

Supplementary Material

*Communication*

# **Stem–Loop Structures in Iron-Regulated mRNAs of *Giardia duodenalis***

**Laura Y. Plata-Guzmán <sup>1</sup>, Rossana Arroyo <sup>2</sup>, Nidia León-Sicaïros <sup>3,4</sup>, Adrián Canizález-Román <sup>3</sup>, Héctor S. López-Moreno <sup>1</sup>, Jeanett Chávez-Ontiveros <sup>1</sup>, José A. Garzón-Tiznado <sup>1</sup> and Claudia León-Sicaïros <sup>1,\*</sup>**

<sup>1</sup> Programa Regional del Noroeste para el Posgrado en Biotecnología de la Facultad de Ciencias Químico-Biológicas, Universidad Autónoma de Sinaloa, Av. de las Américas y Josefa Ortiz (Cd. Universitaria), Culiacán 80030, Mexico

<sup>2</sup> Departamento de Infectómica y Patogénesis Molecular, Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV-IPN), Av. IPN No. 2508, Colonia San Pedro Zacatenco, Mexico City 07360, Mexico

<sup>3</sup> CIASaP Facultad de Medicina, Universidad Autónoma de Sinaloa, Cedros y Sauces Frac. Fresnos, Culiacán 80246, Mexico

<sup>4</sup> Departamento de Investigación, Hospital Pediátrico de Sinaloa, Boulevard Constitución S/N, Col. Jorge Almada, Culiacán 80200, Mexico

\* Correspondence: [claudialeonsicaïros@uas.edu.mx](mailto:claudialeonsicaïros@uas.edu.mx)

**Table S1.** Ends of sequences coding virulence proteins

Annotation or common name	Gene ID	Region	Sequence delineated
<i>Cathepsin B precursor</i>	GL50803_14019	5'CR	cuuuuuccuccaauuu[u]aaaAUGaagcucuucuccucgcccgcugcagcuuucucugc gcccgcugacggucuccgagc
<i>Cathepsin B precursor</i>	GL50803_14019	3'UTR	ugagUAAgcaaugaaucucu <u>UGUAAAC</u> ggaaga[u]cuca
<i>Cathepsin B precursor</i>	GL50803_16160	5'UTR	guauucagcagguuuuuau[a]guuucucaccaguuAUGgaugcauugauugagcc aaacaugcuuuuagcuaaagucuuu
<i>Cathepsin B precursor</i>	GL50803_16160	3'UTR	acuaugcagacguagagUGAauccgcuuuuguccucaacugugucugggugccgg uuaaagcuaaaaaguuacagcgggcauggaccuugaucuccggAUGAACGccgaa[g] agauuaugca
<i>Cathepsin B precursor</i>	GL50803_16779	5'CR	aagcuuacggaaau[u]uaaAUGaagcucuucuccucgcccgcugcagcuuucucugc gcccgcugacggucuccgagcu
<i>Cathepsin B precursor</i>	GL50803_16779	3'CR/UTR	cauugaagagcaaguaauagguggguucuuugagaacUAAcuggugaauccauuu uauAGUAAAUcaugg[c]caggagggu
<i>Cathepsin B precursor</i>	GL50803_10217	3'UTR	gacAUGcgccugucuucauacauagcuaccauugcccuaaugggguagaAGUAAAC guuuuu[c]ugguccaugcauuaaaaa
<i>Cathepsin B precursor</i>	GL50803_15564	5'CR	cuagcauggcuuuuuuuu[a]uugcucaucuuauggucagacggcaccAUGcgcgcg gaauccucugguggcgcugucggu
<i>Cathepsin B precursor</i>	GL50803_16468	3'UTR	uggcagguagcagugaagUAAacguccaaaggcuuuuugggaaaaAGUCAUGcc cggg[u]guaucuuugcugcacggccgc
<i>Cathepsin B precursor</i>	GL50803_17516	5'CR	cuuauacaggaauuugcuccuuugaaggaacagagaaauccu[u]uaaaAUGaggcuu cugauccuucugauugcauccua
<i>Cathepsin B-like cysteine proteinase 3</i>	GL50803_114165	5'CR	ugcuccgggaacggccacggaa[a]uaaaAUGaugguacuucuaaccauucugcuauug cuauauggcacacuuggcagaga
<i>Cathepsin B-like cysteine proteinase 3</i>	GL50803_114165	3'UTR	acuuucacgguaacuagaugacacucauUGAcacacucugaugaacuagauGGUAA AUcuaucc[u]aagcaaagauguugc
<i>Cathepsin L precursor</i>	GL50803_3169	5'CR	aauaguuaacaacugcaaucuaagac[a]ucucacAUGuuagcgcuauugcucuuaa ugucugcuguguuuccaugaucc
<i>Cathepsin L precursor</i>	GL50803_3169	3'UTR	ugcugcaggaagUAGgugcggaugacuucuaacuaauagugaauaguagugaauA GUAAUAgacuag[u]ggacaacuuagg

<i>Cathepsin L precursor</i>	GL50803_16380	3' CR/UTR	uacgugauggu <b>gagUA</b> Aauugcccggguuggugaau <b>gugcucuggccuugGUC</b> <u>AAAA</u> aaug[u]ccuugccaagucauaaa
<i>Cathepsin L precursor</i>	GL50803_14983	3' CR/UTR	agcuacgcgguaguuaa <b>caaaagaaaaagcUA</b> Aaguaaa <b>uucuguggacgcugucua</b> auu <b>AGUGCAA</b> ucagg[u]gacuugg
<i>Cathepsin L precursor</i>	GL50803_17607	3' UTR	uugggUGAcuaaua <b>uagguuacuuaaggagcgcua</b> uagccccu <b>aaacuggccaac</b> cuuugc <b>AUGGUAC</b> ucugc[u]gugu
<i>Cathepsin L-like protease</i>	GL50803_3099	5' CR	guaaaaaacgaaaau <b>aaau[u]uaaa</b> agaaccaua <b>AUG</b> uauacu <b>uuguugguguu</b> ca <b>uacugucga</b> auccucugcauaucc
<i>Cathepsin L-like protease</i>	GL50803_137680	5' UTR/CR	guguaaaaccugaagg <b>acuuuu[a]uaaa</b> aguag <b>uagAUG</b> ugguuguccu <b>auuuu</b> cu ggugaucuuuuugcuugaacucuuug
<i>Cysteine protease</i>	GL50803_112831	5' UTR/CR	cggcuaaccuacccgcggcc <b>aa[a]aaaggaccggAUG</b> uugcugcu <b>uacuuugucu</b> <b>u</b> ccucugcgcggcaucggccgagc
<i>Cysteine protease</i>	GL50803_112831	3' CR/UTR	uacaucuuccacagcccgcuaaccugaaccuuccgcaca <b>agggcgugUAG</b> Gcgccuagc <b>gCCCCGAAC</b> uccccg[u]cguaga
<i>Cysteine protease</i>	GL50803_113656	5' UTR/CR	cggcuaaccuacccgcggcc <b>aa[a]aaaggaccggAUG</b> uugcugcu <b>uaccuugucu</b> uu ccucugcgcggcgucggccgagc
<i>Cysteine protease</i>	GL50803_113656	3' CR/UTR	uacaucuuccacagccccuacaccugaaccuuccgcaca <b>agggcgugUAG</b> Gcgccuagc <b>gCCCCGAAC</b> uccucg[u]cguaga
<i>Cysteine protease</i>	GL50803_114773	5' UTR/CR	cggcuaaccuacccgcagcc <b>aa[a]aaaggaccggAUG</b> uugcugcu <b>uaccuugucu</b> uu ccucugcgcggcaucggccgagc
<i>Cysteine protease</i>	GL50803_114773	3' CR/UTR	ccucaaccugaaccuuccgcaca <b>agggcgugUAG</b> Gcgccuagcg <b>CCCCGAAC</b> ucccc[g] ucguagagggugguagguagaug
<i>Cysteine protease</i>	GL50803_114915	5' UTR/CR	cggcuaaccuauccgcggcc <b>aa[a]aaaggau</b> ccgg <b>AUG</b> uugcugcu <b>uacuuugucu</b> <b>u</b> ccucugcgcggcgucggccgagc
<i>Cysteine protease</i>	GL50803_114915	3' CR/UTR	uacaucuuccacagcccgcuaaccugaaccuuccgcaca <b>agggcgugUAG</b> Gcgccuagc <b>gCCCCGAAC</b> uccccg[u]cguaga

<i>Dipeptidyl-peptidase III</i>	GL50803_3822	5'UTR/CR	agggggga <u>aa</u> uaaa[a]uuucc <b>AUG</b> cugccugagaaguauauuugcccgcagagaugga ccgguuucuaacuugaagccaucg
<i>Dipeptidyl-peptidase III</i>	GL50803_3822	3'UTR	ggaaaUAAaug <b>cgacguuguaaguug</b> auuguaauaacugcuuagaucacaucuaau cuagggga <b>GUAAA</b> Gaugg[u]augg
<i>Dipeptidyl-peptidase I precursor</i>	GL50803_14566	5'CR	aacaauaagaagccggccaga[a]aaaauc <b>AUG</b> cuuuuca <b>cuuggcgcuuccuggg</b> acu cuuauuauacagauacuccagca
<i>Dipeptidyl-peptidase I precursor</i>	GL50803_28651	3'UTR	cugaccaugcauuUA <b>Gaugcacugaguguuu</b> aacugagugcuuau <b>cg</b> <u>CAUAA</u> <u>AG</u> acucaccuau[u]cucuacaauaaua
<i>Dipeptidyl-peptidase III</i>	GL50803_33689	5'CR	gcuagcuuuuacugugaauuuaucauuu[u]ucua <b>AUG</b> ccu <b>aacgaau</b> gcccaucug <b>gagcuuuc</b> acggucaaccugaagc
<i>Dipeptidyl-peptidase III</i>	GL50803_33689	3'UTR	guaacaacauccuuguuuuccucugcuaac <b>UGA</b> cuauacagugucuaauuucucga <b>G</b> <u>CUUAAU</u> cagcaac[u]gcuugugaau
<i>Encystation-specific protease</i>	GL50803_22553	3'CR/UTR	caggaggagagugaggagcuaagca <b>ugcuu</b> ucaaa <b>uccaUGAAGUAAA</b> Ugauuc cacaag[u]acaauuacaaaacagccau
<i>Hypothetical protein</i>	GL50803_6369	5'UTR/CR	ccucuagaga <u>ga</u> [u]gaa <b>nc</b> agagagagacuc <u>au</u> <b>AUG</b> gacuacuggauugcucgu <b>aacucuugggg</b> agaagagugggga
<i>Hypothetical protein</i>	GL50803_113303	3'UTR	uucacggauuuggaugacacuc <u>au</u> <b>UGA</b> cacacucuga <b>ga</b> acuagau <b>GUAAAAU</b> cuaucc[u]aagcaaagauguugucggu
<i>Sentrin specific protease</i>	GL50803_16438	3'UTR	aagaucuggacgagcucug <b>cUAG</b> gccu <b>au</b> gggucuggugucuaaguugugu <b>UUG</b> <u>AAAA</u> cccuac[u]cugcccguguuaggg
<i>Variant-Specific Surface Protein 150 (VSP-150)</i>	GL50803_135831	5'CR	ugugcucaagggua <u>cu</u> [u]cuuauac <b>AUG</b> aauggcuguuacaaggu <b>u</b> gggu <b>ca</b> agaacc <b>aggu</b> agucauauugcaagacagc
<i>Variant-Specific Surface Protein 150 (VSP-150)</i>	GL50803_135831	3'UTR	cuggugguucauauaguagagggaagg <b>caUGA</b> cucaggu <b>AGUGAAU</b> gcua[u]ac aaugugugauaauggguauuagagggg
<i>Variant-Specific Surface Protein 5 (VSP-5)</i>	GL50803_113797	5'UTR	ccccgggcuuugcguuggaagcgcca[c]ccag <b>caggucggcg</b> gccua <b>AUG</b> uucggcag auuuuugcucgcgaucgucaucc
<i>Variant-Specific Surface Protein 5 (VSP-5)</i>	GL50803_113797	3'UTR	ugcuggugguuccucugcagggggaaggcg <b>UAG</b> auguac <b>u</b> agau <b>AGUAAAC</b> c guca[u]cgaugggucugcugcgugucug

<i>Variant-Specific Surface Protein 277 (VSP-277)</i>	GL50803_14307	5'UTR/CR	caggggccugcgacgccaucgugguc[g]gcaac <b>gacggcagc</b> AUGcucucuuacugu ucaaaguguguuggcgccggcuau
<i>Variant-Specific Surface Protein 277 (VSP-277)</i>	GL50803_14307	3'UTR	ccucugcuggugguucguguguagaggggaaggcgUG <b>Acuuaggu</b> <u><b>AGUGAAU</b></u> gc ug[u]gcagugugugauagugggugauau

---

Delineation of the untranslated regions of certain sequences encoding for ribosomal proteins in *G. duodenalis*. At the 5' end: CAATTT element: shading. T/A-rich region: italics. Transcription start site: brackets. Start codon: Uppercase. IRE-like: bold letters. At the 3' end: Stop codon: Uppercase. PolyA signal: Uppercase, italics and underline. Transcription stop site: brackets. IRE-like: bold letters. CR: Coding region.

