Additional File 8: Hairpin loci in *C. elegans* obtained by the filtering protocol "Similar"

List of 41 identified loci in *C. elegans* of genomic hairpins with *L* score >= $1e^{-5}$, not overlapping with exons, having similarity to a known metazoan mature miRNA (at most 3 three mismatches on 19 nt length) and binary filtered on seven individual descriptors: four based on structure (stem length <= 55, loop length <= 40, largest bulge <= 8, max match count >= 17) and three based on sequence complexity (polyNucHairpin <= 8, SCS-mono >= -10, SCS-di <= 0.40). Except for stem length, all individual thresholds fall within the *S*<1 fractions of the descriptors and separately reject at most three *C. elegans* miRNA hairpins. After removal of genomic hairpins that represent known miRNA hairpins, 64 loci remain.

The 84 hairpins of these 64 loci are further filtered for the position of the similarity within the hairpin. The similarity with the mature miRNA may have at most 3 nt overlap with the hairpin loop coordinates and at least 8 nt of stem must separate the similarity from the end of the hairpin. These criteria are set according to the biological model of miRNA maturation from hairpins (*Zeng et al. 2005*). This final filtering step reduced the list to 51 genomic hairpins on 41 loci. Below, the hairpin with the highest *L* score for each locus is given. *L* scores are obtained from the scoring model *Metazoa*.

The 18 descriptors in the figures, from left to right: MFEahl, MFEahl index, Q, max match count, bulgeRatio, GU-match contribution, largest bulge, longest match-stretch, looplength, stem length, dP, SCS-mono, SCS-di, polyA, polyU, polyNucHairpin, GsurplusC, GasurplusCU.

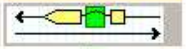
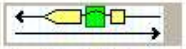
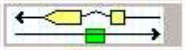
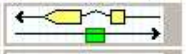
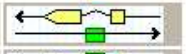
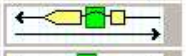
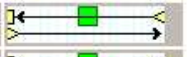
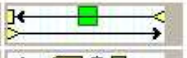
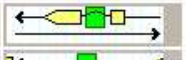
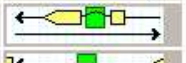
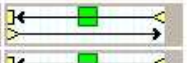
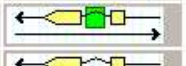
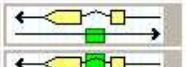
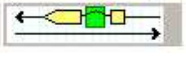
‘Genpos’ denotes the genomic position of the hairpin:

 intergenic

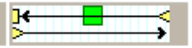
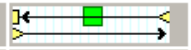
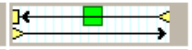
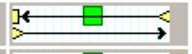
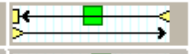
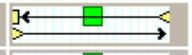
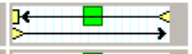
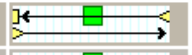
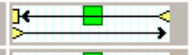
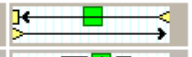
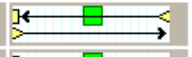
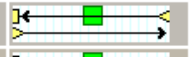
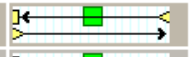
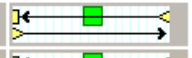
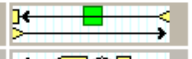
 intronic, on the same strand as the exons

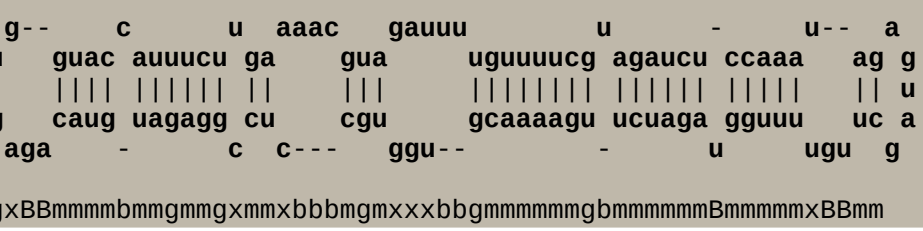
 intronic but opposite to the strand of the exons

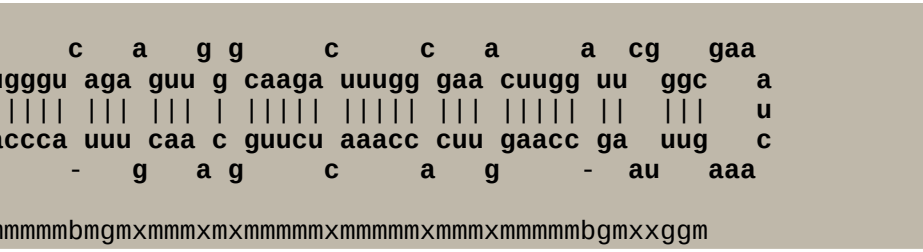
Highest scoring hairpins from the loci obtained by the filtering protocol “Similar” on chromosomes I-III

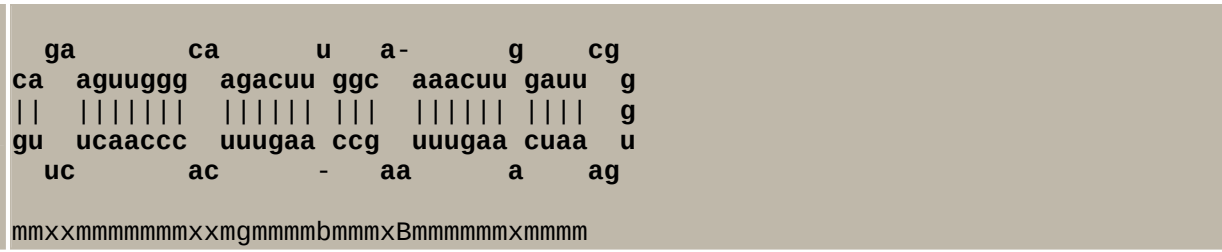
Id	Position	Strand	Lscore	Genpos
1652826	I 103237..103347	-	1.1e-04	
1165306	I 1733470..1733572	+	0.030	
1760974	I 1733608..1733684	-	0.044	
1184736	I 3531370..3531472	+	3.7e-05	
1280658	I 9372758..9372852	+	0.089	
4284206	I 10325772..10325844	+	2.9e-04	
1715134	I 11708848..11708926	-	2.6e-05	
1446507	II 2633349..2633441	+	1.5e-05	
1891777	II 9403659..9403769	+	6.7e-04	
1959550	II 11841755..11841841	+	2.0e-04	
2004295	II 12609234..12609323	-	0.048	
1991116	II 12941938..12942051	+	4.9e-05	
2043412	II 13173538..13173633	+	0.102	
2064086	II 13502012..13502095	+	1.8e-04	
2047661	II 13515555..13515672	+	7.2e-03	
1976550	II 15256703..15256813	+	9.5e-05	
4515471	III 1101498..1101616	+	2.4e-03	
4546124	III 2916984..2917057	+	0.016	
2229988	III 3596986..3597083	+	0.011	
2196306	III 3876769..3876860	+	0.023	
2414207	III 10524895..10524964	+	0.537	

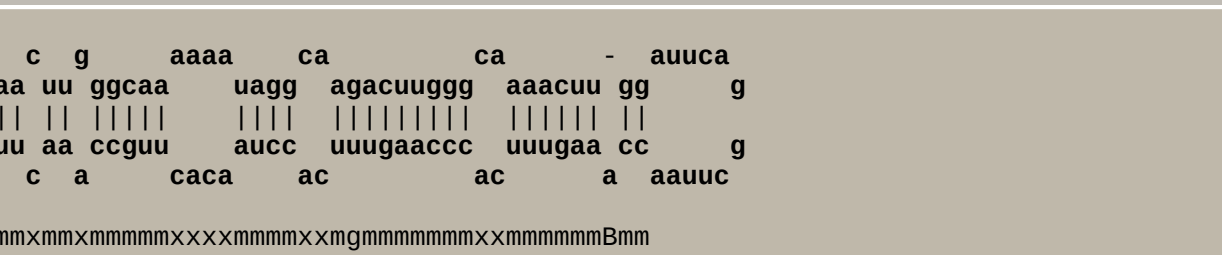
Highest scoring hairpins from the loci obtained by the filtering protocol “Similar” on chromosomes IV-X

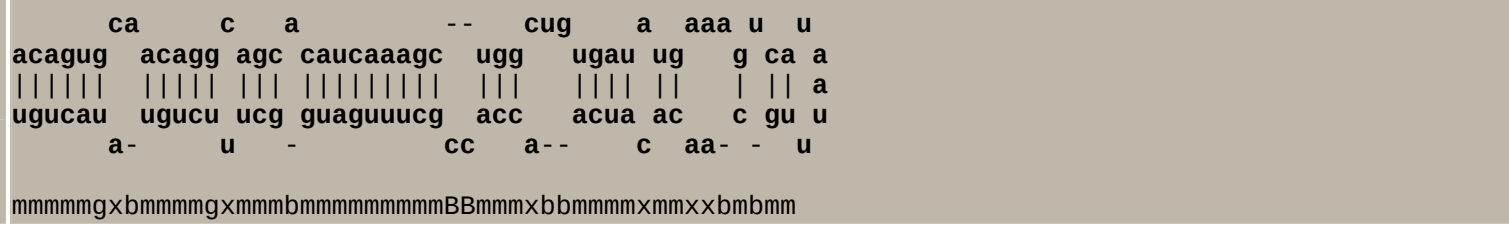
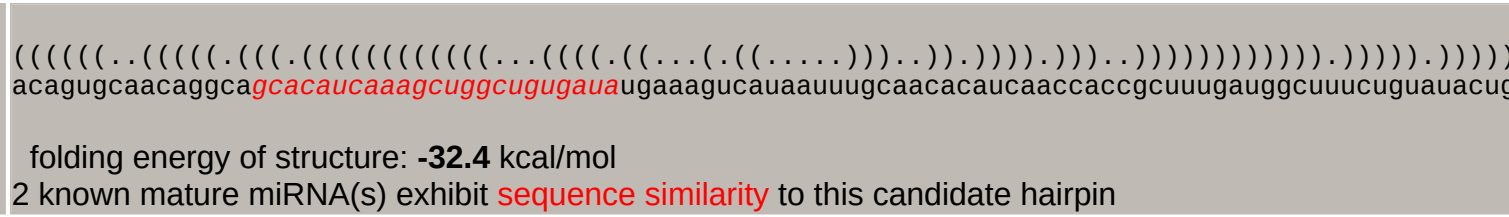
Id	Position	Strand	Lscore	Genpos
2490824	IV 245875..245960	+	0.051	
2875398	IV 12806506..12806596	+	0.018	
2981916	IV 15168041..15168155	+	5.6e-03	
3083314	V 600115..600211	-	8.2e-03	
3130412	V 1304648..1304762	+	6.5e-03	
3148670	V 2851038..2851136	+	0.097	
3179049	V 2851151..2851221	-	0.035	
3240609	V 4703559..4703654	-	5.7e-05	
3227648	V 4844054..4844172	+	1.7e-05	
3411064	V 11066747..11066854	-	0.102	
3450047	V 12302101..12302170	-	2.8e-05	
3561947	V 16348305..16348419	+	0.015	
3714507	V 20590843..20590916	-	2.2e-03	
3729314	X 539532..539646	-	4.4e-04	
3731460	X 624955..625048	-	1.1e-04	
3846390	X 3912837..3912933	-	9.3e-03	
4048106	X 11051158..11051250	+	3.5e-04	
4147031	X 14917325..14917447	+	7.2e-05	
4167805	X 14917516..14917627	-	0.035	
4181655	X 15447707..15447814	-	2.7e-03	

id	1652826
genomic_position	103236-103347 on CEL150_I (-) in Caenorhabditis elegans
L score	0.000113029645737
structure	
sequence	((((.((((.((((((.((....(((.....(((((((.(((((((((((.((.....)).)))))).)))))))))))))...)))).)).)))))))))...)))) gcaugguaccauuucuug*aaaacguagauuuuguuuu*cguagaucuccaaauaguagaucuuguuuugguagaucuugaaaacguggugccu ccggagauguacagagugc folding energy of structure: **-32.8** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

