SUPPLEMENTARY DATA

The stress sigma factor of RNA polymerase RpoS/σ^S is a solvent exposed open molecule in solution

Paola Cavaliere[1]+, Sébastien Brier[2,3]+, Petr Filipenko[4], Christina Sizun[5], Bertrand Raynal[6,7], Françoise Bonneté[8], Fabienne Levi-Acobas[7,9], Jacques Bellalou[7,9], Patrick England[6,7], Julia Chamot-Rooke[2,3], Claudine Mayer[7,10,11] and Françoise Norel[1,7,12]*.

[1] Institut Pasteur, Laboratoire Systèmes Macromoléculaires et Signalisation, Département de Microbiologie, 25 rue du Docteur Roux, 75015 Paris, France

[2] Institut Pasteur, Unité de Spectrométrie de Masse Structurale et Protéomique, Département de Biologie Structurale et Chimie, 28 rue du Docteur Roux, 75015 Paris, France

[3] CNRS USR 2000, 28 rue du Docteur Roux, 75015 Paris, France

[4] Hunter College-CUNY, Department of Computer Science, 695 Park Avenue, Hunter North Building, New York, NY 10065, USA

[5] Institut de Chimie des Substances Naturelles, CNRS UPR2301, Université Paris Saclay, 91190 Gif-sur-Yvette, France

[6] Institut Pasteur, Plateforme de biophysique moléculaire, Centre d'innovation et de recherche technologique, 25 rue du Docteur Roux, 75724 Paris

[7] CNRS UMR3528, 28 rue du Docteur Roux, 75015 Paris, France

[8] Institut des Biomolécules Max Mousseron (IBMM) UMR 5247 CNRS-UM-ENSCM, Université d'Avignon, 301, rue Baruch de Spinoza, F84000 Avignon, France

[9] Institut Pasteur, Plate-forme de Protéines Recombinantes, Département de Biologie Structurale et Chimie, 25 rue du Docteur Roux, 75015 Paris, France

[10] Institut Pasteur, Unité de Microbiologie Structurale, Département de Biologie Structurale et Chimie, 25 rue du Docteur Roux, 75015 Paris, France

[11] Université Paris Diderot, Sorbonne Paris Cité, Paris, France.

[12] Institut Pasteur, Unité de Biochimie des Interactions Macromoléculaires, Département de Biologie structurale et Chimie, rue du Dr. Roux, 75015 Paris, France

