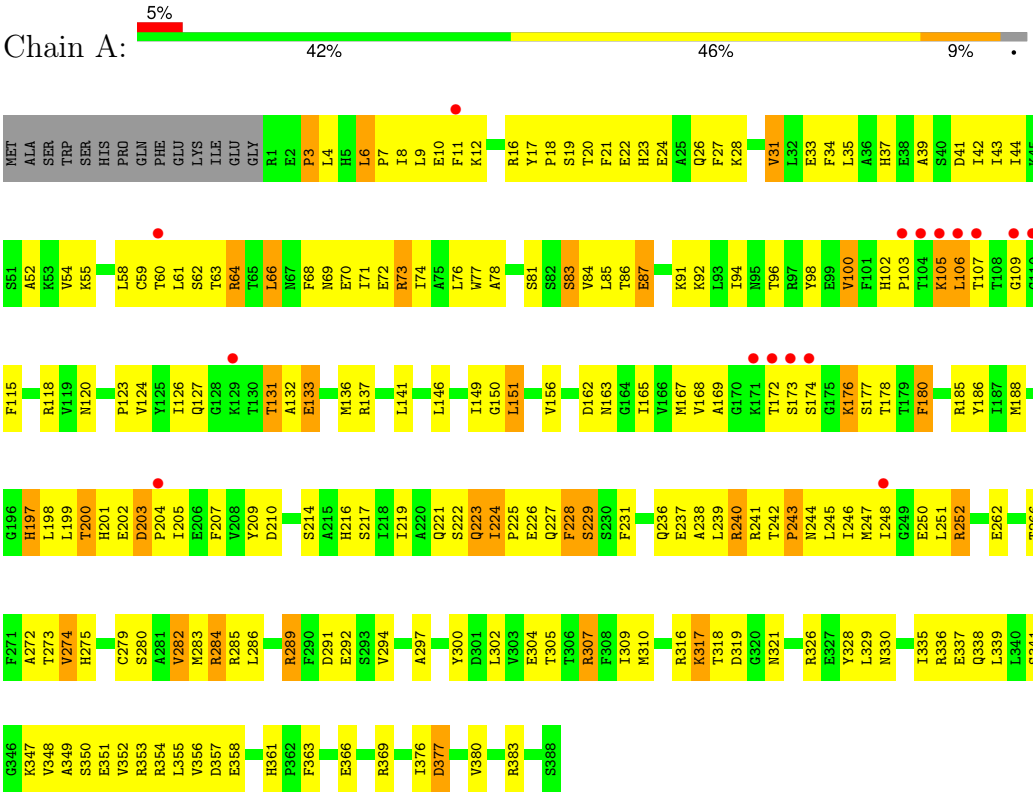
Chain	Residue	Modelled	Actual	Comment	Reference
D	-2	ILE	-	expression tag	UNP A0A0U1QTI9
D	-1	GLU	-	expression tag	UNP A0A0U1QTI9
D	0	GLY	-	expression tag	UNP A0A0U1QTI9
D	1	ARG	-	expression tag	UNP A0A0U1QTI9
D	388	SER	-	expression tag	UNP A0A0U1QTI9
E	-13	MET	-	initiating methionine	UNP A0A0U1QTI9
E	-12	ALA	-	expression tag	UNP A0A0U1QTI9
E	-11	SER	-	expression tag	UNP A0A0U1QTI9
E	-10	TRP	-	expression tag	UNP A0A0U1QTI9
E	-9	SER	-	expression tag	UNP A0A0U1QTI9
E	-8	HIS	-	expression tag	UNP A0A0U1QTI9
E	-7	PRO	-	expression tag	UNP A0A0U1QTI9
E	-6	GLN	-	expression tag	UNP A0A0U1QTI9
E	-5	PHE	-	expression tag	UNP A0A0U1QTI9
E	-4	GLU	-	expression tag	UNP A0A0U1QTI9
E	-3	LYS	-	expression tag	UNP A0A0U1QTI9
E	-2	ILE	-	expression tag	UNP A0A0U1QTI9
E	-1	GLU	-	expression tag	UNP A0A0U1QTI9
E	0	GLY	-	expression tag	UNP A0A0U1QTI9
E	1	ARG	-	expression tag	UNP A0A0U1QTI9
E	388	SER	-	expression tag	UNP A0A0U1QTI9
F	-13	MET	-	initiating methionine	UNP A0A0U1QTI9
F	-12	ALA	-	expression tag	UNP A0A0U1QTI9
F	-11	SER	-	expression tag	UNP A0A0U1QTI9
F	-10	TRP	-	expression tag	UNP A0A0U1QTI9
F	-9	SER	-	expression tag	UNP A0A0U1QTI9
F	-8	HIS	-	expression tag	UNP A0A0U1QTI9
F	-7	PRO	-	expression tag	UNP A0A0U1QTI9
F	-6	GLN	-	expression tag	UNP A0A0U1QTI9
F	-5	PHE	-	expression tag	UNP A0A0U1QTI9
F	-4	GLU	-	expression tag	UNP A0A0U1QTI9
F	-3	LYS	-	expression tag	UNP A0A0U1QTI9
F	-2	ILE	-	expression tag	UNP A0A0U1QTI9
F	-1	GLU	-	expression tag	UNP A0A0U1QTI9
F	0	GLY	-	expression tag	UNP A0A0U1QTI9
F	1	ARG	-	expression tag	UNP A0A0U1QTI9
F	388	SER	-	expression tag	UNP A0A0U1QTI9