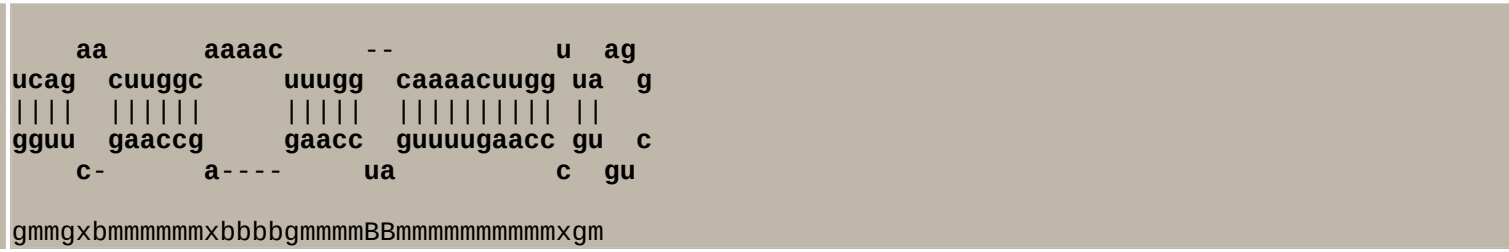
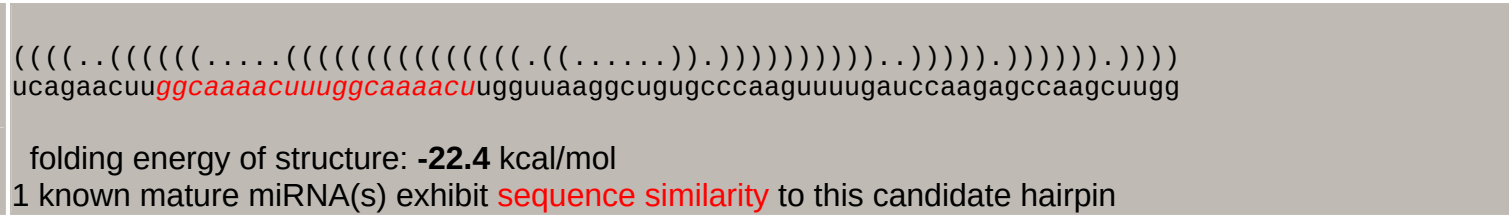
id	1165306
genomic_position	1733469-1733572 on CEL150_I (+) in Caenorhabditis elegans
L score	0.0300871112455
structure	
sequence	(((((((((.(((.(((.(.(((((.(((((.(((.(((((.(((((.((..(((.........))))..)))))))).))).)))))).)))))).).))).))))))))))) acuugggucagaaguug*ggcaagacuuuggcgaaacu*uggauucgggcgaaaucaaaguuuaagccaagguucaccaaacucuuggcaaacguu uacccaagu folding energy of structure: **-40.6** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	1760974
genomic_position	1733607-1733684 on CEL150_I (-) in Caenorhabditis elegans
L score	0.0442810979858
structure	
sequence	((..(((((((..((((((.(((.((((((.((((.......)))).))))))..)))))))))..)))))))..)) cagaaguu*gggcaagacuuuggcaaaacu*uggauucgggugaaaucaaaguuuaagccaaguuucacccaacucuug folding energy of structure: **-27.2** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

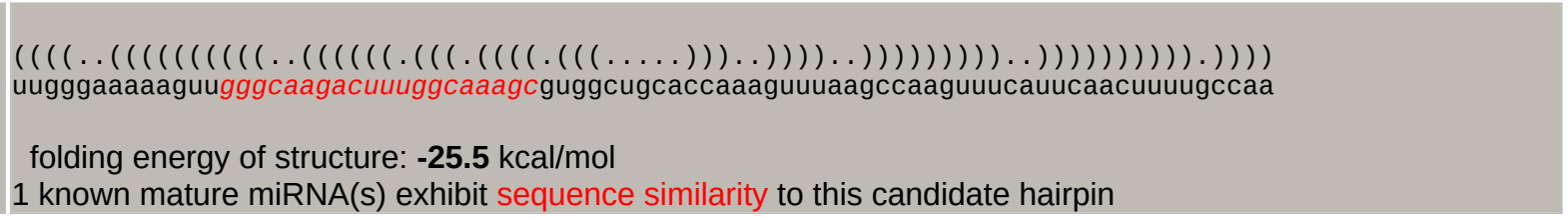
id	1184736
genomic_position	3531369-3531472 on CEL150_I (+) in Caenorhabditis elegans
L score	3.69668875878e-05
structure	
sequence	(((((((.((.((((((....((((..(((((((((..(((((((.............)).)))))))..)))))))))..))))....))))))..)).))))))) ggcuaaacuugggcaaaaaau*aggcaagacuugggcaaaacu*uggauucaggcuuaaccaaaguuucacccaaguuucaccuaacacuugccaaacuuuggcc folding energy of structure: **-38.4** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

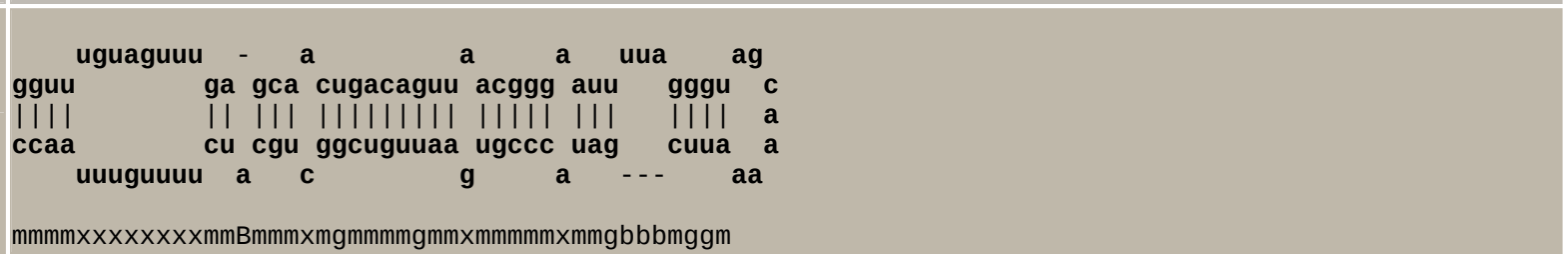
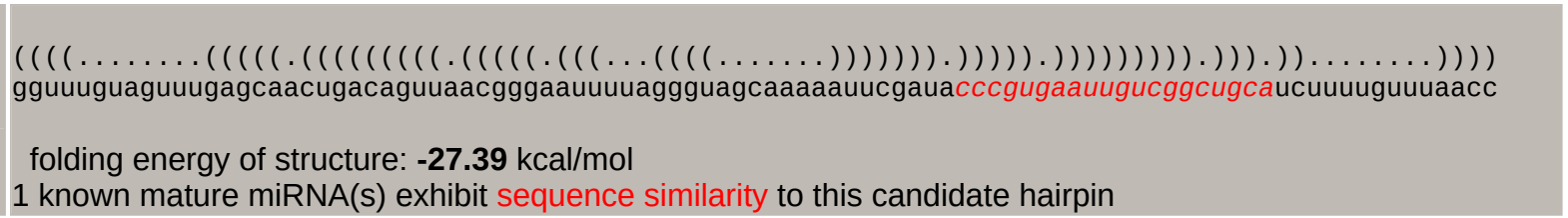
id	1280658
genomic_position	9372757-9372852 on CEL150_I (+) in Caenorhabditis elegans
L score	0.0892294092686
structure	
sequence	

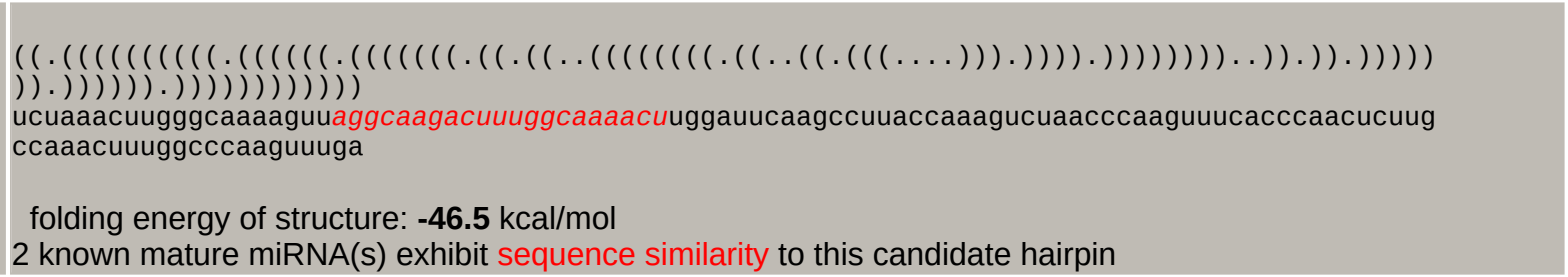
folding energy of structure: **-32.4** kcal/mol
2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	4284206
genomic_position	10325771-10325844 on CEL150_I (+) in Caenorhabditis elegans
L score	0.000290946981613
structure	
sequence	