**Table S2.** Ends of sequences coding ribosomal proteins

Annotation or common name	Gene ID	Region	Sequence delineated
<i>Ribosomal protein L3</i>	GL50803_16525	5'UTR/CR	uuccgggucug <del>aaaag[a]caaggcccg</del> <b>AUGuccaccgcgaaguucucgugcugcc</b> <b>gcaagggca</b> accucggcuaccu
<i>Ribosomal protein L4</i>	GL50803_17547	3'CR/UTR	cgcucgaucccagcggaacgguca <b>ugcccaagUAAaucgcuggCGCAAAA</b> cgu u[u]gcacaaugaggagccucgagga
<i>Ribosomal protein L7a</i>	GL50803_17244	5'UTR/CR	uggguuca[a]uaaaac <b>aggccucgacAUGuccaagguuucuggcagcgacauuaag</b> <b>agggccucg</b> ccguacccgagaa
<i>Ribosomal protein L7a</i>	GL50803_17244	3'CR/UTR	ggcguucgcccgaggau <b>gaggccagcagaugUAAaugccu</b> acauuuuuu <u>A</u> <u>GUGCAU</u> guccuu[u]gcugcuugguc
<i>Ribosomal protein L7Ae</i>	GL50803_2825	5'UTR/CR	uggaaaauguaauaaauu[a]aaacucaauugaa <b>caaggAUGccagaugcacgcgc</b> <b>uguuucccuugg</b> ccucagaggccc
<i>Ribosomal protein L7Ae</i>	GL50803_2825	3'CR/UTR	acaagcuaaacauauuccuu <b>acggagcuaggacacUGAuccggAGUAAA</b> Uggu gg[u]gaagcaguguggccugcccauc
<i>Ribosomal protein L7Ae</i>	GL50803_11287	3'UTR	gcacugguaugac <b>gggagugcguu</b> acagugcaagcaucacuaguaa <b>augcucacc</b> <b>ggua</b> c augcuacacacuacuaguauuug <b>AGUGCA</b> cuaua[u]acgaagg
<i>Ribosomal protein L9</i>	GL50803_17056	5'CR	aaucgcuuucgaugcucuuuu[u]uuagaggucgcgaagca <b>AUGccccucuguuc</b> <b>cuccaugacgau</b> caaggucgccgac
<i>Ribosomal protein L10</i>	GL50803_10428	3'CR/UTR	ggaagggggcccucaacu <b>ugaagcucUGAuccggguccca</b> gcaca <u>AGUAAA</u> Ua auugccaggcauu[u]acuccucugcc
<i>Ribosomal protein L10a</i>	GL50803_1345	3'UTR	uucUAA <b>gaucuuUGAGAA</b> Gguucccgggguuu[u]gcau
<i>Ribosomal protein L12</i>	GL50803_14938	5'CR	uuauaguuaaagcauuuuuuuugau[a]uuuugucagagugguaguuuuccaggc gaca <b>AUGccccagagcugauccuaacgcagagauaguucugu</b> auuuugcgcuca agggcggg
<i>Ribosomal protein L12</i>	GL50803_14938	3'CR/UTR	uugcugaggagcaagagacagag <b>UAAacaaggugccc</b> ugcgugucacggaucggu cccuauugcuugggacuguuuucgcuacggggg <u>CUUAAA</u> Auuu[u]cgau ggccgaggcggc
<i>Ribosomal protein L13</i>	GL50803_14622	5'UTR/CR	uggaagcucgugcaauugc[a]cuuaua <b>AUGgcaaguuaugcuucuuuuu</b> auuu gagauuaauucauuuuuuuugaauu
<i>Ribosomal protein L13a</i>	GL50803_11247	5'UTR/CR	uuuuuuuggauuuagcgcaaauc[u]guuuugcucacggcgagcc <b>AUGucgcgauc</b> <b>guuaucgacugcaaggggc</b> acauac

<i>Ribosomal protein L15</i>	GL50803_8001	5'CR	uc <u>auagga</u> ucauu[u]uuuacccggccugaauucgAUGggugccuacaaguacguu <b>caggagcucugggcg</b> caagaagcaa
<i>Ribosomal protein L15</i>	GL50803_8001	3'CR/UTR	aaccacaccgcucccucc <b>cgccgc</b> uacagaUAGccuugcugcugcgAGUAAACa ucccgucauu[u]acacuagaggaag
<i>Ribosomal protein L17</i>	GL50803_98056	3'UTR	gaaggcgcuccagUGAgucugucgggaguaa <b>augcgcugcu</b> cuuuccaacuucu <u>AAUAAU</u> ugaaaugucuaa <b>uguggggu</b> acauuccgauu[u]cggcccuucccg uuuuuucucggaacu[u]uuucgaccucccgacAUGgcuaa <b>ucugaggcug</b> cagaa
<i>Ribosomal protein L19</i>	GL50803_16431	5'UTR/CR	<b>gcgccugggcggcag</b> acguccucaa
<i>Ribosomal protein L21</i>	GL50803_15520	5'CR	uuuuuuuuu <u>ucauuu</u> [u]caaugcugcccga <b>aaagcg</b> aggacgacucgAUGacgcac <b>agccagggguacaggugu</b> aagac
<i>Ribosomal protein L21</i>	GL50803_15520	3'UTR	cgacaucaccccacucaag <b>uaugacgg</b> aguuuucUAG <b>gcuccgcag</b> aag <b>uaaaug</b> <b>gccug</b> cgggcgggcuAGUGGAggcugggcu[u]ccaccgg <u>uuccucg</u> acgcgcu
<i>Ribosomal protein L23</i>	GL50803_10091	5'UTR/CR	acaacuuuaauuucgcuuuu[u]acucguuuccuaa <b>ugacgg</b> aucgaggagAUGgc <b>gucgaagcguggcagguc</b> ggagg
<i>Ribosomal protein L23</i>	GL50803_10091	3'CR/UTR	cgacggcugagug <b>cguccac</b> UAGgcAGUAAACgaa[u]uc
<i>Ribosomal protein L23A</i>	GL50803_7870	5'CR	aauuucgccgucua[a]uaucucgugccggcgagacAUGuccgaga <b>aggcg</b> caggca cugaaggcccaaaaggcgauuu
<i>Ribosomal protein L24</i>	GL50803_14869	5'UTR/CR	aguguuucucuaccauuuaa <u>uacuuuaa</u> [a]uaauucaggaacaAUGccucgc <u>au</u> <b>agagcauugcu</b> cauucuguccaa
<i>Ribosomal protein L26</i>	GL50803_15046	5'UTR/CR	aaauaa <u>uucg</u> auuugaaauggaaauggcccauuuuuuu <b>gaau</b> caccggcggaagccu <b>gga</b> AUGaagcucaacuccgcccgu <b>caccgcg</b> ucccgucgaagugccgcaaggcuac uucaca
<i>Ribosomal protein L27a</i>	GL50803_16310	5'UTR/CR	acaaaaaag <u>ua</u> auuuuuuuuuuacauuuuccccacgg <u>ucu</u> [a]aaaagggg <u>ccu</u> <b>cgag</b> AUGccgaccagaau <b>cagg</b>
<i>Ribosomal protein L31B</i>	GL50803_16368	5'UTR/CR	ucugaauugucccu <b>gcaa</b> [a]aaacggguccuccga <b>agcgug</b> AUGgguauugua <b>ua</b> <b>ugagcauacga</b> uccgcuca <b>acaa</b>
<i>Ribosomal protein L31B</i>	GL50803_16368	3'CR/UTR	gcagcgaagacggcagcgagUAAacuugugcagcucgcauuuacauaugcuuuu uugcCGUAAGUggg[u]cguccug
<i>Ribosomal protein L34</i>	GL50803_36069	5'UTR/CR	uuuuuuuuuac[u]aaagagcugcccgc <b>ccgag</b> AUGgccgacugcgcgucacuugcc gcuuccacuccacauacaggacu

<i>Ribosomal protein L34</i>	GL50803_36069	3'CR/UTR	agcuaagacggccaagcagaccgccggcaagggg <b>gaugcgaagUAAaucgaucagca</b> uuuacgucucacuuucgggccuuccgaaccccgccugggacucauggggaaucuc ggcuucuggggcu <b>UUGAAG</b> cugcuu[u]ccccggggcgccugccuugau <i>Ribosomal protein L35</i>
	GL50803_6133	5'CR	ucucgcacaccg <b>gaagag[u]caacuugccAUGgcauccuguuugcccuaccagauag</b> <b>auguuau</b> caauguauuuuaaa
<i>Ribosomal protein L35</i>	GL50803_6133	3'CR/UTR	cuccaguucccgaaagcugaaguaugcu <b>guaaaggucUGAgcaccucgguguuagu</b> <b>aaauac</b> ucucuugcuuccuuucaccgcguagaacauagcagaugcggagugcgua uggcgcacgu <b>GUGAAA</b> cucgccagguuu[u]gcaaggguagcaggugggagc <i>Ribosomal protein L35a</i>
	GL50803_5947	3'CR/UTR	gagcgcuugacgggaaccccaa <b>cgcgacgguaaccuacgaguaaaauagucaggcg</b> gcuuguggc <b>GUGAUG</b> cugggggucucgccugugguu[u]acugcuuugc <i>Ribosomal protein L36-1</i>
	GL50803_16114	5'CR	uuuauuuuccaauccauuuuaauucauuucaaauuccg <b>auuu[a]gaaggagcgg</b> uguggu <b>AUGccaggcaaaguauucaaccugaagaaggcgggcg</b> caguaguucgca uuugccgca <i>Ribosomal protein L37</i>
	GL50803_14171	5'CR	ccuccaugccggacauagcauuuuuuuuggaauuagucgc <b>uguu[a]aaaacgggaa</b> cucg <b>AUGuccaagggcacagcgucguucgguaagcgccacacgcgc</b> cauucacggca gcugcc <i>Ribosomal protein L37</i>
	GL50803_14171	3'CR/UTR	gcaaacgccaggaagu <b>accuugccaagUAAauuugcacgaggucgguccgAUGG</b> <b>AUG</b> ccuacacgc[c]ggugcgucgcg <i>Ribosomal protein L37a</i>
	GL50803_5517	5'CR	ucuggaaauuuauaaaaa <u>u[u]aaaau</u> gaaucucugcaagcgaagcggcucaggcg agcu <b>AUGccaagaggacuaaaaagguuggucacccggga</b> aguacggcacucgcu auggga <i>Ribosomal protein L37a</i>
	GL50803_5517	3'CR/UTR	cacgauccgccgcc <b>uccgucuggcccgugaagacgcacagUAAaguaaacagagug</b> <b>ga</b> ccuggcacugcggu <b>UUGCAA</b> cugucugcauagcuguu[u]acucgucc <i>Ribosomal protein L44</i>
	GL50803_9810	5'UTR/CR	aaaugcaauugaauuagcuuug <b>caugcuacu[u]aacagcagcgccgaaggacAU</b> <b>Ggugacauauccggcagagcgca</b> <i>Ribosomal protein SA</i>
	GL50803_7766	3'UTR	cucuucgacgaguacUAAggcggcucucuuaa <b>agcugaaaguguaaacgcuuug</b> <b>cug</b> gguaugcaucccuguuucaccagagagguuucacaacuccaggacgguuccuu cugcgauuccgaggauuu <b>CGUAAA</b> gucuggcuauc[u]ggcacuacaagugu <i>Ribosomal protein S2</i>
	GL50803_8118	5'CR	uaauuuugauuuagacuuguagcc <b>aaaac[u]aacgu</b> ccauggcgua <b>gcAUGcag</b> <b>gcugaaagugcucccaggccagaa</b> <i>Ribosomal protein S4</i>
	GL50803_11359	5'CR	ggauggu <b>auuc[u]gguuu</b> agcagcga <b>AUGgcccugggaguccgcaagc</b> accuuuag cgccuacgcaccgaagcacug



