Supplementary Material

Codon bias shapes bacterial small RNA binding sites within protein-coding sequences

Shira Fisher and Hanah Margalit*

Department of Microbiology and Molecular Genetics
IMRIC, Faculty of Medicine, The Hebrew University of Jerusalem,
Jerusalem, 9112102, ISRAEL

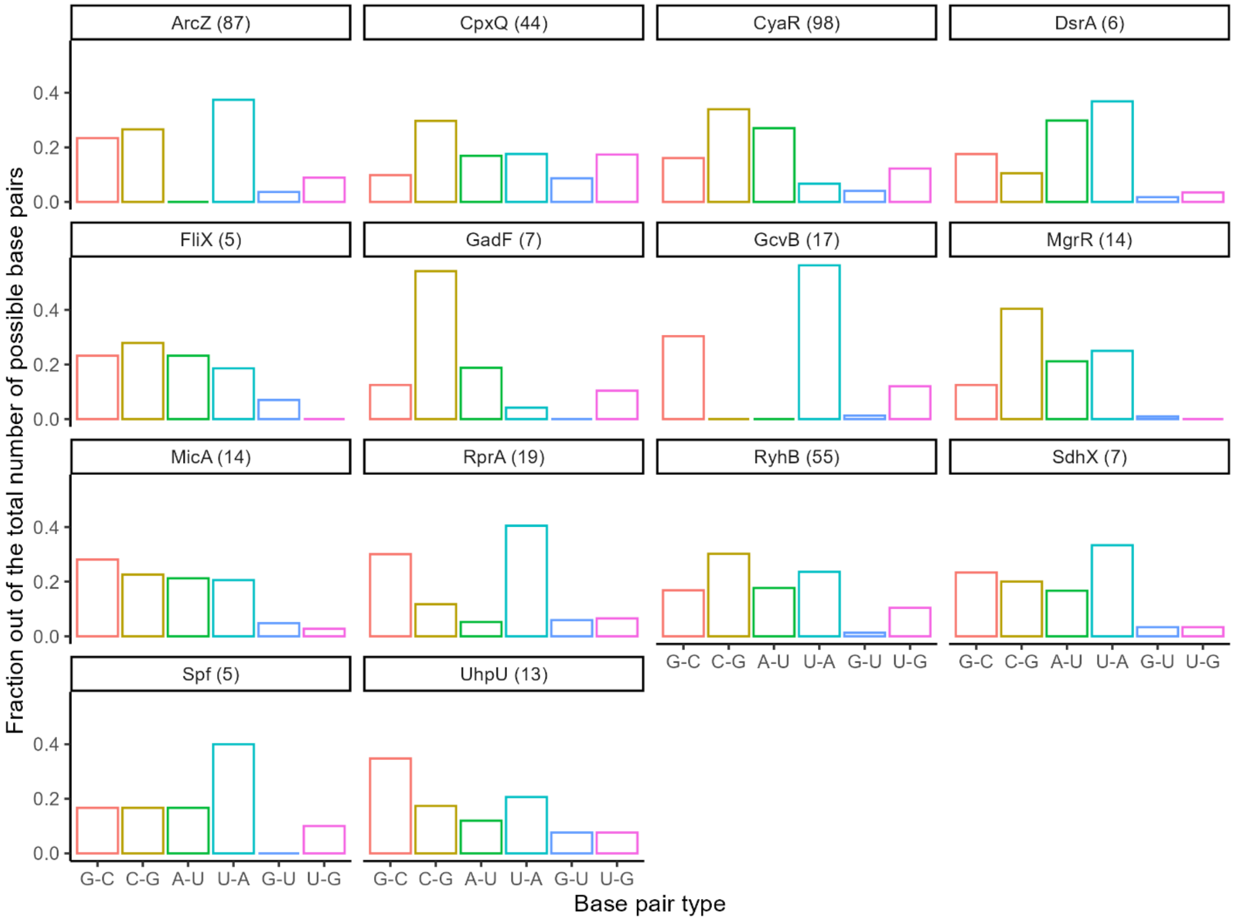


Figure S1: Base pair types involved in sRNA-target interactions. For each sRNA, a histogram is presented, depicting the distribution of base pair types involved in its interactions with the targets (sRNAs with at least five targets were included in this analysis). The Y-axis indicates the fraction of each base pair type out of all base pairing interactions of an sRNA.