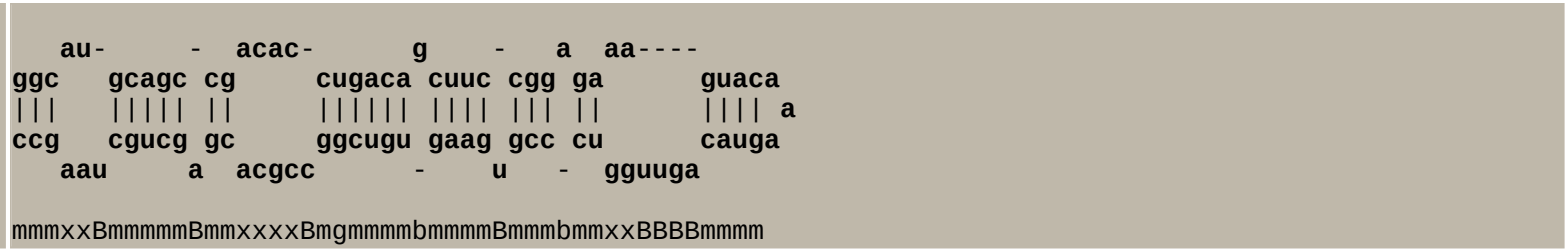
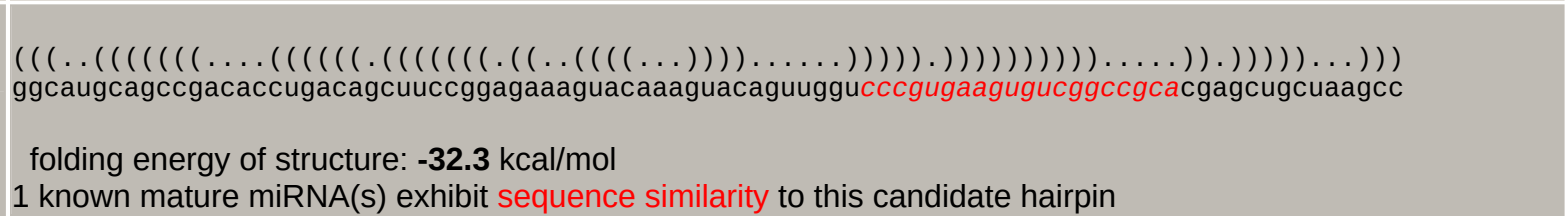
folding energy of structure: **-22.4** kcal/mol
1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	1715134
genomic_position	11708847-11708926 on CEL150_I (-) in Caenorhabditis elegans
L score	2.61725789981e-05
structure	
sequence	 folding energy of structure: **-25.5** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

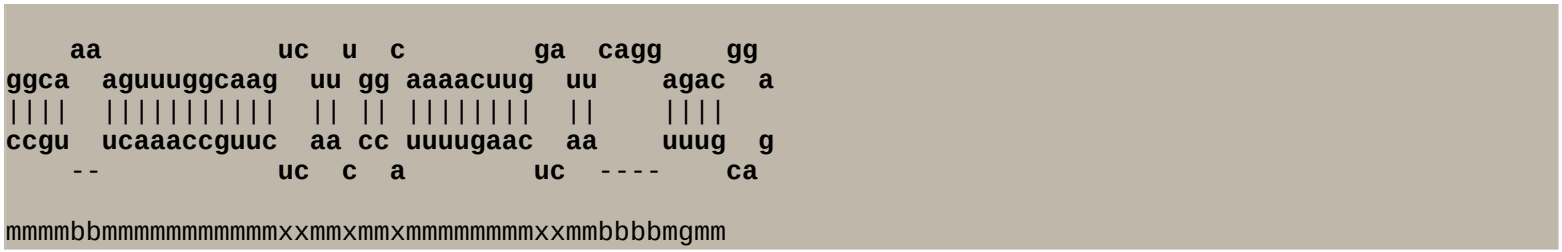
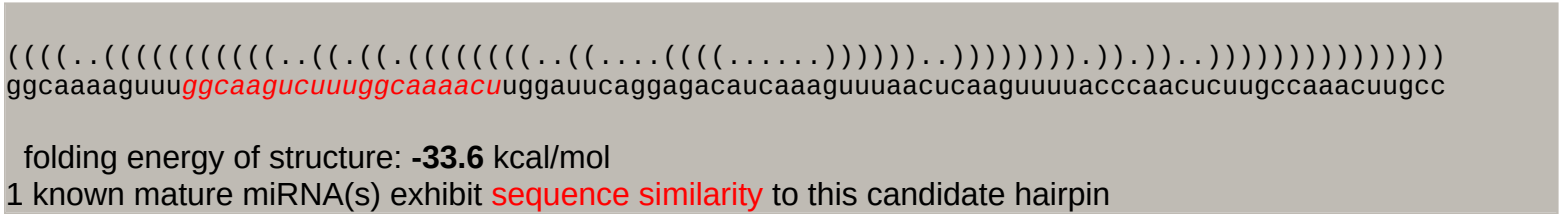
id	1446507
genomic_position	2633348-2633441 on CEL150_II (+) in Caenorhabditis elegans
L score	1.4530952118e-05
structure	
sequence	 folding energy of structure: **-27.39** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

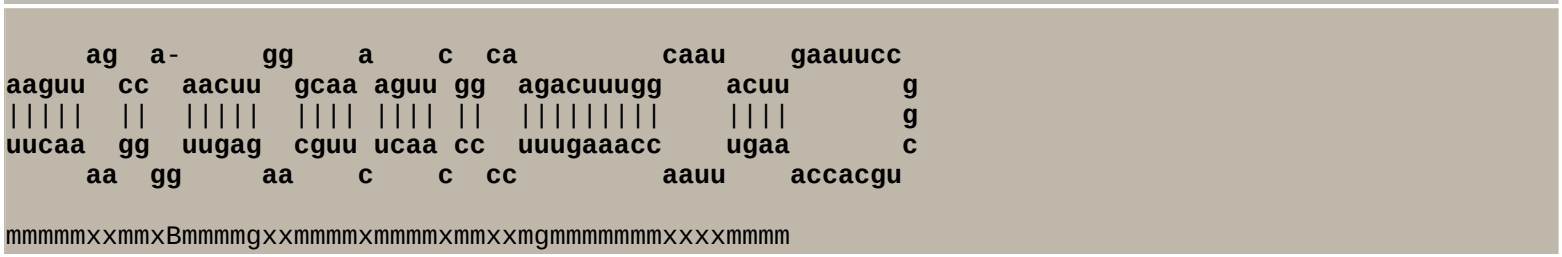
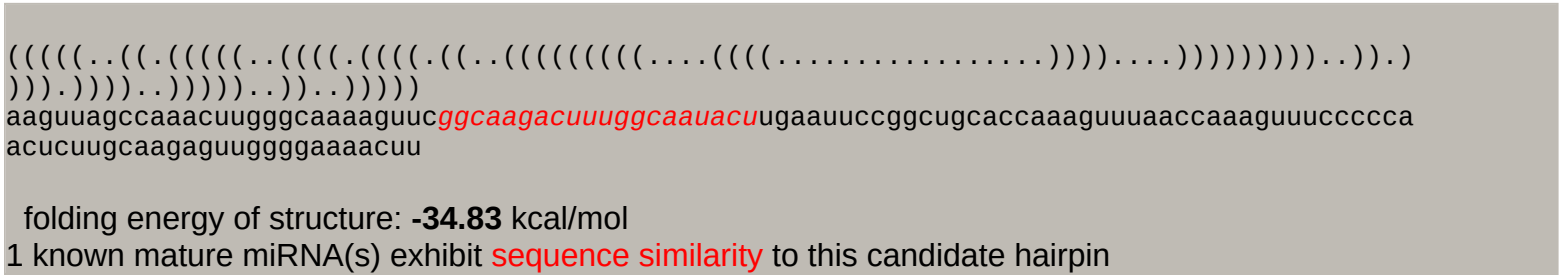
id	1891777
genomic_position	9403658-9403769 on CEL150_II (+) in Caenorhabditis elegans
L score	0.000671559878891
structure	
sequence	

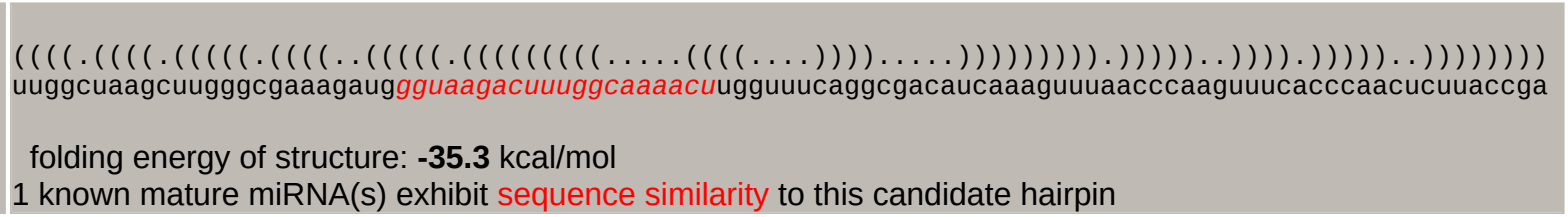
folding energy of structure: **-46.5** kcal/mol
2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