<i>Ribosomal protein S5</i>	GL50803_12981	3'UTR	uauagccguuuccaaccgcUA Aacgccc <b>aggaguaaaccgccccu</b> cugagagaucgg ugucaguccucaaaa <b>AAUCAU</b> auuuuggaugaguugcgguu[u]cuccuugcug
<i>Ribosomal protein S6</i>	GL50803_14620	5'UTR/CR	uggugagaaagcauuuuuuuucauuuucuagu <b>uuc[a]uaaaaauggccccgggguga</b> <b>AUGccaaagggcucauu</b> cugcaag
<i>Ribosomal protein S6</i>	GL50803_14620	3'CR/UTR	aagccugcuaa <b>aggacggguagUA</b> Aacagacgucuuuacuacgggc <b>CUGA</b> <b>AUC</b> uguucgu[u]gcagcguugccuua
<i>Ribosomal protein S8</i>	GL50803_5845	3'CR/UTR	ggaccugagcuucucuuccaaga <b>agaagu</b> ggcaccuaag <b>UA</b> Acaucc <b>AGU</b> <b>GCUU</b> cgugguc[u]ccccacucucg
<i>Ribosomal protein S9</i>	GL50803_4547	5'UTR/CR	uauuuuuuuuucauuu <b>uggcuu</b> ucgccaau[u]cuccccugacgauucc <b>AUGccccg</b> <b>cauuagccgccauagga</b> acucgu
<i>Ribosomal protein S10B</i>	GL50803_10919	3'CR/UTR	aagcaag <b>UA</b> Aa <b>aucugag</b> cccauagcuuccacgcacagaugauccccagcggaug uugg <b>AGUAAAU</b> caagucc[u]gg
<i>Ribosomal protein S11</i>	GL50803_14827	5'UTR/CR	cuugagagagccgagacucugcgcaaggcugc[u]uac <b>gagAUG</b> gcugagaaccag <b>cacgagc</b> gagcguaucacaagcaa
<i>Ribosomal protein S11</i>	GL50803_14827	3'CR/UTR	gagccugucag <b>cgaggcccca</b> agaagcgcucucgcccuc <b>UAG</b> gcagcacu <b>AG</b> <b>UAAAU</b> aaggg[u]cauucacaaggc
<i>Ribosomal protein S12</i>	GL50803_33862	3'CR	gggacgucgacagcuucc <b>uuacua</b> agcuggacgagcag <b>UA</b> Aacgacauacu <b>AG</b> <b>UCGUG</b> cgga[u]cgguagugggauu
<i>Ribosomal protein S13</i>	GL50803_16652	5'UTR/CR	auagcauuu[u]agcgggcacgugacaaagcc <b>AUG</b> ggccgcaugcacucgaagggg <b>aagggua</b> uauc <b>ccgcucug</b> ccau
<i>Ribosomal protein S13</i>	GL50803_16652	3'CR/UTR	guccgaccaggccgcccgcucuccu <b>g</b> cgacguac <b>gcgUG</b> Aguaaacaccgcuacgg ggcucgaacccgcgaccguugg <b>AUUAAA</b> aguccaacgcuc[u]accaacugag
<i>Ribosomal protein S14</i>	GL50803_7878	5'CR	uuuugaauugacuuuagaagg <b>acaaa[u]u</b> ugaggugggcuga <b>AUGucucgcaaa</b> <b>guaaccuacgcgcgggaga</b> aucuuc
<i>Ribosomal protein S15</i>	GL50803_15260	3'CR/UTR	ccgcugcgu <b>UAG</b> ucccccccacc <b>AGUAAA</b> Cauucc[u]cg
<i>Ribosomal protein S15A</i>	GL50803_15228	5'UTR/CR	cgcacaucgcgugccgugcugcugggaagcaga[u]ccaguaggc <b>AUG</b> guucgcau <b>caacguucuccgugacg</b> cacuca
<i>Ribosomal protein S16</i>	GL50803_4652	5'CR	uuuauuuccacuuccgguccca[u]aaaaagcggcuccccg <b>AUG</b> gcgaagggaacag <b>aagagacu</b> uacuggaaaggaac
<i>Ribosomal protein S16</i>	GL50803_4652	3'CR/UTR	uuugguggccguggucgucgucccgucgccagaaguc <b>guaccguUA</b> Agcgggcu <b>ucguccua</b> aguaaacauacugcauccaaauacacgcugucaucucaccacugaccu

<i>Ribosomal protein S20</i>	GL50803_6022	5' CR	ggccuccugggccguuguuguaguaggaggcuugggugucguagccauaugacug gccguaccgauuuuu <u>GUGAU</u> Caccauaagcgugcu[u]ccauguccugcc uaaaauga <u>uuucu</u> [u]uaaguacgaaaacgcggccaccgagggcuAUGgcagagagc <b>g</b> <b>ugucgacugggaggcg</b> ugaauuu
<i>Ribosomal protein S21</i>	GL50803_7082	5' CR	uaaaugauuuuuuguccuaca <u>a</u> ]uuggaccgaaacgacugagccAUGcaau <b>cg</b> <b>augacaucggg</b> augaagaacagc
<i>Ribosomal protein S21</i>	GL50803_7082	3' CR/UTR	cggucucauggaac <b>gccuuUA</b> <u>Gguugccgcauacacccggccaaguaaacgc</u> cucuga uacuugcuguauuccugcgugauuuacacagccuucggguucagagcgagcaccu ggcugcuagguc <u>GUAAC</u> Cggccucaugacagcaucguuu[u]ggggaacc cgugcgcuUA <b>Agcacggcgugu</b> aa <b>acugccacuggcug</b> ugcauuuu <u>AGUU</u> <u>UA</u> cucuucuaagcgcu[u]gagccacacuccauacagaauuucgugugucugaa uuuuuuuucugaauuucguuguc <u>a</u> ]uuuagcggggcuagcgAUGccc <b>gagauca</b> <b>caguuaaggugaggaagg</b> uccuca
<i>Ribosomal protein S23</i>	GL50803_14699	3' UTR	gcgaggaaggcgccacggcacgccccg[g]gccc <b>cagga</b> AUG <b>cg</b> cagug <b>cacugga</b> <b>caccggcg</b> uag <b>cggcug</b> accgc
<i>Ribosomal protein S24</i>	GL50803_10367	5' CR	auauccaagacaaa[u]aaggagcgucacccgagcucguucguAUGcagag <b>ucaagaa</b> <b>accgcguguguuu</b> uggccgcu
<i>Ribosomal protein S29a</i>	GL50803_39483	5' UTR/CR	agucuucgacUGAgucugcaugcacaaggacgagggcuuu <b>acuguaaaggAGU</b> <u>GAA</u> caccgauuu[u]auuguuacgcu
<i>Ribosomal protein P1B</i>	GL50803_17337	5' CR	
<i>Ribosomal protein P1B</i>	GL50803_17337	3' UTR	

---

Delineation of the untranslated regions of certain sequences encoding for ribosomal proteins in *G. duodenalis*. At the 5' end: CAATTT element: shading. T/A-rich region: italics. Transcription start site: brackets. Start codon: Uppercase. IRE-like: bold letters. At the 3' end: Stop codon: Uppercase. PolyA signal: Uppercase, italics and underline. Transcription stop site: brackets. IRE-like: bold letters. CR: Coding region.

**Table S3.** Ends of sequences coding cytoskeleton proteins

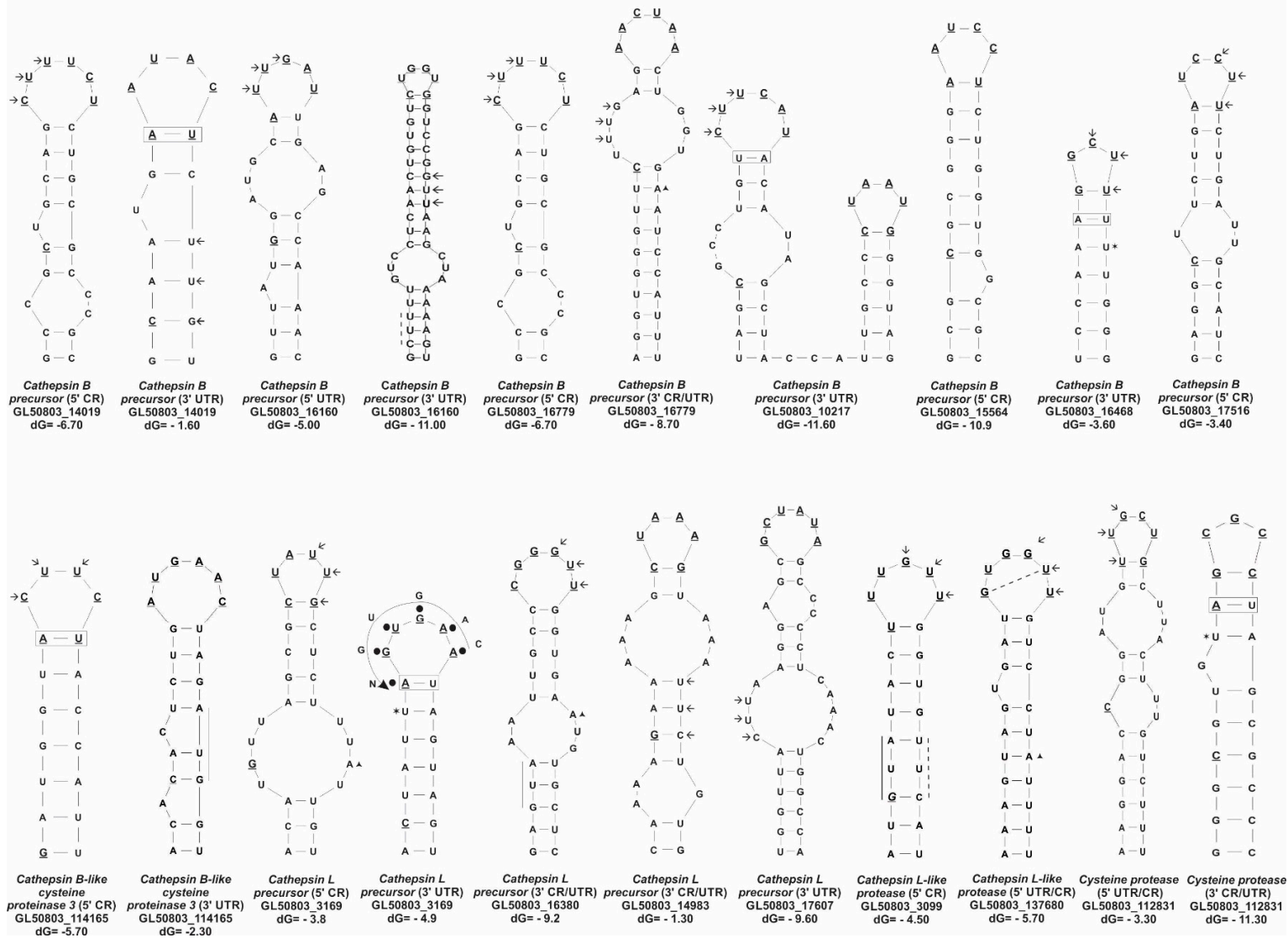
Annotation or common name	Gene ID	Region	Sequence delineated
<i>Actin-related protein</i>	GL50803_8726	3' CR/UTR	ccgaaaccaguugccuu <b>ugagguuUGAaaggccucgugucu</b> <u>AUGGUAU</u> au acuauu[u]ccgaugaauaaucgcu
<i>Actin-related protein</i>	GL50803_16172	3' CR/UTR	ugaaaucuu <b>aaugaaagugauaaugUGAcacucuuacaaaagacucuu</b> agagc ggaucaucguugcccucauacuugucuc <u>UGUAAGC</u> gauccugccuuu[u] acagcacuu
<i>Actin-related protein</i>	GL50803_11039	3' CR/UTR	gaaaaggcaacaa <b>agUA</b> <u>Aguacauaagagu</u> aa <u>uaacu</u> aucuaucuaagaua uauaguguggggugcauguuuuuu <u>AGUGAA</u> uccaacagaacaaaugcggg accuuguauacauuaagccuauaaaauauggaaucauuu[u]aguuuuaua
<i>Hypothetical protein</i>	GL50803_17591	3' UTR	cuggacuggaacacaac <b>UGAuuuuuaggauu</b> <u>AGAAUA</u> Aaccguggaagau[ u]cugcuugcuauuuuuuggcuuugcu

Delineation of the untranslated regions of certain sequences encoding for ribosomal proteins in *G. duodenalis*. At the 5' end: CAATTT element: shading. T/A-rich region: italics. Transcription start site: brackets. Start codon: Uppercase. IRE-like: bold letters. At the 3' end: Stop codon: Uppercase. PolyA signal: Uppercase, italics and underline. Transcription stop site: brackets. IRE-like: bold letters. CR: Coding region.