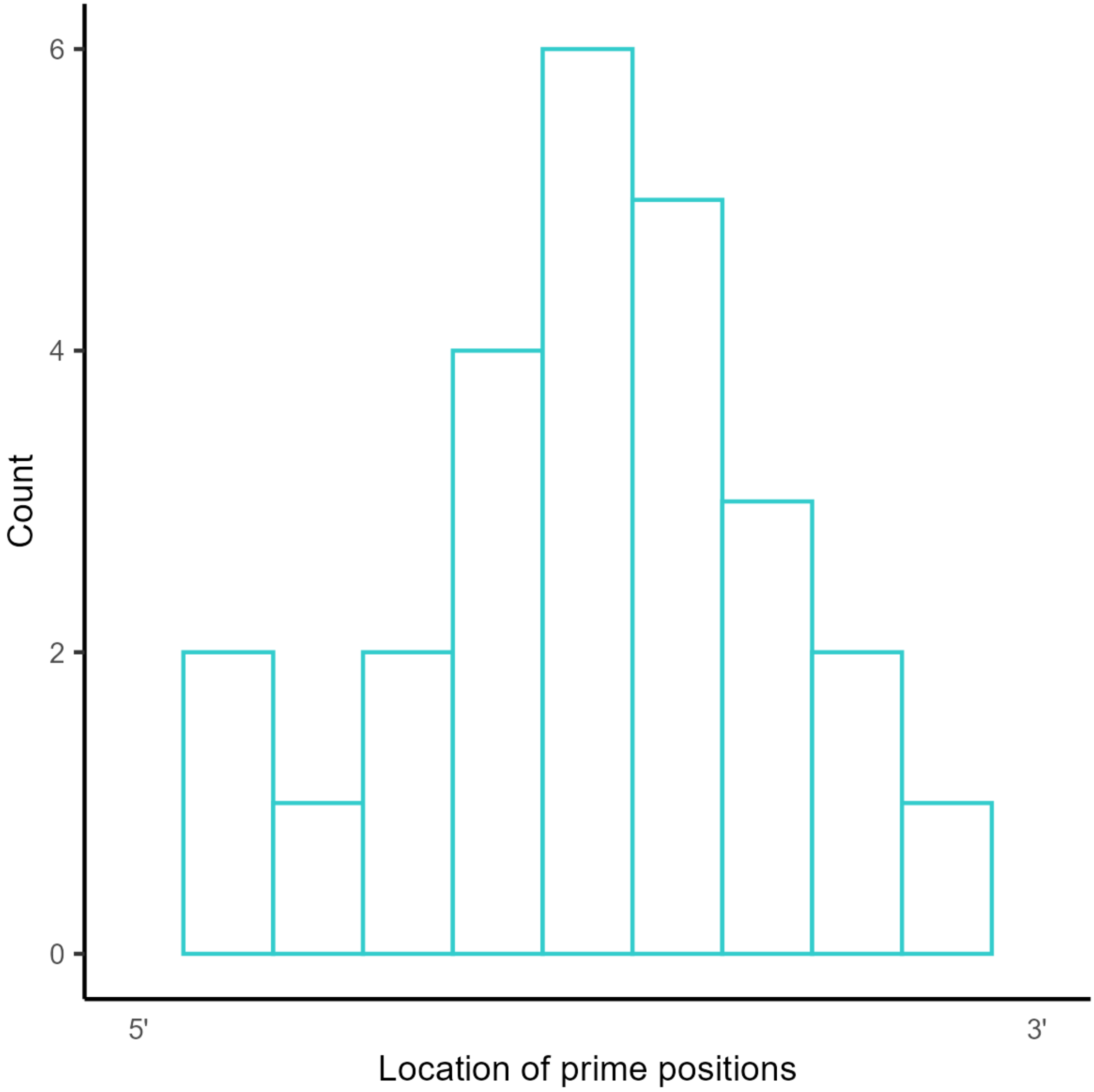
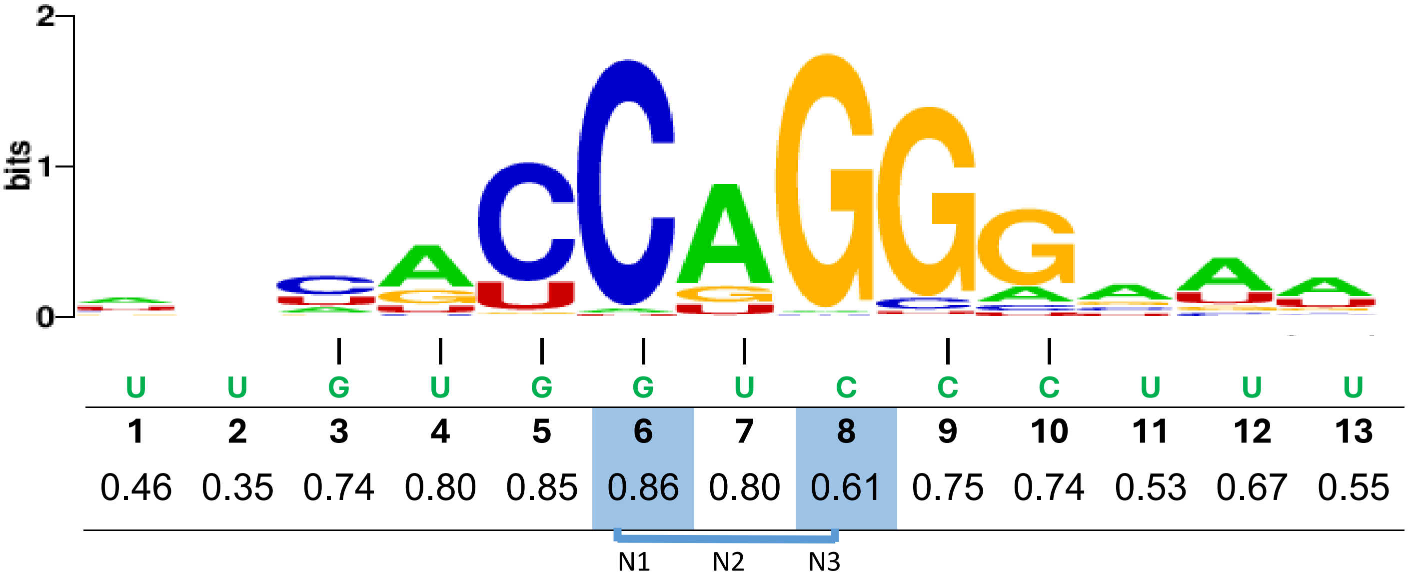


