SUPPLEMENTARY INFORMATIONS

Structural insights into the dimerization and the DNA binding of the response regulator ComE from *Streptococcus pneumoniae*

Marion Boudes[1], Dyana Sanchez[1], Marc Graille[1,2], Herman van Tilbeurgh[1], Dominique Durand[1,2] and Sophie Quevillon-Cheruel[1,2,*]

[1] Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université de Paris-Sud XI, UMR8619, Bât 430, 91405 Orsay, France.

[2] Centre National de la Recherche Scientifique, Orsay, 91405, France

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

* To whom correspondence should be addressed. Tel: +33 1 69 15 31 56; Fax: +33 1 69 85 37 15; Email: sophie.quevillon-cheruel@u-psud.fr

Correspondence may also be addressed to: dominique.durand@u-psud.fr

Present addresses: Marion Boudes, Department of Biochemistry and Molecular Biology, School of Biomedical Sciences, Monash University, Clayton, VIC 3800, Australia; Marc Graille, Laboratoire de Biochimie, Ecole Polytechnique, CNRS UMR7654, 91128 Palaiseau, France.

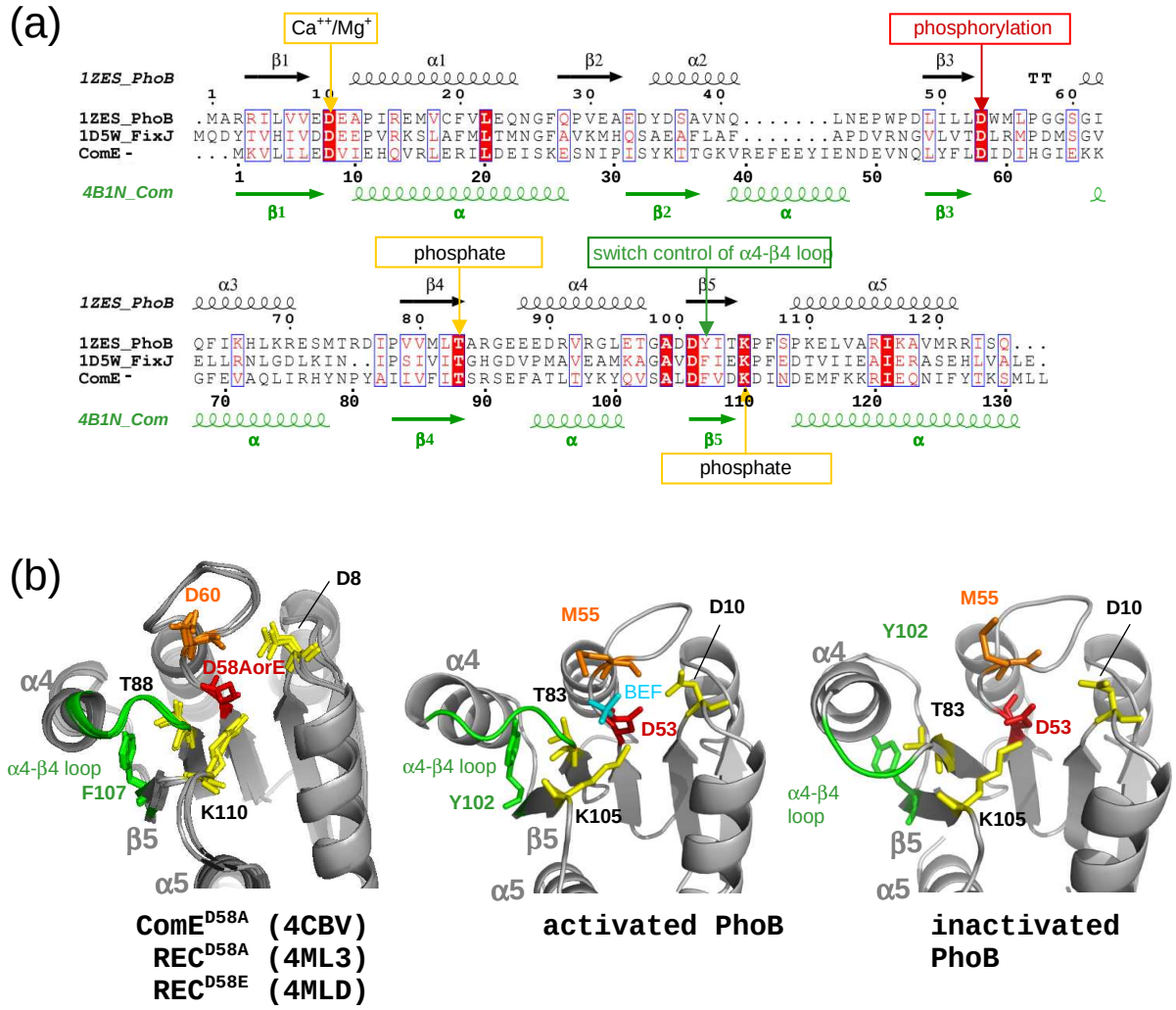

Figure S1. Structural comparison of ^{ComE}REC domain with activated and inactivated homologues **(a)** Structure-based sequence alignment of ComE REC domain. The multialignment was generated by Clustalw2 (1). The figure was generated using ESPRIPT (2). The secondary structure elements of ComE REC are in green on the bottom of the multialignment. That corresponding to the REC domain of PhoB (PDB ID: 1ZES) is represented on top. The functional