**Table S4.** Ends of sequences coding metabolic proteins

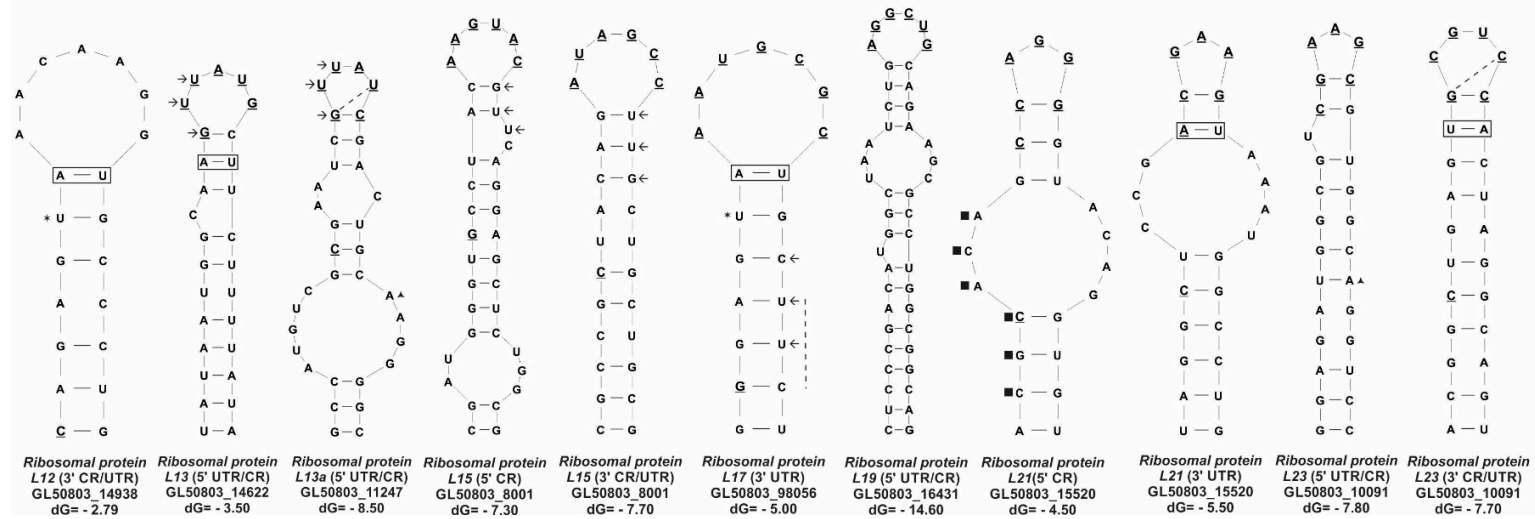
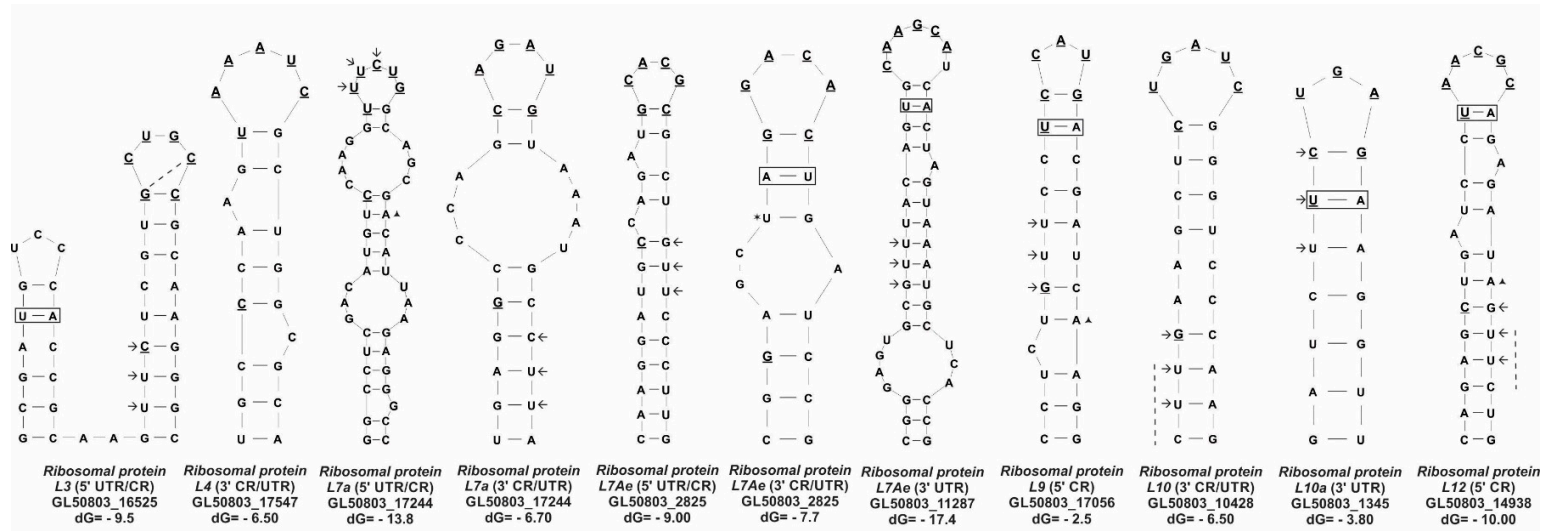
Annotation or common name	Gene ID	Region	Sequence delineated
<i>Glucosamine-6-phosphate deaminase</i>	GL50803_8245	5'CR	ccucgcaaaa[u]aaaaAUGccguccauccacgucuccaaa <b>ugcgcu</b> gauccagcca <b>uaaagcuugcccaccguauc</b>
<i>Pyruvate-flavodoxin oxidoreductase</i>	GL50803_17063	5'CR	cgaagugaaauacaacauuuugccccaauaa[a]uuuaaAUG <b>ucugugguccacgc</b> <b>gccuauagacggc</b> ugucacgcg
<i>Pyruvate-flavodoxin oxidoreductase</i>	GL50803_17063	3'CR	cuacaagaaguacagcgaaugcaggcgac <b>ccgccaaggcggag</b> UAAgc <u>AGUGA</u> <u>A</u> cuuuc[u]uucaugucugggcgg
<i>Thioredoxin reductase</i>	GL50803_9827	5'CR	acgagcauu[u]uuuaAUGuccacucagcgccacgucaggauccggaaucaucggcgg <b>aggcccugcaggccu</b> cacugc
<i>Thioredoxin reductase</i>	GL50803_9827	3'CR/UTR	aaggcggcacucgacgccgagcgcug <b>gcuugccaugcaggagUAAGUAA</b> Augg cagcgggcccug[u]ggcuccaguucagccgaaaggauagcuggggggaggggaugu

Delineation of the untranslated regions of certain sequences encoding for ribosomal proteins in *G. duodenalis*. At the 5' end: CAATTT element: shading. T/A-rich region: italics. Transcription start site: brackets. Start codon: Uppercase. IRE-like: bold letters. At the 3' end: Stop codon: Uppercase. PolyA signal: Uppercase, italics and underline. Transcription stop site: brackets. IRE-like: bold letters. CR: Coding region.

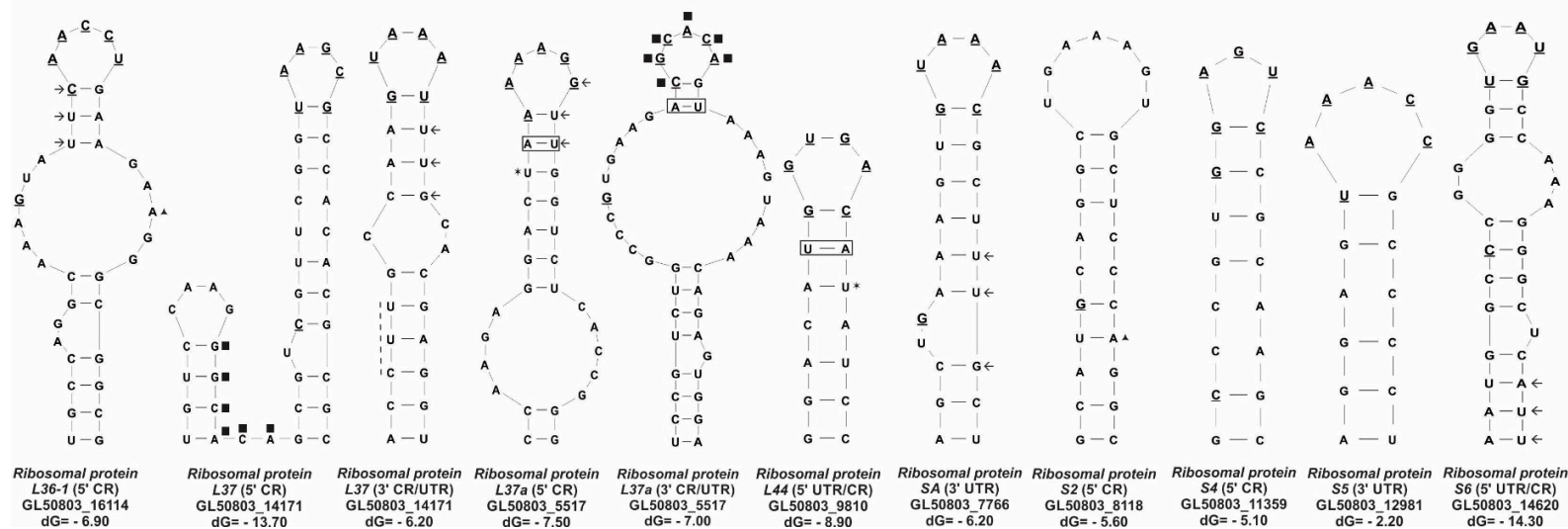
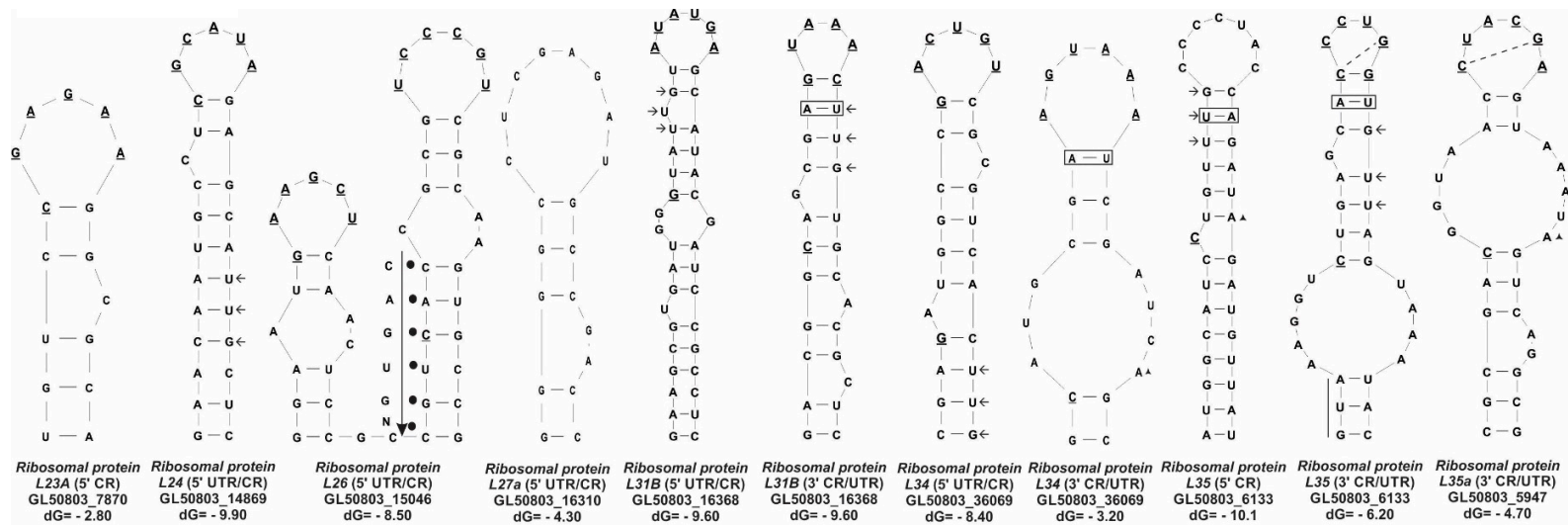


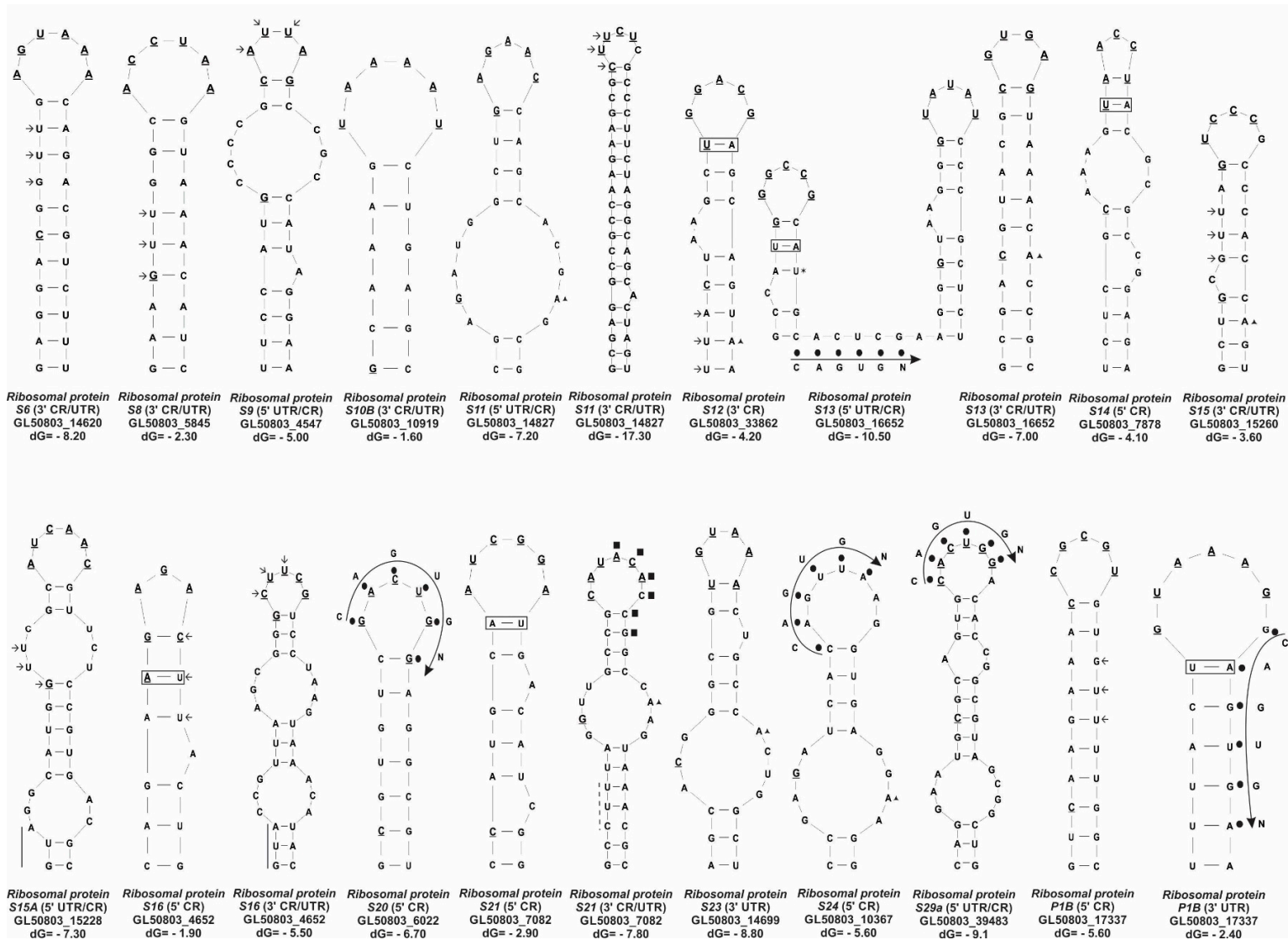


**Figure S1.** Prediction of stem-loop structures of *G. duodenalis* mRNAs coding for virulence proteins. The six nucleotides that form the apical loop of the canonical IREs are underlined. The dotted line in the loop indicates the residues at position 1-5 that form a pseudotriloop (G-C/C-G or G-U). Nucleotides U-A/A-U in rectangle indicates conserved nucleotides in the upper stem. Filled black circles indicate similar nucleotides to the CAGUGN consensus human motif and a large, curved arrow indicates the 5' to 3' or 3' to 5' orientations of the sequence. The GUU/UUG protozoa-specific motif is indicated in small arrows. Filled black squares show residues similar to that reported for the *tvcp4*-IRE loop. Asterisk: residue U conserved on canonical IREs. Small triangles show the residue A conserved in protozoa-specific motif. The residue C or G underlined indicates the unpaired residue at position 8. dG, Gibbs free energy or free enthalpy; in this case, a reduction in dG (negative dG) is a necessary condition for the spontaneity of the structure formation.

