Supplementary Information

POST-TRANSCRIPTIONAL REGULATION BY DISTAL SHINE-DALGARNO SEQUENCES IN THE *GRPE-DNAK* INTERGENIC REGION OF *STREPTOCOCCUS MUTANS*

Sara R. Palmer and Robert A. Burne*

Department of Oral Biology, University of Florida, College of Dentistry,

Gainesville, Florida 32610 USA

Running title: Control of molecular chaperone expression

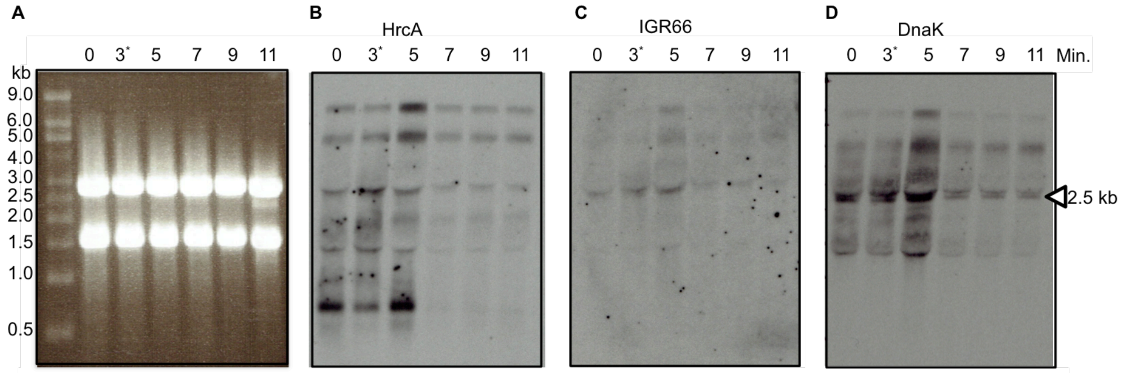
*Corresponding Author:
Department of Oral Biology
University of Florida
1395 Center Drive
P.O. Box 100424
Gainesville, FL 32610
Email: rburne@dental.ufl.edu

Table S1. Strains and Plasmids used in this study

Strains	**Source**	
UA159	*S. mutans* clinical isolate	Lab stock
SP50	pSP01-2 (Full length IGR66 pLDH and CAT	This study
SP51	pSP02-pLDH-CAT deletion of IGR	This study
SP52	pSP03 Δ5' loop (-244bp 5' end)	This study
SP53	pSP04 Δ5' loop (-214bp 5' end)	This study
SP54	pSP05 Δ81bp from 5' end	This study
SP55	pSP06 Δ48bp from 5' end	This study
SP56	pSP07 Δ5' loop (-301 bp from 5' end)	This study
SP57	pSP08 Δ3' loop (-229-350 bp from 3' end)	This study
SP58	pSP09 1-86 of 5' end intact (-86-350 bp from 3' end)	This study
SP59	pSP10 Partial deletion of 3' loop (-302-350bp of IGR 66)	This study
NC-RBS #1 & #2 G to C	NC-RBSs GAAGGAAUUAAGAAAGG to CACCAAATTAACAAACC	This study
NC-RBS #1 & #2 G to A	NC-RBSs GAAGGAAUUAAGAAAGG to AAAAAAAUUAAAAAAAA	This study
NC-RBS #1 G to A	NC-RBS #1 GAGG to AAAA	This study
NC-RBS #2 G to A	NC-RBS #2 GAAAGGA to AAAAAAA	This study
Δ 3'SL	Deletion of 3'SL bp 265-358 of *igr66*	This study
Plasmids		
pGEMT	Cloning vector	Promega
pJL105	*S. mutans* integration vector -*phnA-mtlA* locus (Kan^r, Spec^r and Erm^r)	Burne lab
pSP01	pJL105 with *Pldh-igr66-cat*	This study
pSP02	Q5-mutagenesis of pSP01 primers: SP31F-A and SP31R-B	This study
pSP03	Q5-mutagenesis of pSP01 primers: SP32F-E and SP31R-B	This study
pSP04	Q5-mutagenesis of pSP01 primers: SP34F-F and SP31R-B	This study
pSP05	Q5-mutagenesis of pSP01 primers: SP35F-G and SP31R-B	This study
pSP06	Q5-mutagenesis of pSP01 primers: SP37F-I and SP31R-B	This study
pSP07	Q5-mutagenesis of pSP01 primers: SP31F-A and SP38R-J	This study
pSP08	Q5-mutagenesis of pSP01 primers: SP35F-G and SP31R-B	This study
pSP09	Q5-mutagenesis of pSP01 primers: SP31F-A and SP36R-H	This study
pSP10	Q5-mutagenesis of pSP01 primers: SP31F-A and SP39R-K	This study