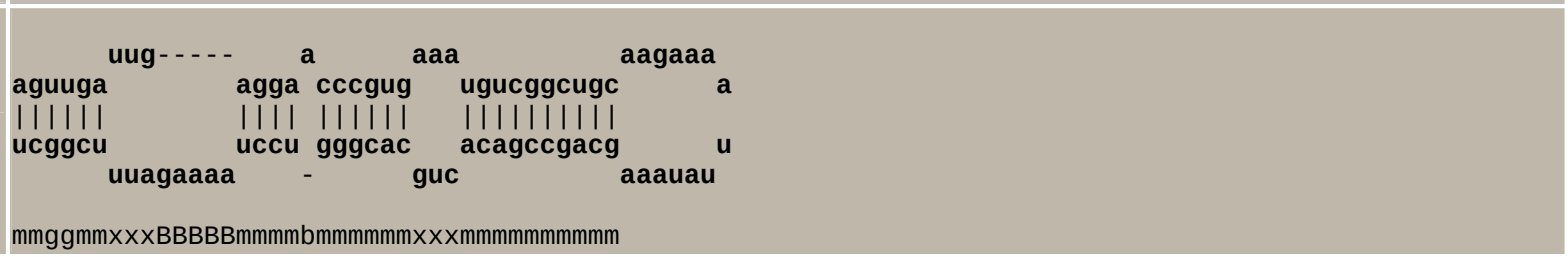
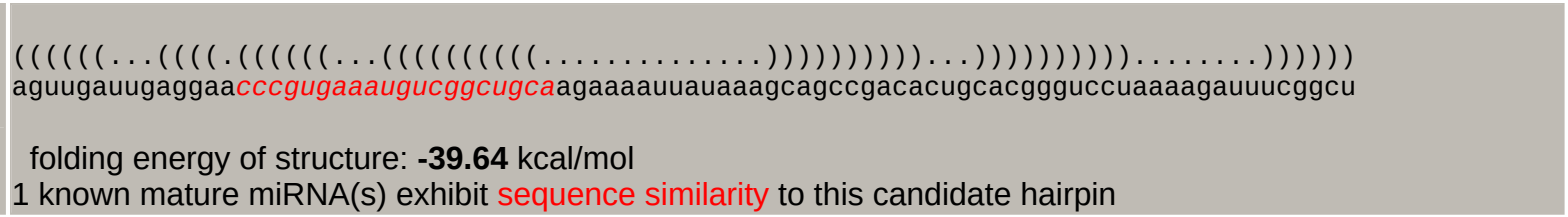
id	1959550
genomic_position	11841754-11841841 on CEL150_II (+) in Caenorhabditis elegans
L score	0.000204673248929
structure	
sequence	

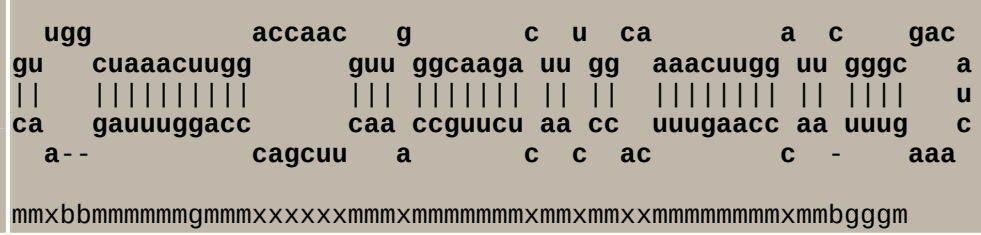
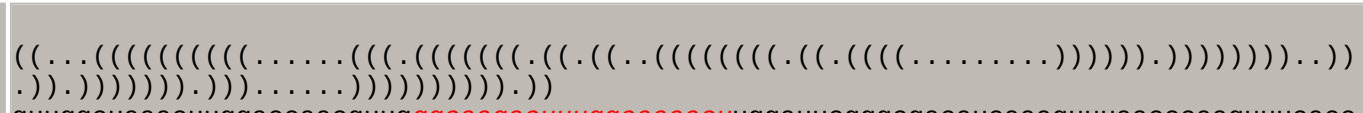
folding energy of structure: **-32.3** kcal/mol
1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

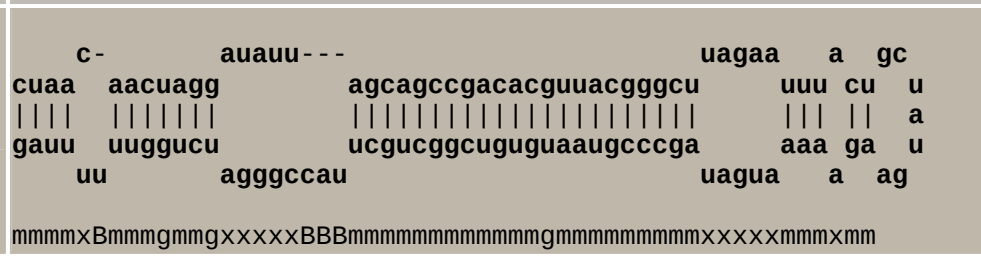
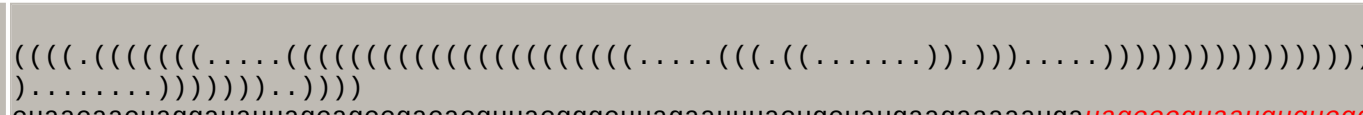
id	2004295
genomic_position	12609233-12609323 on CEL150_II (-) in Caenorhabditis elegans
L score	0.0481020993305
structure	
sequence	

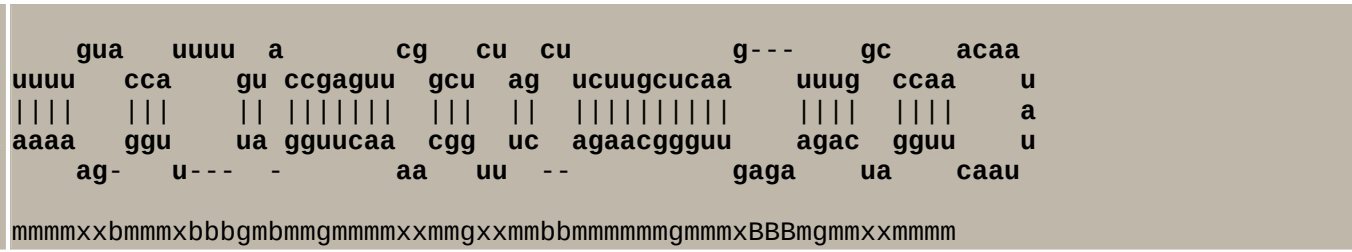
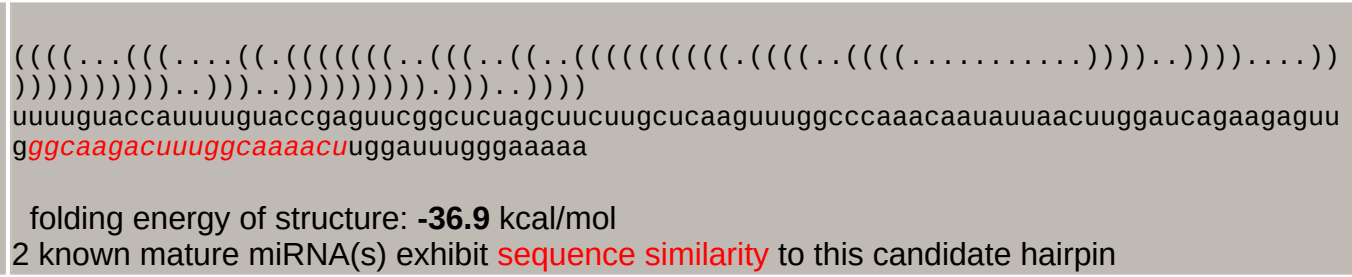
id	1991116
genomic_position	12941937-12942051 on CEL150_II (+) in Caenorhabditis elegans
L score	4.86890627252e-05
structure	
sequence	

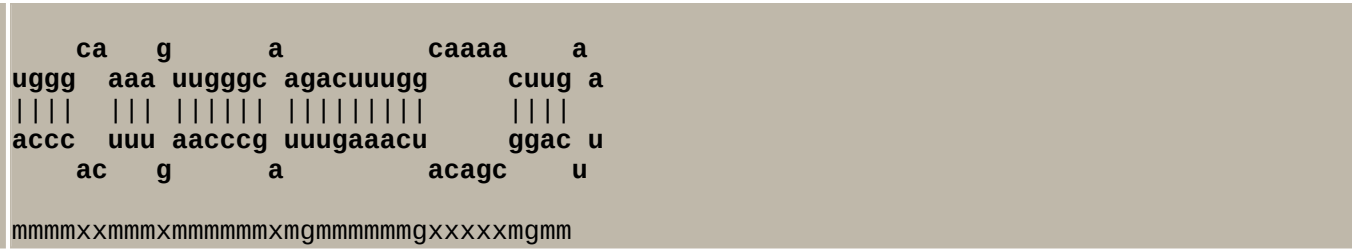
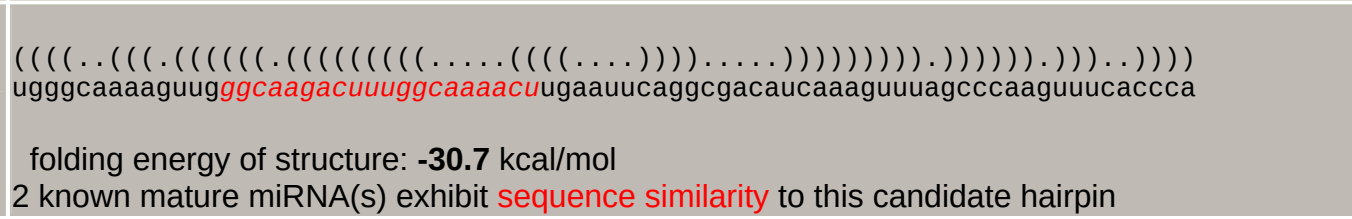
id	2043412
genomic_position	13173537-13173633 on CEL150_II (+) in Caenorhabditis elegans
L score	0.102327407922
structure	
sequence	

