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 ⓘ

Jun 18, 2024 – 07:19 PM EDT

PDB ID	:	4CBV
Title	:	X-ray structure of full-length ComE from Streptococcus pneumoniae.
Authors	:	Boudes, M.; Durand, D.; Graille, M.; van Tilbeurgh, H.; Quevillon-Cheruel, S.
Deposited on	:	2013-10-16
Resolution	:	3.39 Å(reported)

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

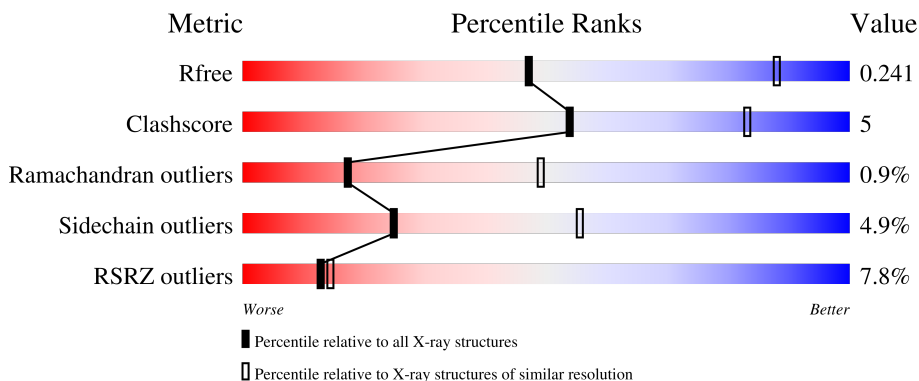|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1   Overall quality at a glance   ⓘ

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

The reported resolution of this entry is 3.39 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.
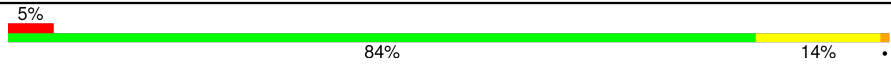
Metric	Percentile Ranks	Value
Rfree		0.241
Clashscore		5
Ramachandran outliers		0.9%
Sidechain outliers		4.9%
RSRZ outliers		7.8%

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}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	256	76% 20% • •
1	B	256	75% 22% • •
1	C	256	6% 80% 16% • •
1	D	256	15% 84% 13% • •
1	E	256	19% 81% 15% • •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	256	5% 84% 14%

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 12732 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 COME.

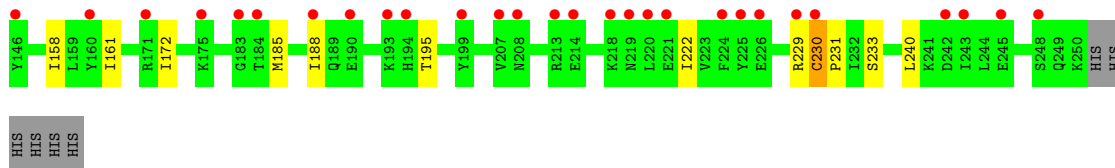| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	251	Total	C	N	O	S	Se	0	0	0
			2123	1368	355	395	1	4			
1	B	251	Total	C	N	O	S	Se	0	0	0
			2123	1368	355	395	1	4			
1	C	249	Total	C	N	O	S	Se	0	0	0
			2104	1356	350	393	1	4			
1	D	250	Total	C	N	O	S	Se	0	0	0
			2113	1362	352	394	1	4			
1	E	248	Total	C	N	O	S	Se	0	0	0
			2096	1351	349	392	1	3			
1	F	256	Total	C	N	O	S	Se	0	0	0
			2173	1398	370	400	1	4			

There are 42 discrepancies between the modelled and reference sequences:

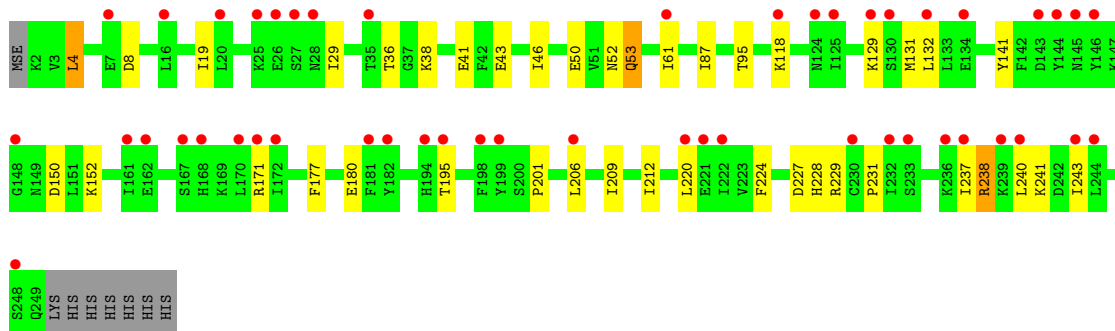| Chain	Residue	Modelled	Actual	Comment	Reference
A	251	HIS	-	expression tag	UNP Q79CK7
A	252	HIS	-	expression tag	UNP Q79CK7
A	253	HIS	-	expression tag	UNP Q79CK7
A	254	HIS	-	expression tag	UNP Q79CK7
A	255	HIS	-	expression tag	UNP Q79CK7
A	256	HIS	-	expression tag	UNP Q79CK7
A	58	ALA	ASP	engineered mutation	UNP Q79CK7
B	251	HIS	-	expression tag	UNP Q79CK7
B	252	HIS	-	expression tag	UNP Q79CK7
B	253	HIS	-	expression tag	UNP Q79CK7
B	254	HIS	-	expression tag	UNP Q79CK7
B	255	HIS	-	expression tag	UNP Q79CK7
B	256	HIS	-	expression tag	UNP Q79CK7
B	58	ALA	ASP	engineered mutation	UNP Q79CK7
C	251	HIS	-	expression tag	UNP Q79CK7
C	252	HIS	-	expression tag	UNP Q79CK7
C	253	HIS	-	expression tag	UNP Q79CK7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	254	HIS	-	expression tag	UNP Q79CK7
C	255	HIS	-	expression tag	UNP Q79CK7
C	256	HIS	-	expression tag	UNP Q79CK7
C	58	ALA	ASP	engineered mutation	UNP Q79CK7
D	251	HIS	-	expression tag	UNP Q79CK7
D	252	HIS	-	expression tag	UNP Q79CK7
D	253	HIS	-	expression tag	UNP Q79CK7
D	254	HIS	-	expression tag	UNP Q79CK7
D	255	HIS	-	expression tag	UNP Q79CK7
D	256	HIS	-	expression tag	UNP Q79CK7
D	58	ALA	ASP	engineered mutation	UNP Q79CK7
E	251	HIS	-	expression tag	UNP Q79CK7
E	252	HIS	-	expression tag	UNP Q79CK7
E	253	HIS	-	expression tag	UNP Q79CK7
E	254	HIS	-	expression tag	UNP Q79CK7
E	255	HIS	-	expression tag	UNP Q79CK7
E	256	HIS	-	expression tag	UNP Q79CK7
E	58	ALA	ASP	engineered mutation	UNP Q79CK7
F	251	HIS	-	expression tag	UNP Q79CK7
F	252	HIS	-	expression tag	UNP Q79CK7
F	253	HIS	-	expression tag	UNP Q79CK7
F	254	HIS	-	expression tag	UNP Q79CK7
F	255	HIS	-	expression tag	UNP Q79CK7
F	256	HIS	-	expression tag	UNP Q79CK7
F	58	ALA	ASP	engineered mutation	UNP Q79CK7

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: COME

Chain A:

- Molecule 1: COME

Chain B:

- Molecule 1: COME

Chain C:

- Molecule 1: COME

Chain D:

Y146 I158 L159 Y160 I161 R171 I172 K175 G183 T184 M185 I188 Q189 E190 K193 H194 T195 Y199 V207 N208 R213 E214 K218 N219 L220 E221 I222 V223 F224 Y225 E226 R229 C230 P231 I232 S233 L240 K241 D242 I243 L244 E245 S248 Q249 K250 HIS HIS

HIS HIS HIS HIS

● **Molecule 1: COME**

Chain E:

19%

81%   15%

MSE K2 V3 L4 E7 D8 L16 I19 L20 K25 E26 S27 N28 I29 T35 T36 G37 K38 E41 F42 E43 I46 E50 V51 N52 Q53 I61 I87 T95 K118 N124 I125 K129 S130 M131 L132 L133 E134 Y141 F142 D143 Y144 N145 Y146 K147

G148 N149 D150 L151 K152 I161 E162 S167 H168 K169 L170 R171 I172 F177 E180 F181 Y182 H194 T195 F198 Y199 S200 P201 L206 I209 I212 L220 E221 I222 V223 F224 D227 H228 R229 C230 P231 I232 S233 K236 I237 R238 K239 L240 K241 D242 I243 L244

S248 Q249 LYS HIS HIS HIS HIS HIS

● **Molecule 1: COME**

Chain F:

5%

84%   14%

M1 D8 V9 I10 E11 R15 L16 L20 I31 K38 V39 R40 E47 N48 D49 I59 D60 I61 G68 F69 L74 I75 R76 N79 A82 T88 T97 V102 K110 D111 I112 M116 I125 N145 D150 Y155 I158 I161

