

# Computational prediction of RNA secondary structures directing Selenocysteine incorporation in Archaea

**Andreu Bofill Pumarola**

Biologia Humana 2011-2015

Treball de Fi de Grau.

Supervisor: Roderic Guigó.

Computational Biology of RNA Processing, Center for Genomic Regulation.



## ABSTRACT

---

The UGA codon is most often translated to a stop codon, serving as a termination signal to the protein synthesis. However, in a few cases this codon can also encode the 21<sup>st</sup> amino acid, selenocysteine, which is an analog of cysteine containing selenium instead of sulfur. Complex and specific translational machinery is needed to recognize a stem-loop structure named SECIS (SEC Insertion Sequence) element, present in the mRNA of Selenoproteins and essential for selenocysteine insertion. Selenoproteins are present in the three domains of life: Eukarya, Bacteria and Archaea; but not in all their species. However, the SECIS element presents several differences between these three domains in terms of position in the mRNA sequence, stem length, secondary structure and conserved patterns. As UGA is normally a translational stop signal, Selenoproteins are normally misannotated. For this reason, dedicated annotation programs had to be developed. Seblastian (1) is a computational tool for prediction of Selenoprotein genes, based in the identification of eukaryotic SECIS as first step. It allows a quick and accurate detection of Selenoproteins in eukaryotic genomes. The annotation and information of SECIS elements in archaea species is not much extensive. In this project, I created a secondary structure alignment of Archaea SECIS which serves as a model to identify these elements in genomes. I contributed to the development of a computational tool for prediction of Selenoprotein genes in Archaea, based in the identification of the SECIS element as first step. This tool will allow the correction of missannotation of Selenoproteins in Archaea domain, and possibly the discovery of novel Selenoproteins.

## INTRODUCTION

---

Selenoproteins are a particular class of proteins that contain the amino acid selenocysteine (Sec), known as the 21st amino acid (2). Sec is a non-standard amino acid analogous to the cysteine (Cys), containing an atom of selenium instead of sulfur.

Selenoproteins are classified with selenoprotein families, which contain homologous sequences from different organisms. Some members of the selenoprotein families are homologs containing Cys instead of Sec (UGU UGC). However, the protein sequence is still conserved enough and for this reason, we expect a very similar folding to the Sec-homologs and consequently, a similar function.

Sec is the major biological form of selenium. However, other selenium-compounds are described in cells, such as selenomethionine (3). The most important difference with selenoproteins is that in this last the insertion of Sec is highly specific, and occurs in very few positions in its proteome. In contrast, other non-specific Se-containing amino acids are scattered throughout the proteome, depending on the selenium concentration (4).

The biochemical role of many selenoproteins is still not clear, but it is described that generally, they play a role in an extensive variety of redox reactions including reduction of reactive oxygenspecies (Glutathione peroxidases: GPx), reduction of thioredoxin (Thioredoxin reductases: TrxR) or oxido-reduction of iodothyroxine (Iodothyronine deiodinases: DIO) (5). It is important to remark that this large variety of redox reactions participates in a great number of cellular specific processes such as signaling pathways, thyroid hormone metabolism, reduction of oxidized proteins, redox regulation of transcription factor and apoptosis (6).

The selenol group of Sec seems to be really important for the selenoprotein function, being generally part of the enzymatic active site. The difference between selenocysteine and cysteine lies in the physiologic pH. At this pH, the selenol group of Sec ( $pK_a = 5.2$ ) is almost fully ionized, being a more reactive form compared to the thiol group of Cys ( $pK_a = 8$ ), which is protonated and consequently, less reactive (7). This difference between both amino acids implies that selenium compounds are