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.

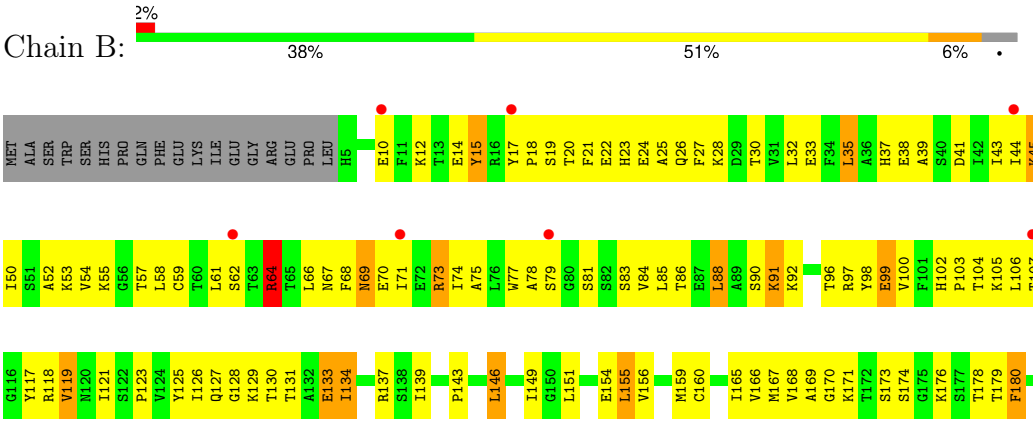
• Molecule 1: Type IV secretion system protein DotB

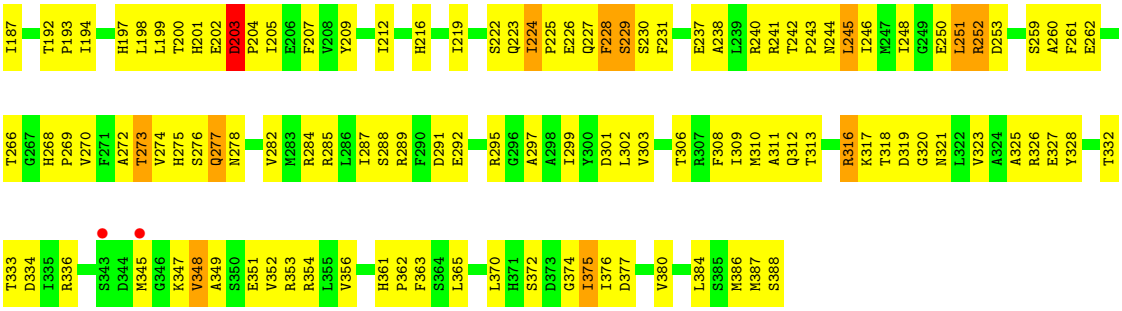
Chain A:



• Molecule 1: Type IV secretion system protein DotB

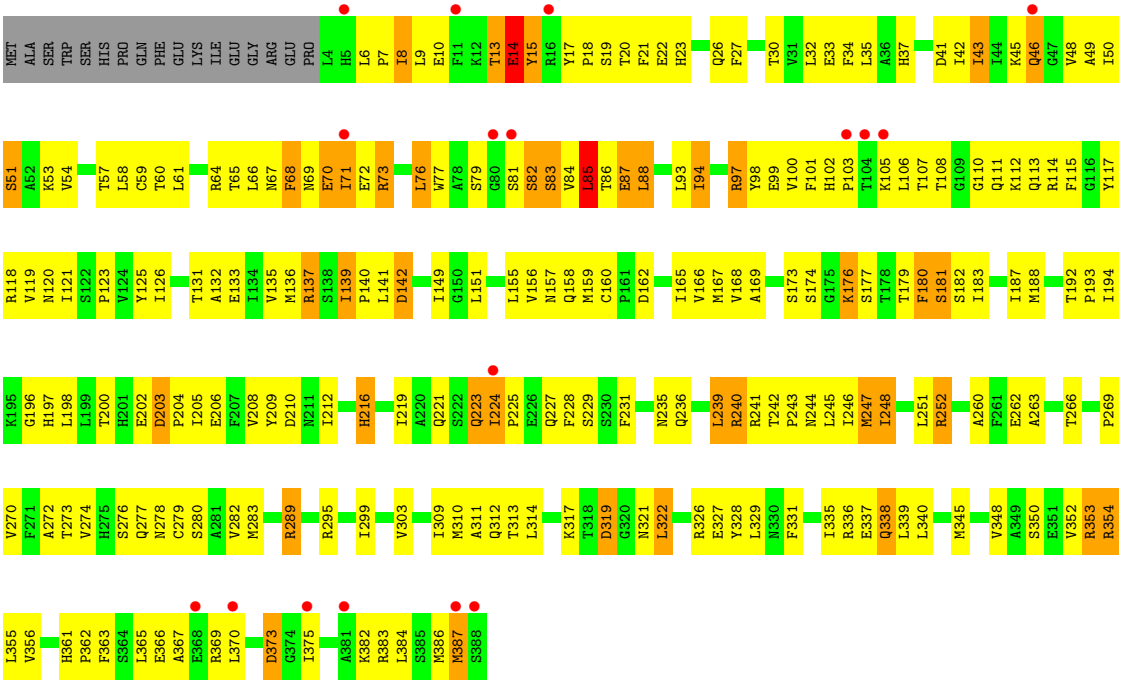
Chain B:





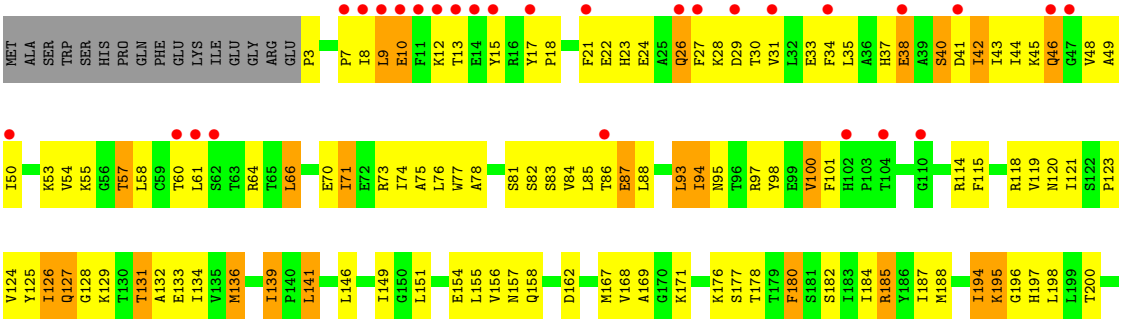
• Molecule 1: Type IV secretion system protein DotB