+ These authors contributed equally to the manuscript

*To whom correspondence should be addressed: Françoise Norel

Email: francoise.norel@pasteur.fr Tel: (33) 140613122

Supplementary Table S1. Bacterial strains and plasmids used in this study

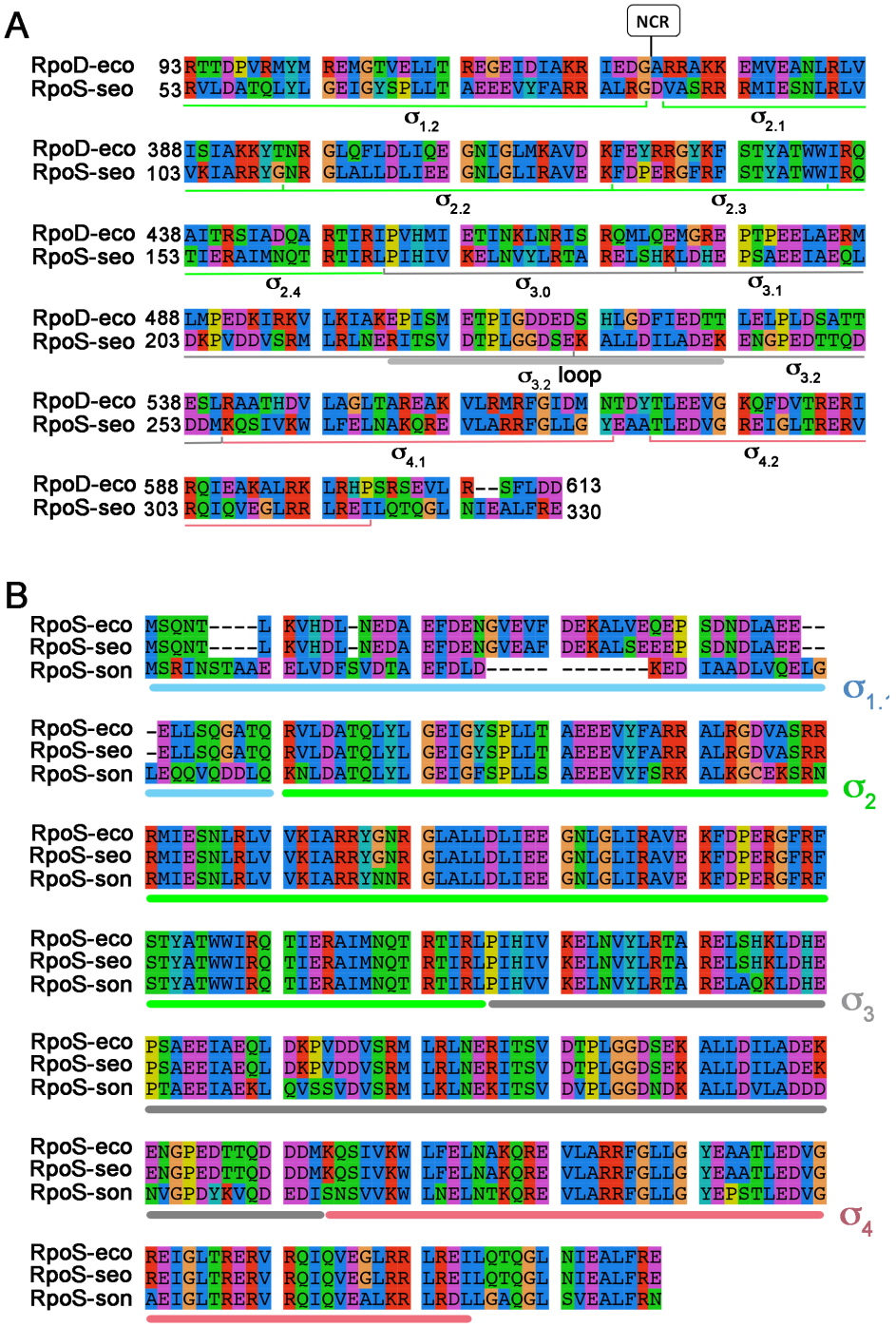
Strain or Plasmid	Characteristics	Source or reference
Strain		
BL21 (DE3)	*Escherichia coli F– ompT hsdSB (rB– mB–) gal dcm (DE3)*	Merck-Millipore
Plasmids		
pETM11	Vector with his-tag and TEV cleavable site, used for expression of σ^S	G. Stier, EMBL Heidelberg
pET-MCN-EAVNH	Vector with his-tag and TEV cleavable site, used for expression of σ^S_{1-162}	G. Stier, EMBL Heidelberg
pVFD969	pETM11 expressing $(\text{his-tag})_6$-σ^S	[28]
pVFE66	pET-MCN-EAVNH expressing $(\text{his-tag})_6$-σ^S_{1-162}	This study[a]

[a]: A DNA fragment was PCR-amplified from total DNA of *S*. Typhimurium ATCC14028 using primers RpoS-NdeI (5'-TCGACGCATATGagtcagaatacgctgaaagttcat-3') and D2-XhoI (5'-ACTTAGCTCGAGttatgcggtttggttcatgatcgcccgt-3') (Sigma-Aldrich, France) and was cloned between the NdeI and XhoI sites of pET-MCN-EAVNH. The resulting plasmid was confirmed to be correct by DNA sequencing (Beckman Coulter Genomics, France).

