



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 01:49 PM GMT

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)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

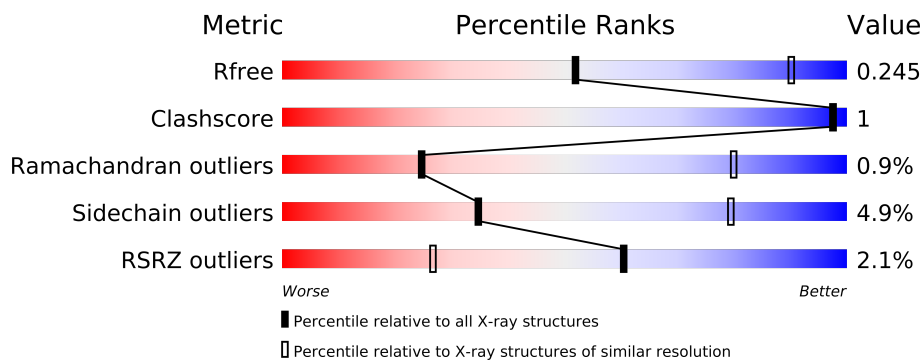
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	66092	1017 (3.52-3.28)
Clashscore	79885	1214 (3.50-3.30)
Ramachandran outliers	78287	1177 (3.50-3.30)
Sidechain outliers	78261	1177 (3.50-3.30)
RSRZ outliers	66119	1017 (3.52-3.28)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	256	
1	B	256	
1	C	256	
1	D	256	
1	E	256	
1	F	256	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 12732 atoms, of which 0 are hydrogen and 0 are deuterium.

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...*

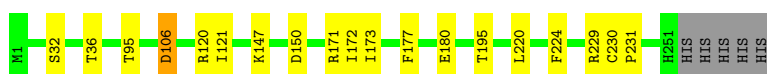
Chain	Residue	Modelled	Actual	Comment	Reference
C	254	HIS	-	EXPRESSION TAG	UNP Q79CK7
C	255	HIS	-	EXPRESSION TAG	UNP Q79CK7
C	256	HIS	-	EXPRESSION TAG	UNP Q79CK7
C	58	ALA	ASP	ENGINEERED MUTATION	UNP Q79CK7
D	251	HIS	-	EXPRESSION TAG	UNP Q79CK7
D	252	HIS	-	EXPRESSION TAG	UNP Q79CK7
D	253	HIS	-	EXPRESSION 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 and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: 



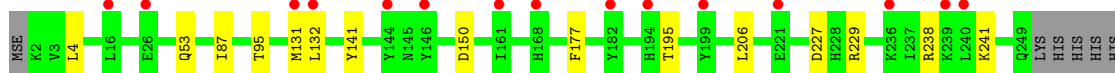
- Molecule 1: COME

Chain D: 



- Molecule 1: COME

Chain E: 



HIS  
HIS

- Molecule 1: COME

Chain F: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	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
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.74 (at 3.40Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, $R_{free}$	0.196 , 0.223 0.212 , 0.245	Depositor DCC
$R_{free}$ test set	1897 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	79.7	Xtriage
Anisotropy	0.336	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 75.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtriage
Outliers	0 of 37652 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	12732	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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	0.42	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 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2123	0	65	3	0
1	B	2123	0	65	3	0
1	C	2104	0	65	0	0
1	D	2113	0	65	3	0
1	E	2096	0	65	0	0
1	F	2173	0	65	1	1
All	All	12732	0	390	10	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 1.

The worst 5 of 10 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:D:138:VAL:HG23	1:D:139:VAL:HG22	1.82	0.61
1:D:134:GLU:HA	1:D:138:VAL:HG21	1.87	0.56
1:D:139:VAL:HG23	1:D:141:TYR:CE1	2.46	0.50
1:B:226:GLU:O	1:B:228:HIS:N	2.50	0.45
1:A:224:PHE:CE1	1:A:231:PRO:CD	3.00	0.45

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	Distance(Å)	Clash(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%)	27	81
1	B	249/256 (97%)	235 (94%)	12 (5%)	2 (1%)	27	81
1	C	247/256 (96%)	231 (94%)	12 (5%)	4 (2%)	14	67
1	D	248/256 (97%)	239 (96%)	8 (3%)	1 (0%)	43	90
1	E	246/256 (96%)	232 (94%)	12 (5%)	2 (1%)	27	81
1	F	254/256 (99%)	242 (95%)	10 (4%)	2 (1%)	27	81
All	All	1493/1536 (97%)	1415 (95%)	65 (4%)	13 (1%)	25	80

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	51	VAL
1	B	227	ASP
1	C	134	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
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%)	33	78
1	B	237/238 (100%)	221 (93%)	16 (7%)	22	69
1	C	235/238 (99%)	224 (95%)	11 (5%)	36	81
1	D	236/238 (99%)	228 (97%)	8 (3%)	49	88
1	E	234/238 (98%)	221 (94%)	13 (6%)	30	75
1	F	242/238 (102%)	233 (96%)	9 (4%)	45	85
All	All	1421/1428 (100%)	1352 (95%)	69 (5%)	35	79

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

Mol	Chain	Res	Type
1	C	87	ILE
1	C	230	CYS
1	F	125	ILE
1	C	95	THR
1	C	177	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates 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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	251/256 (98%)	-0.07	0 100 100	35, 60, 104, 150	0
1	B	251/256 (98%)	0.02	1 (0%) 90 68	35, 68, 109, 184	0
1	C	249/256 (97%)	0.22	3 (1%) 75 39	74, 118, 163, 206	0
1	D	250/256 (97%)	0.49	9 (3%) 41 16	86, 140, 185, 212	0
1	E	248/256 (96%)	0.59	15 (6%) 21 8	119, 158, 192, 220	0
1	F	256/256 (100%)	0.24	3 (1%) 75 39	94, 125, 159, 171	0
All	All	1505/1536 (97%)	0.25	31 (2%) 60 27	35, 118, 179, 220	0

The worst 5 of 31 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	221	GLU	4.6
1	E	131	MSE	4.4
1	D	225	TYR	4.0
1	C	1	MSE	3.6
1	D	144	TYR	3.6

### 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 carbohydrates 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.