Figure S2: Location of the prime positions on the sRNA. The distribution of the location of the prime positions on the binding site on the sRNA is shown. sRNAs with at least five targets were included in this analysis.

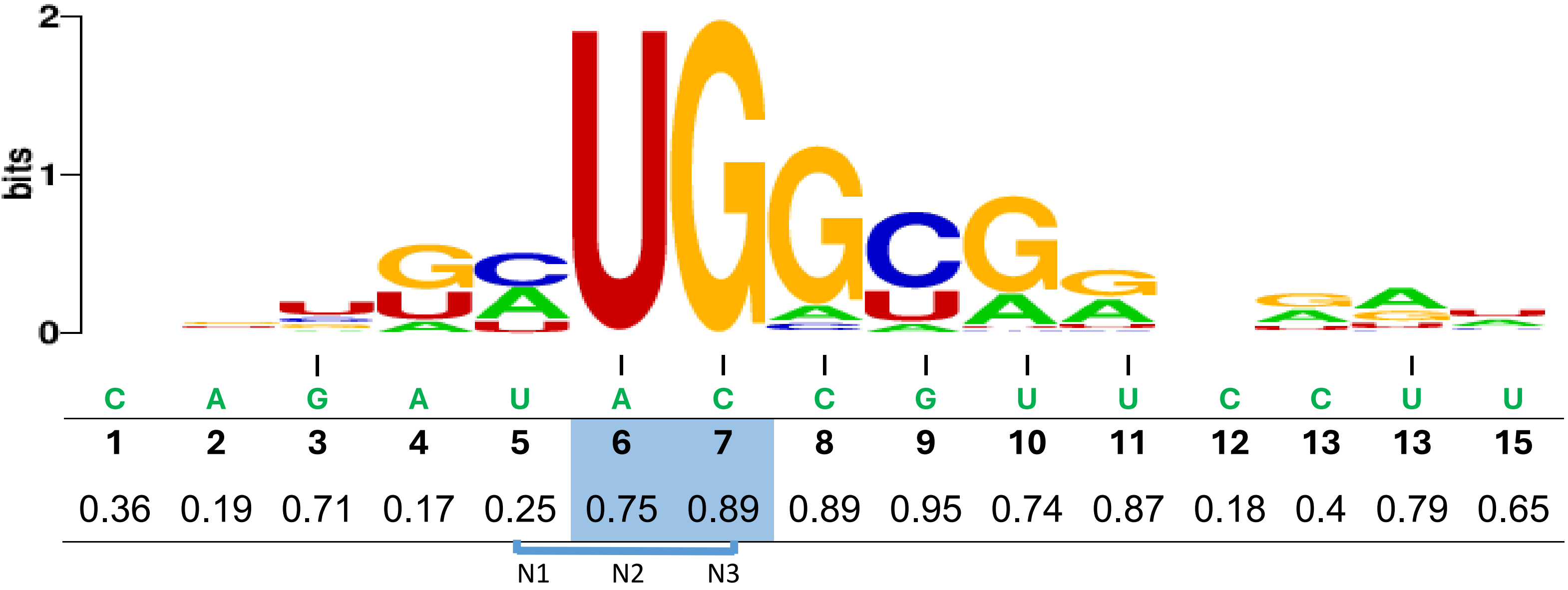
ArcZ



U	U	G	U	G	G	U	C	C	C	U	U	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
0.46	0.35	0.74	0.80	0.85	0.86	0.80	0.61	0.75	0.74	0.53	0.67	0.55

N1 N2 N3

CpxQ



C	A	G	A	U	A	C	C	G	U	U	C	C	U	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**13**	**15**
0.36	0.19	0.71	0.17	0.25	0.75	0.89	0.89	0.95	0.74	0.87	0.18	0.4	0.79	0.65