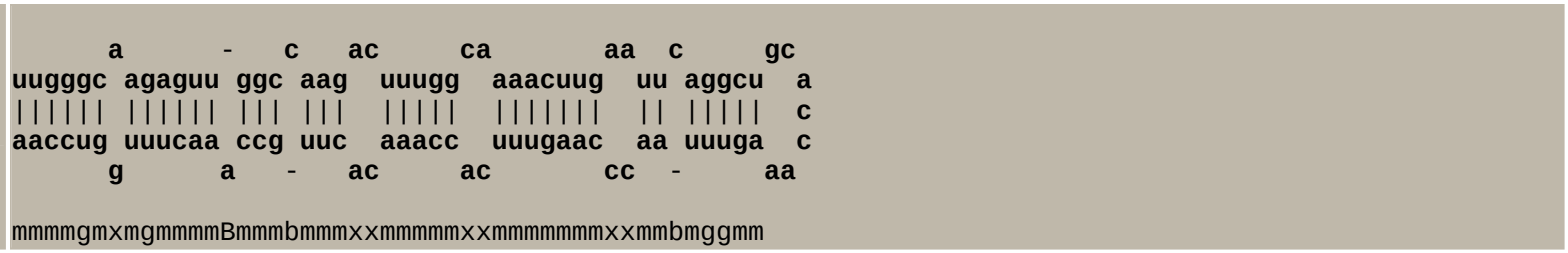
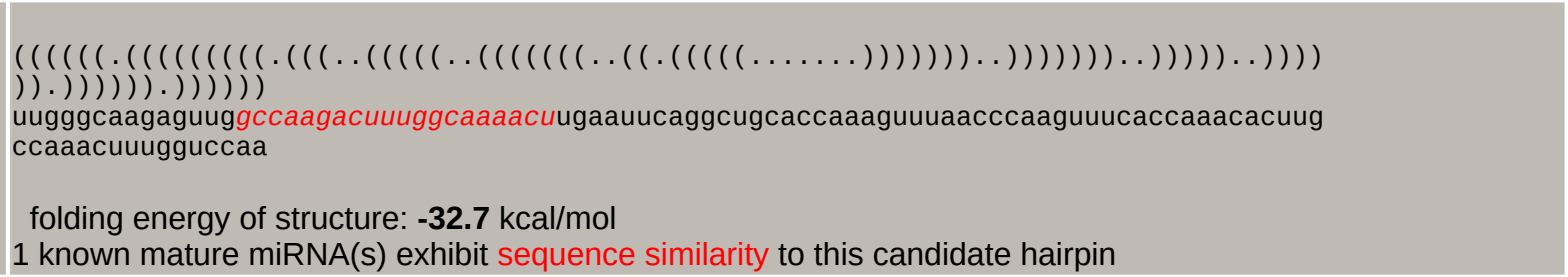
id	2064086
genomic_position	13502011-13502095 on CEL150_II (+) in Caenorhabditis elegans
L score	0.000178408489523
structure	
sequence	

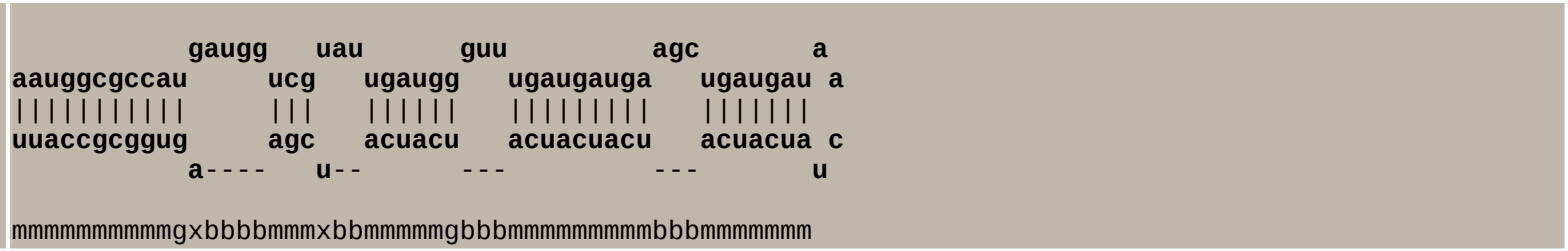
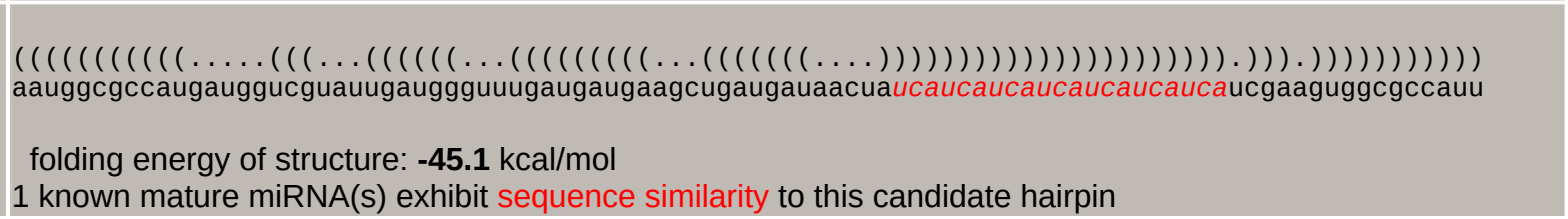
id	2047661
genomic_position	13515554-13515672 on CEL150_II (+) in Caenorhabditis elegans
L score	0.00717506840493
structure	
sequence	 guuggcuaaacuuggaccaacguug*ggcaagacuuuggcaaaacu*uggauucgggcgacaucaaaguuuaacccaaguuucacc caacucuugccaaacuucgacccagguuuagaac folding energy of structure: **-39.0** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

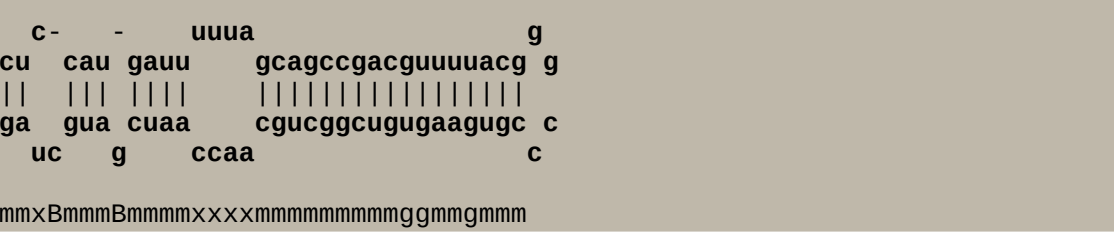
id	1976550
genomic_position	15256702-15256813 on CEL150_II (+) in Caenorhabditis elegans
L score	9.45151064191e-05
structure	
sequence	 cuaacaacuaggauauuagcagccgacacguuacgggcuuagaauuuacugcuaugaagaaaaauga*uagcccguaaugugucggcug*c uuaccgggaucuggguuuuuag folding energy of structure: **-49.92** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

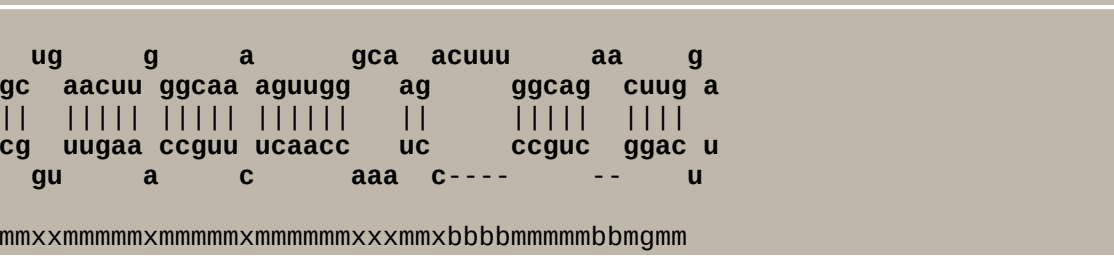
id	4515471
genomic_position	1101497-1101616 on CEL150_III (+) in Caenorhabditis elegans
L score	0.00242345988778
structure	
sequence	 folding energy of structure: **-36.9** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

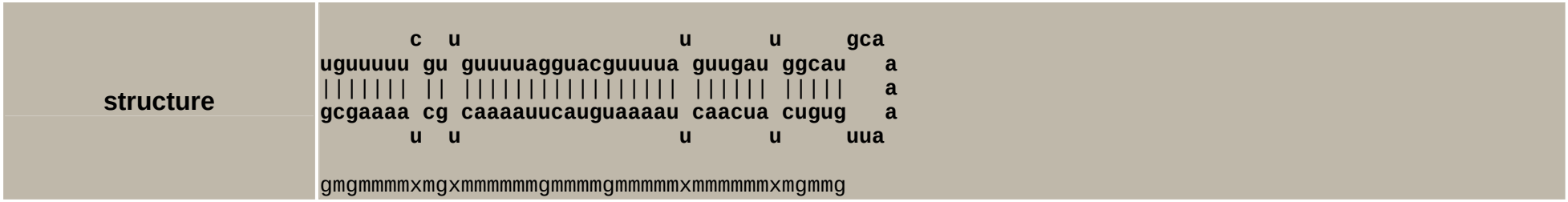
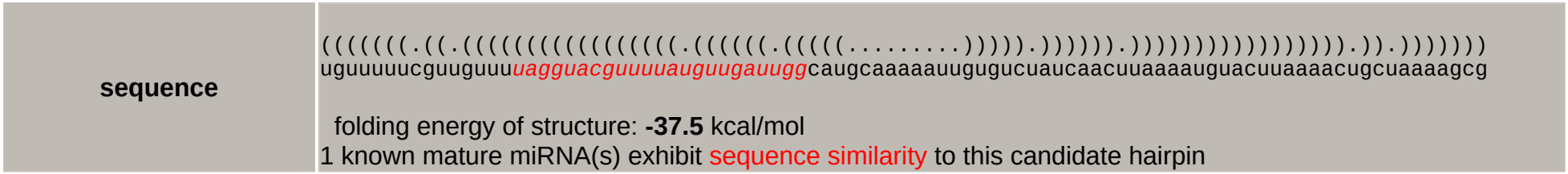
id	4546124
genomic_position	2916983-2917057 on CEL150_III (+) in Caenorhabditis elegans
L score	0.0155916136128
structure	
sequence	 folding energy of structure: **-30.7** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

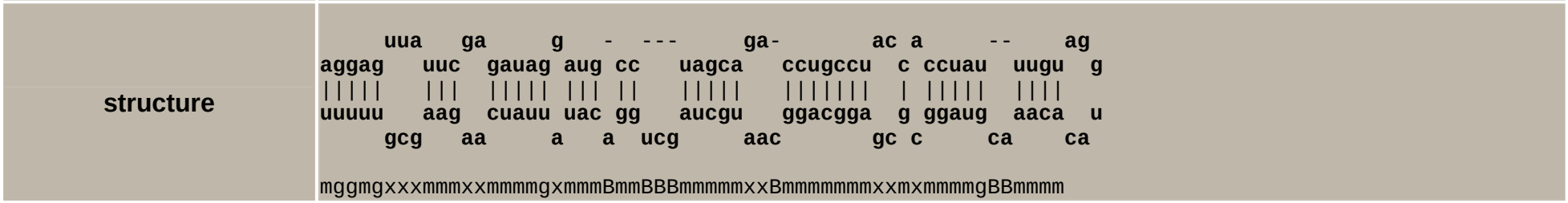
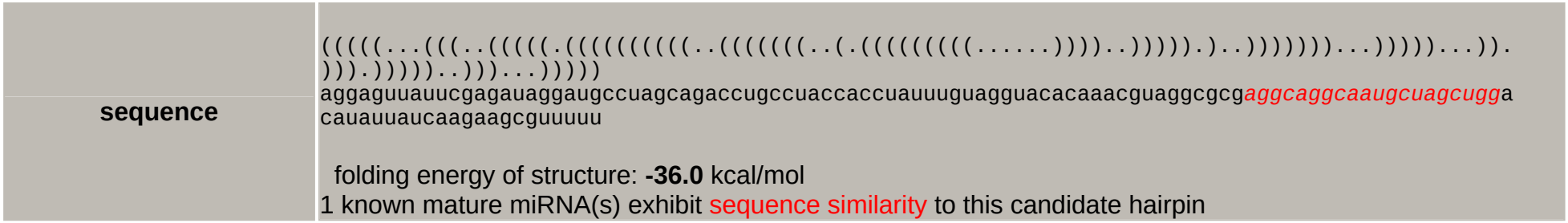
id	2229988
genomic_position	3596985-3597083 on CEL150_III (+) in Caenorhabditis elegans
L score	0.0114687741443
structure	
sequence	