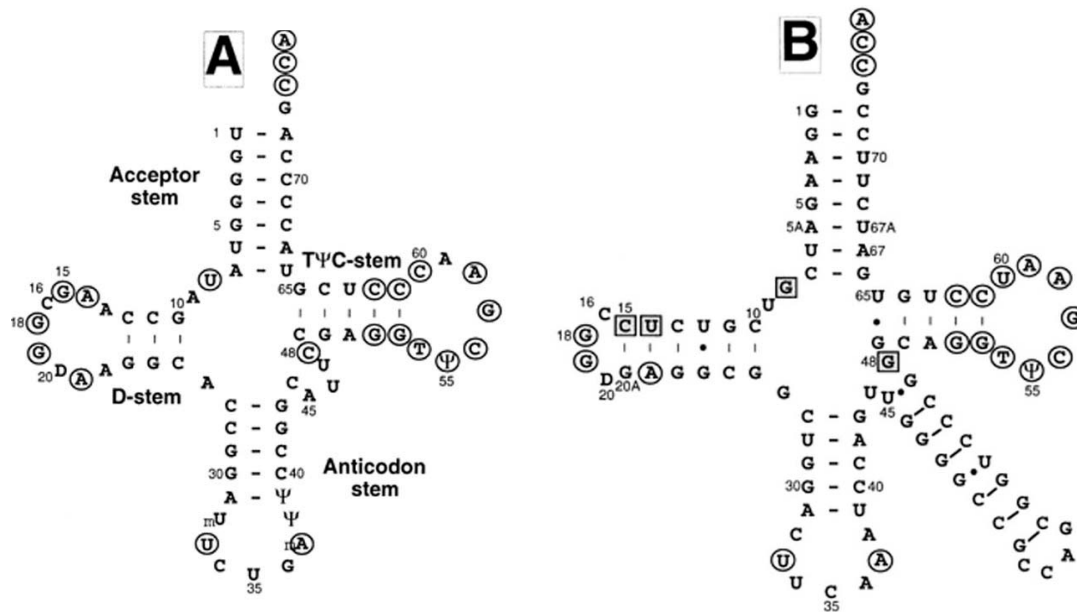
more effective antioxidants than their sulfur analogues (8). Probably this fact is crucial for understanding the relevance and the conservation of selenoproteins in some species. Other studies carried out in human cells suggest that most reductant property of selenium compounds can decrease the incidence of certain cancer, being more potent cancer-preventive agents than sulfur compounds (9–11). Some of these studies also emphasize the important function of selenium suggesting that insufficient selenium levels in humans are associated with some disease like diabetes, immune or cardiovascular disorders (11).

Selenoprotein synthesis is an evolutionarily conserved process. The three domains of life, Eukarya, Bacteria and Archaea, but not all species within domains, possess the ability to recode UGA codons from a translation termination signal to a Sec codon (12). Nevertheless, all selenoprotein-containing organisms have some common features, which are the presence of UGA-Sec codon and SECIS element in selenoprotein mRNA, and the existence of a specific tRNA for Sec (tRNA<sup>sec</sup>). However, the domains of life present several differences in the mechanism of synthesis.

Synthesis and co-translational insertion of Sec requires the recoding of an in-frame UGA codon, which generally terminates protein synthesis as a stop codon (13–15). Due to this major function of UGA codon as a stop signal, Selenoprotein genes are generally misannotated or mispredicted in genomes. There is another exception to the genetic code where a stop codon recodes to another amino acid, such as pyrrolysine. Known as the 22<sup>nd</sup> amino acid, pyrrolysine is inserted through the recoding of a stop codon (UAG) in several bacterial and archaeal organisms (16,17).

A specific and unique tRNA molecule with an anti-codon complementary to UGA mediates the synthesis and incorporation of Sec into the protein sequence. Its structure is very atypical, having a long extra-arm (Fig 1B). Hence, Sec is synthesized in its own tRNA using serine as an intermediate (8,18).