N1 N2 N3

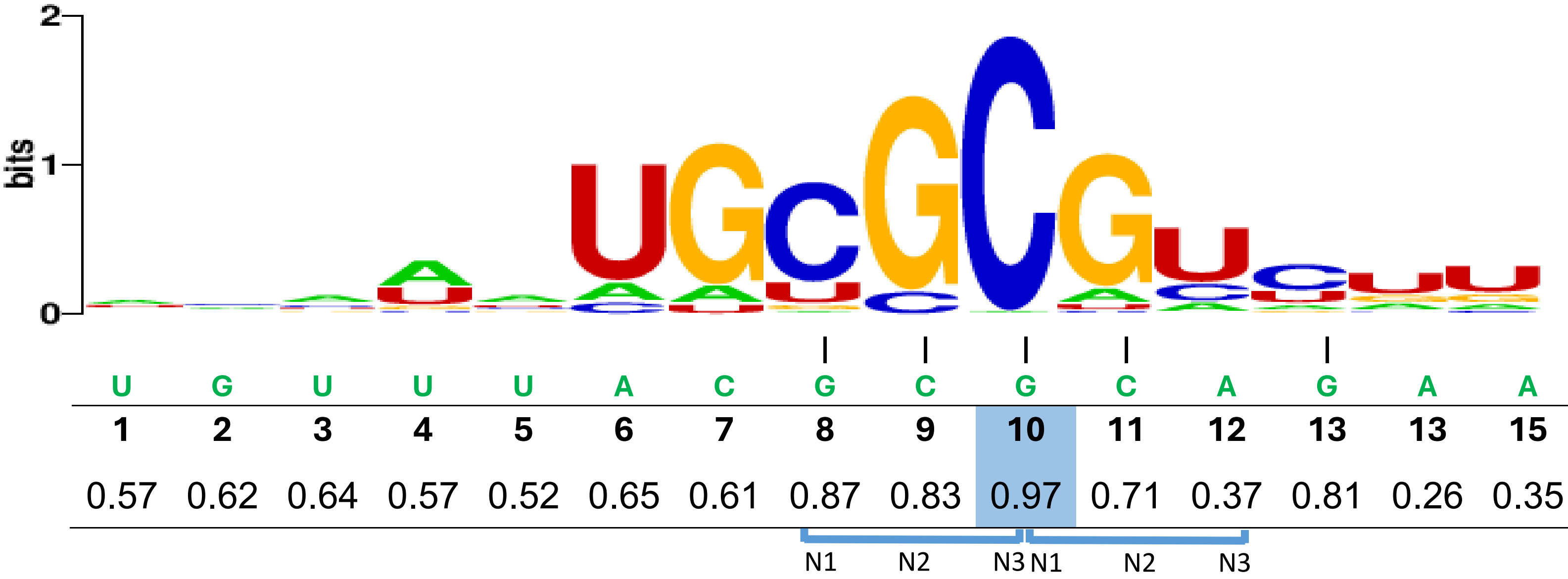
CyaR

										|				
G	G	A	C	C	A	U	G	U	C	G	A	U	C	
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**13**	**15**
0.51	0.28	0.49	0.67	0.66	0.60	0.56	0.48	0.25	0.58	0.86	0.15	0.55	0.36	0.29
								N1	N2	N3				

MgrR

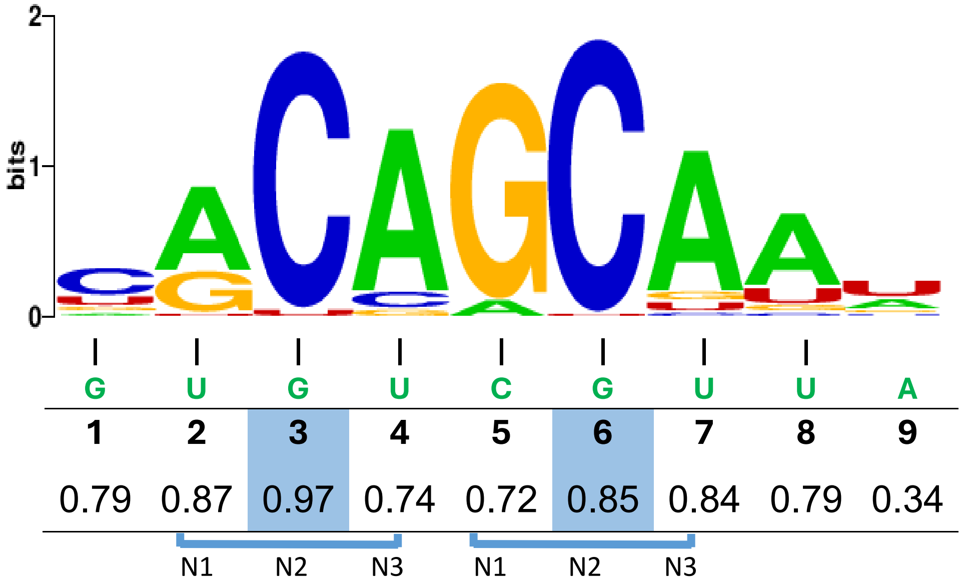
|	|	|	|	|	|	|	
C	G	U	A	C	C	U	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	1	1	1	0.89	0.56

