wwPDB X-ray Structure Validation Summary Report (i)

May 29, 2024 – 10:22 am BST

PDB ID	:	2Y1V
Title	:	Full length structure of RrgB Pilus protein from Streptococcus pneumoniae
Authors	:	El-Mortaji, L.; Contreras-Martel, C.; Manzano, C.; Vernet, T.; Dessen, A.; DiGuilmi, A.M.
Deposited on	:	2010-12-10
Resolution	:	2.39 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 (1)) were used in the production of this report:

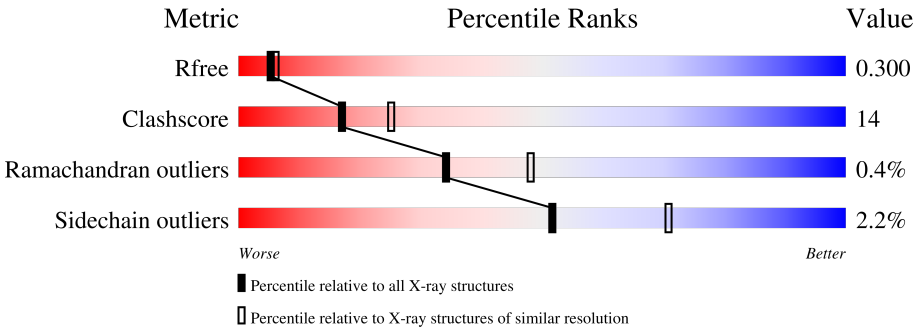
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	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 2.39 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)

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%

Mol	Chain	Length	Quality of chain
1	A	605	76% 22% .
1	B	605	77% 22% .
1	C	605	69% 30% .

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 14167 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 CELL WALL SURFACE ANCHOR FAMILY PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	604	Total 4576	C 2866	N 763	O 942	Se 5	0	0	0
1	B	604	Total 4576	C 2866	N 763	O 942	Se 5	0	0	0
1	C	604	Total 4576	C 2866	N 763	O 942	Se 5	0	0	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	29	GLY	-	expression tag	UNP Q97SC2
B	29	GLY	-	expression tag	UNP Q97SC2
C	29	GLY	-	expression tag	UNP Q97SC2

- Molecule 2 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total 3	Ni 3	0	0
2	B	3	Total 3	Ni 3	0	0
2	C	3	Total 3	Ni 3	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	169	Total 169	O 169	0	0
3	B	141	Total 141	O 141	0	0

Continued on next page...

Continued from previous page...

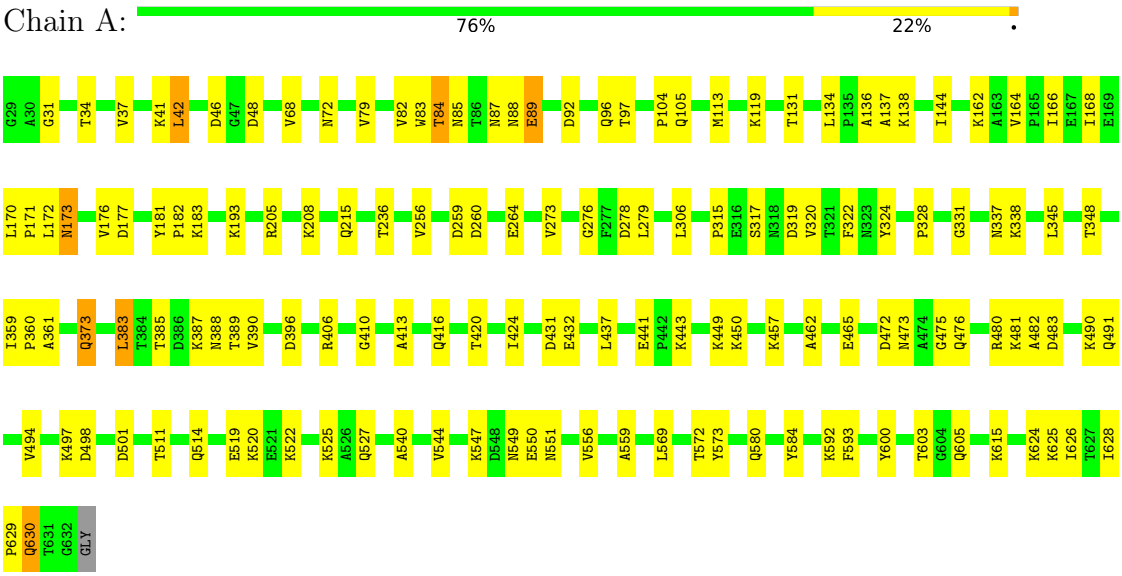
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	120	Total	O	0	0
			120	120		

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. 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. 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.