Table S2. Primers used in this study

Promoter IGR66 fusion to CAT		
SP27F	LDH promoter with BamHI site	AAA**GGATCC**TGTGACGGTAAGACCACCATT
SP27R	LDH promoter (IGR66 overhang)	GGACAAATTGTTAAGTATATATTCTATACATTTTCATT
SP28F	IGR66 with pLDH overhang	GAATATATACTTAACAATTTGTCCGAAACGACAGTAAA
SP28R	IGR66 with CAT overhang	ATTAAAGTTCATAATAATGTTACCAAGAACGCCGAAA
SP29F	CAT with IGR66 overhang	GGTAACATTATTATGAACTTTAATAAAATTGATTTAGACAA
SP29R	CAT with Sph1 site	AAA**GCATGC**TTATAAAAGCCAGTCATTAGGCCTA
IGR66 deletion constructs- Loop-out mutagenesis		
SP31F-A	SP51, SP57, SP58, SP59	CGGCGTTCTTGGTAACATTATT
SP31R-B	SP51-SP56	AAGTATATATTCTATACATTTTCATT
SP32R-D	SP57	TTTGCTATTTCCCTTAGTAGTG
SP32F-E	SP52	GAGGAATTAAGAAAGGATTTGAA
SP34F-F	SP53	GGGAAATAGCAAAAAAGAAAAATA
SP35F-G	SP54	ACTACAGCGATATTTCCCGAT
SP36R-H	SP58	GTAGTTTGCTTCGCTCACTGTA
SP37F-I	SP55	GAGCTCAGCTTTGCCTACAGT
SP38F-J	SP56	CTAGCACTTATTGAATGCGTCGT
SP39R- K	SP59	GCGGTTTATCTATTTTCCGAAGT
IGR66 NC-RBS mutation		
SP81F	NC-RBS G’s to C’s	ACAAACCATTTGAACCCGAACTAATAC
SP81R	NC-RBS G’s to C’s	TAATTGGTGTTTAAATATTTTTCTTTTTTGCTATTTCC
SP86F	NC-RBS G’s to A’s	AAAAAAAAATTTGAACCCGAACTAATAC
SP86R	NC-RBS G’s to A’s	AATTTTTTTTTAAATATTTTTCTTTTTTGCTATTTCC
SP101F	NC-RBS-1 G’s to A’s	AGAAAGGATTTGAACCCGAACTAATAC
SP102R	NC-RBS-1 G’s to A’s	TAATTTTTTTTTAAATATTTTTCTTTTTTGCTATTTCC
SP103F	NC-RBS-2 G’s to A’s	AAATTTGAACCCGAACTAATAC
SP103R	NC-RBS-2 G’s to A’s	TTTTTTAATTCCTCTTTAAATATTTTTCTTTTTTG
SP83F-new	Δ3’ SL	CTTGGTAACATTATTATGAACTTTAATAAAATTG
SP83R	Δ3’ SL	AAATCCTTTCTTAATTCCTCTTTAAATATTTTTC
SP80F	888-910-bp in *hrcA*	GGGTTTGGAATTTTGACCGTAA
SP80R	924-944-bp in *dnaK*	CGTGTTGAACCGCCGACTAA
SP105F	NC-RBS-2 G’s to A’s MM	GAGGAATTAAAAAAAAATTTGAACCCGAA
SP105R	NC-RBS-2 G’s to A’s MM	TTCGGGTTCAAATTTTTTTTTAATTCCTC
Real-time primers		
RT-Cat-F	RT-PCR of *cat*	GTTACAATAGCGACGGAGAGTT
RT-Cat	RT-PCR of *cat*	GGAGTCCAAATACCAGAGAATGT
RT-16S F	RT-PCR 16S RNA	ACGTGTGTGAGAGTGGAAAG
RT-16S R	RT-PCR 16S RNA	GCTCCCTTTACGCCCAATAA
RT-DnaK F	RT-PCR for DnaK	GAGCCTTTCTCGTGCTAAGT
RT-DnaK R	RT-PCR for DnaK	TGAACCGCCGACTAAGATAAC
Probes for Northern Blot		
IGR66 Fwd	IGR66-217 bp probe	GAGCTCAGCTTTGCCTACAGT
IGR66 Rev	IGR66-5’ Biotin label	5'Biotin-GTGCTAGGCGGTTTATCTATT
DnaK Fwd	DnaK for Northern probe	GTTTTGGCAACTGCTGGAG
DnaK Rev	DnaK for Northern Blot Probe	CAAGTGACAATGGCGTTACG
HrcA Fwd	HrcA for Northern probe	GTGATTACCCAGCGTCAAAAGGAT
HrcA Rev	HrcA for Northern probe	GCCAGTCAGATTCGCCAAAA
5’RACE primers		
dnaK_GSP1		TGCTGCAGCTGTTGGTTCGTT
dnaK_GSP2		GTTCTACTTCAAGACCAGCGAT
dnaK_GSP3		GCATCGTTAAAGTAAGCAGGAA



Cultures were placed in 37°C water bath and aliquots were removed at indicated times. * Indicates100 µg/ml rifampicin was added

Figure S1. Northern blot analysis of *dnaK* operon transcripts at 37°C.

A) Ethidium bromide stained agarose gel with 5 µg of total RNA loaded per lane. Northern analysis using a probe against (B) *hrcA*, (C) *igr66* or (D) *dnaK*. Samples were incubated at 37°C. Cells were collected at time 0, then at the 3’ time point rifampicin was added to a final concentration of 100 µg/ml. Aliquots were collected at the 3, 5, 7, 9 and 11 minute time points. RNA was prepared and Northern blots were performed as detailed in the methods section. The location of the 2.5 kb transcript recognized by the DnaK probe is indicated on the right of Panel D with an open arrow.

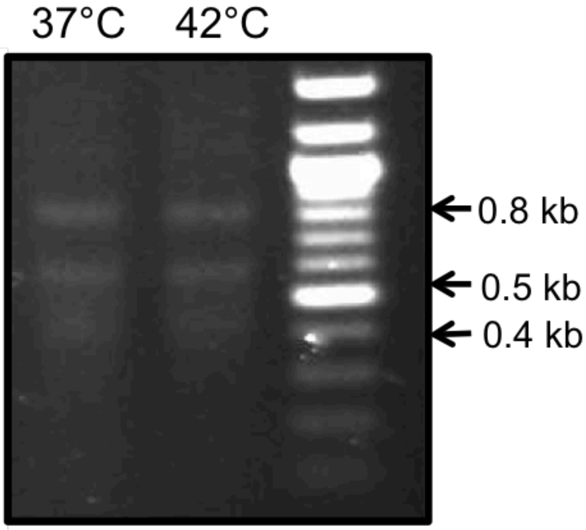


Figure S2. Agarose gel (1.3%) of 5’-RACE products using a *dnaK* gene-specific primer and RNA isolated from wild-type *S. mutans* UA159 grown at 37°C or heat shocked at 42°C for 3 minutes.