residues are pointed by arrows. **(b)** Structural comparison of the ComE REC domain with activated (PDB ID: 1ZES) and non-activated (PDB ID: 1B00) states of PhoB. The phosphorylable Asp are in red, the phosphate and cation binding residues are in yellow, the α4-β4 loop and Phe or Tyr switch control are in green. The residues involved in the stabilization of the Mg^{2+} are in orange.

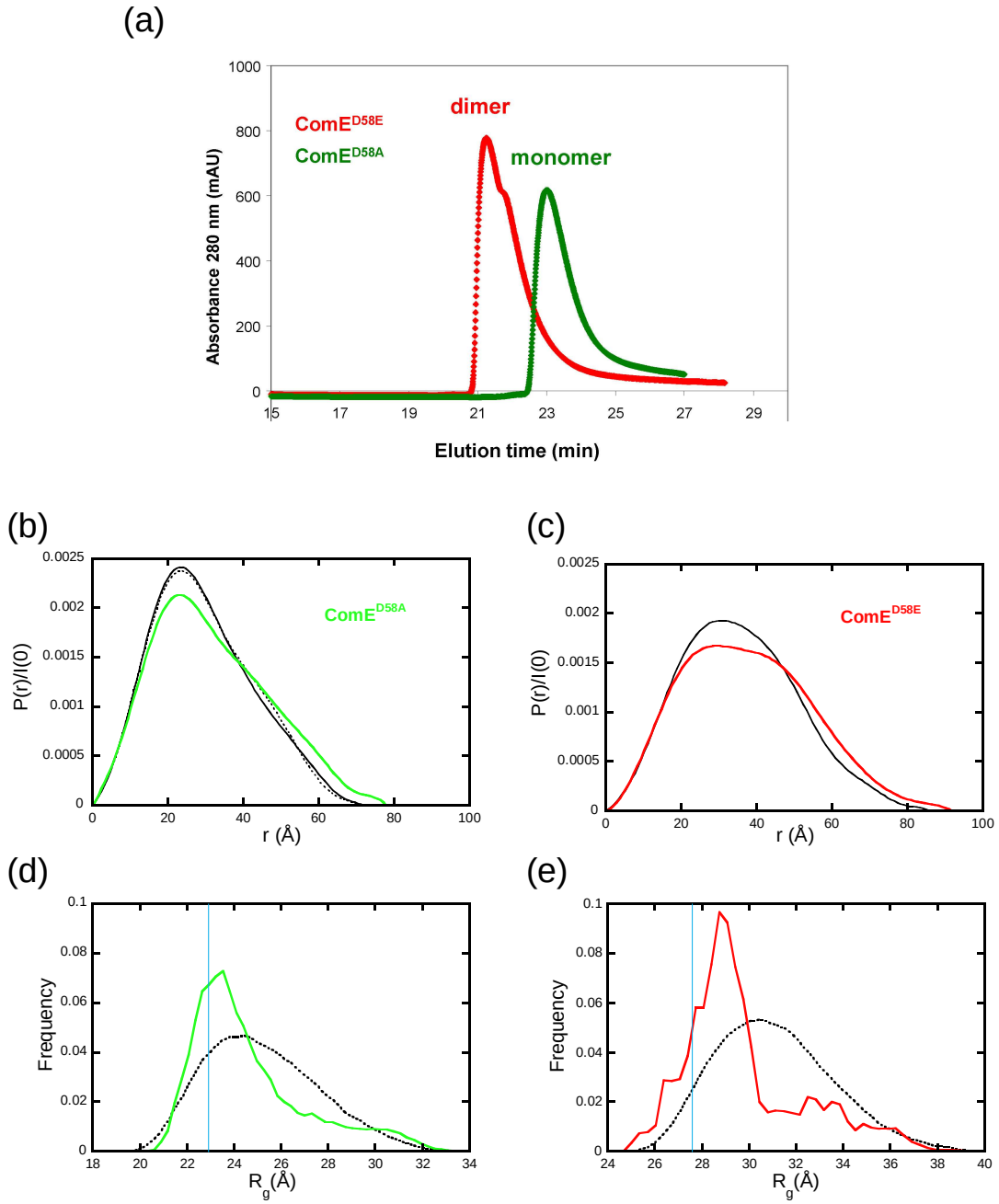

Figure S2. (a) Elution profiles on exclusion chromatography of ComE^{D58A} (green) and ComE^{D58E} (red) on the gel filtration chromatography coupled upstream to the scattering measurements of the beam-line SWING of synchrotron SOLEIL. **(b)** Comparison of the distance distribution function P(r) for ComE^{D58A} (green) to the calculated P(r) using the crystal subunit A (continuous black line) and B (dotted line). **(c)** Distance distribution function P(r) for ComE^{D58E} (red line) compared to the calculated one using the crystal dimer (black line). **(d)** Distribution of radius of gyration values of ComE^{D58A} conformations from the ensemble selected by the program EOM so that the average scattering curve fits the experimental data; corresponding distribution for the complete pool of 10000 randomly generated conformations (dotted line). See section Material and Methods for details. The vertical blue bar corresponds to the value of R_g for the crystal subunits (R_g=22.9Å). **(e)** Distribution of radius of gyration values of ComE^{D58A} conformations.