Sec biosynthesis initiates with the recognition of tRNA<sup>sec</sup> by seryl tRNA synthetase that attaches serine to tRNA<sup>sec</sup>. The next step in eukaryotes (Fig 2) and Archaea (Fig 3) is the phosphorylation of the complex, named seryl-tRNA<sup>[ser]sec</sup>, by phosphoseryl tRNA kinase to yield Phosphoseryl-tRNA<sup>[ser]sec</sup> (19). At this moment, Selenocysteine Synthase (SecS or also named SepSecS in eukaryotes and



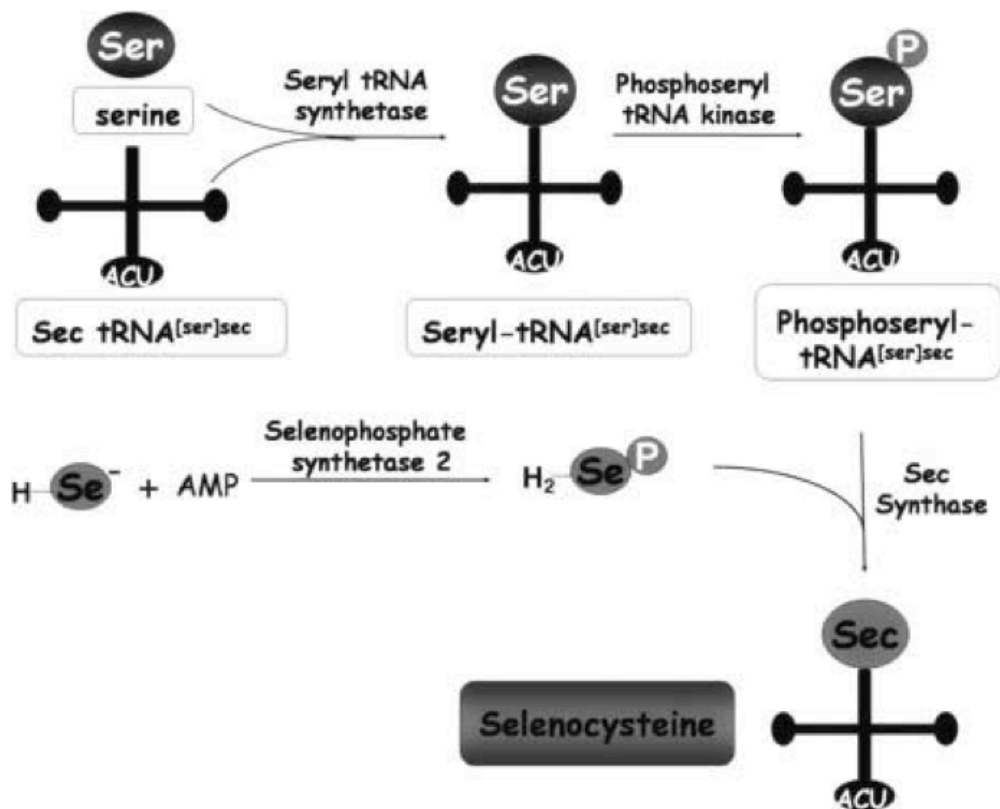
**Fig 1 | (A).** Structure and sequence of *E. coli* tRNA<sup>sec</sup>. **(B).** Structure and sequence of *E. coli* class I tRNA (left structure). Adapted from [Commans S. and Böck A., 1999](18).

archaea) uses selenium donor to convert the phosphoseryl group into selenocysteinyl ones to produce finally the Selenocysteine-tRNA<sup>sec</sup> (20). To do this step, Selenophosphate synthetase (SPS2, called SelD in prokaryotes) is needed to generate monoselenophosphate, the active selenium donor, from selenide. Curiously SPS2 family is itself a Selenoprotein family in many of selenoprotein-containing organisms. This fact makes SPS2 unique, being part of the Sec-machinery, and a selenoprotein family itself (4,8,21).