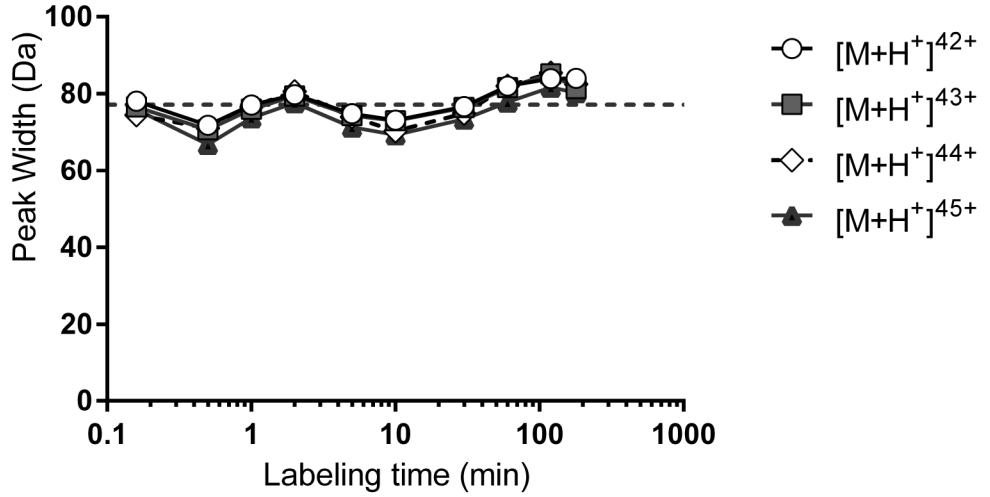
Supplementary Table S2. Chemical shifts of the N-terminal region of $\sigma^{S}_{1\text{-}162}$ at 298 K, obtained from HNCA, HN(CO)CA and HNCO spectra in 50 mM pH 7.5 sodium phosphate buffer, supplemented with 300 mM NaCl, 50 mM arginine and 5 mM potassium glutamate. $^{1}H_{N}$ chemical shifts were referenced with DSS.

Residue	$^{1}H_{N}$ (ppm)	^{15}N (ppm)	^{13}CO (ppm)	$^{13}C\alpha$ (ppm)
-6Asn	-	-	-	53.2
-5Leu	8.11	122.1	-	55.5
-4Tyr	8.04	119.6	-	57.8
-3Phe	8.04	121.3	-	57.8
-1Gly	-	-	174.2	45.5
0His	8.28	120.3	-	56.8
6Leu	-	-	175.2	55.3
7Lys	8.25	122.7	176.2	56.1
8Val	8.11	121.5	-	62.4
9His	-	-	174.9	56.1
10Asp	8.29	121.9	175.2	54.2
11Leu	8.24	122.6	177.2	55.4
12Asn	8.49	119.2	175.4	53.4
13Glu	8.38	121.4	176.2	57.0
14Asp	8.32	120.8	175.9	54.4
15Ala	8.07	124.0	177.5	52.6
16Glu	8.29	119.7	176.1	56.5
17Phe	8.15	120.8	175.1	57.5
18Asp	8.25	123.0	176.0	53.9
19Glu	8.38	122.5	176.5	57.2
20Asn	8.48	118.6	175.8	53.5
21Gly	8.23	109.1	174.0	45.5
22Val	7.95	119.2	176.2	62.3
23Glu	8.52	124.9	175.9	56.5
24Ala	8.25	125.3	177.2	52.2
25Phe	8.20	119.9	175.2	57.7
26Asp	8.22	122.7	176.3	53.8
27Glu	8.51	122.5	176.8	57.2
28Lys	8.28	121.5	176.6	56.6
29Ala	8.09	124.3	177.9	52.6
30Leu	8.12	120.9	177.6	55.3
31Ser	8.22	116.2	174.5	58.3
32Glu	8.39	122.7	176.2	56.4
33Glu	8.31	121.6	176.2	56.3
34Glu	8.47	123.7	-	54.2
38Asn	-	-	175.0	53.5
39Asp	8.33	120.9	176.3	54.6
40Leu	8.08	122.1	177.3	55.2
41Ala	8.24	124.8	178.2	52.7
42Glu	8.34	120.2	177.2	57.3
43Glu	8.44	120.9	-	57.4

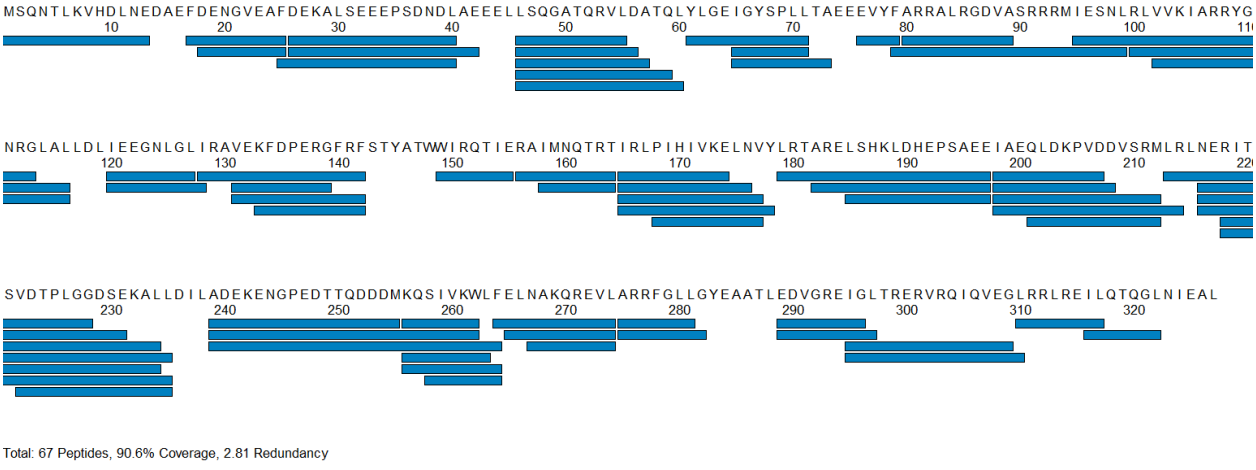
52Gln	8.17	124.3	-	55.3
54Val	7.99	113.5	-	61.6