MicA



U	G	U	U	U	A	C	G	C	G	C	A	G	A	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**13**	**15**
0.57	0.62	0.64	0.57	0.52	0.65	0.61	0.87	0.83	0.97	0.71	0.37	0.81	0.26	0.35

RprA



G	U	G	U	C	G	U	U	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.79	0.87	0.97	0.74	0.72	0.85	0.84	0.79	0.34

RyhB

A	U	G	A	C	C	U	U	C	G	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
0.60	0.74	0.93	0.93	0.73	0.78	0.60	0.67	0.69	0.47	0.64

N1 N2 N3

Spf

U	U	U	C	C	A	G	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0.99	0.95	0.83	0.75	0.23	0.03	0.99	0.42

N1 N2 N3 N1 N2 N3

Figure S3: Overall conservation of positional base pairing across all targets of an sRNA. For each sRNA, we counted for each position of the binding site the number of putatively fulfilled base pairing interactions across all targets and across all genomes in all families (except for enterobacteriaceae, which were analyzed separately). Presented are the fractions of genomes putatively fulfilling a positional base pairing interactions out of all genomes included in an analysis across all targets. The reading frame corresponding to the dominant alignment of the prime position is marked for each sRNA (N1, N2, N3). MicA shows two dominant reading frame adjustments of its prime position and therefore both were marked. For the sRNAs ArcZ, CyaR, MicA, RprA, SpF, RyhB, the orthologues and binding site region were identified by Infernal and for the other sRNAs by BLAST. The sequence LOGO illustrates the nucleotide frequencies at each position of the binding site across all *E. coli* target genes.

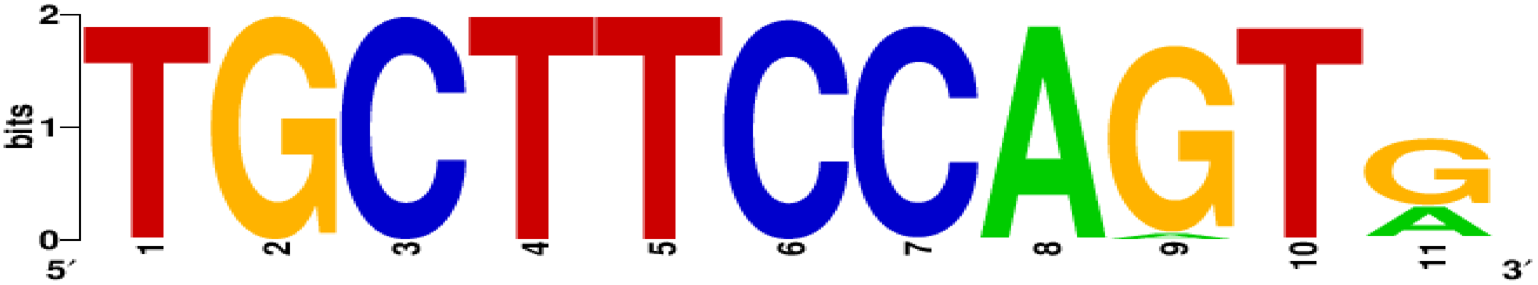


Figure S4: Conservation of RyhB across all genomes in all families (except for enterobacteriaceae). Presentation is by a logo, where the height or the letters represent the information content in each position in bits, based on equal background frequency of the nucleotides (0.25).

ArcZ

		|	|	|	|	|	|	|	|		|	
U	U	G	U	G	G	U	C	C	C	U	U	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
0.44	0.42	0.71	0.85	0.95	0.91	0.83	0.89	0.84	0.79	0.64	0.71	0.55