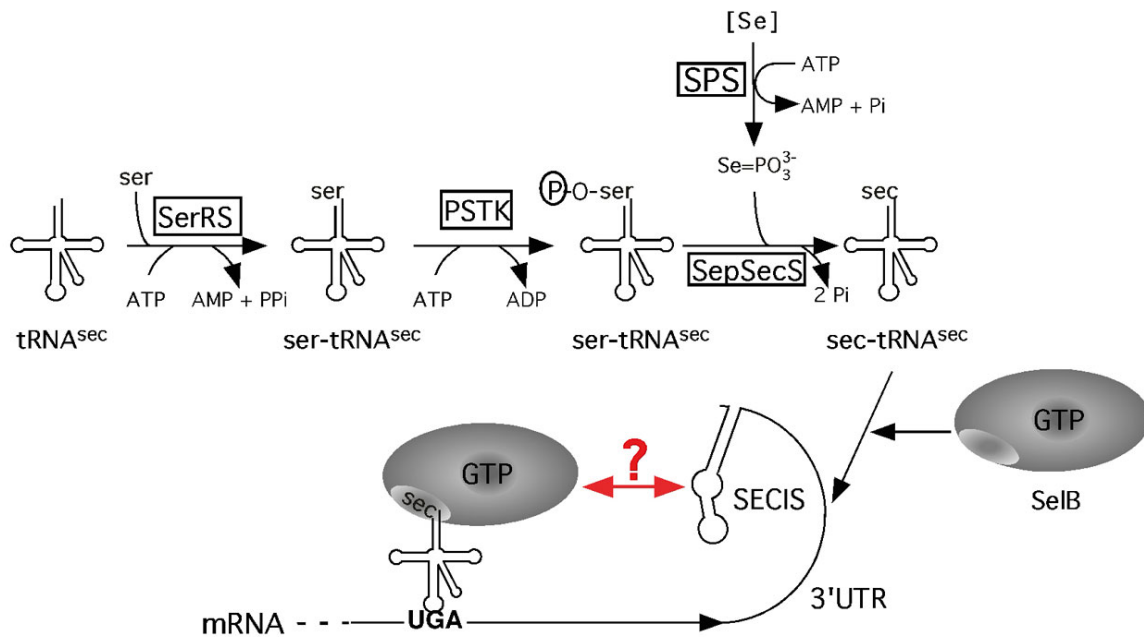
Sec specific factors are needed for Sec insertion into the selenoprotein sequence. The eukaryotic Sec specific mechanism of Sec insertion has been well characterized (reviewed in Krol A, 2002 and Lescure A. et al, 2002 (7,22)). Although the Sec synthesis in Archaea species is very similar to Eukaryotes the Archaeal Sec insertion system has some particularities (Fig 3). Archaeal SelB is an archaeal elongation factor for delivery of selenocysteinyl-tRNA to the ribosome. Apparently archaeal SelB seems not to bind the SECIS element, like in Bacteria, where SelB performs both functions, recruiting of tRNA<sup>sec</sup> and binding SECIS element. These results were very interesting because it represented the first indication that the archaeal mechanisms of Sec synthesis and incorporation would be closer to Eukarya than Bacteria (22). Archaeal SelB lacks part of the C-terminal extension existing in bacterial SelB, which is responsible for binding the bacterial SECIS (bSECIS)

element. For this reason, the function of binding SECIS element in Archaea species could not be attributed to any protein or other factor (4,23,24).

The presence of tRNA<sup>sec</sup> and Sec specific factors is essential for the Sec synthesis but it is not enough for the recoding of UGA codon and, consequently, for the Sec incorporation to the protein sequence. Selenocysteine Insertion Sequence (SECIS) is needed to recode the UGA codons to Sec. This element is an RNA stem-loop present in selenoprotein mRNA, whose secondary structure is normally highly conserved within domains. Although SECIS elements generally present low sequence similarity, short regions are extremely conserved. Some of this regions are suggested to be part of the tRNA<sup>sec</sup> binding, permitting the recognition and recode of UGA codon (7).