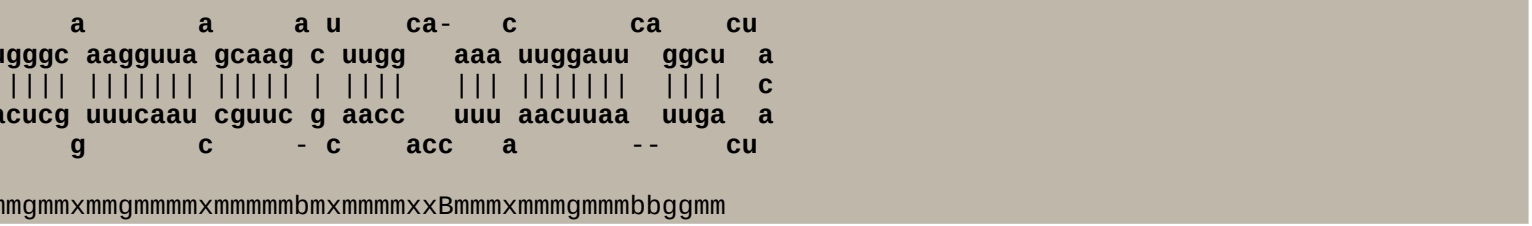
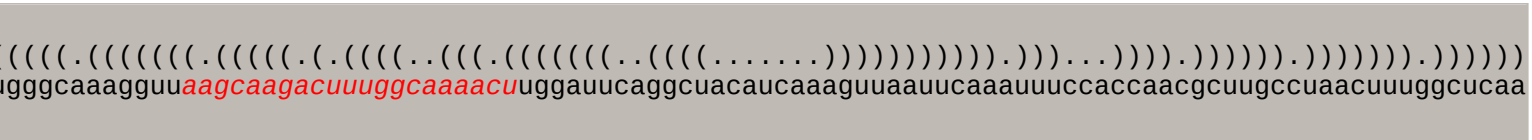
id	2196306
genomic_position	3876768-3876860 on CEL150_III (+) in Caenorhabditis elegans
L score	0.0225865595926
structure	
sequence	

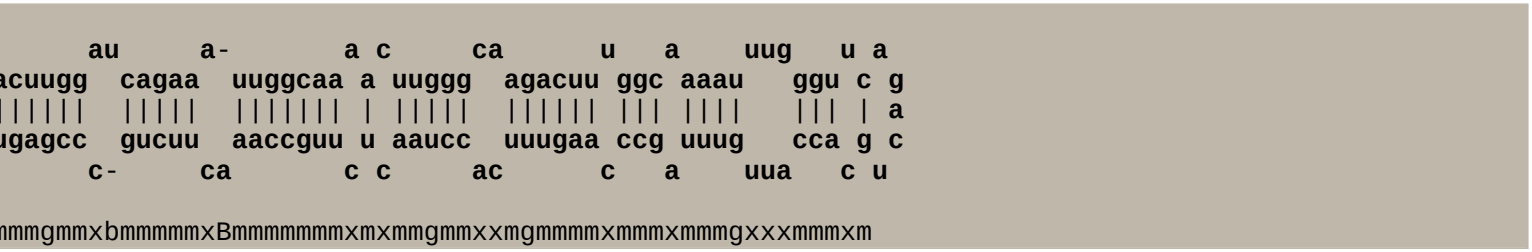
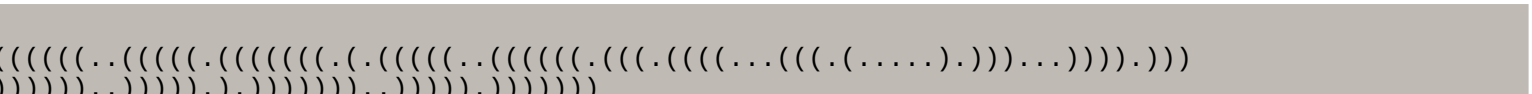
id	2414207
genomic_position	10524894-10524964 on CEL150_III (+) in Caenorhabditis elegans
L score	0.53734455228
structure	
sequence	(((.(((((((....((((((((((((((((....))))))))))))))))....)))))..))))..))) acuccaugauuuuuagcagccgacguuuuacggg*cccgugaagugucggcugca*accaaucgaugcuagu folding energy of structure: **-30.8** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	2490824
genomic_position	245874-245960 on CEL150_IV (+) in Caenorhabditis elegans
L score	0.0511826610145
structure	
sequence	(((..(((((.(((((.((((((...((.....(((((..((((....))))))))).))...))))))..)))))..)))))..))) ggcugaacuugggcaaaaguug*ggcaagacuuuggcagaacu*uggauucaggcugccccuaaaccaacucuugccaaaguuuggcc folding energy of structure: **-38.8** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

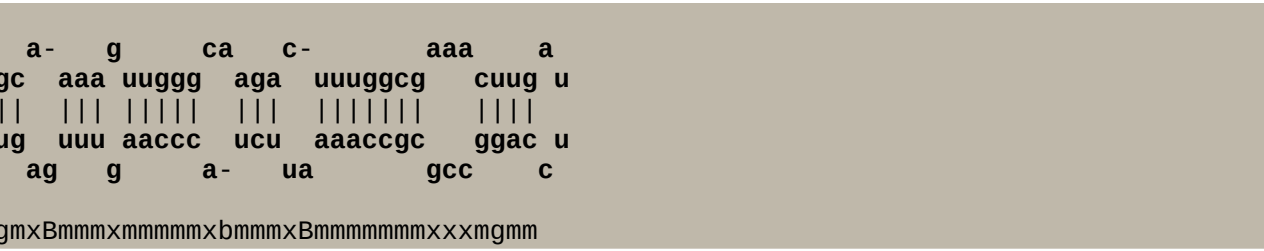
id	2875398
genomic_position	12806505-12806596 on CEL150_IV (+) in Caenorhabditis elegans
L score	0.0177509172603
structure	
sequence	

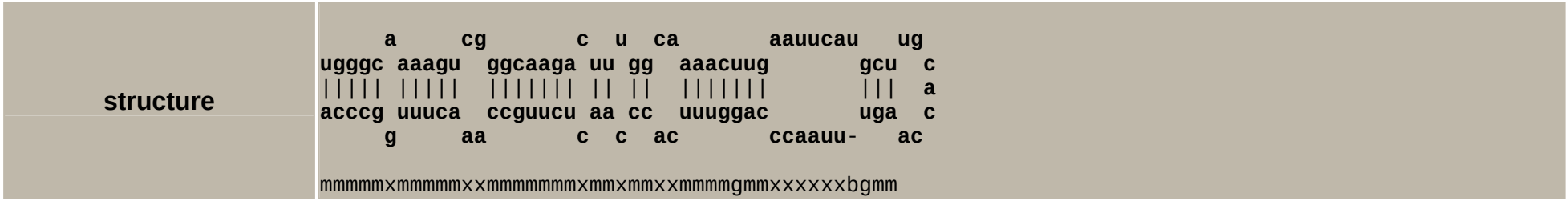
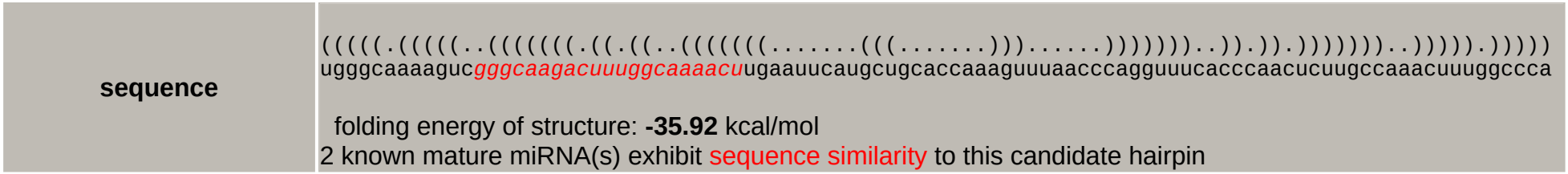
id	2981916
genomic_position	15168040-15168155 on CEL150_IV (+) in Caenorhabditis elegans
L score	0.00561533942048
structure	
sequence	

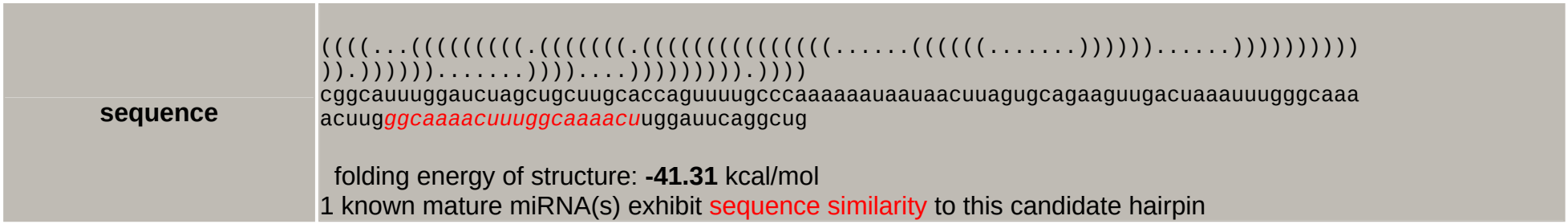
id	3083314
genomic_position	600114-600211 on CEL150_V (-) in Caenorhabditis elegans
L score	0.00822253846695
structure	
sequence	 uugggcaaaggu*aagcaagacuuuggcaaaacu*uggauucaggcuacaucaaaguuaauucaaauuuccaccaacgcuugccuaacuuuggcucaa folding energy of structure: **-29.5** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