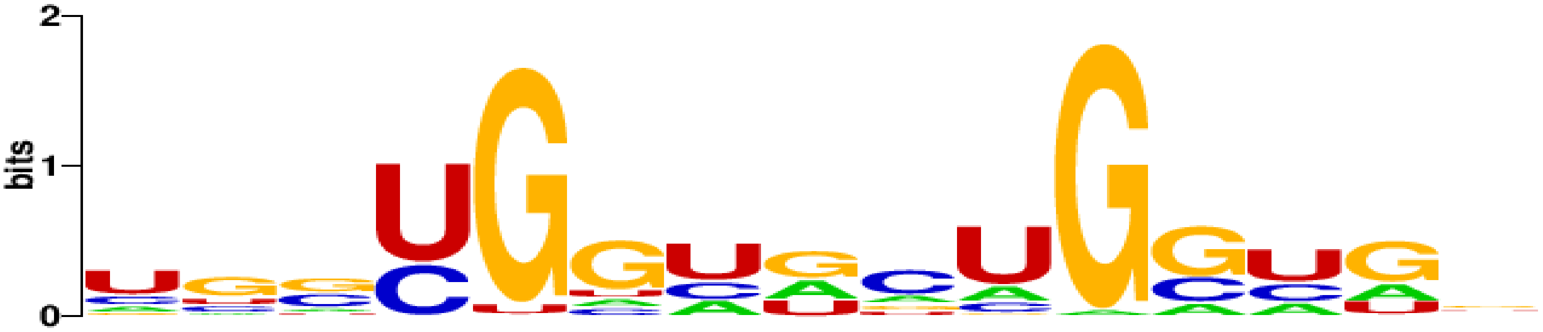
N1 N2 N3

CpxQ

		|			|	|	|	|	|	|			|	
C	A	G	A	U	A	C	C	G	U	U	C	C	U	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**13**	**15**
0.35	0.28	0.81	0.35	0.29	0.91	0.93	0.92	0.93	0.81	0.92	0.19	0.43	0.77	0.56

N1 N2 N3

CyaR



				|			|	|		|			|	
A	G	G	A	C	C	A	U	G	U	C	G	A	U	C
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**13**	**15**
0.46	0.31	0.48	0.62	0.84	0.67	0.57	0.71	0.77	0.19	0.93	0.19	0.55	0.79	0.42

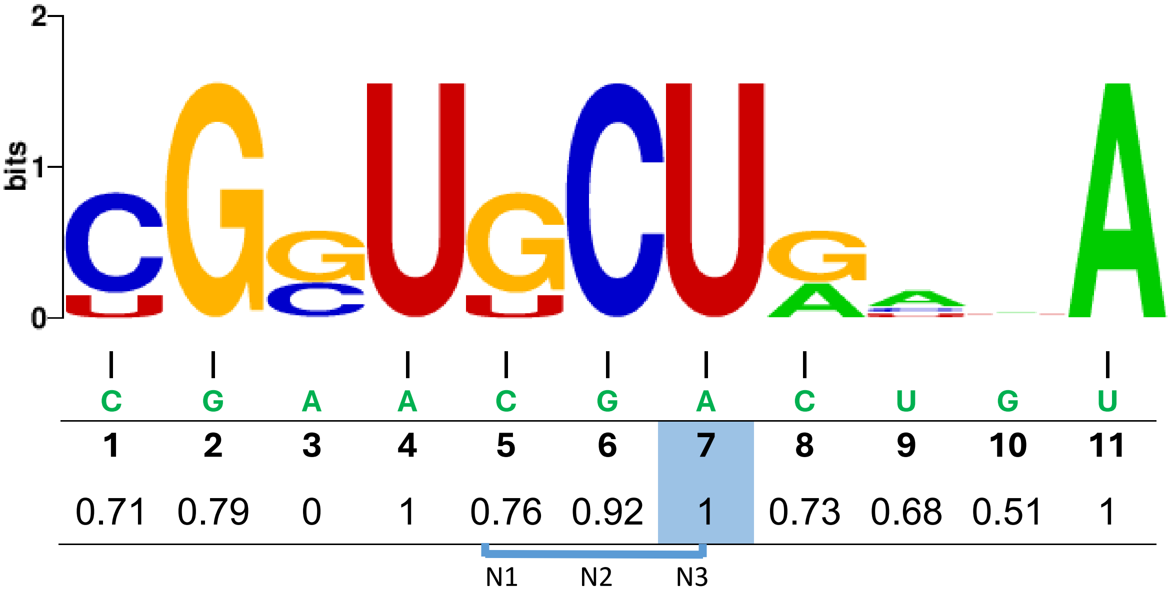
N1 N2 N3

DsrA

|	|	|		|	|	|			|	
U	U	U	A	A	G	C	A	A	U	G
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	0.99	0.83	0.66	0.75	0.95	0.83	0.55	0.4	0.83	0.61

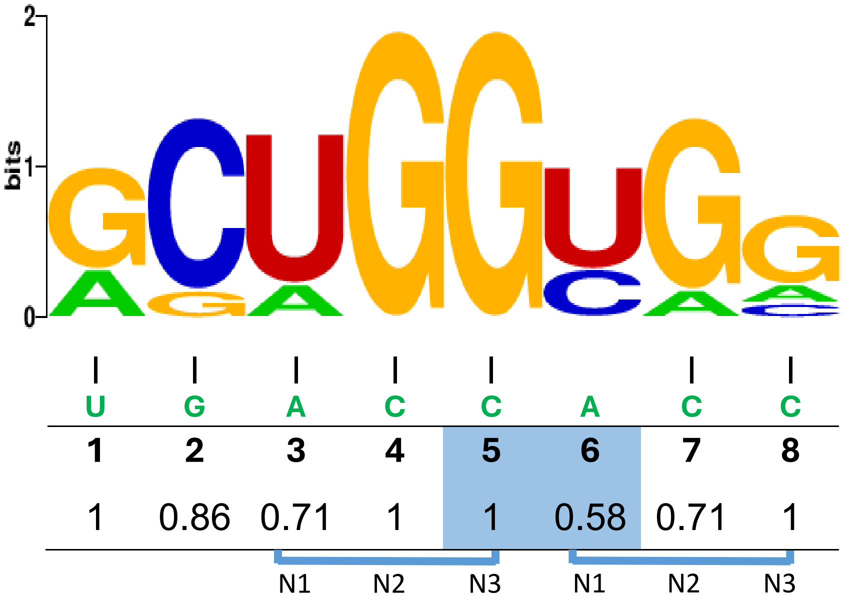
N1 N2 N3

FliX



C	G	A	A	C	G	A	C	U	G	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
0.71	0.79	0	1	0.76	0.92	1	0.73	0.68	0.51	1