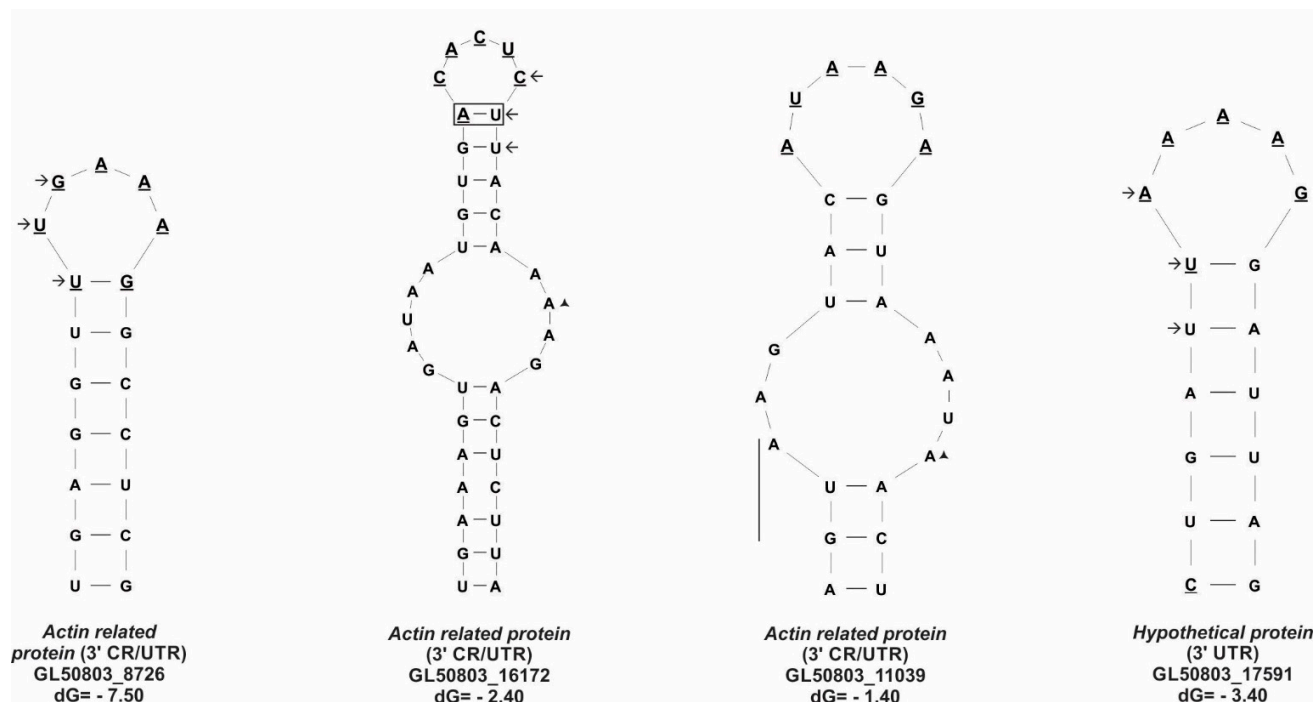




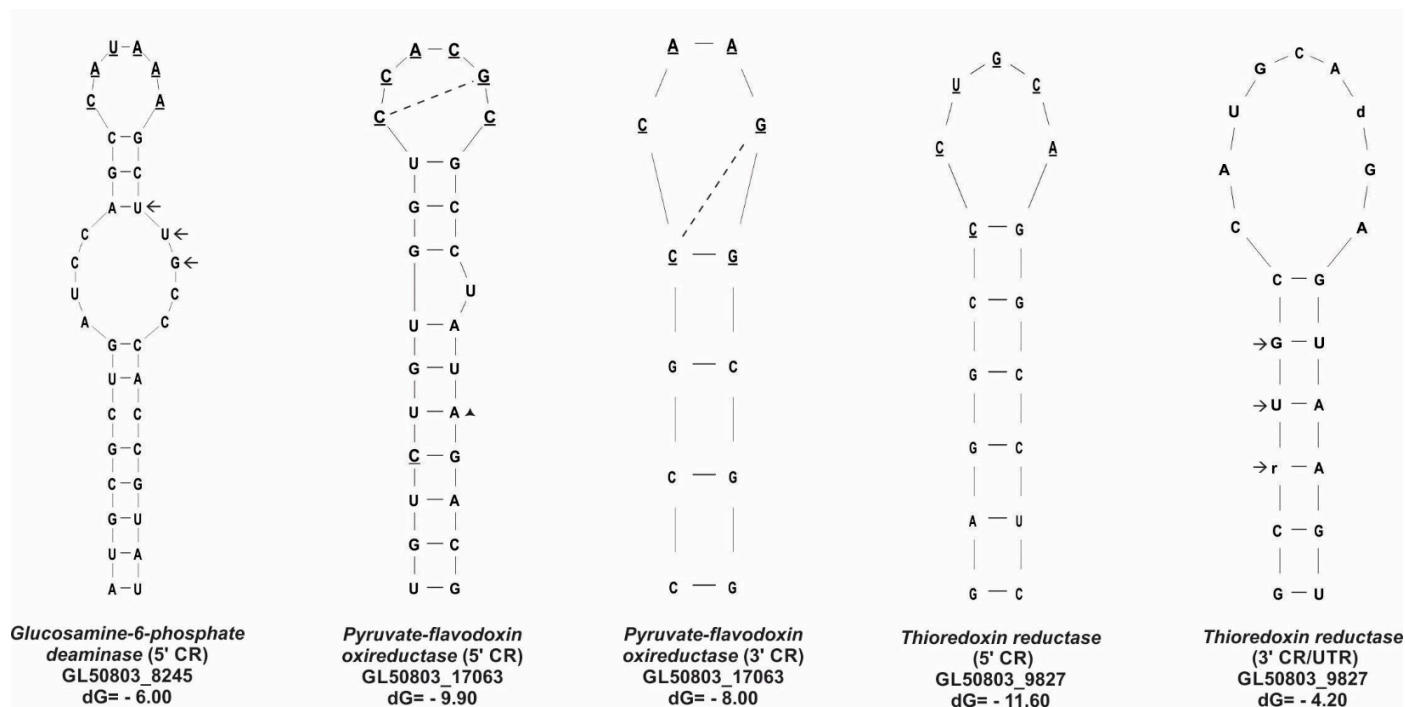




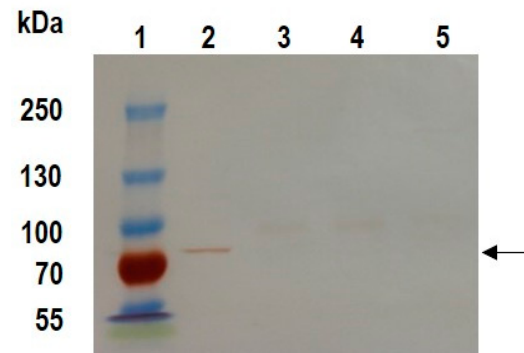
**Figure S2.** Prediction of stem-loop structures of *G. duodenalis* mRNAs coding for ribosomal proteins. The six nucleotides that form the apical loop of the canonical IREs are underlined. The dotted line in the loop indicates the residues at position 1-5 that form a pseudotriloop (G-C/C-G or G-U). Nucleotides U-A/A-U in rectangle indicates conserved nucleotides in the upper stem. Filled black circles indicate similar nucleotides to the CAGUGN consensus human motif and a large, curved arrow indicates the 5' to 3' or 3' to 5' orientations of the sequence. The GUU/UUG protozoa-specific motif is indicated in small arrows. Filled black squares show residues similar to that reported for the *tvc4*-IRE loop. Asterisk: residue U conserved on canonical IREs. Small triangles show the residue A conserved in protozoa-specific motif. The residue C or G underlined indicates the unpaired residue at position 8. dG, Gibbs free energy or free enthalpy; in this case, a reduction in dG (negative dG) is a necessary condition for the spontaneity of the structure formation.



**Figure S3.** Prediction of stem-loop structures of *G. duodenalis* mRNAs coding for cytoskeleton proteins. The six nucleotides that form the apical loop of the canonical IREs are underlined. The dotted line in the loop indicates the residues at position 1-5 that form a pseudotriloop (G-C/C-G or G-U). Nucleotides U-A/A-U in rectangle indicates conserved nucleotides in the upper stem. The GUU/UUG protozoa-specific motif is indicated in small arrows. Small triangles show the residue A conserved in protozoa-specific motif. dG, Gibbs free energy or free enthalpy; in this case, a reduction in dG (negative dG) is a necessary condition for the spontaneity of the structure formation.



**Figure S4.** Prediction of stem-loop structures of *G. duodenalis* mRNAs coding for metabolic proteins. The six nucleotides that form the apical loop of the canonical IREs are underlined. The dotted line in the loop indicates the residues at position 1-5 that form a pseudotriple loop (G-C/C-G or G-U). Nucleotides U-A/A-U in rectangle indicate conserved nucleotides in the upper stem. The GUU/UUG protozoa-specific motif is indicated in small arrows. Small triangles show the residue A conserved in protozoa-specific motif. dG, Gibbs free energy or free enthalpy; in this case, a reduction in dG (negative dG) is a necessary condition for the spontaneity of the structure formation.



**Figure S5.** Immunodetection of a cytoplasmic protein in *G. duodenalis* by an anti-human IRP-1 antibody. Western blot assays using cytoplasmic extracts of parasites grown in different iron conditions and interacting with the anti-human IRP-1 antibody. Lane 1. Molecular size markers. Lane 2. Cytoplasmic extracts under iron-restricted condition. Lane 3. Cytoplasmic extracts in normal iron concentrations. Lane 4. Cytoplasmic extracts in high iron concentrations (200  $\mu$ M). Lane 5. Cytoplasmic extracts in high iron concentrations (300  $\mu$ M).

**Table S5. IRE-like structures predicted by SIRES web server of some mRNA from *G. duodenalis***

<b>Annotation or common name</b>	<b>Gene ID</b>	<b>IRE Sequence</b>
<i>Pyruvate-flavodoxin oxidoreductase</i>	GL50803_114609	TGCCTGCCACGCGCAGAGCTACGTGAGCAAG
<i>Cathepsin L precursor</i>	GL50803_3169	GAAAGGTTCGCAACCAGGGGTCTGCGGGTGC T
<i>Cathepsin L-like protease</i>	GL50803_3099	GTAATTCCCACTGTAGAATTGGAGGAAAAC
<i>Hypothetical protein</i>	GL50803_113303	CCTTCCCCCTTGCGGGAGGGTGACGACCAGA
<i>Ribosomal protein L7a</i>	GL50803_17244	TCTTATGCGCGAGGAGCCGTCCGCTGACCGC
<i>Ribosomal protein L12</i>	GL50803_14938	GGTCTTCCGCCCAAAGTTGTTGGTGAGAAGA
<i>Actin-related protein</i>	GL50803_8726	AGCAGAGCCAGGTGGGAGCATCTAGCATGCCG

**Figure S6. Attached material showing the sequences of some mRNAs presented in our study and analyzed by SIRES web server.**

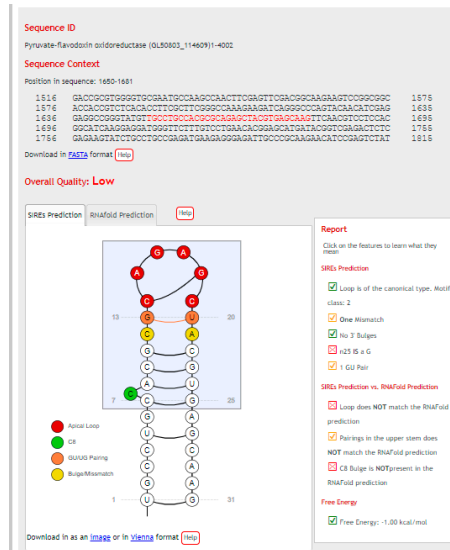
**>Pyruvate-flavodoxin oxidoreductase (GL50803\_114609)1-4002**