Table S1. Data Collection and Structure Refinement Statistics

Data collection	ComE^{D58A} SeMet	REC^{D58A}	REC^{D58E}
Space group	$C222_1$	$P6_3$	$P6_3$
Unit cell parameters	a=88.9Å b=135.0Å	a=91.5Å b=91.5Å	a=90.9Å b=90.9Å
	c=461.4Å	c=134.9Å	c=140.3Å
	α=ß=γ=90°	α=ß=90° γ=120°	α=ß=90° γ=120°
Redundancy	3.77 (3.30)	9.22(9.03)	19.5(18.5)
Resolution range (Å)†	45 - 3.4 (3.6 - 3.4) †	45.7-3.2(3.3-3.2) †	45.3-2.9(3.1-2.9) †
Completeness (%)†	95.6 (87.6) †	99.7 (98.1)	99.8(99.1)
I/σ(I)	10.9 (2.8)	15.0 (2.9)	21.74(3.55)
R_{sym} (%)§	10.8 (50.6)	14.4(97.5)	16.7(94.8)
Refinement			
Resolution range (Å)	45 - 3.4 (3.5 - 3.4)	45.7-3.2(3.5-3.2)	45.4-2.9 (3.1-2.9)
R/R_{free}	19.6/22.3 (22.1/24.5)	27,2/32.2(37.30/38.16)	19.5/23.9(29.5/35.8)
Geometry statistics			
r.m.s. deviation bonds (Å)	0.008	0.004	0.003
r.m.s. deviation angles (°)	0.970	0.878	0.699
Average B-factor (Å^2)	114.5	98.6	70.2
Average B-factor (Å^2) dimer AB/CD/EF	(69.0/130.6/143.7)		
Ramachandram plot	(Molprobity)	(Molprobity)	(Molprobity)
Most favored (%)	94.8	92.0	96.7
Additionally allowed (%)	4.3	7.8	3.3

† Values in parentheses refer to the highest resolution shell (3.6-3.4Å) for ComE^{D58A}, (3.3-3.2) REC^{D58A} domain, and (3.1-2.9) for REC^{D58E}.
§ $R_{sym} = \Sigma_h\Sigma_i |\langle I\rangle_h - I_{h,i}| / \Sigma_h\Sigma_i I_{h,i}$, where $\langle I\rangle_h$ is the mean intensity for reflection I_h and $I_{h,i}$ is the intensity of an individual measurement of reflection I_h.
The ComE^{D58A}, REC^{D58A} and REC^{D58E} R_{free} were calculated using respectively 1897 (5.0%), 553 (5.0%) and 739 (5.0%) of the reflections, which were set aside from the refinement.

Table S2. Oligomeric states of ComE, and characteristic paramaters calculated from the SAXS experimental curves.

	quaternary state	Rg Guinier (Å)	Rg P(r) (Å)	Dmax (Å)	MW(kDa) from Porod volume	Theoretical MW(kDa) [b]
REC^{D58A}	monomer	17.4 +/- 0.6	17.6 +/- 0.2	50 +/- 2	19.8 +/- 2	17.1
REC^{D58E}	dimer	21.4 +/- 0.3	21.0 +/- 0.2	61 +/- 3	34.2 +/- 3	34.4
LytTR	monomer	16.0 +/- 0.2	16.0 +/- 0.2	51 +/- 2	13.9 +/- 1.5	14.4
ComE^{D58A}	monomer	25.0 +/- 0.5	25.0 +/- 0.5	78 +/- 3	31.9 +/- 3	30.7
ComE^{D58E}	dimer	28.6 +/- 0.2	29.2 +/- 0.3	92 +/- 4	64 +/- 6	61.5
Crystal subunit A [a]		*22.8*	*22.9*	*71*		
crystal subunit B [a]		*22.8*	*23.0*	*71*		
crystal dimer [a]		*27.5*	*27.5*	*87*		

Oligomeric states, characteristic dimensions R_g and D_{max} deduced from experimental curves, molar mass M obtained from the whole I(q) curve using the macromolecule volume and the method developed by Craievich's team.
[a] Four (five) histidine residues are missing in the crystal structure of the ComE subunit A (B) respectively.
[b] The theoretical masses were derived from the sequence taking into account the His-Tag.

SI REFERENCES

1. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947-2948.
2. Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305-308.