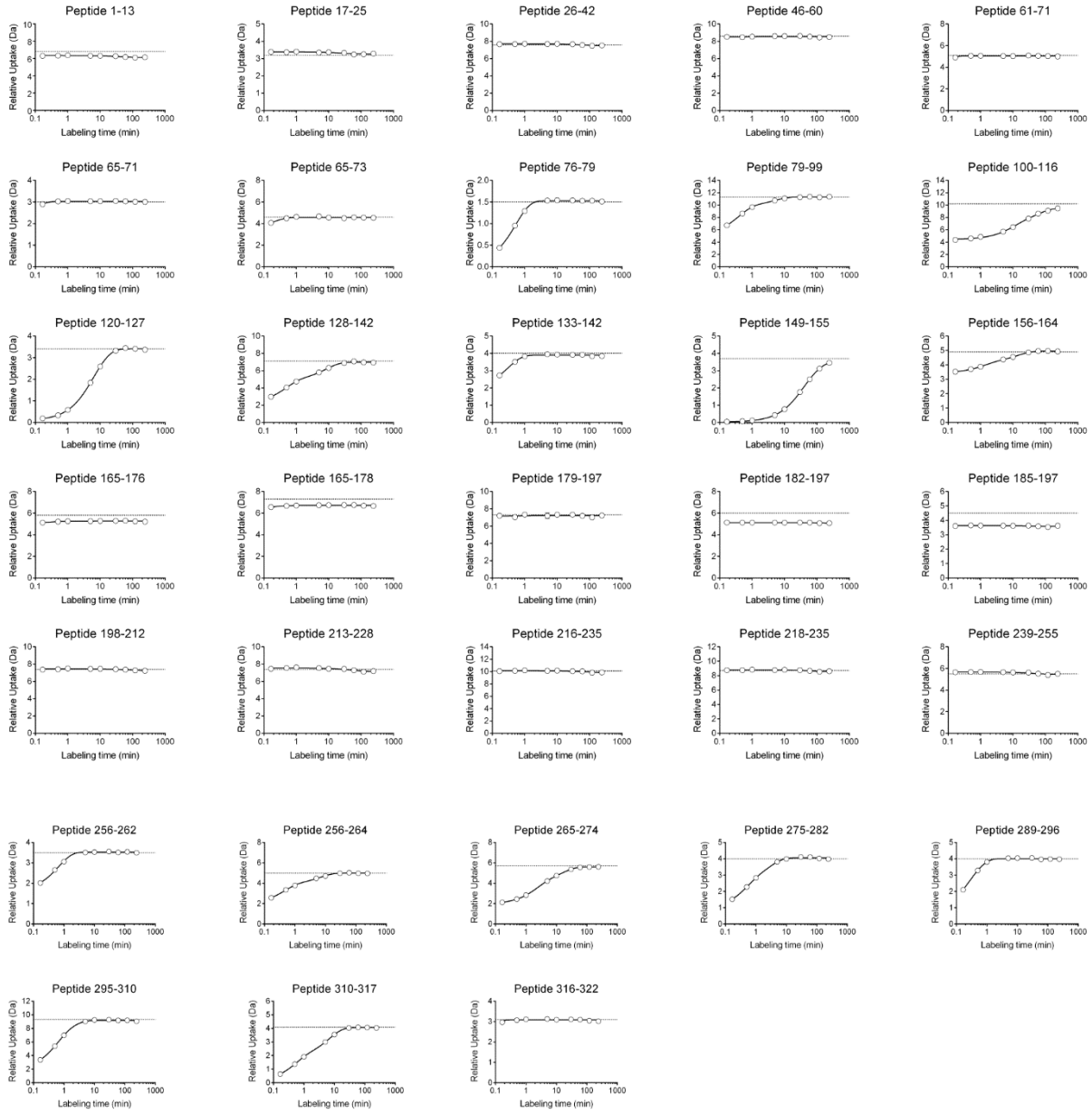
Supplementary Figure S1. Sequence alignments of sigma factors.(A) sequence alignment of σ^S from *S*. Typhimurium ATCC14028 (RpoS-seo) and σ^{70} from *E. coli* MG1655 (RpoD-eco). Sequences are presented with the one-letter amino acid code with numbers indicating amino acid positions relative to the start of σ. Conserved regions of σ are labeled below the sequences. The non-conserved region of σ^{70} is indicated by NCR. Residues of σ^S forming a loop in the structure of $E\sigma^S$ [26] are shown (the $\sigma_{3.2}$ loop). (B) sequence alignment of σ^S from *S*. Typhimurium (RpoS-seo), *E. coli* MG1655 (RpoS-eco) and *Shewanella oneidensis* (RpoS-son). Protein sequences are from KEGG (http://www.genome.jp/kegg/kegg2.html).