**Fig 2** | Sec biosynthesis pathway in mammalian cells. Eukaryotic Sec biosynthesis initiates with the attachment of Serine into Sec tRNA<sup>[sec]</sup> by Seryl tRNA synthetase. The phosphoseryl tRNA kinase phosphorylates the complex and finally this phosphoseryl group is substituted into selenocysteinyl by Sec synthase (SecS) yielding the Selenocysteine-tRNA<sup>sec</sup>. From [Papp et al, 2007](8).

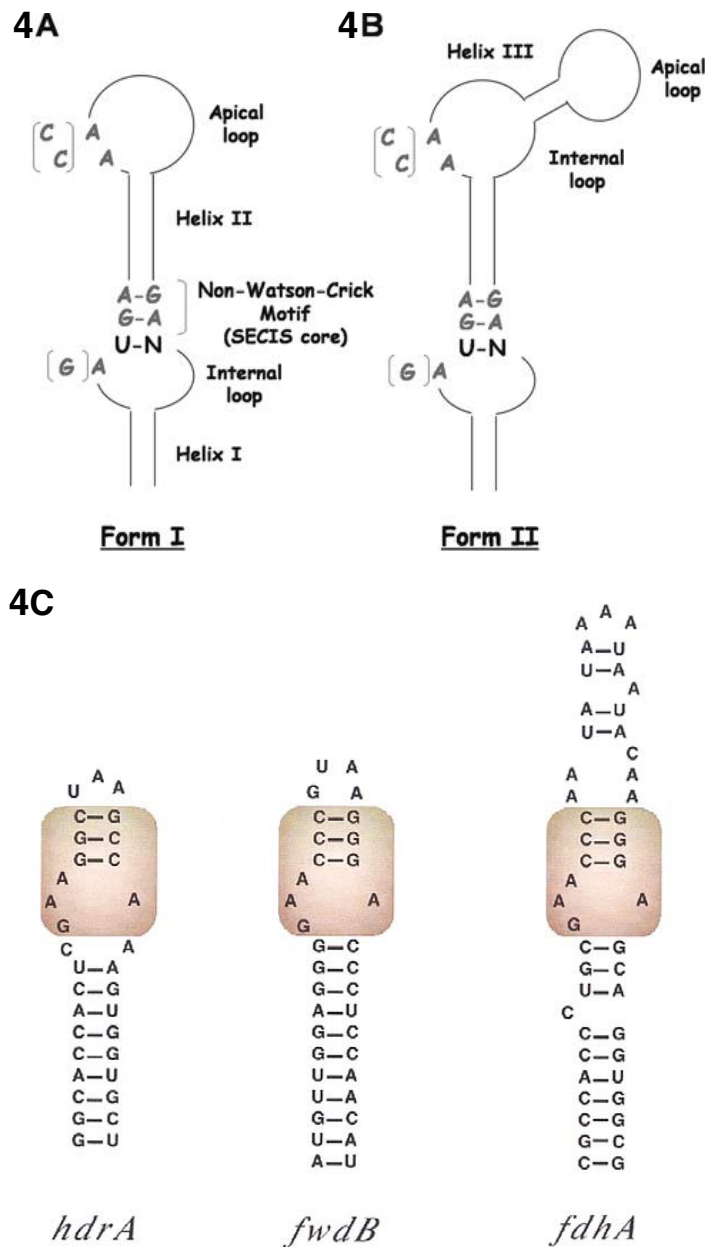


**Fig 3** | Model of Sec biosynthesis and incorporation in Archaeal genome of *M. maripaludis*.  
From [Stock and Rother, 2009](23).

The best-characterized SECIS is the eukaryotic one (eSECIS). The eukaryotic SECIS, like archaeal ones, is located within the 3' Un-translated Region (3'UTR) (25,26). eSECIS distance to UGA is quite variable, between 5200 and 50 nucleotides (4,27). It presents a short consensus sequence, where the most conserved region is a quartet of non-Watson-Crick AG base pairing known as the SECIS core, which is preceded by another A/G residue. Agreeing with this strong conservation, some studies suggest that SBP2 protein is binding to this region (28).