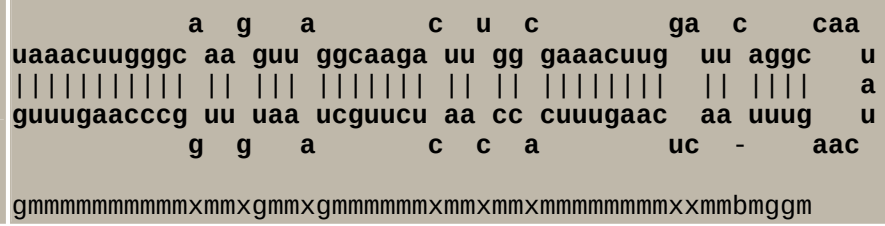
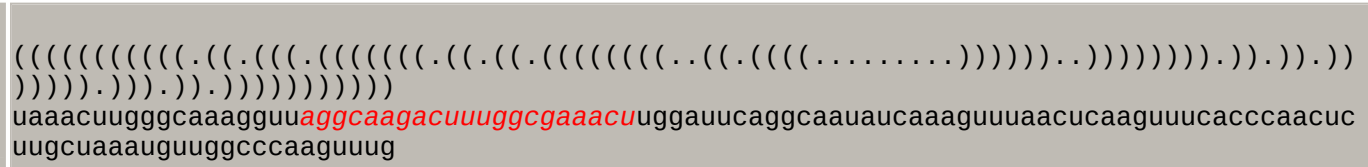
id	3130412
genomic_position	1304647-1304762 on CEL150_V (+) in Caenorhabditis elegans
L score	0.00653196564705
structure	
sequence	 gacuuggaucagaaauuggcaaaacuug*ggcaagacuuuggcaaaauu*uggguucagacugcaccauuguuuagcc caaguuucaccuaacucuugccaaacuucugcccgaguu folding energy of structure: **-42.4** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

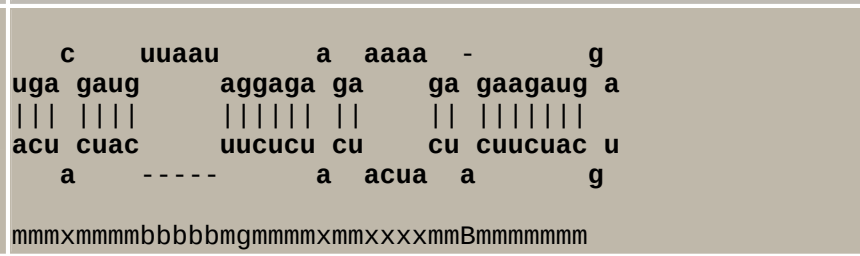
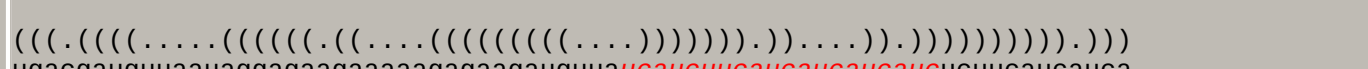
id	3148670
genomic_position	2851037-2851136 on CEL150_V (+) in Caenorhabditis elegans
L score	0.0966069330178
structure	
sequence	(((((((((.(((.(((((((..((((.(((((((((....(((((.....)))))...)))))))))..)))).)))))))).))))))))))) uuggaucagaaguug*ggcaagacuuuggcaaaacu*uggauucaggcugcaccaaaguuuuacccaaguuuuaaccaacuc uugccaaacuuugauccaa folding energy of structure: **-40.9** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

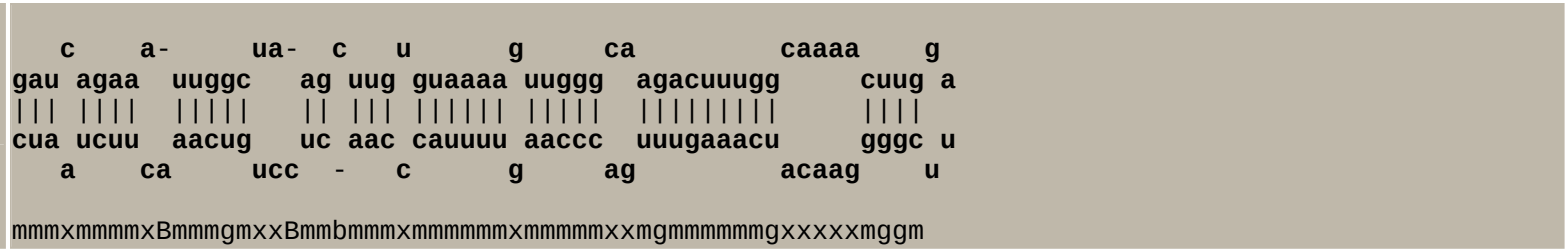
id	3179049
genomic_position	2851150-2851221 on CEL150_V (-) in Caenorhabditis elegans
L score	0.0349718376886
structure	
sequence	(((.(((.((((((..(((.(((((((...((((....)))).)))))))..))).)))))).)))..))) ggcaaaaguug*ggcaagacuuuggcgaaacu*uggauucaggccgcgccaaaauucuacccaaguuugaguc folding energy of structure: **-24.4** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	3240609
genomic_position	4703558-4703654 on CEL150_V (-) in Caenorhabditis elegans
L score	5.74977015012e-05
structure	
sequence	 folding energy of structure: **-35.92** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

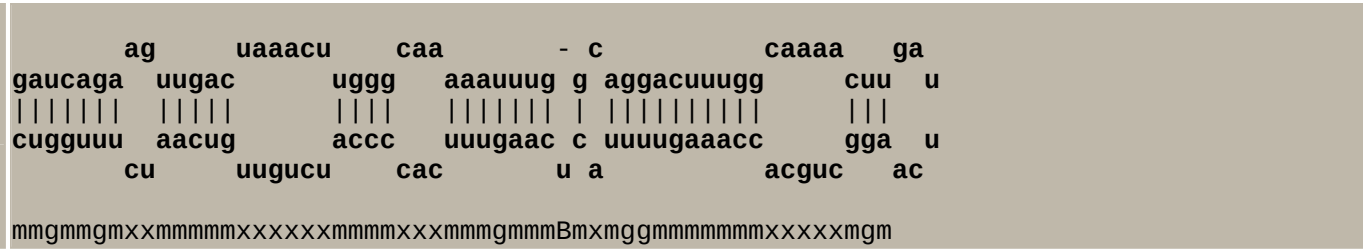
id	3227648
genomic_position	4844053-4844172 on CEL150_V (+) in Caenorhabditis elegans
L score	1.70886568918e-05
structure	
sequence	 folding energy of structure: **-41.31** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

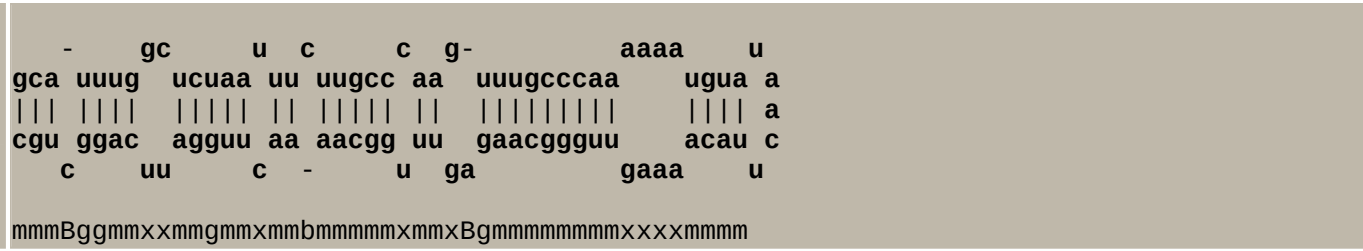
id	3411064
genomic_position	11066746-11066854 on CEL150_V (-) in Caenorhabditis elegans
***L* score**	0.102327407922
structure	
sequence	 uaaacuugggcaaagguu*aggcaagacuuuggcgaaacu*uggauucaggcaauaucaaaguuuaacucaaguuucacccaacuc uugcuaaauguuggcccaaguuug folding energy of structure: **-44.7** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

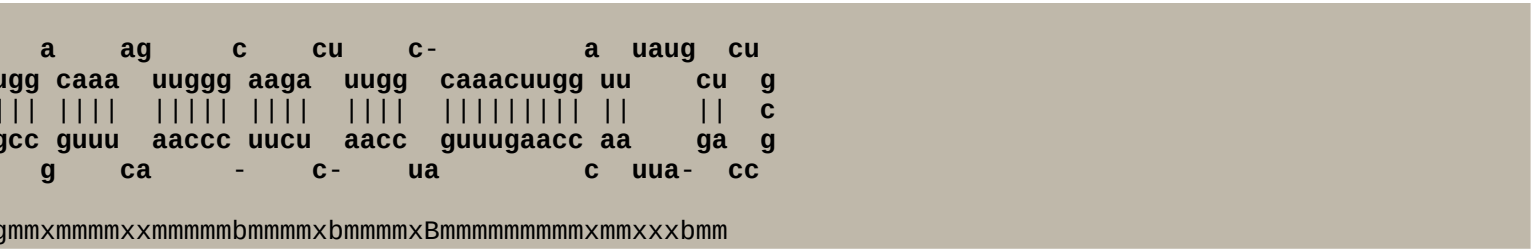
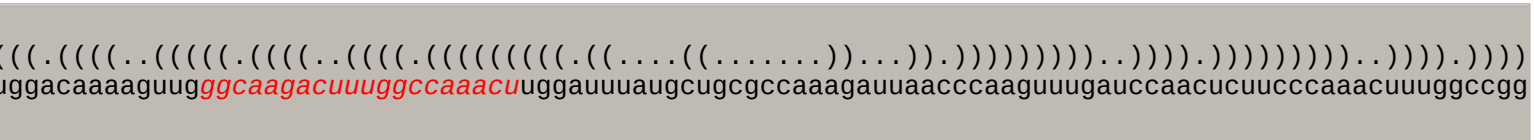
id	3450047
genomic_position	12302100-12302170 on CEL150_V (-) in Caenorhabditis elegans
***L* score**	2.76249354171e-05
structure	
sequence	 ugacgauguuaauaggagaagaaaaagagaagauguua*ucaucuucaucaucaucauc*ucuucaucauca folding energy of structure: **-19.6** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