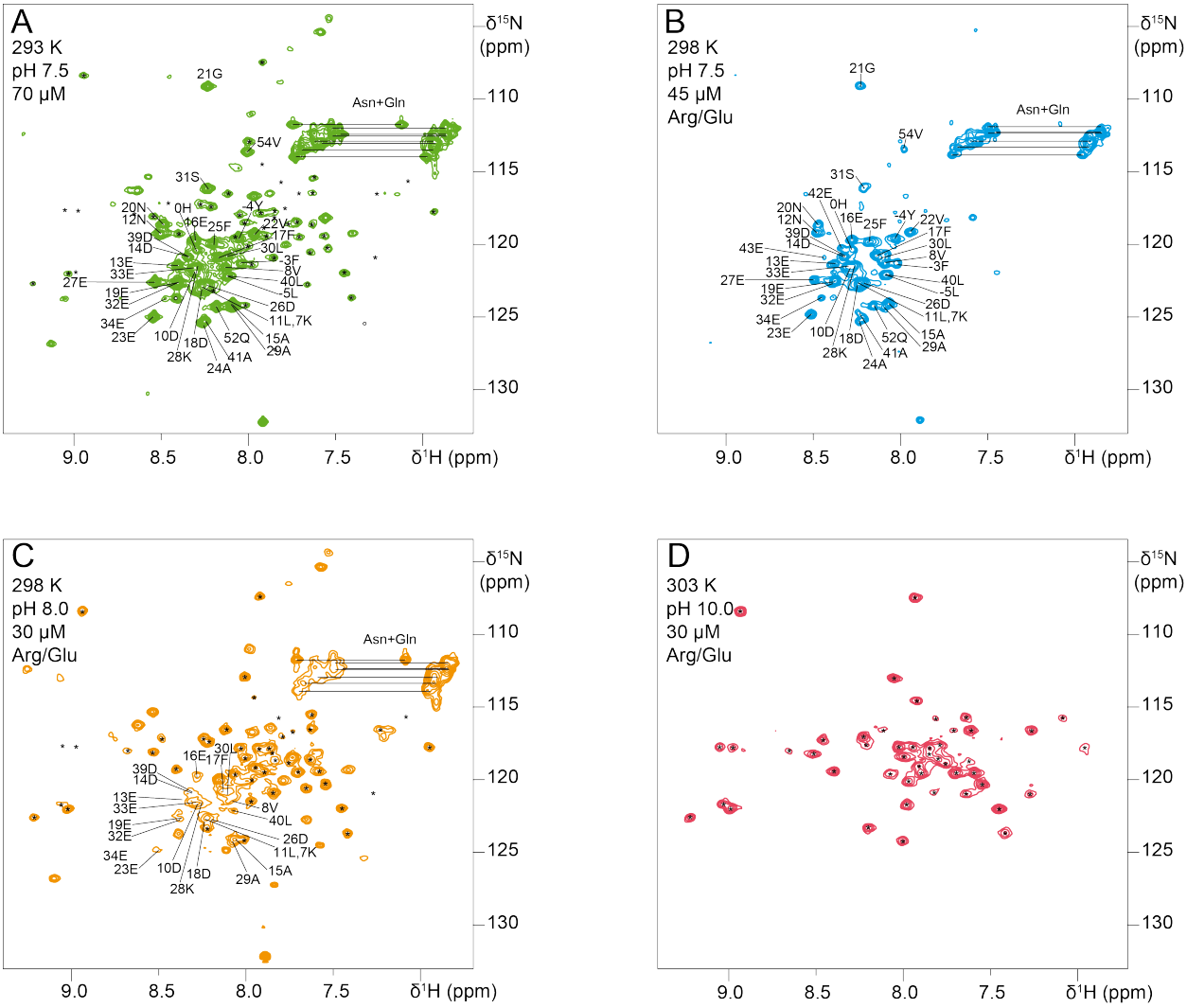
Supplementary Figure S2. Comparison of the peak width for the most intense charge states of free full-length σ^S. The peak width in Da has been calculated by multiplying the peak width in m/z (at 50% of maximum intensity) by the charge state of the measured peak. No difference between states was observed. The average peak width value is reported as a dashed line.