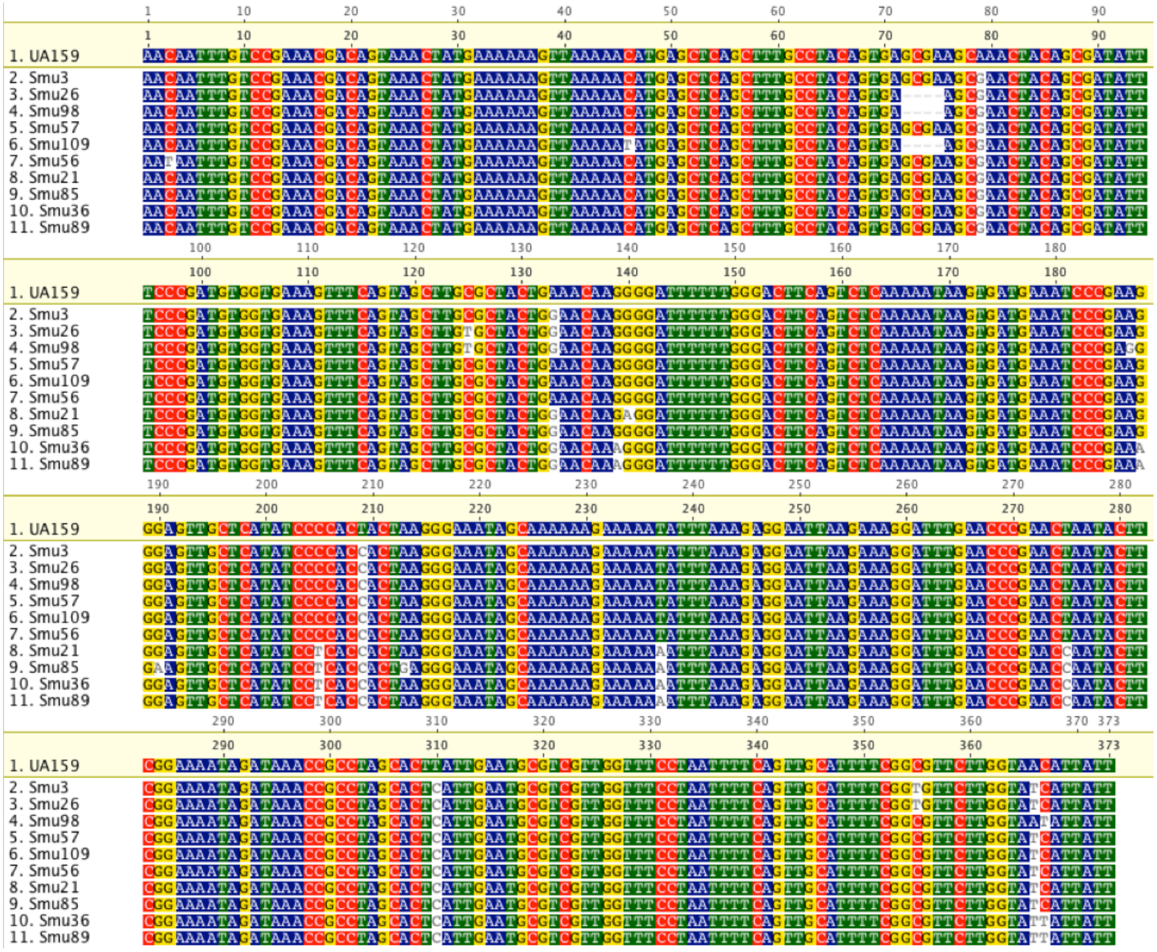


Figure S3. Sequence alignment of *grpE-dnaK intergenic* regions from genomically diverse *S. mutans* clinical isolates. Sequence alignment was made with the Geneious 8.0.5 program using the ClustalW algorithm. Sequence #1 is *S. mutans* UA159 and #2 through #11 correspond to selected isolates of *S. mutans* for which whole genome sequence is available (See Palmer *et al.* 2013 for details on isolates.)

Chromosomal integration of *cat* fusions into the mannitol PTS locus

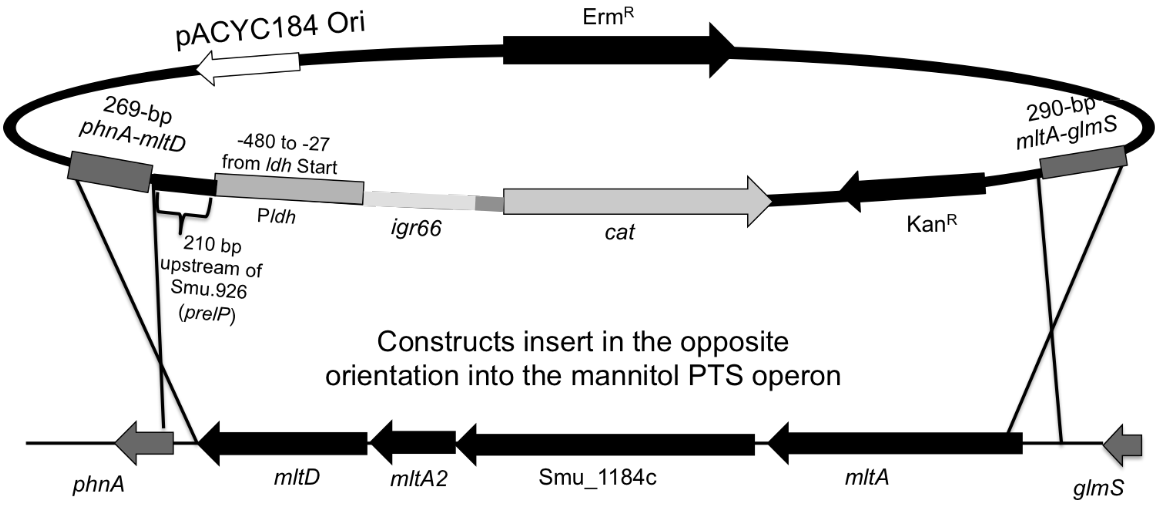


Figure S4. Schematic diagram of the vector used to integrate reporter-gene fusions in single copy in the *S. mutans* chromosome. Gene fusion-constructs were inserted into the chromosome by double cross-over recombination, replacing the mannitol PTS operon. See the methods section for more details.