Two types of eSECIS exist (see Fig 4A-B), classified according to their apical loop structure. The only difference between two types lies in the existence of an extra short stem-loop on type II eSECIS top. However, both forms have the quartet at the same location, at the base of helix two. A conserved motif containing two consecutive unpaired A/C residues is located within the second loop, being the apical loop one in type I, and the second internal loop in type II eSECIS. The two types, having the same topology, are functionally equivalent. Some studies have pointed out that SECIS RNA structure itself is important for SBP2 binding function, even more than sequence (7,29).

**Fig 4** | Secondary structure models for different SECIS elements. **(A,B)**. Form I and II of eukaryotic SECIS element. From [Papp et al, 2007](8). **(C)**. Three different models of archaeal SECIS elements identified in the mRNA of three selenoprotein genes in *M. jannaschii*. Modified from [Krol A. 2002](7).



The archaeal SECIS element (aSECIS) (Fig 4C) is quite different compared with the eukaryotic and bacterial ones. The most studied archaeal species is *M. jannaschii*, where seven selenoprotein genes were identified in its genome (26). Curiously and surprisingly, aSECIS elements location is not equal to bacterial one. It is not found



adjacent to the UGA codon (30), but located in the non coding region of the gene. For six out of seven selenoproteins genes in *M. jannaschi*, the SECIS was found in the 3' UTR, while in one of them, corresponding to the formate dehydrogenase A (Fdha), the aSECIS element was found upstream of the initiation codon, in the 5' UTR.

The stem length and the sequence of aSECIS present a great variation. However, SECIS in known archaeal selenoprotein genes presented specific regions with a significant high conservation, which draws a possible aSECIS consensus. Until now, the aSECIS conservation regions consist on the presence of a GAA\_A internal loop followed by three C-G base pair just before the apical loop.

As an overview, we can see how, the knowledge of archaeal mRNA sequences and secondary structure is not much extensive. However, some conserved regions are reported for aSECIS element. This report, coupled with the fact that aSECIS are indispensable for the recode of archaeal Selenocysteine and consequently for the selenoprotein synthesis, makes the search for this archaeal mRNA-conserved elements a good starting point for selenoprotein identification.

## PROBLEM APPROACH AND OBJECTIVES

---

In a previous study from Kryukov G. et al. 2004 (31), a set of aSECIS elements in archaeal genomes was identified from selenoproteins in two archaeal genomes. Then, this set was used to perform structural alignment. All aSECIS sequences and a structural alignment was performed were aligned using their predicted secondary structure. a structural alignment was performed. From this alignment, they developed a simple model of aSECIS element with primary sequence conservation limited to the unpaired GAA\_A region. The two analyzed genomes were: *M. jannaschii* and *M. kandleri*, with exactly eight and seven selenoprotein genes respectively.

From the *M. jannaschii* genes set, they predicted two different aSECIS element in coenzyme F420-reducing hydrogenase, d-subunit and heterodisulphide reductase, subunit A. In our project, we noticed that these two aSECIS elements are in fact in the 3' UTR of the same gene, at 500 nucleotides away one for the other, so we treat this case as only one gene.

This project aims to expand the knowledge of the archaeal SECIS. The first objective is to improve the current model of archaeal SECIS. This most robust aSECIS model can be used for accurate identification of SECIS element in archaeal genomes. The ultimate goal of this project is to develop in a near future, a new version of Seblastian program able to identify selenoproteins in archaeal species based on the improved model.