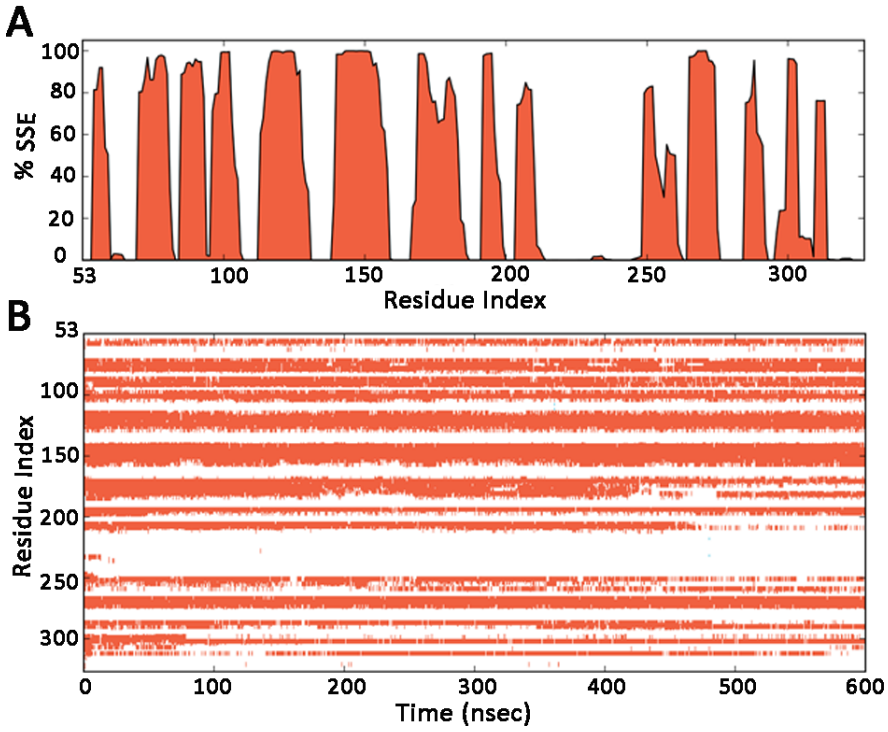
Supplementary Figure S3. Peptide map of the full-length free σ^S factor upon pepsin digestion. Each blue bar represents one unique peptide. A sequence coverage of 90.6% was achieved.