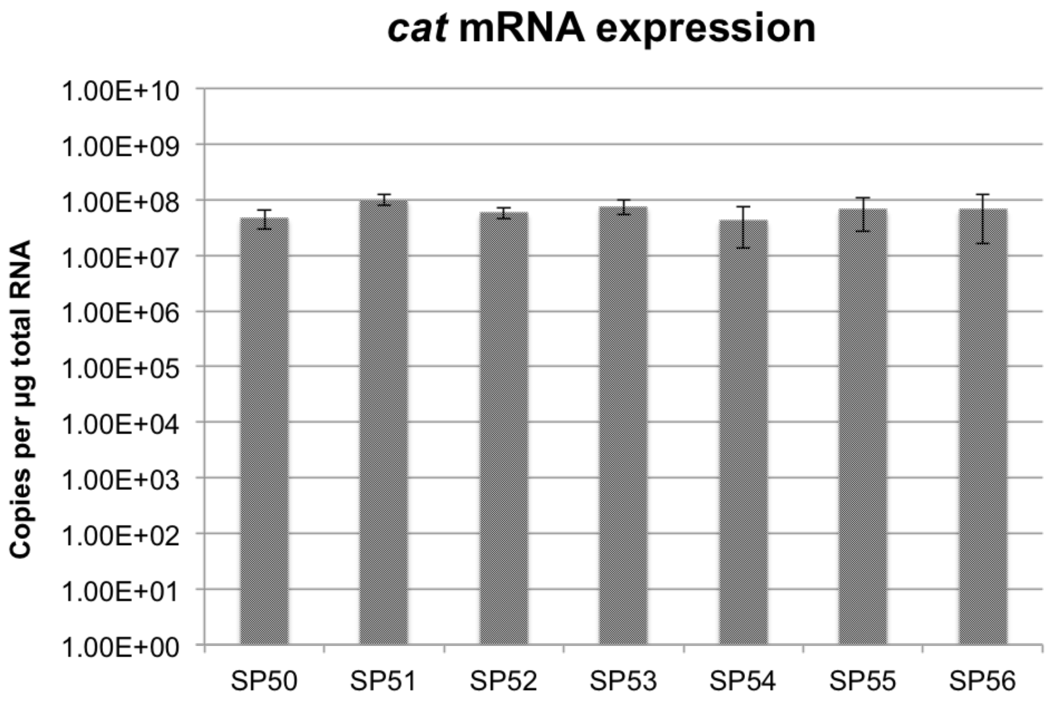


Figure S5. ***cat*** **mRNA expression in various constructs.** qRT-PCR results for *cat* gene mRNA expression in selected constructs, displayed as copies per µg of total RNA. See Figure 4 for methodological details.

Perfect RBS
5’AAGGAGGUGA 3’

3’AUUCCUCCACUAGGUUGGCG 5’ ***E. coli*** **K-12 16S rRNA 3’ terminus**
3’UCUUUCCUCCACUAGGUCGGC 5’ ***S. mutans*** **UA159 16S rRNA 3’ terminus**

Figure S6. The 3’ termini of the 16S rRNA of *E. coli* and *S. mutans* . The sequence of a canonical Shine-Dalgarno (SD) sequence is on top. The sequences of the 3’ termini of the 16S rRNAs from *E. coli* and *S. mutans*; the regions that are predicted to base pair with SD sequences in the mRNA are underlined.