H168 I172 G183 H194 T195 L206 E214 I215 I222 Y225 R229 C230 P231 I232 S233 R234 L235 R238 K241 H251 H252 H256

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	88.89Å 135.00Å 461.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.95 – 3.39 44.45 – 3.39	Depositor EDS
% Data completeness (in resolution range)	96.3 (37.95-3.39) 96.3 (44.45-3.39)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.74 (at 3.40Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.196 , 0.223 0.211 , 0.241	Depositor DCC
R_{free} test set	1898 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	79.7	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 100.7	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	0.91	EDS
Total number of atoms	12732	wwPDB-VP
Average B, all atoms (Å^2)	114.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.90% 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 (i)

5.1 Standard geometry (i)

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	0.42	0/2163	0.65	0/2903
1	B	0.40	0/2163	0.64	0/2903
1	C	0.39	0/2143	0.59	0/2877
1	D	0.40	0/2152	0.57	0/2888
1	E	0.41	0/2135	0.57	0/2867
1	F	0.39	0/2218	0.58	0/2978
All	All	0.40	0/12974	0.60	0/17416

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 (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	2123	0	2132	34	0
1	B	2123	0	2132	25	0
1	C	2104	0	2112	21	0
1	D	2113	0	2125	19	0
1	E	2096	0	2100	18	0
1	F	2173	0	2167	19	1
All	All	12732	0	12768	134	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 5.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

The worst 5 of 134 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:A:49:ASP:HB2	1:A:53:GLN:HG3	1.51	0.92
1:A:224:PHE:HE1	1:A:231:PRO:HD2	1.43	0.83
1:A:93:PHE:HA	1:A:96:LEU:HD23	1.67	0.76
1:C:224:PHE:HB2	1:C:228:HIS:HB2	1.72	0.71
1:C:36:THR:HG21	1:C:41:GLU:HB3	1.75	0.68

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:F:116:MSE:CE	1:F:116:MSE:CE[3_655]	0.63	1.57

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	249/256 (97%)	236 (95%)	11 (4%)	2 (1%)	19	51
1	B	249/256 (97%)	235 (94%)	12 (5%)	2 (1%)	19	51
1	C	247/256 (96%)	231 (94%)	12 (5%)	4 (2%)	9	34
1	D	248/256 (97%)	239 (96%)	8 (3%)	1 (0%)	34	67
1	E	246/256 (96%)	232 (94%)	12 (5%)	2 (1%)	19	51
1	F	254/256 (99%)	242 (95%)	10 (4%)	2 (1%)	19	51
All	All	1493/1536 (97%)	1415 (95%)	65 (4%)	13 (1%)	17	49

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	51	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	227	ASP
1	C	134	GLU
1	A	229	ARG
1	F	49	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	237/238 (100%)	225 (95%)	12 (5%)	24	54
1	B	237/238 (100%)	221 (93%)	16 (7%)	16	45
1	C	235/238 (99%)	224 (95%)	11 (5%)	26	57
1	D	236/238 (99%)	228 (97%)	8 (3%)	37	65
1	E	234/238 (98%)	221 (94%)	13 (6%)	21	51
1	F	242/238 (102%)	233 (96%)	9 (4%)	34	62
All	All	1421/1428 (100%)	1352 (95%)	69 (5%)	25	55

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	229	ARG
1	E	241	LYS
1	F	206	LEU
1	B	233	SER
1	B	229	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	48	ASN
1	F	145	ASN
1	B	228	HIS
1	D	145	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	176	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, 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	247/256 (96%)	-0.05	0	100	100	35, 60, 104, 150	0
1	B	247/256 (96%)	0.09	1 (0%)	92	92	35, 67, 109, 184	0
1	C	245/256 (95%)	0.41	15 (6%)	21	22	74, 116, 163, 206	0
1	D	246/256 (96%)	0.87	39 (15%)	1	2	86, 140, 185, 212	0
1	E	245/256 (95%)	1.03	48 (19%)	1	1	119, 158, 192, 220	0
1	F	252/256 (98%)	0.47	12 (4%)	30	31	94, 125, 159, 171	0
All	All	1482/1536 (96%)	0.47	115 (7%)	13	14	35, 118, 179, 220	0

The worst 5 of 115 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	221	GLU	7.9
1	D	225	TYR	5.6
1	D	144	TYR	5.3
1	E	199	TYR	5.1
1	E	168	HIS	4.9

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.

wwPDB X-ray Structure Validation Summary Report

6.5 Other polymers ⓘ

There are no such residues in this entry.