

saRNA Design and Promoter Analysis

Bruktawit Maru

December 9, 2019

Promoter Analysis

1) Obtain promoter region for POLA2 from either:

- a) UCSC genome browser
- b) Ensembl genome browser
- c) dbEST and dbTSS – databases containing info on Expressed Sequence Tags (EST) and Transcriptional Start Site (TSS)
- d) Data tracks from genome-wide location analysis of RNA polymerase (RNAPII) occupancy and histone markers to predict TSS

2) Use MethPrimer to identify CpG Islands within promoter region

Rules to Improve Chance of Identifying saRNA target

1) Location-related rules:

- a) Target should be -100 to -1000bp away from TSS (-200 to -500bp even better!)
- b) saRNA should have A/T rich sequence at the 3' end of sense RNA strand

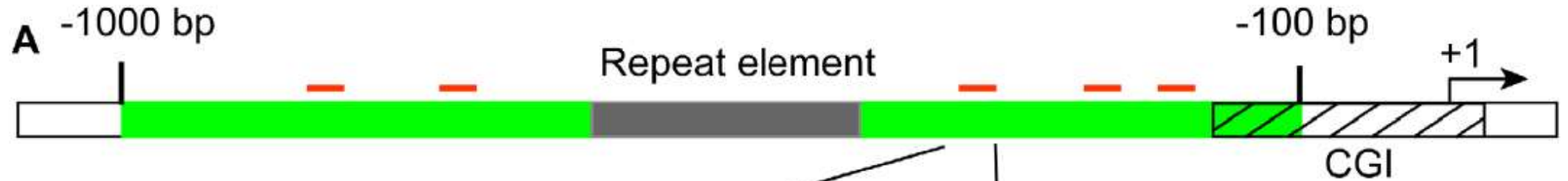
2) Duplex-related rules:

- a) Size of saRNA should be 19nt
- b) 2nt overhangs on the 3' ends of the RNA strands
- c) GC content between 40-60%
- d) Asymmetric thermodynamic stability in base pairing at the 5' end – calculate delta G of the first 4 base pairs at either end of the duplex by expanded nearest-neighbour model for RNA duplexes with Watson-Crick base pairing (Xia et al, 1998)

saRNA Design Rules with Scoring System

Rule	Criterion	Score
Distance from TSS	-100 to -1000	
CpG Islands	Avoid	
Target Size	19 nt	
Target GC Content	40-60%	
Consecutive nucleotides (≤ 5)	Avoid	
Simple repeats (di or tri-nucleotides)	Avoid	
CpG in target (≤ 1)	Avoid	
1 st and 2 nd nucleotides	Preferably G or C	
18 th nucleotide	A or T	
19 th nucleotide	A	
20-23 th nucleotide	Preferably A or T	
Asymmetric Thermodynamic Stability	Delta G of 4 bases on either end asymmetric	

A hypothetical promoter with a CpG island (CGI, hatched box) and a repeat element (grey box) are shown



B

1 19

agtg **CCA**ACTCATTCT**CCA**AGTA ataa DNA-S

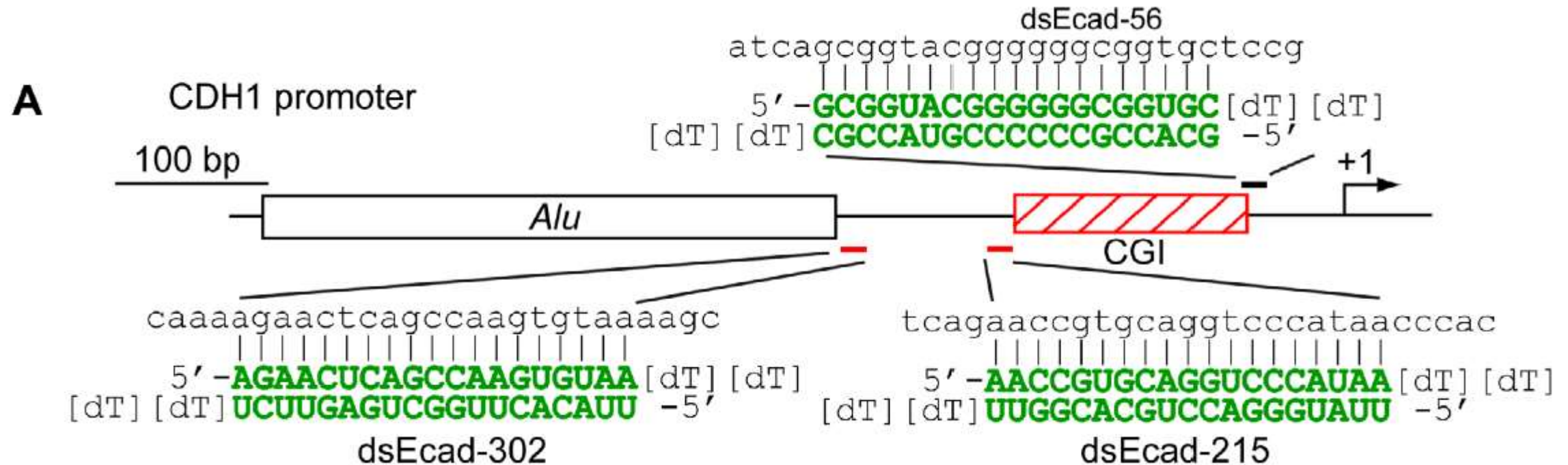
C

5' **CCA**ACUCAUUCUCA**AGUA** [dT] [dT] RNA-S

[dT] [dT] **GGUUG**AGUAAGAG**GUUCAU** 5' RNA-AS

-2.34 -1.02 $\Delta G^{\circ}37(\text{kcal/mol})$

Schematic representation of the CDH1 promoter



- Two saRNAs (marked in red) targeting sequences -302 (dsEcad-302) and -215 (dsEcad-215) relative to the TSS were generated according to design rules.
- dsEcad-56 (marked in black) deviated from the design criteria.