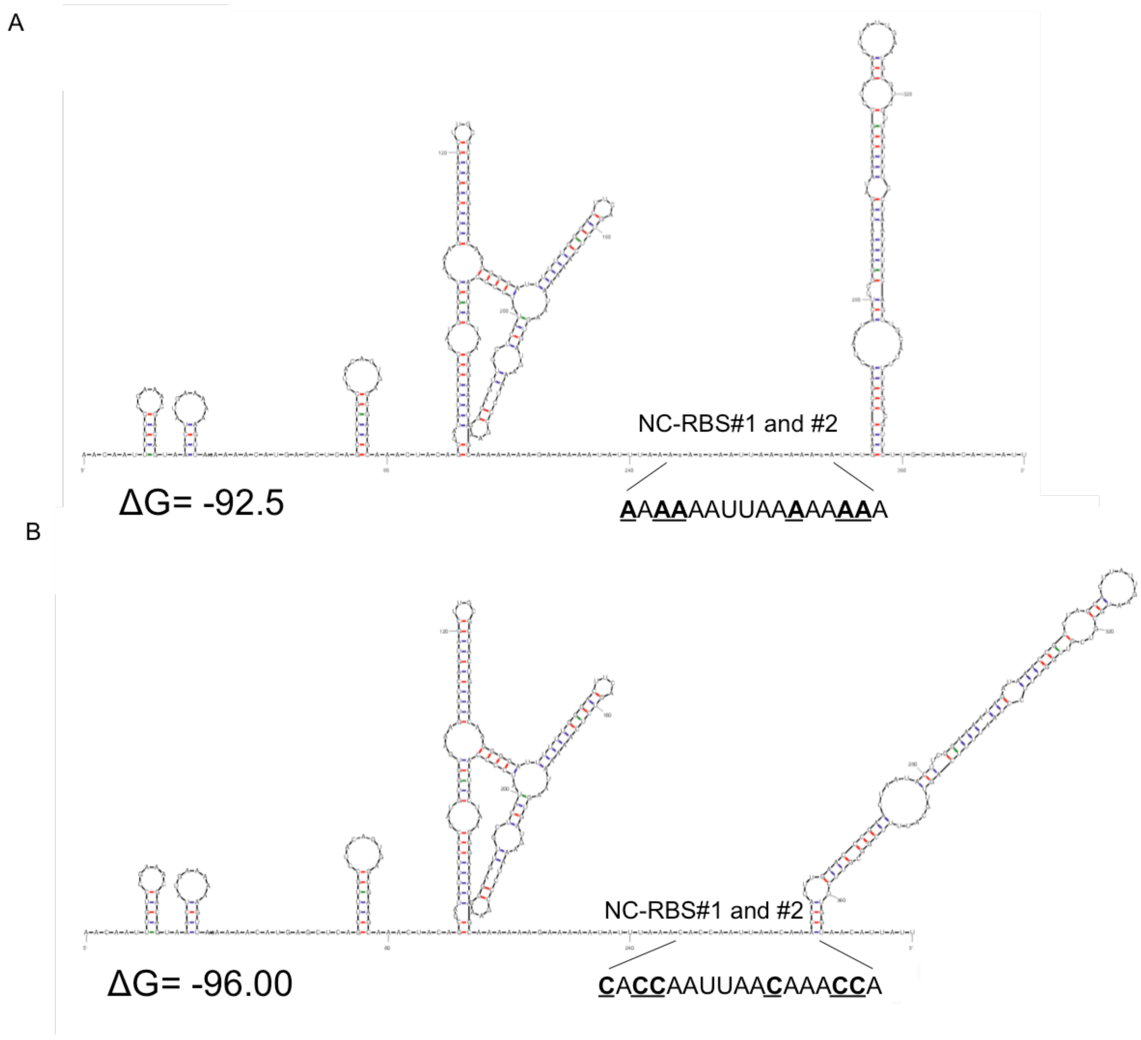


Figure S7. Predicted secondary structure of *igr66* point mutants. Secondary structures were predicted using the latest version of the mfold Web Server (http://mfold.rna.albany.edu). The predicted secondary structures shown are those that result from mutating the guanines in the NC-RBSs to adenines (Panel A) or cytosines (Panel B).

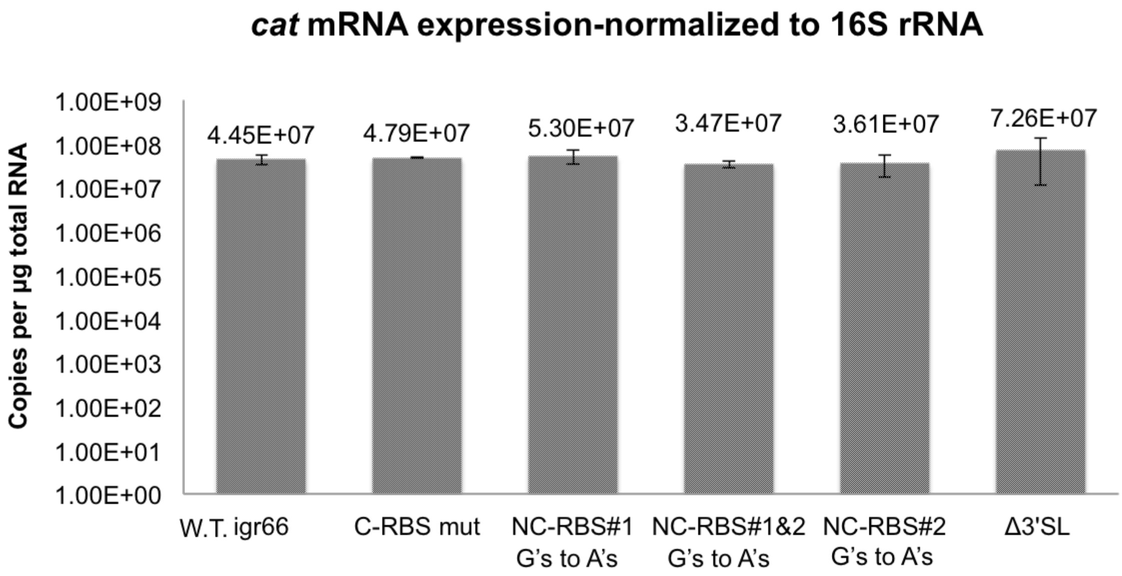


Figure S8. mRNA expressed from constructs containing *igr66* and mutant derivatives. qRT-PCR was performed to detect *cat* gene mRNA and the data are displayed as copies per µg of total RNA after normalization to 16S rRNA. See Figure 5 for details.