CCAATGGGCTAGCGAAAACCCTGGTGCCCCCTATGGTCTGGAGCAAGGCCCAAACGGGTA  
ACTTAATTTAAAGCTTAAACGCCTGTGGTCAAGACATTTTCAGAGGCCACAAAATTAAAAT  
ATGACGACCGCCCTGTCTCTCAGGAGCAGCTGTGCATCGACGGCGCGACTGCGACGTCC  
ATGATGGTGTACAGGCTGATCGACTTCACGGCCATCTTCCCCATCACCCCGTCCTCTGTC  
ATGGCCGAGATGATCGAGCAGTACGCCGCGCAAGGCAAGAAGAACATCTTCGGAAACGTC  
GTCAGTGTCCGCCAGGCCTGGTTCGGAGCTTGGAGCCATTGGCACGTGTCACGGAGCAGCG  
ACCGCAGGCGCCCTCACGGCGACCTTCAGCTCGTCGCAGGGGGTCTCTCTCATGATCTCT  
AACCTCTACCACATGGCTGGCGAGCGCATCCCCTTCGTCCTGCACGTCGCCACGCGCACC  
GTCGGGATGGCAGCCACCTCTCTCGCTACAGATCATACGGACGTCTACGCAATCGAAGGC  
ACGAACATGTGCGCCCTGTCCAGCTGCAACGTCCAGGAGTGCTACGACATGTCTCTGGTT  
GCCCACACGACTGCTACGAATGTCCAGGCCGCCTTCTGCCATTTCTTCGAGGGATATCGT  
GTCTCCCAACGATTCGAGACGATCCAGGTCTTCAAGGACGACGCAGACATCAAGGACTTC  
CTCAGCGAGGACCAGCTTGCCAAGTGGAGGAGCGAGAACGCCATGGACACTGATCGCCCG  
ATCGCCTCGAACATCGGTCTCGGAGGCGAGATCCACATGCAGATAGTTGAGTCCCAGGGT  
GCTGTCTTCGACTCTATCCCGTACCGTATCGAGGAGGCCTTCGACCTGCTCCACCGGAAG  
ACTGGCCGCGGTTACCACTGCTTCGAGTACTTCGGCCACCCGACCCCGATGTTGTTGTC  
GTCGTCATGGGCGCAACCAGCCTCACTTTTGA AAAACGTCGTCGACGCGCTTGTTAACCGC  
GAGAACCGCCGCGTTGGCTTTGTCAGAGTCCGCCTCTTCAGGCCGTTTCAGCATCGAGCTC  
TTCAGCAACGCTATTCCCAAGTCTGCGAAAGTTGTTGTGGCCCTCGATCGCGCACCAGAG  
CTCCTGAACGCCGGCGGTACGCTCTACCGCGAGACTCTTGCCGCTCTCATGAAGACTGGC  
CGCCTTGCGCGGGCAAGAACTCTGTCCGCCTCGTTGCTGGCGGGCGCTACTCGTACTAC  
GGATGGGAGCTCGGGCCCGCAGACGTGGTCACTCTCTATGAGACGTTTCGACGTCAGCACG  
CGCCACGTTCGAGAAGATGCCGACGGAATTCGTGCTCAGCATCAAGGACGATATCCGCAAG  
CGCTCTCTCCGCCCCGCGCTCCCGGAGCACGTCCCGAAGAGCCTCCTTCAGAGGGCACC  
ACGGAGTGCATCCTCTGGGGTCTCGGCTCAGACGGCACCATTGGTGCCTGCCGCAACGCC  
ATGAAGATCCTCTCCGACCGCGTGGGGTGCGAATGCCAAGCCAACTTCGAGTTCGACGGC  
AAGAAGTCCGGCGGCACCAACCGTCTCACACCTTCGCTTCGGGCCAAAGAAGATCAGGGCC  
CAGTACAACATCGAGGAGGCCGGGTATGTTGCCTGCCACGCGCAGAGCTACGTGAGCAAG



TTCAACGTCCTCCACGGCATCAAGGAGGATGGGTTCTTTGTCCTGAACACGGAGCATGAT  
ACGGTCGAGACTCTCGAGAAGTATCTGCCTGCCGAGATGAAGAGGGAGATTGCCCCGAAAG  
AACATCCGAGTCTATGCCGTCAACGCAAATAAGGTCGCGCAGTCGGTCGGCCTTGCGCGA  
CGCATAAACACGATCATGATCCTCTTCTTCCTCAAGCTCGGGCTCAGCAAGCTTCTTGAC  
TTCGATGTCGCCTGCGAGGACATGAAGGCTGCCGCGAGGATTACCTACGCTAAGCAGAAG  
GCAGAGGTCATCGAGGCGAACGTGAAGGCCATTGACGTTGCTCGCTCCATTATAGAGCAG  
TGCCACATAGAGTACGATAAGGCGCGCTGGGTCAACGCTGACCCGAGCGAGTCGACTGCA  
AACCAAACCTACGGCCCAGAGCCTGACAAGTATGTCAAGGACATCATCTTGCTGCCGTC  
ACGAGGAAGTTCGCCAACATCGGCACCAAGGATGTGATGAAGTACAACCTCTGGGAAGATT  
CCAGGCGGCTACTCCAAGTATGAGAAGCTCGGCTCTGCCGCCAAGGTTCCCATGTGGAAG  
AAGGAGAACTGCATCCAGTGCGGTTTTCTGCTCTGTCCAGTGCCCGCACGCCGCCGTCCGC  
TGCTTCGTCATGAATGAGGACGCAAAGCCTGACAACGTCCCCGAGGACTTCGAGATGCTC  
GACATGAAGGGCAAGCTCGCTGCCGTTAACAGCGATTTCCTCGAAGATCAAGTTCCGCGTC  
CAGGTTTCGCCGCTCGACTGCAGGGGATGCGGAGTCTGTGTAGAGGCCTGTCCCAAGGAT  
GCCCTTGCCATGACCCCCATCGACACTGTCCTCGACAAGCAGCAGAAGCTCTTCGACTGG  
GCCTACGATTCTTTCAGTTCACCCAGGGCACCACAAACATCGACACGGAGAGCTGCAAC  
CTCCGCGACCTGCAGCTCCGCTTCCCGTACATGGAGTTCCCTGGCTCCTGCCCCGGCTGC  
GGTGAGAGTAACACGATGAAGATGCTGGCGACCCTCTATGGAGATTTCGATGGTCGCCTCC  
ATCGGCGTCGGGTGCTCGCTTGTCTGGATGCATTTTCGGCTACATGCGTCCGTTCAACCTT  
GACAGCGACTCCCGCGGTATTGCCGCCTGCTCGTCGCTCTTTGAGGACAACCTCTGTTTTC  
GGCTGGGGTGTCGCTCTTGATCGCGATGCCAAGCGTGCAAACCTTCGCGAGTACCTTCAA  
GCCAACGTCGAGAAGATCACTGACGAGTCCCTGAAGACTACCGTCCAGGGCTGGCTCGAC  
GCGTTCGACGACCGCAAGGCCTCCATCCGCACGTCTCGTGAGCTCCGCAGCACGGTCGGC  
ATCGAAAGGCTCGACAAGGGCGACTACAAGAACAACCCTGAGTATATCATTCCGACTGTG  
GACAATCTTGCTTCCAAGTTCCAGCAGTCCGGCGTCGATCTGGATGTTGCGAAGCACATC  
CTCCTCAACAACAACCTTATCAAGGCCTCCCATTTGGCTTGTGGGCGGTGACGGCTGG  
GCCTACGACATTGACTTTGCTGGCTTCGACCACATCGTCGCCTCTCGCGATCGCGTGCGT  
GTGCTCGTTCTCGTCAACCACTGCTTCGGTAACACGGGTGGCCAGAAGTCGAAGGCCACG  
TCGATCGGCAGCGTCATGAAGAACGCGTACTCCGGCGTCAACGTCAAGGACAAGAAGATG  
GGCGTCATGCTCATGAGCACATAACGCGACGTCTACGTGGCACAGATCTCCCTCGGCTAC  
AACCGTAACCAAGGCCCTCCAGGTCTTCAGGGAGGCAGAGCACTTTGACGGCCCCGCAGTT  
GTTCTCGCCTACTGCCCCGTGCATCTCGCACATGGTCCAGGGCGGTCTCACTCACCAGGAG

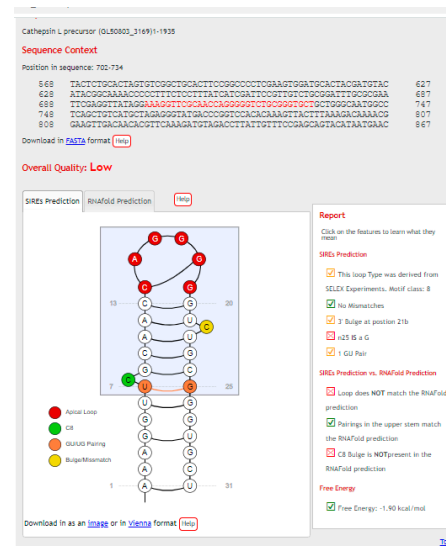
AGGCAGCAGAAGCTCGCCGTCGAGACGGGTGTCTGGCCTGTCTTCCGCTACAACCCGTCC  
AAGCCCATCGGTGACCGCATGAAGCTGGAGTACAAGGGGCCAATCAAGCCCGTCAAGGAC  
TTCATGAACACGGAGCAGAGGTTCTGGCGCCCTTGCAAAGAGGGACCCTGCCAGGTTTCGAG  
GAGATGGCTGCGGCTCTCCAGTCCGACGTTGAGGAGCACTGGAAGGTTCTTCAGGGGGCTC  
AAGGCCATCGACGTGCCCACGGACCCGATTGTTTTCGGGTAAAAAAGTGAATCATCTCGC  
GATGCTTTTTTCGGGGGGTCTTGTACGTCTCTTACGGTATCCACCTTGCTCTCCACGCTCT  
TCTGTTGCTGCTCCTCGTGGCCCAGTCTCCTTCTTGTCTTCCA



### >Cathepsin L precursor (GL50803\_3169)1-1935

AAATACTTCAAAAAAGTGTTATGCCAATATGGCATAAAATATTA AAAACTGTGGAAATTC  
ATTCTGAATTTCTCATTGAGGTAACCTTAATAGTTACAAACTGCAATCTAAGACATCTCAC  
ATGTTAGCGCTATTGCTCTTTATTGTCTGCTGTGTTTCCCATGATCCGAGTAACATGGCC  
TTTTATATTGACCCCGATAAAATCACATTTGAGACAATTCATGAAAACACAAAGACCCTG  
TTTTCTCTTTTCAACAGTCGTTTTACTTTAGACTATCCCTCAGATGAGGCTCACAACAAA  
GCCTTGCTGTCTTTCGCGAAAAATCTAGCACACCACAATGCGACCCATGCGAGGGACACG  
GACCACCTGGACGAATTGTTGAGGTATACAGATGGATATGCAACATATGGGCTGACATCA  
GCCTTTGCCAGAGACCCAACTGAGCCCATGAAGGGATTTGTCCAGGACACCGGTGTCCCC

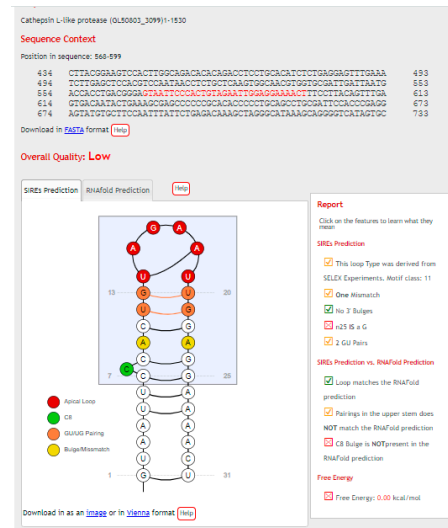
GAGTACGACTTGAGCAGCTTTAAGGACAGAATACGCTCAAGATTGAATCTGTTCAAGGAA  
GCAGCCGCTTATCAGCTCAAGGTCATCTACTCTGCACTAGTGTCGGCTGCACTTCCGGCC  
CCTCGAAGTGGATGCACTACGATGTACATACGGCAAAACCCCCTTTCTCCTTTATCATCG  
ATTCCGTTGTCTGCGGATTTGCGCGAATTCGAGGTTATAGGAAAGGTTGCAACCAGGGG  
GTCTGCGGGTGCTGCTGGGCAATGGCCTCAGCTGTCATGCTAGAGGGTATGACCCGGTCC  
ACACAAAGTTACTTTAAAGACAAAACGGAAGTTGACAACACGTTCAAAGATGTAGACCTT  
ATTGTTTCCGAGCAGTACATAATGAACAACTCTCGCAGCTCTATCAACAACTTCTGTGAA  
GGTGAAACTACAGGATCACATCGATGGATTATGCCAATGAAAACATAGTTAAGACAGTG  
GAGTCTGAAAAGAATTTTCCACTAGAATCGACTCACTTGGAGAAGAATCCATCAAGAATT  
GACGTTATAGCTAATTTCAAAGTGTCCAAGGCCTTCTTGCCTTACAACACAGAGACAGTG  
ACAGGGACCAACTGCAAAAGGTCTATCATTAAGCTGCTGGACAAGATGCAGACCTCGTCG  
CCAATGACAAACGGGGCCGTCAATCAAATCAAGTCTTACCTTGACGGGGGATCCCTGTT  
GCCGCGGCCATGTTTACGGGGGCAGACGGGCCTCTGTCCTCACGAGACTTCAGCAACTAC  
AAGGGCGGCATTTTCAATAAGCCGTGCAGCAAAACAGGGCTTGACCACCAAGTCATGTTT  
GCGGGATATGGATATTACCAGGGAGTTGAGGTCTGGGTCATGCGCAACTCCTGGGGAGAA  
CAGTGGGGGTCATACGGCCACTTCTACACACCGATAGGCAACAACGTTCTTTGCTCTGAG  
ACGTACGCATATACAGACATACCACAGTACTTTCCATTAGACAAAGCTGATAAGGACAAG  
CCATACACAGAAAGGCAAAAAACGAAAGCTTCTACCTACTGGGCGAGAAAATTAAAACGT  
GGAGATGGTGTAGCCCTCGACTCAGACCCCTCTACGATACCACCAGCCAGTCCAGGCAAT  
GATAATGACATATATGTTCCGAGAGGGAAGATCATTTCTGCTTCAAGCAGTGGAAGTACTAGT  
AGTAGCAGCAGTAGCACTACTATCACGGGCAAGTTTAACTTTCAATCCCGAGTATTATC  
ATGATTGTGACGGGCATCCTGGGGAGCGCAGGCGTTGCAGGCATTATCGCAAAGTTCACC  
TGCTGCAGGAAGTAGGTGCGGATGACTTCTAACTATTAGTGAATAGTAGTGAATAGTAAT  
AGACTAGTGGACAACCTTAGGATTGATTTTTTTTCTTTGCAATCTTCTCTCTACAAACAAT  
TTTAAATTGGAATTT