## MATERIALS AND METHODS

---

### **Infernal. A SECIS prediction method**

Infernal (Inference of RNA Alignment) (32,33) is a bioinformatic program that 'builds probabilistic profiles of the sequence and secondary structure of an RNA family called covariance models (CMs) from structurally annotated multiple sequence alignments given as input' and uses these profiles to search homologous RNAs in nucleic acid sequence databases (33). Infernal also uses CMs to generate new multiple sequence alignment.

Considering the relevance of both secondary structure and sequence on SECIS, Infernal can be used to obtain a CM model of aSECIS element and search for homology in archaeal genomes. Initially we obtained the starting set of 13 aSECIS elements sequence from the previous literature (from only two species: *M. jannaschii* and *M. kandleri*) (31) and we build a first rough alignment with *cmalign* Infernal program. At this point we evaluated the first alignment using RALEE, which is an editor of RNA alignment (34).

The program Selenoprofiles, a homology-based annotation pipeline for the identification of selenoproteins (35), was used to identify selenoprotein genes in archaeal genomes. A total of 59 selenoprotein genes were identified. Since the aSECIS is located in the UTR region of the gene, we extended our predictions, which only included the coding sequence, 2000 nucleotides upstream and downstream. This set of selenoproteins genes was used as database to collectt additional SECIS sequences, using the Infernal program *cmsearch* to search it using the current model.

Infernal uses two different score types: a bit-score, giving an indication of how well the hit fits the model; and an E-value, which is a probabilistic approach 'expressing how many alignments with the same or better bit-score are expected by chance searching the current target' (1). In all these procedures we decided to use a filter based on bit-score because it is not dependent on the target size.

We needed to parse the search results and select the acceptable aSECIS elements to align these with *cmalign*. At this point we needed an acceptance criteria to ensure that those predictions that we choose as aSECIS elements truly are the real ones

existing in each one of the selenoprotein genes. First of all, we need to select only this hits which were outside the coding region. After that, the main conditions we chose were the existence of the unpaired bases GAA\_A forming an internal loop, and the presence of the consensus RNA structure that seems to be really important for aSECIS recognition (7,29). The presence of three G/C base pair after GAA\_A loop was another criteria used to discern the good aSECIS elements. However, the absence of the third G/C base pair was not necessarily an excluding criteria. Finally the candidate bit-score was also taken into account.