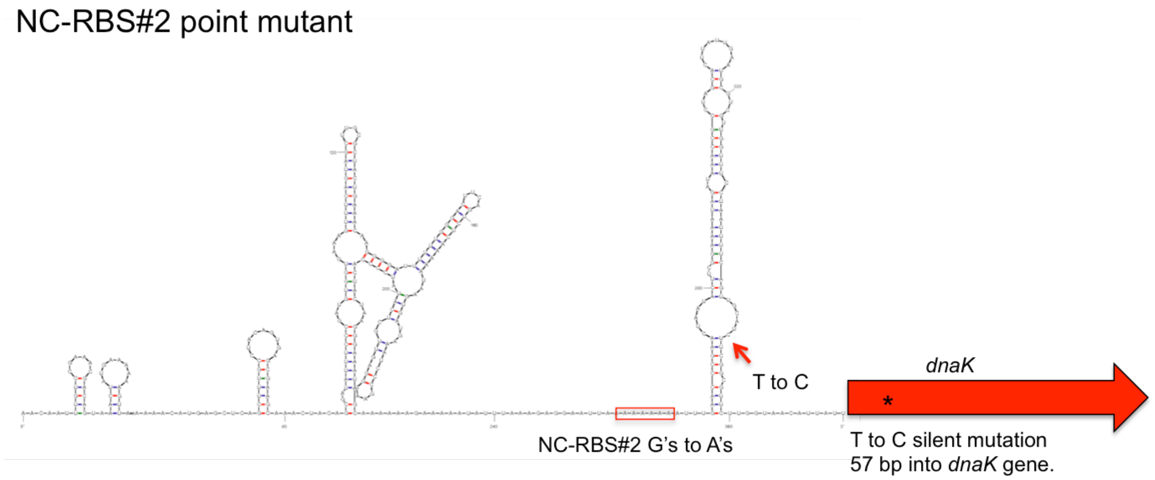
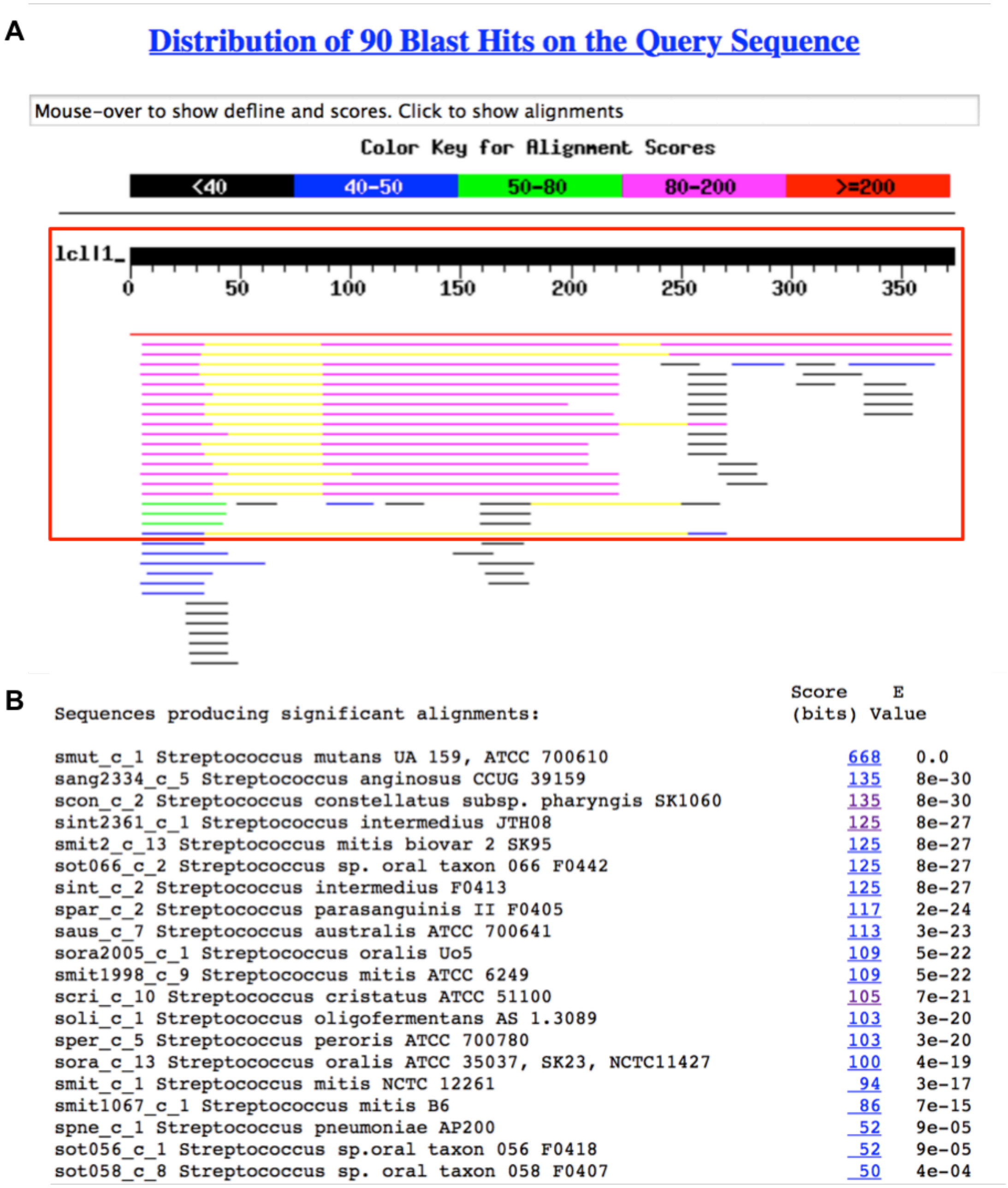
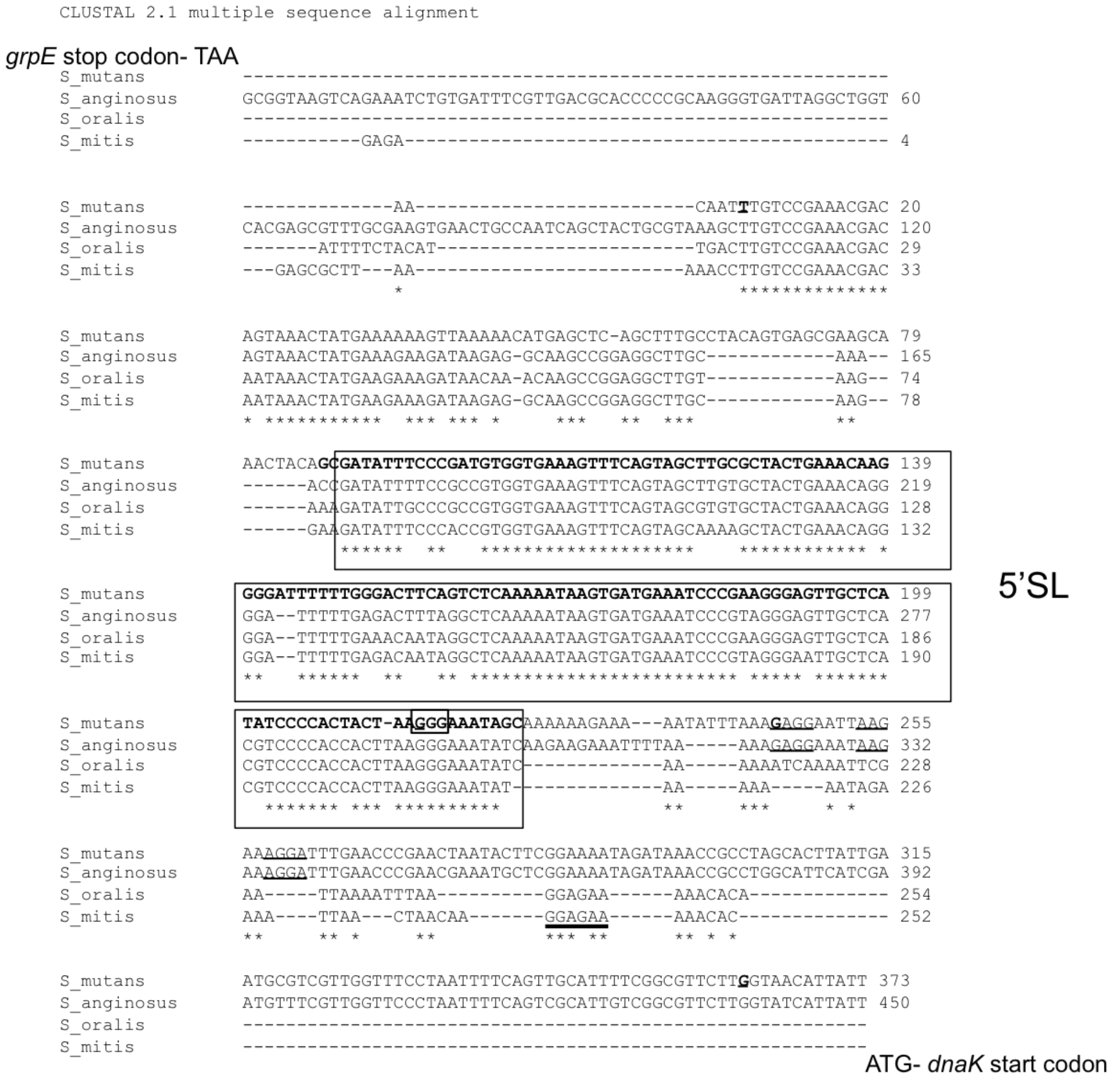