GadF



U	G	A	C	C	A	C	C
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0.86	0.71	1	1	0.58	0.71	1

MgrR

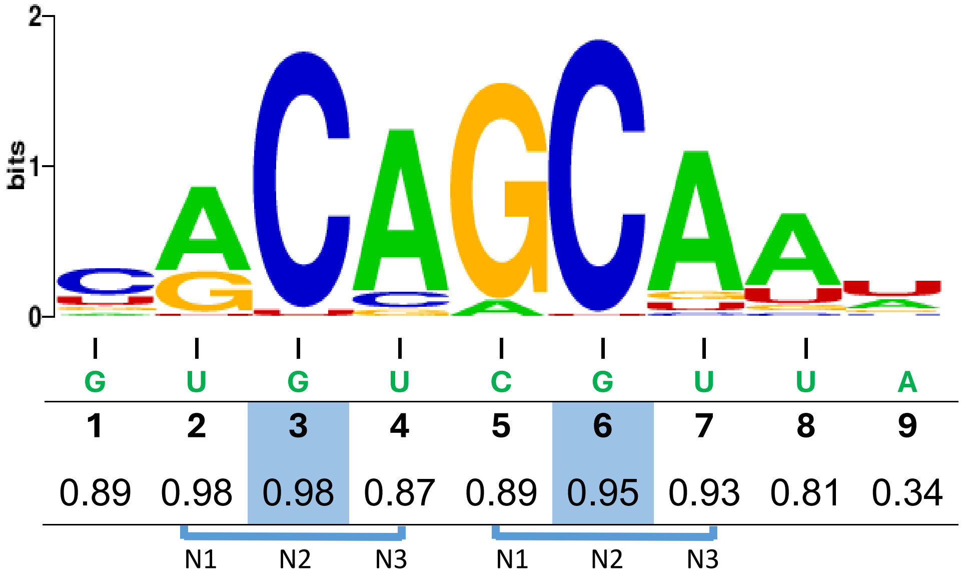
C	G	U	A	C	C	U	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0.91	0.89	1	1	1	0.86	0.53

MicA

U	G	U	U	U	A	C	G	C	G	C	A	G	A	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**13**	**15**
0.59	0.54	0.69	0.62	0.58	0.59	0.57	0.92	0.93	1	0.81	0.46	0.78	0.31	0.43