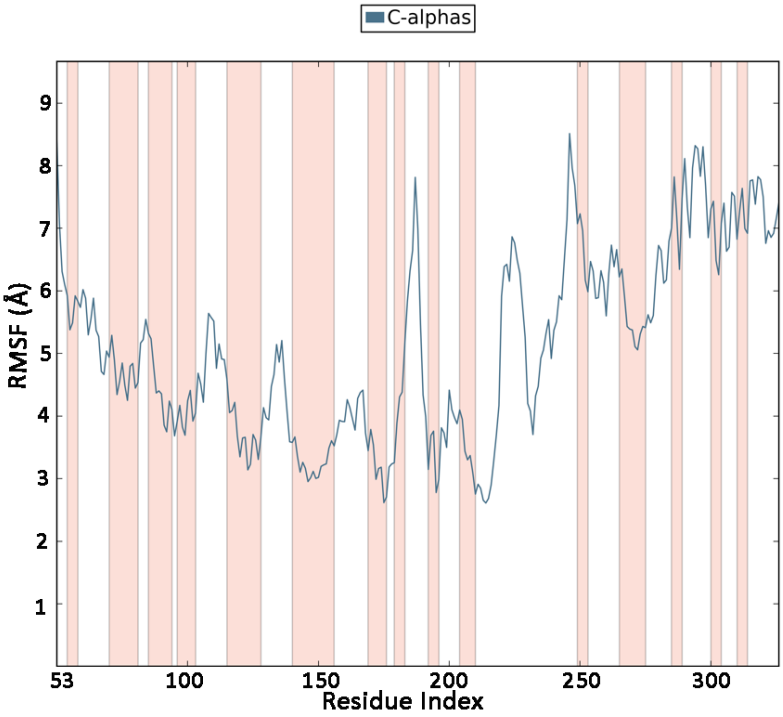
Supplementary Figure S4. Deuterium uptake curves for all σ^S peptides. The gray dashed lines correspond to the maximum deuterium uptake level measured after 6 h at 60°C. All kinetics were fitted to a tri-exponential expression using GraphPad Prism 6.0 (GraphPad Software, San Diego, California, USA). Each dot corresponds to the average of three independent experiments.



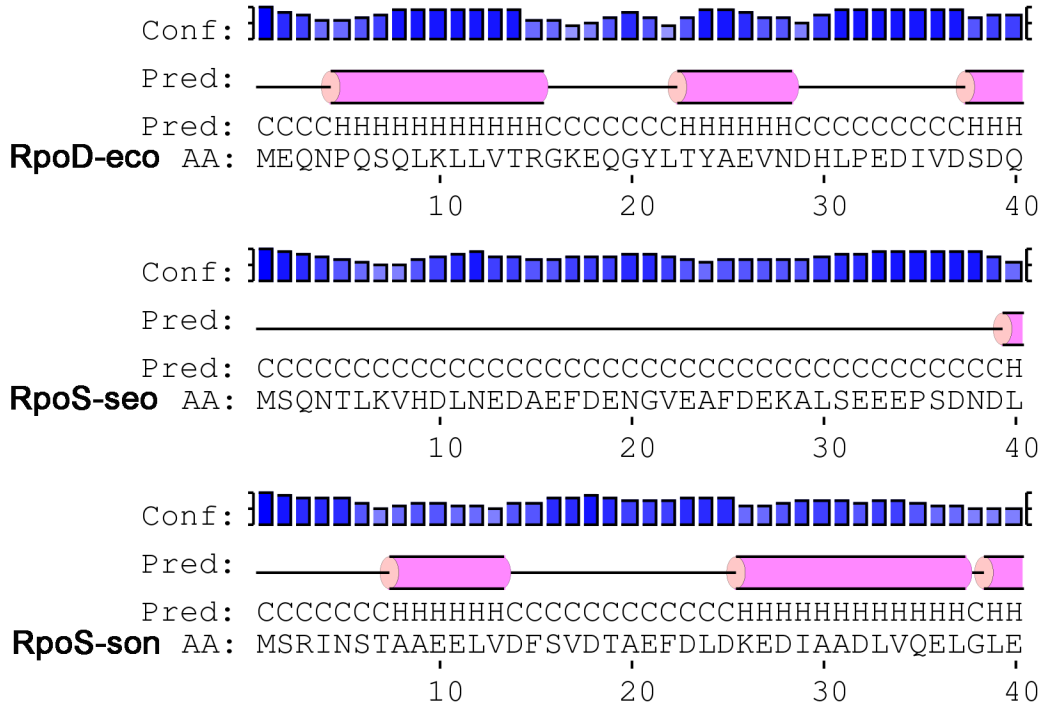
Supplementary Figure S5. Structural analysis of ^{15}N-labeled $\sigma^{S}_{1\text{-}162}$ by NMR spectroscopy. ^{1}H-^{15}N HSQC (A) or ^{1}H-^{15}N HMQC (B, C and D) spectra were acquired under different experimental conditions (temperature, pH, addition or not of 50 mM Arg and 5 mM Glu as stabilizing agents), as indicated. Spectra A and B were acquired at 800 MHz ^{1}H frequency, spectra C and D at 600 MHz, with 128-160 points in the ^{15}N dimension. The numbers of scans were 256, 32, 512 and 128 for spectra A, B, C and D, respectively. Total measurement times were 12 h, 16 min, 6 h 40 min and 1 h 20 min with recycling delays of 1 s, 0.15 s, 0.2 s and 0.2 s for spectra A, B, C and D, respectively. Spectra A-D display amide (HN) and Asn-Gln NH_2 side chain signals. Low temperature and low pH (293 K and pH 7.5, panel A) favor HN signals of disordered regions, while at high temperature and high pH (303K and pH 10.0, panel D) only HN signals in slow exchange with water are observed. Chemical shifts were assigned under the conditions of panel B (298 K, pH 7.5, 50 mM Arg and 5 mM Glu) for ~40 sharp signals with narrow ^{1}H chemical shift dispersion (8.0-8.5 ppm). They include residues from the N-terminal His-tag that were left over after Tev cleavage and assigned negative numbers, since $\sigma^{S}_{1\text{-}162}$ numbering was started at Met1 ("1M"). They belong to domain $\sigma_{1.1}$ and are also present without stabilizing agents (panel A). They are still present at higher pH albeit much broader (panel C), but become undetectable at 303K and pH 10 due to exchange-broadening (panel D). Analysis of the corresponding backbone chemical shifts $^{1}H_N$, ^{15}N, $^{13}C\alpha$, ^{13}C' with Talos+ [50] predicts random coil structure, which corroborates the highly dynamic and disordered state of $\sigma_{1.1}$ determined by HDX-MS analysis. The subset of ~45 amide signals marked with stars (*) in panel D display large ^{1}H chemical shift dispersion (7.0-9.5 ppm). They correspond to exchange-protected amides engaged in stable secondary structure elements and presumably belong to domain σ_2. They were not assigned, but are also observed at lower temperature, where they appear broader due to more efficient nuclear relaxation (panels A, C and D).