Figure S9. Location of point mutations in the NC-RBS#2 markerless mutant, and the resulting predicted secondary structure. A red box indicates the location of point mutations in NC-RBS #2. The location of additional mutations is also indicated, see text for details. The *dnaK* gene is shown as a red arrow. The construct was found to contain a silent point mutation at position 57 in the *dnaK* structural gene.



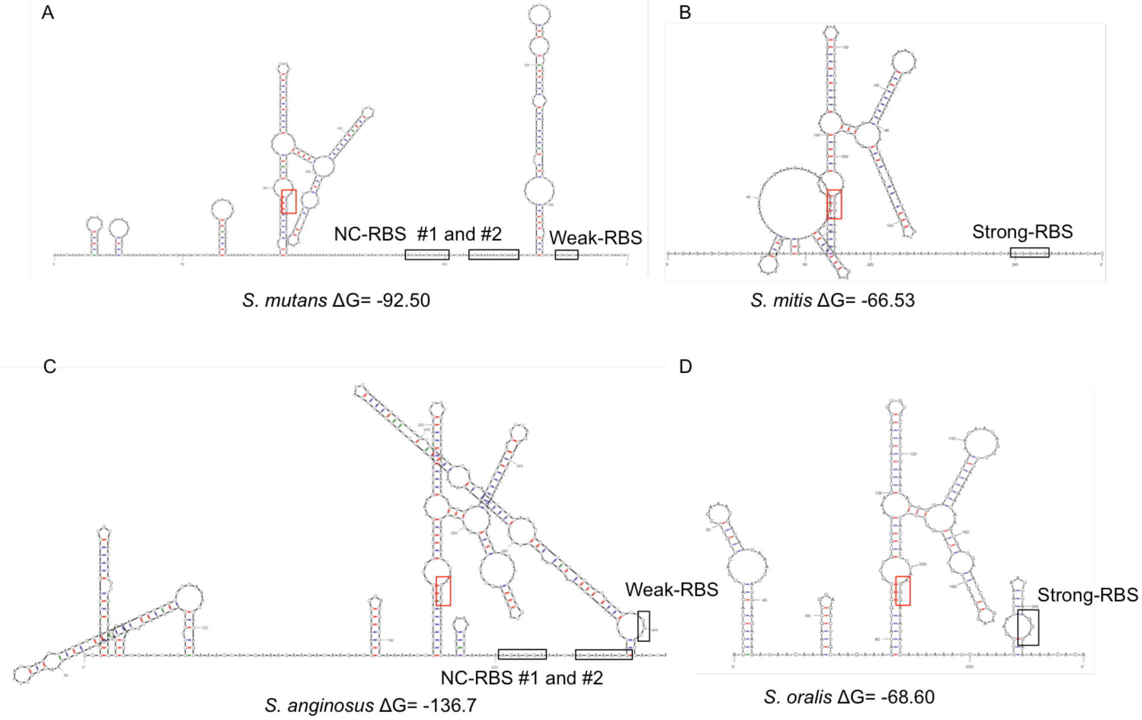
```
Sequences producing significant alignments:                         Score    E
                                                                    (bits) Value

smut_c_1 Streptococcus mutans UA 159, ATCC 700610                      668   0.0
sang2334_c_5 Streptococcus anginosus CCUG 39159                        135   8e-30
scon_c_2 Streptococcus constellatus subsp. pharyngis SK1060            135   8e-30
sint2361_c_1 Streptococcus intermedius JTH08                           125   8e-27
smit2_c_13 Streptococcus mitis biovar 2 SK95                           125   8e-27
sot066_c_2 Streptococcus sp. oral taxon 066 F0442                      125   8e-27
sint_c_2 Streptococcus intermedius F0413                               125   8e-27
spar_c_2 Streptococcus parasanguinis II F0405                          117   2e-24
saus_c_7 Streptococcus australis ATCC 700641                           113   3e-23
sora2005_c_1 Streptococcus oralis Uo5                                  109   5e-22
smit1998_c_9 Streptococcus mitis ATCC 6249                             109   5e-22
scri_c_10 Streptococcus cristatus ATCC 51100                           105   7e-21
soli_c_1 Streptococcus oligofermentans AS 1.3089                       103   3e-20
sper_c_5 Streptococcus peroris ATCC 700780                             103   3e-20
sora_c_13 Streptococcus oralis ATCC 35037, SK23, NCTC11427             100   4e-19
smit_c_1 Streptococcus mitis NCTC 12261                                 94   3e-17
smit1067_c_1 Streptococcus mitis B6                                     86   7e-15
spne_c_1 Streptococcus pneumoniae AP200                                 52   9e-05
sot056_c_1 Streptococcus sp.oral taxon 056 F0418                        52   9e-05
sot058_c_8 Streptococcus sp. oral taxon 058 F0407                       50   4e-04
```