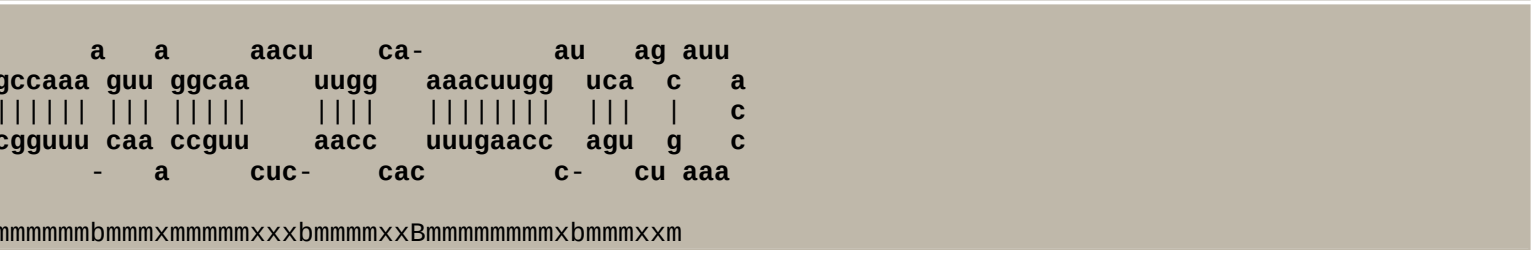
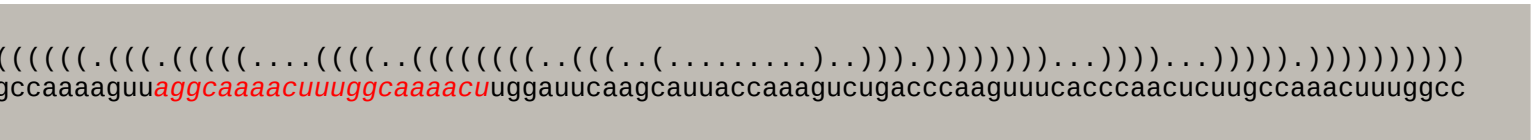
id	3561947
genomic_position	16348304-16348419 on CEL150_V (+) in Caenorhabditis elegans
L score	0.0153853123148
structure	
sequence	(((.((((.(((((..((.(((.((((((.(((((..(((((((((.....(((....))))).....)))))))))..)))))).)))))).)))))...)))))..)))).))) gaucagaaauuggcuaagcuuguguaaaaguug*ggcaagacuuuggcaaaacu*uggauucggggaacaucaaaguuugacccaaguuuua cccaacuccugucaaacuucuaauc folding energy of structure: **-35.7** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

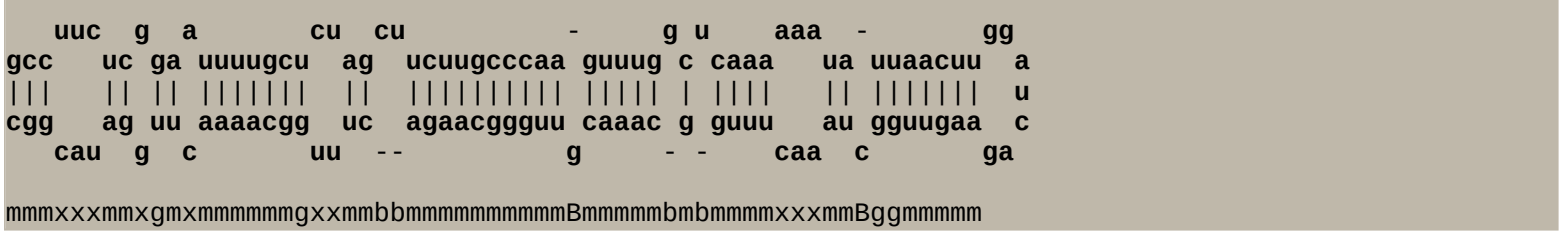
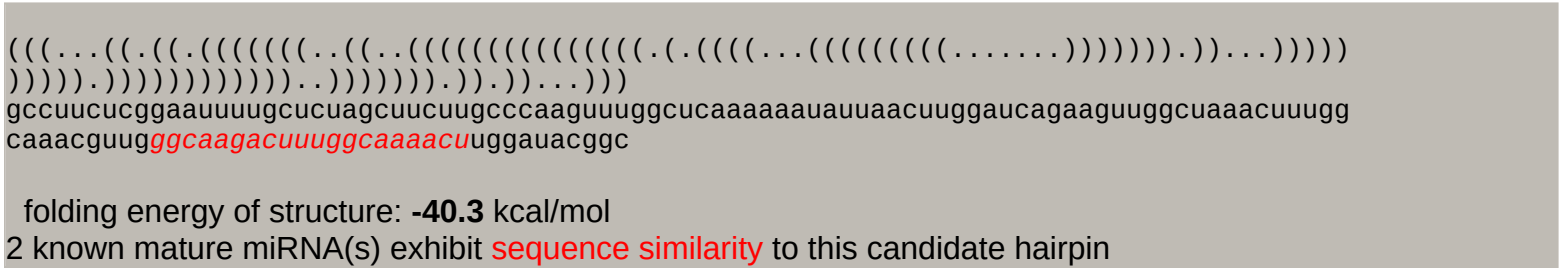
id	3714507
genomic_position	20590842-20590916 on CEL150_V (-) in Caenorhabditis elegans
L score	0.00222552655899
structure	
sequence	(((((.(((.((...(((.((((((((((....(((...)))...)))))))))).)))...)).)).))))) aaagcucgca*ucaucaucaucaucaucauca*cggaggcucucauaaugaugaugacgauuuggacagugguuuu folding energy of structure: **-21.4** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	3729314
genomic_position	539531-539646 on CEL150_X (-) in Caenorhabditis elegans
L score	0.000440348779386
structure	
sequence	(((((((..(((((.....((((...(((((((.((((((((((.....(((.....))).....)))))))))).).))))))))...))))).....)))))..))))))) gaucagaaguugacuaaacuugggcaaaaauuu*ggcaggacuuuggcaaaacu*uugauucaggcugcaccaaaguuuuacucaag uuucaccccaucuguugucaaucuuugguc folding energy of structure: **-34.7** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

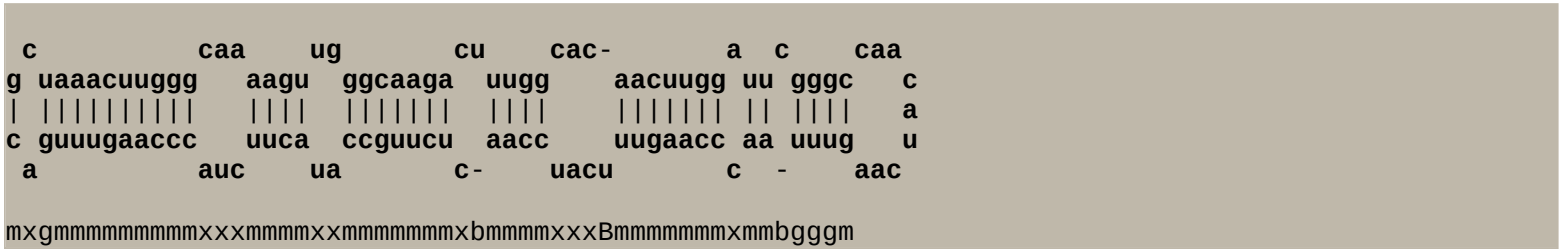
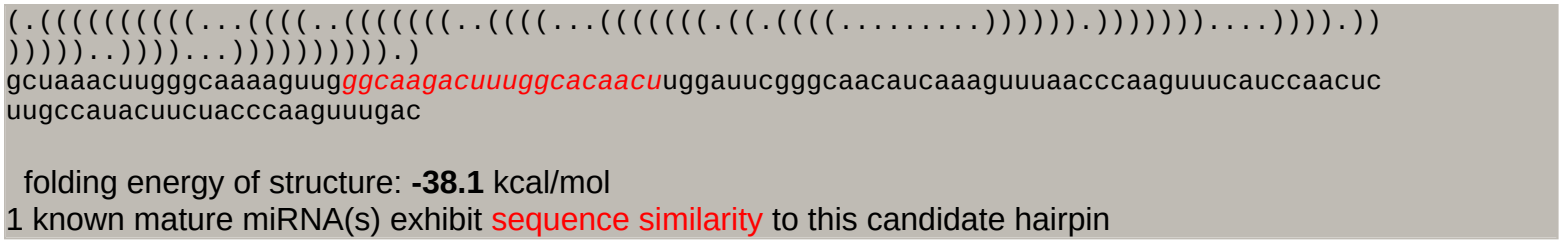
id	3731460
genomic_position	624954-625048 on CEL150_X (-) in Caenorhabditis elegans
L score	0.000109205356242
structure	
sequence	(((((((..(((((.((.(((((.((.((((((((((....((((.....))))....)))))))))..)).)))))))..)))))..)))).))) gcauuuggcucuaauuucuugcccaaguuugcccaaaaaauguauuaacuacaaaaguug*ggcaagaguuuggcaaaacu*uggauucaggcugc folding energy of structure: **-30.4** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	3846390
genomic_position	3912836-3912933 on CEL150_X (-) in Caenorhabditis elegans
L score	0.00931295735244
structure	
sequence	 folding energy of structure: **-31.6** kcal/mol 1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	4048106
genomic_position	11051157-11051250 on CEL150_X (+) in Caenorhabditis elegans
L score	0.000347862126048
structure	
sequence	 folding energy of structure: **-33.3** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	4147031
genomic_position	14917324-14917447 on CEL150_X (+) in Caenorhabditis elegans
***L* score**	7.22734328364e-05
structure	
sequence	

folding energy of structure: **-40.3** kcal/mol
2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	4167805
genomic_position	14917515-14917627 on CEL150_X (-) in Caenorhabditis elegans
***L* score**	0.0351579811802
structure	
sequence	

folding energy of structure: **-38.1** kcal/mol
1 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin

id	4181655
genomic_position	15447706-15447814 on CEL150_X (-) in Caenorhabditis elegans
***L* score**	0.00266635069795
structure	**ag a u caa ca- aaa- gg** **caga uuggcaa a uuggg aaauuggg agacuuuggc acu a** **|||| |||||| | |||| ||||||| ||||||||| ||| u** **guuu aaccguu u aaccc uuugaccc uuugaaacug uga u** **ca c c ac- cag cagc ac** mmgmxxmmmmmmmxmxmmmmmxxbmmmgmmmmxxBmgmmmmmmgmxxxBmmm
sequence	((((..(((((((.(.(((((...(((((((..((((((((((...(((.......))).....))))))))))...))))))))..)))))).).)))))))..)))) cagaaguuggcaaaauuugggcaaaaauu*gggcaagacuuuggcaaaacu*uggauucagucgacgucaaaguuugacccca guuucacccaacucuugccaaacuuug folding energy of structure: **-38.3** kcal/mol 2 known mature miRNA(s) exhibit sequence similarity to this candidate hairpin