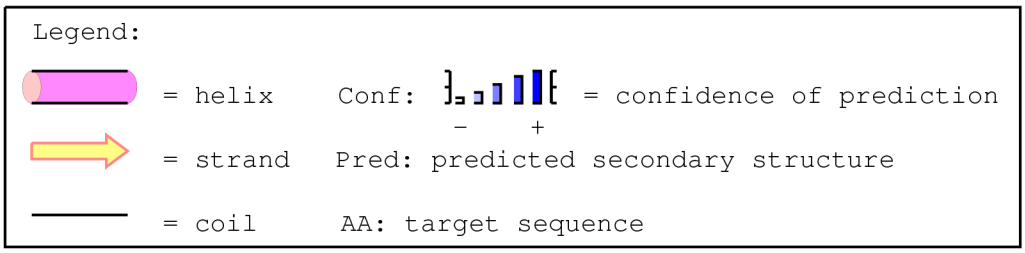


Supplementary Figure S6. A) The secondary structure element (SSE) distribution in σ^S is reported *versus* the residue index as monitored through the simulation. Regions with high values of SSE percentage retained their secondary structure along the simulation. Note that the N-terminal domain of σ^S ($\sigma_{1.1}$; residues 1 to 52) has not been traced in the Eσ^S structure [26]. B) The SSE for each residue in the protein was plotted *versus* the time course of the simulation. For some regions, (i.e. around residues 210 to 220), the secondary content decreases for the last trajectories, suggesting an unstable/unstructured region.



Supplementary Figure S7. Diagram of the Root Mean Square Fluctuations (RMSF) of Cα atoms obtained for σ^S. Peaks indicate areas of the protein that fluctuate the most during simulation. Typically, the N- and C-terminal regions fluctuate more than any other part of the protein, while secondary structures as α-helices and β-strands are usually more rigid and thus fluctuate less than the loop regions. The pink highlights in the plot correspond to the α-helices regions that persist over 70% of the entire simulation.





Supplementary Figure S8. Secondary structure prediction of the N-terminal region of σ^{70} from *E. coli* MG1655 (RpoD-eco), σ^{S} from *S.* Typhimurium ATCC14028 (RpoS-seo) and σ^{S} from *Shewanella oneidensis* (RpoS-son) using the PSIPRED protein structure prediction server (http://bioinf.cs.ucl.ac.uk/psipred/). Sequences are presented with the one-letter amino acid code with numbers indicating amino acid positions relative to the start of σ. Protein sequences are from KEGG (http://www.genome.jp/kegg/kegg2.html).

Supplementary Reference

50. Shen, Y., Delaglio, F., Cornilescu, G. and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR.* **44**(4), 213-223.