### >Cathepsin L-like protease (GL50803\_3099)1-1530

TGTAGTGTCTGATGAGGAACTACGTAAGCAGCGTGCAATACGACAGCGTGAAAAAACTGA  
 GCGGAAACAACGAAAATAGGAATTGTGTAAATAACGAAAATAATATTTAAAAGAACCATA  
 ATGTATACTTTGTTGGTGTTCACTGTCGAATCCTCTGCATATCCTATTAAGACACGC  
 TCTTACACAGGAACCTTTTCGGCATGCACTAAGGACAATGCAGTTGTTTCGTAGATTGCAA  
 TCTAACGATCTGTTTCTGCTGTTGGCTAATCCAATTAAGTGGCCGCAGAGGACAACCTA  
 CCACAATCTGTTGACTTAAGGGAGTACGGCCTTATGACTCCTGTCAGAAACCAGGGTAAG  
 TGCGGCTCGTGCTGGGCGTTTGCAACCGTTGCAGCCTTCGAGGCGTGTGCCCTGTATGAT  
 GCGGTCCTTTACCCTTACGGAAGTCCACTTGGCAGACACACAGACCTCCTGCACATCTCT  
 GAGGAGTTTGAAATCTTGAGCTCCACGTCCAATAACCTCTGCTCAAGTGGCAACGTGGTG  
 CGATTGATTAATGACCACCTGACGGGAGTAATCCCACTGTAGAATTGGAGGAAAACCTT  
 CCTTACAGTTTGAGTGACAATACTGAAAGCGAGCCCCCGCACACCCCCTGCAGCCTGCG

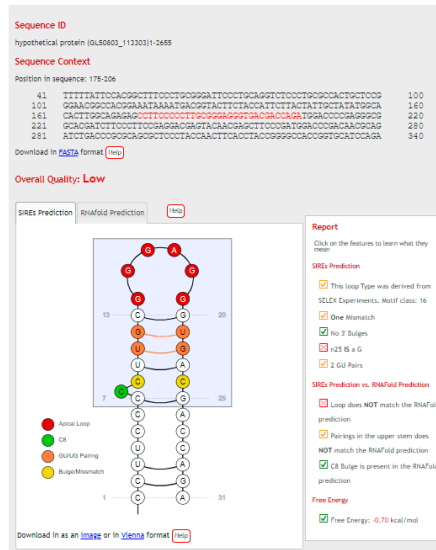
ATTCCACCCGAGGAGTATGTGCTTCCAATTTATTCTGAGACAAAGCTAGGGCATAAAGCA  
 GGGGTCATAGTGCTTTACAACAGAGACAATCCCTATAATGACTCAATTGCACAGCTGGTG  
 AAAACGTATCTGTCACTGGGGGTTCTGTGATGGCTACGATTAACACGTGGGCTAATGGA  
 AGGAAAGGTCACCGGTTAATTAAGCGTTATCGCGGTGGCATTGAGCGCACCGTGCCGT  
 AATACAAGCCTGGATCACGAGGTGCTCATTGTTGGATATGGACGGTATCGAGGTACTGCC  
 GTTTGGATAGTTCGCAATTCTTGGGGAGCTGACTGGGGAGTCTATGGGTATGCATATGTG  
 GCCCAGGGCGGTGACAGCTACTGCTTGGAGCACAGTGCCATTGCTGTTCTTCCCCGGACA  
 CTCGATCCGACCGCCAAGTACGCATCCTTCTCTACGCATCCACCTCCGTCTCCGCTCCTC  
 CGGCCGATATATGAGGGCACTATTGTACGCTGTGAAAACGGGCTCGATGCTGATCCACCG  
 GAGGGTTGCAGCCCATTCACTAGGGGCAGCTCGCCCGCGCTCGGGGCCAGCATCGCCCTC  
 AACTCACAACCATAGCAGCTATAGGGAAATTTACTACCAGCCCAGCCCCTTAATCAAC  
 CTTTCGTCGTCATGATCGCGTCGAAAATGCGAGTGTGAGGAACAAGGTCCTGAAGCAATA  
 ATCACCGAAGTGCTCGGAATTATCGGATAAGACAGTCATTATCTCGGCGAGATTCTTCTG  
 CATCATCATCTGGTTGAAGCGGCGCTCCTCCGAATTCTTCTTGAGGATGGGGAGATACTT  
 CGCCACGCTCTCTGCGAAGGACTTGATCTT



**>Hypothetical protein (GL50803\_113303)1-2655**

CTGCAGAAGAGTAAAATGCAATGTAAATAGCCTCCTCAATTTTTTATTCCACGGCTTTCC  
CTGCGGGATTCCCTGCAGGTCTCCCTGCGCCACTGCTCCGGGAACGGCCACGGAAATAAA  
ATGACGGTACTTCTACCATTCTTACTATTGCTATATGGCACACTTGGCAGAGAGCCTTCC  
CCCTTGCGGGAGGGTGACGACCAGATGGACCCCGAGGGCGGCACGATCTTCCCTTCCGAG  
GACGAGTACAACGAGCTTCCCGATGGACCCGACAACGCAGATCTGACCCGCGCAGCGCTC  
CCTACCAACTTCACCTACCGGGGCCACCGGTGCATCCAGATCATAGACCAGGGCTCGTGC  
GGCTGCTGCTACGCGGCGGCGGCGGTGGAGATGGTCACTGCCCGGCGGTGCCTCCAGCTC  
AACGACAGCAGGCTTGTGAGCCTCGAGGACCTCGTCACGTGCGACCACCAAGTATCTG  
AACATCCAGAACAACGGGTGCCGCGGCGGCAATCCCCTTGCGTCGCTGAAGTTCGGAGAG  
ACGACGGGCATGGTGTACGACACCTGCGAGGACTACTGGAACCGGACGTATCCCTACCCC  
ACGGAGACGTGCAAGACCGTCTGCAAGGATAAGCGCCCCAAGGACCGCACCATTAAGAAC  
AAGGCGCCCTACAGGCTCAGCGGCGTGGACGCGATGATGCGCGACATATACCAGAACGGG  
CCCATTGCGGTCTCCATGTACCTGGCCAACGACTTCCCCTCCAAGGACAAGAAGGGCATC

TACAGCTCCGGCCCCAATACGAAGCTGGGGGGAGGCCACGCCGTGATGATAGTCGGGTGG  
GGCGAGGAGAACGGCGTGCCCTACTGGGACTGTGCGAACACGTACGGCACCAATTGGGGC  
GACCAAGGCTACTTCAAGATCAAGCGCGGGTCCAACGAGCTCAAGATCGAGACGTGGCCC  
GGTTCAGCGCTGCCTATAGACACGACTCCTAAGCCAGAGCCCCGAAACGGCACAATCACT  
GTAGAGACGGCCAAAACGTACATTGCCGGATACGAGGTTGCAATACCGTACGAGAAGGTA  
GCCACCCCTGCAGAGCTCCAGCTGAAGTCAGGAGCAGGAGCGGTGACGTCGCTCGTTAAG  
CTGACGACCTCTAGTGGCACCGCCCGCGTCACCATACTCGAGACACGTCTGCAGGGCAA  
TACTACCTGACAACCAGCACTGGAGACGCAACGTCTGCGCGCTTCTCGTTGGCGTCCTAC  
TTCACGCTGGCATTTCGACAAGGCTAGCTATCAGGGACAAGAGGGAGAGAGCGTCGCACTC  
CGTTTCGCTGCACCGCTCCAGGTGCCCTGCAAGTTGATGGACGGCGCGTCTGTGCTGGTG  
GAGCTCAGTAAGGGGGCGCAATCCTGCATGGCCCCGTCGAGCATGAGCGCTGGCAAGCAT  
GCGCTGACCCTAACCACGTCCGACACCGTTCCCGTCTATCCGCAAGCACAGTGCTCACA  
ATAGAGAAGAAGCACAAATGATGAAGCAGTCATCAGATATCTAAGCCGGTCGCCGGCGCA  
CAGGTGACTGCGTTGGCCGAGGGCCTGCAGATTTCTGTATACCTCAACGGGGACGGGGCGCA  
AAGCTGCTTCTCCAGGCGTATTGCGGAGAGCGACTGCTGGATGTCATCGCTGCTGGCCTG  
CCTCCCTCGGGAAGCATCGAGGTCCGCCTGCCGTCTTCTATGGCAACTGCTCAGCTGCG  
TACCTCCGATTGCGCTCAGAGACGTCCCCCTTCGCGTACGCCGATGTGGGACCGCTGCAT  
GTCACGTCGTACAGCTACCCTGCAGATAATCTCCCGGCCCGTCTGATAAGAAGGATCCC  
GTGCCCCGACCCGGAACCACAGCCCAACCCAGATCCCGAACCTGAGCCCGAACCTGAGCCT  
ATTGGCGACGTGCCCCTGCTTATTACGGAGCCGTCTACGCAGAGTGTCTGGGTCCCCGGC  
GGTGCAGTTACAATCCGATGGACCAGTAACCTAACAGCCGGTACTGAGATGACGATCCTT  
TTGTACGAAAAGGTCGGATCGAAAAGCTACCTGCGCCACACGTTACCCCGCTCTGCCCCT  
AACACGGGGCTCTATGCCGACACTCTTCCGGCCTCTGTCCCCTCTGGCCCTAACTACTTC  
GTCCGCATGCGCTCCAACAGTCCCATGTTACGAACACTTCGGATGACTTTGAAGTCAAG  
GCGAGCAGATACCTGGTCAGAGACGTTCCAGACGCCATTACTATGCGGGATCCTCTGAGC  
TTCATATCCATAGACACGGGTTTGTGTTTCCCTTGCCTCGCAGAAGGTTCTCGGTCACG  
CTAATGCAGGCGGGAAAGGAGGTTTGTGCACTTGACGGAGTCGAGGGAGTGGCAGGACTC  
AACACGATCTATCTAGATAGCTGTGCTGAGAGGATGCAGGAGCTCCGAGACAAACAATAC  
TACCTAAAACATATGCTCTCAGAAAATGGAGTGCACGGTAACGCCGACTTTCACGGTATTG  
GATGACACTCATTGACACACTCTGATGAACTAGATGGTAAATCTATCCTAAGCAAAGATG  
TTGTCGGTTACACTGGACAGACTGTCATAAGCTCCTTCCCTTCATAGCCGACCGACTTTA  
CCGAGGCTCCGAGAA

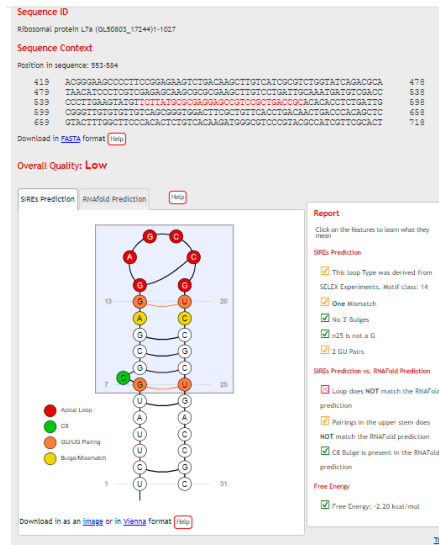


**>Ribosomal protein L7a (GL50803\_17244)1-1027**

TAGTATGTTTAAACACAATCGTAAACCGTTACTTCACCTAAATTCTATACCTGATCCCAGA  
 TTTAAAAAATAAAAACTAATTTAAATTTATTTCACTGGGTTCAATAAAACAGGCCTCGAC  
 ATGTCCAAGGTTTCTGGCAGCGACATTAAGAGGGCCCTCGCCGTACCCGAGAACAAAGAGC