Figure S10. Blastn results using *igr66* against Oral Microbial Genomes annotated at the Human Oral Microbiome Database (www.homd.org). (A) Multi-sequence alignment with areas of significant similarity indicated by a lines of the top 32 sequence hits. (B) The top 20 best hits with corresponding Blast scores and E-values. The HOMD database used contains 46,296 sequences; 1,195,875,138 total letters.



```
CLUSTAL 2.1 multiple sequence alignment

S_mutans         ------------------------------------------------------------
S_anginosus      GCGGTAAGTCAGAAATCTGTGATTTCGTTGACGCACCCCCGCAAGGGTGATTAGGCTGGT 60
S_oralis         ------------------------------------------------------------
S_mitis          -----------GAGA--------------------------------------------- 4


S_mutans         --------------AA--------------------------CAATTTGTCCGAAACGAC 20
S_anginosus      CACGAGCGTTTGCGAAGTGAACTGCCAATCAGCTACTGCGTAAAGCTTGTCCGAAACGAC 120
S_oralis         -------ATTTTCTACAT------------------------TGACTTGTCCGAAACGAC 29
S_mitis          ---GAGCGCTT---AA-------------------------AAACCTTGTCCGAAACGAC 33
                               *                               **************

S_mutans         AGTAAACTATGAAAAAAGTTAAAAACATGAGCTC-AGCTTTGCCTACAGTGAGCGAAGCA 79
S_anginosus      AGTAAACTATGAAAGAAGATAAGAG-GCAAGCCGGAGGCTTGC------------AAA-- 165
S_oralis         AATAAACTATGAAGAAAGATAACAA-ACAAGCCGGAGGCTTGT------------AAG-- 74
S_mitis          AATAAACTATGAAGAAAGATAAGAG-GCAAGCCGGAGGCTTGC------------AAG-- 78
                 * ***********  *** *** *     ***   **  ***             **

S_mutans         AACTACAGCGATATTTCCCGATGTGGTGAAAGTTTCAGTAGCTTGCGCTACTGAAACAAG 139
S_anginosus      ------ACCGATATTTTCCGCCGTGGTGAAAGTTTCAGTAGCTTGTGCTACTGAAACAGG 219
S_oralis         ------AAAGATATTGCCCGCCGTGGTGAAAGTTTCAGTAGCGTGTGCTACTGAAACAGG 128
S_mitis          ------GAAGATATTTCCCACCGTGGTGAAAGTTTCAGTAGCAAAAGCTACTGAAACAGG 132
                          ******  **   ********************    ************ *

S_mutans         GGGATTTTTTGGGACTTCAGTCTCAAAAATAAGTGATGAAATCCCGAAGGGAGTTGCTCA 199
S_anginosus      GGA--TTTTTGAGACTTTAGGCTCAAAAATAAGTGATGAAATCCCGTAGGGAGTTGCTCA 277
S_oralis         GGA--TTTTTGAAACAATAGGCTCAAAAATAAGTGATGAAATCCCGAAGGGAGTTGCTCA 186
S_mitis          GGA--TTTTTGAGACAATAGGCTCAAAAATAAGTGATGAAATCCCGTAGGGAATTGCTCA 190
                 **   ******  **   ** ************************* ***** *******

S_mutans         TATCCCCACTACT-AAGGGAAATAGCAAAAAAGAAA---AATATTTAAAGAGGAATTAAG 255
S_anginosus      CGTCCCCACCACTTAAGGGAAATATCAAGAAGAAATTTTAA-----AAAGAGGAAATAAG 332
S_oralis         CGTCCCCACCACTTAAGGGAAATATC-------------AA-----AAAATCAAAATTCG 228
S_mitis          CGTCCCCACCACTTAAGGGAAATAT--------------AA-----AAA-----AATAGA 226
                   ******* *** **********               **     ***     * *

S_mutans         AAAGGATTTGAACCCGAACTAATACTTCGGAAAATAGATAAACCGCCTAGCACTTATTGA 315
S_anginosus      AAAGGATTTGAACCCGAACGAAATGCTCGGAAAATAGATAAACCGCCTGGCATTCATCGA 392
S_oralis         AA-----TTAAAATTTAA----------GGAGAA------AAACACA------------- 254
S_mitis          AAA----TTAA---CTAACAA-------GGAGAA------AAACAC-------------- 252
                 **     ** *     **          *** **      ** * *

S_mutans         ATGCGTCGTTGGTTTCCTAATTTTCAGTTGCATTTTCGGCGTTCTTGGTAACATTATT 373
S_anginosus      ATGTTTCGTTGGTTCCCTAATTTTCAGTCGCATTGTCGGCGTTCTTGGTATCATTATT 450
S_oralis         ----------------------------------------------------------
S_mitis          ----------------------------------------------------------
```

Figure S11. ClustalW (Version 2.1) of *grpE-dnaK* intergenic regions from streptococci that share the highest degree of conservation with *S. mutans* UA159 *igr66*. Bold sequences indicate the location of predicted secondary structures (See Figure 4). The box indicates the location of highly conserved sequences in the predicted 5’-SL of *igr66* from *S. mutans.*



S. mutans ΔG= -92.50

S. mitis ΔG= -66.53

S. anginosus ΔG= -136.7

S. oralis ΔG= -68.60

Figure S12. Predicted secondary structures of the *grpE-dnaK* intergenic regions from selected streptococci. Predicted structures were determined using the latest version of the mfold Web Server (http://mfold.rna.albany.edu). Predicted secondary structures of the *grpE-dnaK* intergenic regions from: *S. mutans* (A*), S. mitis* (B), *S. anginosus* (C), and *S. oralis* (D). The location of a strong RBS, weak RBS, and/or NC-RBS is indicated for each predicted structure in black boxes. The location of a processing site in *S. mutans igr66* that was identified using 5'RACE is highlighted with a red box. See text for additional details.