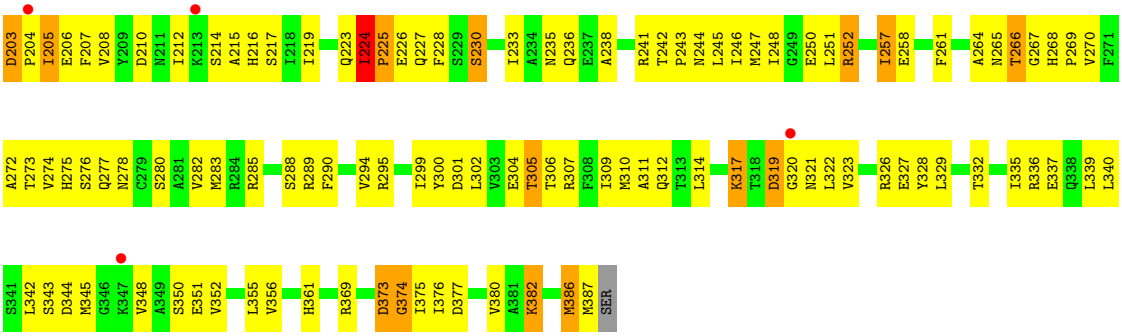
Chain C:



• Molecule 1: Type IV secretion system protein DotB

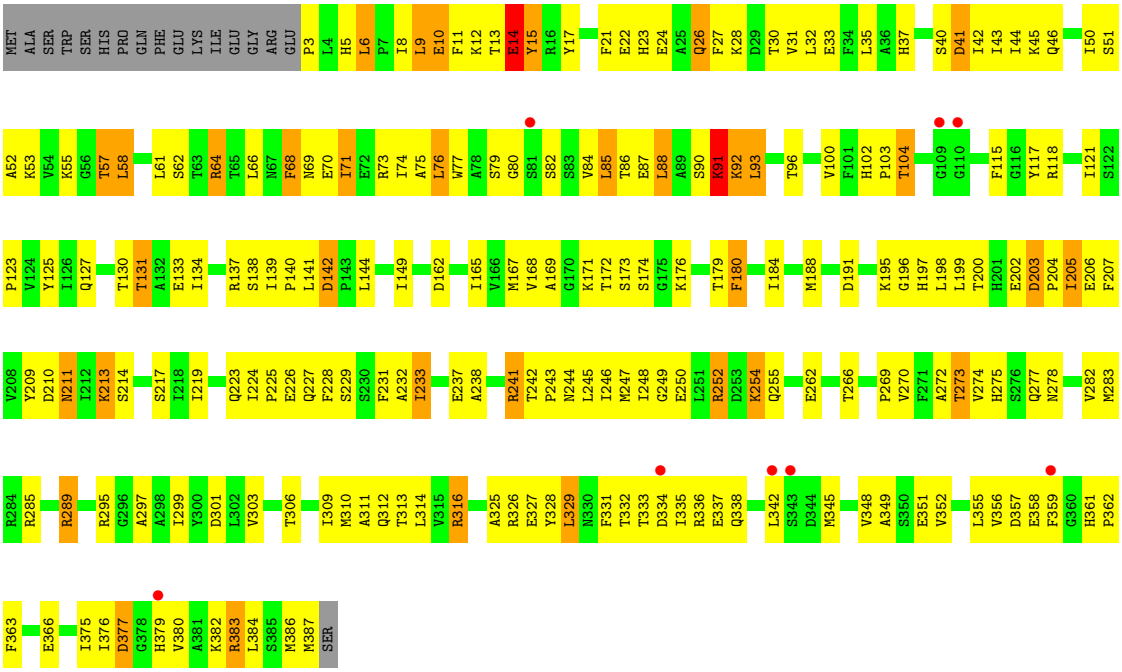
Chain D:





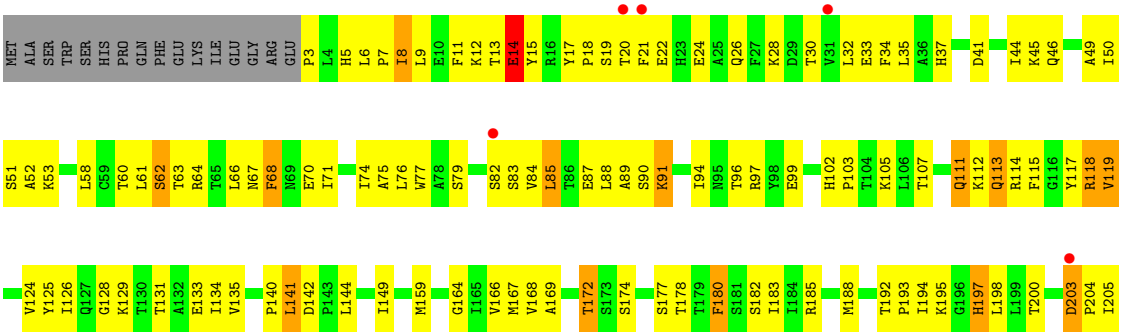
• Molecule 1: Type IV secretion system protein DotB

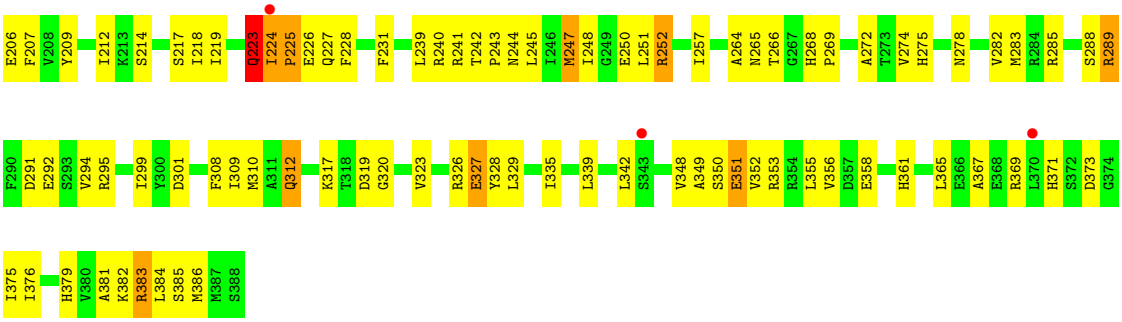
Chain E:



• Molecule 1: Type IV secretion system protein DotB

Chain F:




E206
F207
V208
Y209
I212
K213
S214
S217
I218
I219
Q223
I224
P225
E226
Q227
F228
F231
L239
R240
R241
T242
P243
N244
L245
I246
M247
I248
G249
E250
L251
R252
I257
A264
N265
T266
G267
H268
P269
A272
T273
V274
H275
N278
V282
M283
R284
R285
S288
R289
P290
D291
E292
S293
V294
R295
I299
Y300
D301
F308
I309
M310
A311
Q312
K317
T318
D319
G320
V323
R326
E327
Y328
L329
I335
L339
L342
S343
V348
A349
S350
E351
V352
R353
R354
L355
V356
D357
E358
H361
L365
E366
A367
E368
R369
L370
H371
S372
D373
G374
I375
I376
H379
V380
A381
K382
R383
L384
S385
M386
M387
S388

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	83.04Å 93.56Å 109.92Å 103.90° 101.98° 99.94°	Depositor
Resolution (Å)	40.56 – 2.75 45.71 – 2.75	Depositor EDS
% Data completeness (in resolution range)	97.8 (40.56-2.75) 97.9 (45.71-2.75)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.82 (at 2.77Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.255 , 0.296 0.253 , 0.290	Depositor DCC
R_{free} test set	3866 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	70.0	Xtriage
Anisotropy	0.314	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 40.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	18021	wwPDB-VP
Average B, all atoms (Å^2)	76.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.61% 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	0.64	0/3081	0.61	2/4169 (0.0%)
1	B	0.64	0/3040	0.59	0/4117
1	C	0.63	0/3064	0.60	1/4147 (0.0%)
1	D	0.65	0/3036	0.61	1/4112 (0.0%)
1	E	0.68	0/3050	0.62	3/4129 (0.1%)
1	F	0.70	0/3077	0.63	2/4163 (0.0%)
All	All	0.66	0/18348	0.61	9/24837 (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	B	0	2
1	C	0	3
1	D	0	5
1	F	0	2
All	All	0	13

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	3	PRO	N-CA-CB	7.53	112.33	103.30
1	F	3	PRO	N-CA-CB	6.92	111.61	103.30
1	A	210	ASP	CB-CG-OD2	6.89	124.50	118.30
1	D	3	PRO	N-CA-CB	6.86	111.53	103.30
1	E	289	ARG	NE-CZ-NH2	-6.06	117.27	120.30

There are no chirality outliers.