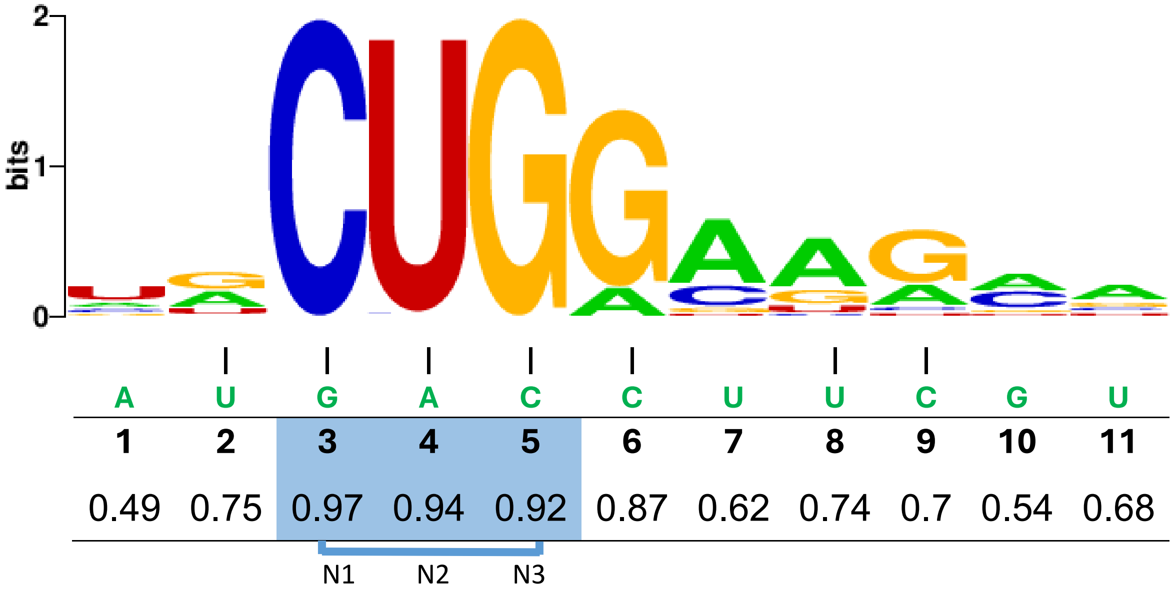
N1 N2 N3 N1 N2 N3

RprA



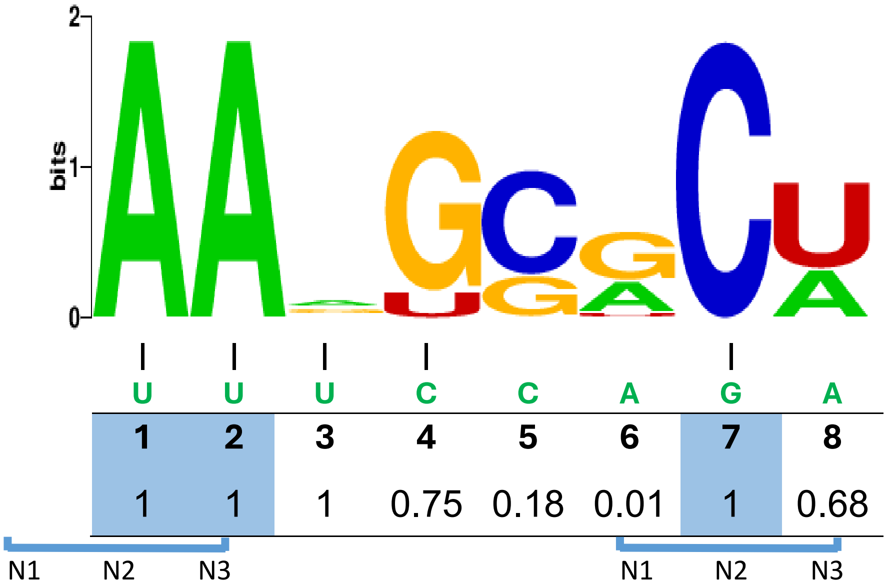
G	U	G	U	C	G	U	U	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.89	0.98	0.98	0.87	0.89	0.95	0.93	0.81	0.34

RyhB



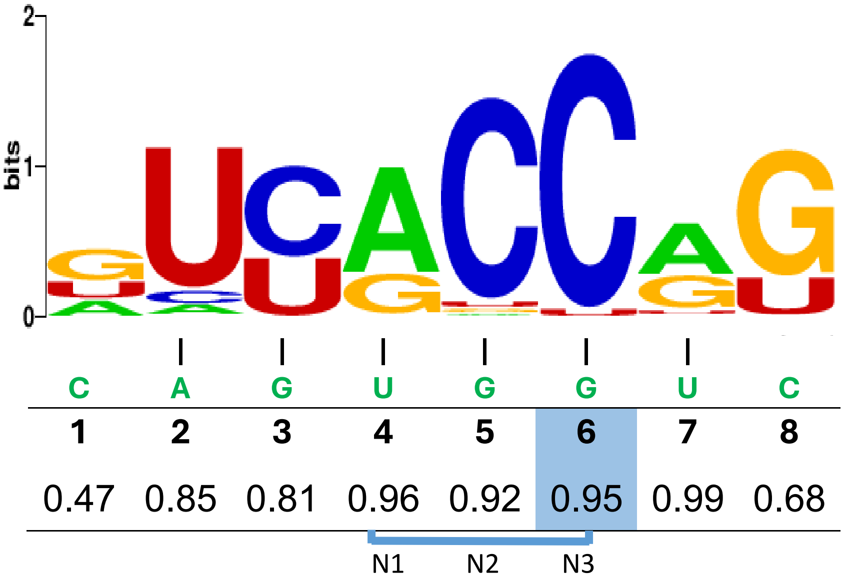
A	U	G	A	C	C	U	U	C	G	U
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
0.49	0.75	0.97	0.94	0.92	0.87	0.62	0.74	0.7	0.54	0.68

Spf



U	U	U	C	C	A	G	A
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	0.75	0.18	0.01	1	0.68

UhpU



C	A	G	U	G	G	U	C
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0.47	0.85	0.81	0.96	0.92	0.95	0.99	0.68

Figure S5: Overall conservation of positional base pairing across all targets of an sRNA. For each sRNA, we counted for each position of the binding site the number of putatively fulfilled base pairing interactions across all targets and across all genomes in enterobacteriaceae family. Presented are the fractions of genomes putatively fulfilling a positional base pairing interactions out of all genomes included in an analysis across all targets. The reading frame corresponding to the dominant alignment of the prime position is marked for each sRNA (N1, N2, N3). MicA shows two dominant reading frame adjustments of its prime position and therefore both were marked. For the sRNAs ArcZ, CyaR, DsrA, MicA, RprA, SpF, RyhB, the orthologs and binding site region were identified by Infernal and for the other sRNAs by blast. The sequence LOGO illustrates the nucleotide frequencies at each position of the binding site across all E. coli target genes.