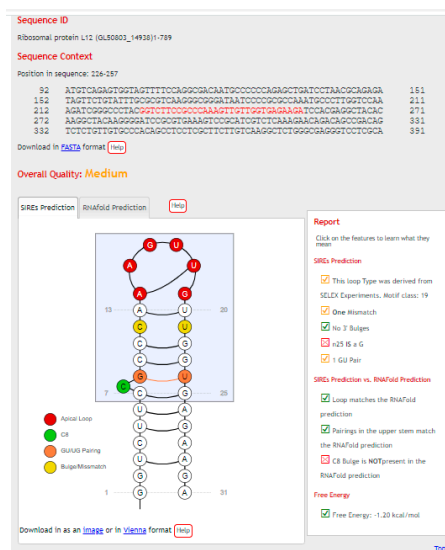


CGCAGCAAGTGC GACTTCGACCTGACTCCGTTCTGTCAGGTGGCCTCGCCAGGTCCGCATC  
CAGAGACAGAAGGCAGTCCCTCCAGAGGCGCCTCAAGGTTCCCCCAACTGTCAATCAGTTC  
ATGAATCCGATCTCGAGGAACCTCACAAACGAGATTTTCAACCTTGCTCGCAAGTACTCC  
CCCGAATCGAAGGAGGAGCACAAGGCGCGCCTGCTCCAGATCGCCGACGCAAAGGCCAAC  
GGGAAGCCCCTTCCGGAGAAGTCTGACAAGCTTGTCATCGCGTCTGGTATCAGACGCATA  
ACATCCCTCGTCGAGAGCAAGCGCGCGAAGCTTGTCCTGATTGCAAATGATGTCGACCCC  
CTTGAAGTATGTTCTTATGCGCGAGGAGCCGTCCTGACCCGCACACACCTCTGATTGCG  
GGTTGTGTGTTGTACGCGGTGGACTTCGCTGTTCACCTGACAACTGACCCACAGCTCGT  
ACTTTGGCTTCCCACACTCTGTCAACAAGATGGGCGTCCCGTACGCCATCGTTTCGCACTAA  
GGGCGATCTGGGCAAGCTCGTCCATCTGAAGAAGACGACCAGCGTCTGCTTCACCGACGT  
GAACCCAGAGGACAAGCCCACCTTTGATAAGATCCTCGCGGCCGTGGCCCATGAAGTTGA  
TTATGCAAAGGCCATGAAGACGTACGGAGGCGGCGTTTCGCCGCGAGGATGAGGCCAGCA  
GATGTAAATGCCTTACATTTCTTTAGTGCATGTCCTTTGCTGCTTGGTCCTTTGTGCTCG  
GGCTGTGCACCTTGGTAGGCCTGGTCTTGTTCGCCGTCCGCAGCGCAAGCCCAGTGACA  
CGACGCC



>Ribosomal protein L12 (GL50803\_14938)1-789

GTGGGGAGTATGCCCAGACACTTTTTACGTCTTTATTTCAATTTGTCCAATTAATAATATA  
 AATTATAGTTAAAGCATTTTAAATTTGATATTATGTCAGAGTGGTAGTTTTCCAGGCGACA  
 ATGCCCCCAGAGCTGATCCTAACGCAGAGATAGTTCTGTATTTGCGCGTCAAGGGCGGG  
 ATAATCCCCGCGCCAAATGCCCTTGGTCCAAGATCGGGCCCTACGGTCTTCCGCCCAA  
 GTTGTGGTGAGAAGATCCACGAGGCTACACAAGGCTACAAGGGGATCCGCGTGAAAGTC  
 CGCATCGTCTCAAAGAACAGACAGCCGACAGTCTCTGTTGTGCCACAGCCTCCTCGCTT  
 CTTGTCAAGGCTCTGGGCGAGGGTCCTCGCACGATTCCCAAGGGTCAGCCCCCTTCTGCAC  
 ACAGGCACAGTGAAGTTCGACACGGTCCTAGACATCGCAAAGGAGCTCCGCGCTAACTCC  
 TTCGCTCTGAAGTACGCCGGCACGGTCCTCGAGGTCCTGGGTTCTGCCCGCTCCGTTGGC  
 TGCAAGGTGGAGTACAAGGGCACTGTATATTCTCCCGCTGAGATCACGGAGATGGTGAAG  
 GCCGGGGACATAGAGATCCCAGAATATAATGTACCCTACCAGCTTGCTGAGGAGCAAGAG  
 ACAGAGTAAACAAGGTGCCCTGCGTGTACGGATCGGTCCCTATCTGCTTGGGACTGTTT  
 TCGCCTACGGGGGCTTAAAATATTTGATGGCCGAGGCGGCGGAGTCCTTTGCCGATGCA  
 GTCGTCCGT



**>Actin related protein (GL50803\_8726)1-3147**

CCTGCTCTAAATTTTCATCCAAATTTTCATTTTATAAATAAATTCTGCGACAAGTTAATTTA  
AAAGAAAAGGACATATTCTTTGGAAAGCTTTTGTATTTTGGAAATTTGACTGAAATTATAA  
ATGTCTAGACCCCTGACCTCTCTTCGTGCACACTCCGGAAGCACCTTTCAGCATTTCGGCG  
GACTTGCGTAACGAGCTCGAAGAGGCCATTCTCGCATATACACATGGATACGCCGACAAT  
CTTCGCCAGTTTTTCTTGCAATGATAAGGCCAAGACGGGGCAGCTCTCTGACATTCTG  
TTCCGCCGGGCGCTTCGCTCCATTGGTATCCACGGACCAGTTGAGCTGCTTAATCGTGCA  
ATAGAAGAGGCCACTGACCTTAAGAGTGGACTAATAGATTATGAGGCGTTTTTGGGGCTC  
TTTGCGTTTAAGGGCGACATACCCCCGAGGCTGAACGAGGAACATCAACCATTGTTTCAT  
AGCCGATACGTGGGTAATTACGCAGAGCTTGACTCTCAGAAAAGCAAGAGACCTTCCTCG  
CGTTCAGCCACAAATCAGAGACAGCATCATGACGAACCCCTCAACCTTGCTGATGTTGAA  
GGTGATCCTCTACCTCAGGAGGAAGTGATTGCCGGCGCACCTTCGAGCAGAAGCACGTTT  
AATCTGCAGGAGGAGATGCGCAAGGCAGCGGATGACTTAAAGCGCATGCAATCCGAAACA  
AATACAGAATTTCAAGTTGGATTGAAGCTCATTGCGACCCAAGTACTTCAAAAGGGCTCT  
TGTGTTAAGTTTTCCGAACGAAATTTGACAGAAACCACAAGGGAAATATAGGGGTCAA  
GACTTCCTTCTGGGTGTTTCGATACCTAGGGATACCATCTGATCTGATTAGCAATGATGAT  
CTCGTCGAAATTTTCCAATCGTACTGCACAGAGGACGGCTCCATCTCGGTCATTGAGTTT  
GAGGCTATGGTGAACGCCGGTGAAGACACCAAGGACGTTACCCGGACTATGCACGATCGT  
CAAGCAAGGTCCACCAGAGACCCGACCCATGATGTGGCTCTTCTGCTGGAGGAAGCCCAG  
ACTGTTAGTGCAAGTGGGATCTCAGGTGAAAGGAAACCCTTACGTCGCACTCAGGGCACT  
CCAGGAAGGAATCTAACTGTTTCCAATCTTGCCACAAACCAGGGATCTGATCCTGCTTGC  
TCCTCAGCGCCTAGTGTTGCTGCAAGCACAATAACCGACCATCTAAGTATAGTGATGAG  
GAAATAGTAGGAACGCTCAAGCGAAAAATTGCATTGATATGCTCTAACAATACTCCTCT  
TTTGGTGATTGCATGGCTGATGCCCAAGACTACAACAGAAAAATTACACCAGACTCGCTG  
CGCAGGTTTGTTAGCTCTCAGGTGCCAAATTTAGATAAGAGATACGTCAATCTACTCGTC  
TCATACCTGTGCACACAGGAGAATCTCAGACAGAGTGCAAATTCGCTTGGGATAAGCATA  
GCGGAAATACAGCAGAGCCAGGTGGGAGCATCTAGCATGCCGGAACCAGCCTACATATCG  
GCAGTAGTTTTCAATGACCTTTACAAGCTTCCTACCGCTACAGACGGCGGGAAGAGACGC  
ATAGATGATGCTATCTGGACATGCATCACAACAGAGGCCAATCTGAGTAACATGGCTCTT  
TCTAAGGCAGAGCTTATGGCCATATCTGTTAACAACCAGAGGTCGCAGGGGCTATACACT

GATGACCCCCCTACGCCCTGTGAGCTCCAGACCCATGCTTCGGTGTTTGGTAACCTGAAC  
ATGGCTGAGGGCCGCCAGGGTGGGTCTCAGGGTGCCGCCATGGCCCAGGAGATCGCGCAG  
CATCAGCGCCACACTAAGTCTGGAAATTACTTTGATACCACCGCGGCCCAAATGATGAC  
ATCACGTCCTACGTCAACTGCAGTAATGTCAATATTGCGAACCCCGACACCCCCTCTACT  
CGCCGTCTCAGACAGTCTGGTCGTCAAGCCGAGCTTCTAACAAGCCAGCTCCAGAACCCC  
GCCTCGTCTACTCAGGCAGAGCGCGCCAGCGAGATAATACGTACTAAAAGACAGCTCCTT  
GAGGCAATGAACAATAAGCGGAGCGCACTGGCAGAGGCCTTAATCATGGCAGCAGATCGA  
GCGACAGATCCAATATCATTAGAACTATCTCAAGGACTTTACTCAGGAGTCTCTGACT  
AATAACCAAATACAATGCATTCTTGAGGGCTCTCTGGATTCTAACGGCAAGATCCAAAAG  
GAGATCATACTCAAGACTATGGGCCTCGATAACGAGTTTATAGCATATGCAGATTCCGGT  
TTGGAGAGTTCGAAGCTGAAGAACGATCGTGTTTCTGACAGTATGGTGTGCATCTCGGCC  
GTTGCGGAGCAAATGAATGCTTCGATAAGCGGAGCGCCACGCGATATGATCTCATCCAAG  
CGTGCCATGCAGGCCAGACAAAAGCAGGAGTCATCTATAATGAGGCAGCTGTCTGCTGCA  
GACCTAGATCTTATCAAACCTCGTACTCCAAACATGAGCACAGTGGCAAAGCATCAGGAT  
AGCTCTGCACTCGCCAAGGTTGTATTGACGTAGAGAAGAATATCACTGAGGATTATGAG  
GGCTGCCTAGCAGAAGCTACCGAGGCCATTGTTGGGGGACACAGAAGGCGCCACCTTGGC  
GGTGACCTCAATAAGAAGGACAACTTGTTATGAGCAATGGAGGAGTATACTCACTGCGA  
GACAGACAACTAGAAGAAGCAGAAGCAAACACCCGACGAATCAAGAACACCCAAAGGTCT  
GTCATGGAGCACTCCACCCCGCCGATGCTATTCCACGACGTGGACGCGGGTGCAATCAG  
GCTGTATCGCAGCTAAGCAGCACAAATTGGTAACACACTTAAACAAATTGATGAAGAATAT  
CTTGTAAGATTATCAATCTGGCCAAATTGATGAAACTGCCCCTGCTGAAGGAGGAAGTGTC  
CCGAAACCAGTTGCCTTTGAGGTTTGAAAGGCCTCGTGTCTATGGTATATACTATTTCCG  
ATGAATAAATCGCTTTGTTTTTGTATGCATGAAAATTTGACTTAAAATTTCGCCCTGTG  
ATTTGCGAGGTCTATGAAACAGGCTGC

Sequence ID

Actin related protein (GL3083\_8724)-1-3147

Sequence Context

Position in sequence: 1511-1543

1477

0CTGCGCAGGTTTGTAGCTCTCAAGTGCAGAAATTAGATAAGAGATACGTCAATCTACT

1486

1487

CGTCCAGACCGTGTGACACAGGAGATCTCAGACAGATGCGAATTCCTTGGGATAG

1496

1497

CATAGCGAAGATACGCGAGACCGAGTGGAGCATCTAGCATGCCGAGACCGGCTACAT

1506

1507

ATCGCAGTAGTTTTCAATGACTTTACAAAGTCTTACCGCTACAGACGCGGGAAGAG

1516

1517

AGCGTAGATGATGATCTCTGGACATCATCAACAGAGGCCAATCTGATGACATGAC

1526

Download

in

FASTA

format

new

Overall Quality: Low

SREs Prediction

RNAfold Prediction

new

10

20

1

25

1

31

Legend

Normal Loop

CE

Unfolded Pairing

Bulge/Mismatch

Download

in

as an

image

or in

Yarns

format

new

Report

Click on the features to learn what they mean

SREs Prediction

☒ This loop type was derived from MRE Experiments, MRE class: 16

☒ no structure

☒ 3 Bulge at position 21b

☒ G is not a G

☒ 1 GU Pair

SREs Prediction vs. RNAfold Prediction

☐ Loop does NOT match the RNAfold prediction

☒ Pairings in the upper stem does NOT match the RNAfold prediction

☐ CE Bulge is NOT present in the RNAfold prediction

Free Energy

☒ Free Energy: -3.00 kcal/mol