5 of 13 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	243	PRO	Peptide
1	B	203	ASP	Peptide
1	B	78	ALA	Peptide
1	C	105	LYS	Peptide
1	C	13	THR	Peptide

5.2 Too-close contacts (i)

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	3027	0	3036	282	1
1	B	2986	0	2977	371	0
1	C	3009	0	3010	364	0
1	D	2982	0	2961	385	1
1	E	2995	0	2976	328	0
1	F	3022	0	3035	282	0
All	All	18021	0	17995	1938	1

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 54.

The worst 5 of 1938 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:224:ILE:HG23	1:D:225:PRO:CD	1.45	1.46
1:D:46:GLN:NE2	1:D:131:THR:HG21	1.42	1.29
1:F:75:ALA:HB1	1:F:85:LEU:CD2	1.63	1.27
1:E:204:PRO:HB2	1:E:206:GLU:OE1	1.34	1.27
1:E:224:ILE:CD1	1:E:225:PRO:HD3	1.65	1.27

All (1) 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:A:318:THR:CG2	1:D:374:GLY:CA[1_656]	2.19	0.01

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	386/402 (96%)	364 (94%)	18 (5%)	4 (1%)	15	27
1	B	382/402 (95%)	363 (95%)	16 (4%)	3 (1%)	19	34
1	C	383/402 (95%)	358 (94%)	21 (6%)	4 (1%)	15	27
1	D	383/402 (95%)	366 (96%)	12 (3%)	5 (1%)	12	21
1	E	383/402 (95%)	358 (94%)	21 (6%)	4 (1%)	15	27
1	F	384/402 (96%)	356 (93%)	24 (6%)	4 (1%)	15	27
All	All	2301/2412 (95%)	2165 (94%)	112 (5%)	24 (1%)	15	27

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	3	PRO
1	A	203	ASP
1	B	203	ASP
1	B	229	SER
1	C	14	GLU

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	334/351 (95%)	289 (86%)	45 (14%)	4 5
1	B	329/351 (94%)	289 (88%)	40 (12%)	5 7
1	C	332/351 (95%)	287 (86%)	45 (14%)	3 5
1	D	325/351 (93%)	278 (86%)	47 (14%)	3 4
1	E	327/351 (93%)	285 (87%)	42 (13%)	4 6
1	F	335/351 (95%)	303 (90%)	32 (10%)	8 14
All	All	1982/2106 (94%)	1731 (87%)	251 (13%)	4 7

5 of 251 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	240	ARG
1	F	51	SER
1	D	100	VAL
1	F	5	HIS
1	F	178	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	277	GLN
1	F	379	HIS
1	F	26	GLN
1	F	157	ASN
1	B	275	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 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	388/402 (96%)	0.24	19 (4%)	29	36	48, 71, 107, 135	0
1	B	384/402 (95%)	0.20	10 (2%)	56	65	49, 76, 112, 128	0
1	C	385/402 (95%)	0.29	17 (4%)	34	41	50, 71, 100, 117	0
1	D	385/402 (95%)	0.52	32 (8%)	11	13	51, 79, 131, 156	0
1	E	385/402 (95%)	0.18	8 (2%)	63	72	47, 71, 99, 144	0
1	F	386/402 (96%)	0.22	8 (2%)	63	72	40, 66, 99, 120	0
All	All	2313/2412 (95%)	0.27	94 (4%)	37	44	40, 72, 113, 156	0

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	34	PHE	5.7
1	D	61	LEU	5.6
1	E	109	GLY	5.4
1	D	29	ASP	5.3
1	A	109	GLY	5.2

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.