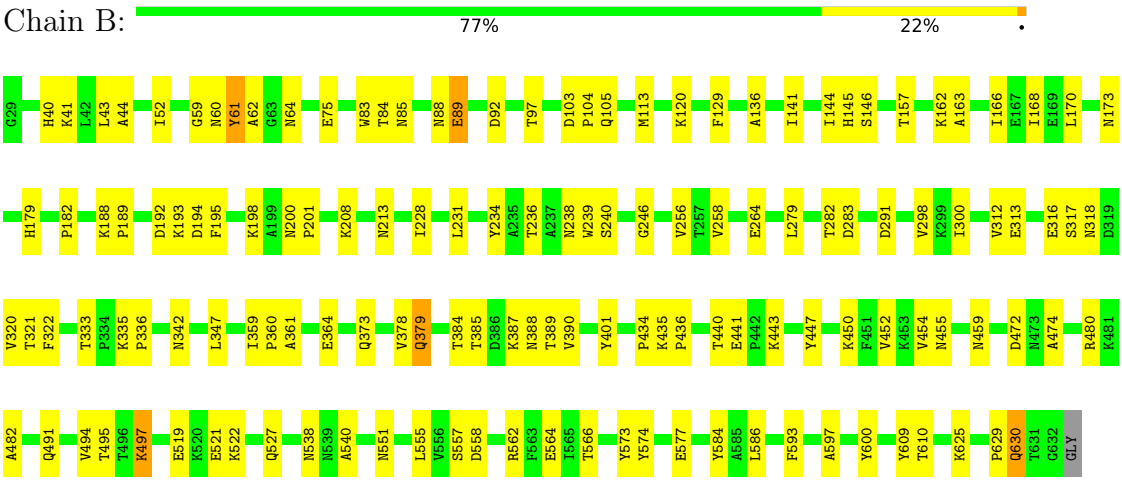
• Molecule 1: CELL WALL SURFACE ANCHOR FAMILY PROTEIN

Chain A:



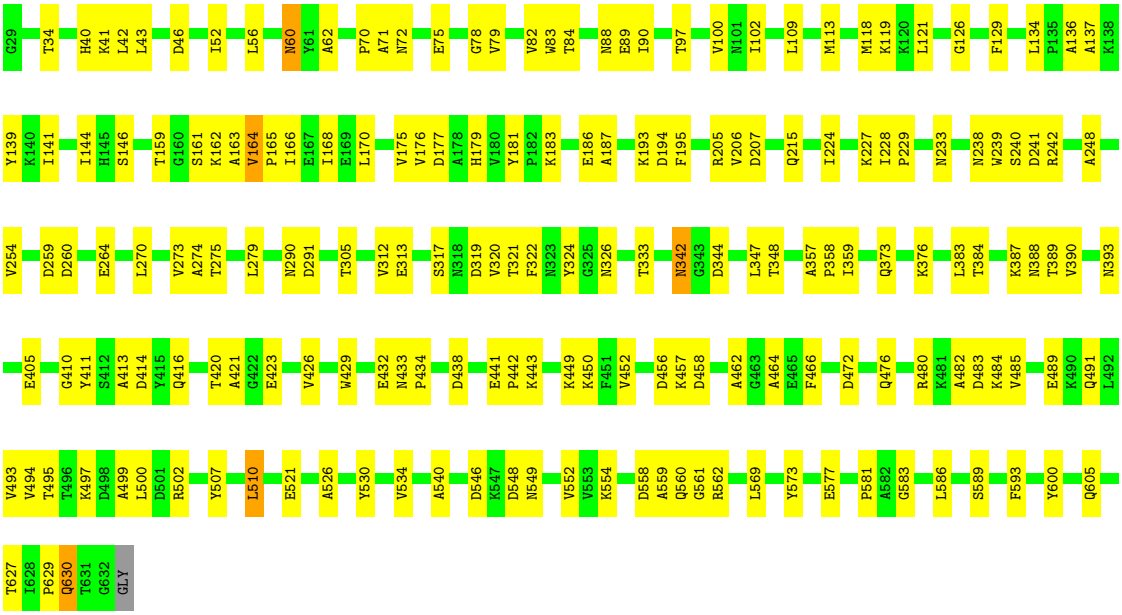
• Molecule 1: CELL WALL SURFACE ANCHOR FAMILY PROTEIN

Chain B:



• Molecule 1: CELL WALL SURFACE ANCHOR FAMILY PROTEIN

Chain C:


G29 T34 H40 K41 L42 L43 D46 I52 L56 N60 Y61 A62 P70 A71 N72 E75 G78 V79 V82 W83 T84 N88 E89 I90 T97 V100 N101 I102 L109 M113 M118 K119 K120 L121 G126 F129 L134 P135 A136 A137 K138
Y139 K140 I141 I144 H145 S146 T159 G160 S161 K162 A163 V164 P165 I166 E167 I168 E169 L170 V175 V176 D177 A178 H179 V180 Y181 P182 K183 E186 A187 K193 D194 F195 R205 V206 D207 Q215 I224 K227 I228 P229 N233 N238 W239 S240 D241 R242 A248
V254 D259 D260 E264 L270 V273 A274 T275 L279 N290 D291 T305 V312 E313 S317 N318 D319 V320 T321 F322 N323 Y324 G325 N326 T333 N342 G343 D344 L347 T348 A357 P358 I359 Q373 K376 L383 T384 K387 N388 T389 V390 N393
E405 G410 Y411 S412 A413 D414 Y415 Q416 T420 A421 G422 E423 V426 W429 E432 N433 P434 D438 E441 P442 K443 K449 K450 F451 V452 D456 K457 D458 A462 G463 A464 E465 F466 D472 Q476 R480 K481 A482 D483 K484 V485 E489 K490 Q491 L492
V493 V494 T495 T496 K497 D498 A499 L500 D501 R502 Y507 L510 E521 A526 Y530 V534 A540 D546 K547 D548 N549 V552 V553 K554 D558 A559 Q560 G561 R562 L569 Y573 E577 P581 A582 G583 L586 S589 F593 Y600 Q605
T627 I628 P629 Q630 T631 G632 GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	74.62Å 74.62Å 340.53Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	113.53 – 2.39 19.85 – 2.39	Depositor EDS
% Data completeness (in resolution range)	98.7 (113.53-2.39) 93.7 (19.85-2.39)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	2.25 (at 2.38Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.187 , 0.234 0.239 , 0.300	Depositor DCC
R_{free} test set	8066 reflections (10.32%)	wwPDB-VP
Wilson B-factor (Å^2)	41.6	Xtriage
Anisotropy	0.395	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 38.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.55$, $< L^2 > = 0.39$	Xtriage
Estimated twinning fraction	0.000 for -h,-k,l 0.097 for h,-h-k,-l 0.002 for -k,-h,-l	Xtriage
Reported twinning fraction	0.592 for H, K, L 0.408 for K, H, -L	Depositor
Outliers	0 of 78179 reflections	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	14167	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 3.15% 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

Bond lengths and bond angles in the following residue types are not validated in this section: NI

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	0.47	0/4646	0.62	0/6315
1	B	0.44	0/4646	0.57	0/6315
1	C	0.40	0/4646	0.57	0/6315
All	All	0.44	0/13938	0.59	0/18945

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	4576	0	4509	108	0
1	B	4576	0	4509	120	0
1	C	4576	0	4510	153	0
2	A	3	0	0	0	0
2	B	3	0	0	0	0
2	C	3	0	0	0	0
3	A	169	0	0	12	0
3	B	141	0	0	15	0
3	C	120	0	0	33	0
All	All	14167	0	13528	381	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 14.

The worst 5 of 381 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)
3:A:2008:HOH:O	1:B:387:LYS:HB2	1.45	1.15
1:B:84:THR:HG21	1:B:88:ASN:HA	1.24	1.15
1:B:256:VAL:HG21	1:B:279:LEU:HD11	1.16	1.13
1:A:96:GLN:HB3	3:A:2012:HOH:O	1.48	1.12
1:C:168:ILE:HB	3:C:2026:HOH:O	1.47	1.11

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	602/605 (100%)	566 (94%)	34 (6%)	2 (0%)	41	55
1	B	602/605 (100%)	573 (95%)	25 (4%)	4 (1%)	22	32
1	C	602/605 (100%)	569 (94%)	31 (5%)	2 (0%)	41	55
All	All	1806/1815 (100%)	1708 (95%)	90 (5%)	8 (0%)	34	48

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	475	GLY
1	B	61	TYR
1	A	173	ASN
1	B	173	ASN
1	B	379	GLN

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	489/484 (101%)	475 (97%)	14 (3%)	42	62
1	B	489/484 (101%)	481 (98%)	8 (2%)	62	79
1	C	489/484 (101%)	478 (98%)	11 (2%)	52	71
All	All	1467/1452 (101%)	1434 (98%)	33 (2%)	52	71

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	483	ASP
1	C	510	LEU
1	C	630	GLN
1	A	497	LYS
1	A	490	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	459	ASN
1	C	476	GLN
1	C	630	GLN
1	B	64	ASN
1	B	60	ASN

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)

Of 9 ligands modelled in this entry, 9 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.