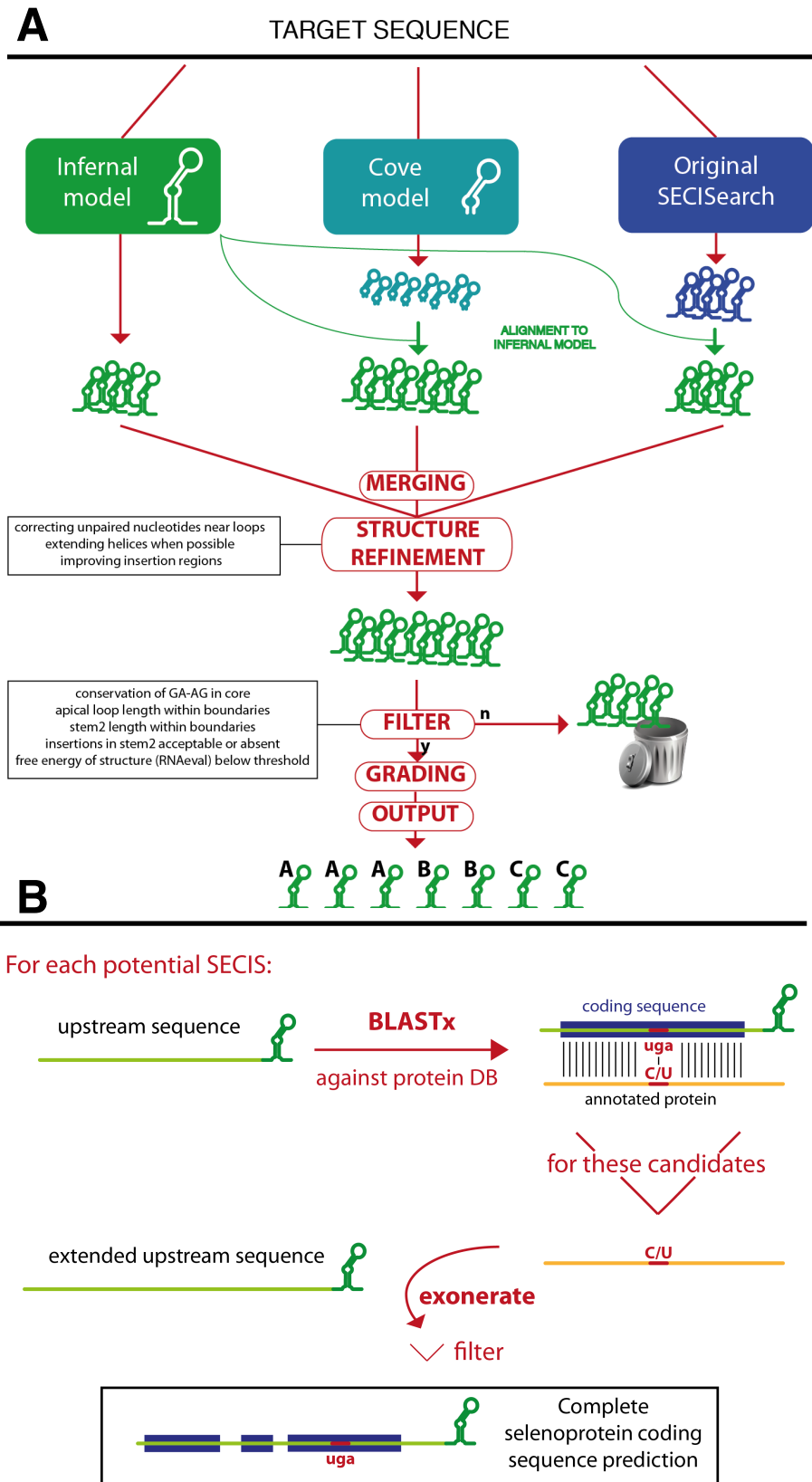
The resulting secondary structure alignment was again analyzed with RALEE to identify and extract the sequences that follow the consensus secondary structure assigned. Knowing that each gene in our database should have one SECIS element in the selected region, we could be more permissive with the results screening scores. The subset that we accepted as aSECIS, is added to the initial model set and a new Infernal model is build.

Then we iterated this procedure and for each round we inspected manually the alignment to add or exclude sequences. In every iteration the number of sequence within the model was increased, making the model more sensitive.

### **SECISearch3**

This program (Fig 5A) is focused only in eukaryotic selenoproteins and SECIS elements allowing the searching and identification of eSECIS elements in a nucleotide database (1). "It is based on Infernal model consisting of a curated, secondary structure based alignment of 1122 eSECIS elements". This method combines three sources of predictions: Infernal, Covels and the Original SECISearch. In our work to develop aSeblastian, we are going to use only the Infernal method. For more information of the others methods see Mariotti et al. 2013 (1). Then, SECISearch3 use RNAfold to structurally refine the new predictions resulting from these methods. The last step in this program is to remove all the improbable eSECIS candidates with a eSECIS filter. This filter is very specific for eukaryotic SECIS, and "checks the following characteristics: presence of kink-turn in the core, length of apical loop, length of stem2, bending of stem2 (insertions in 5' portion versus insertions in 3' portion), free energy of the structure". In aSeblastian, is important to find some specific characteristics of aSECIS that allows filtering our Infernal aSECIS candidates.

**Fig 5 |** Sebastian program Workflow. **(A)** Extended view of SECISearch3 operation ending with the SECIS potential candidates outputting. **(B)** The second part of program Sebastian: BLASTx and Exonerate parts. Extracted from [Mariotti et al, 2013].



## **Sebastian**

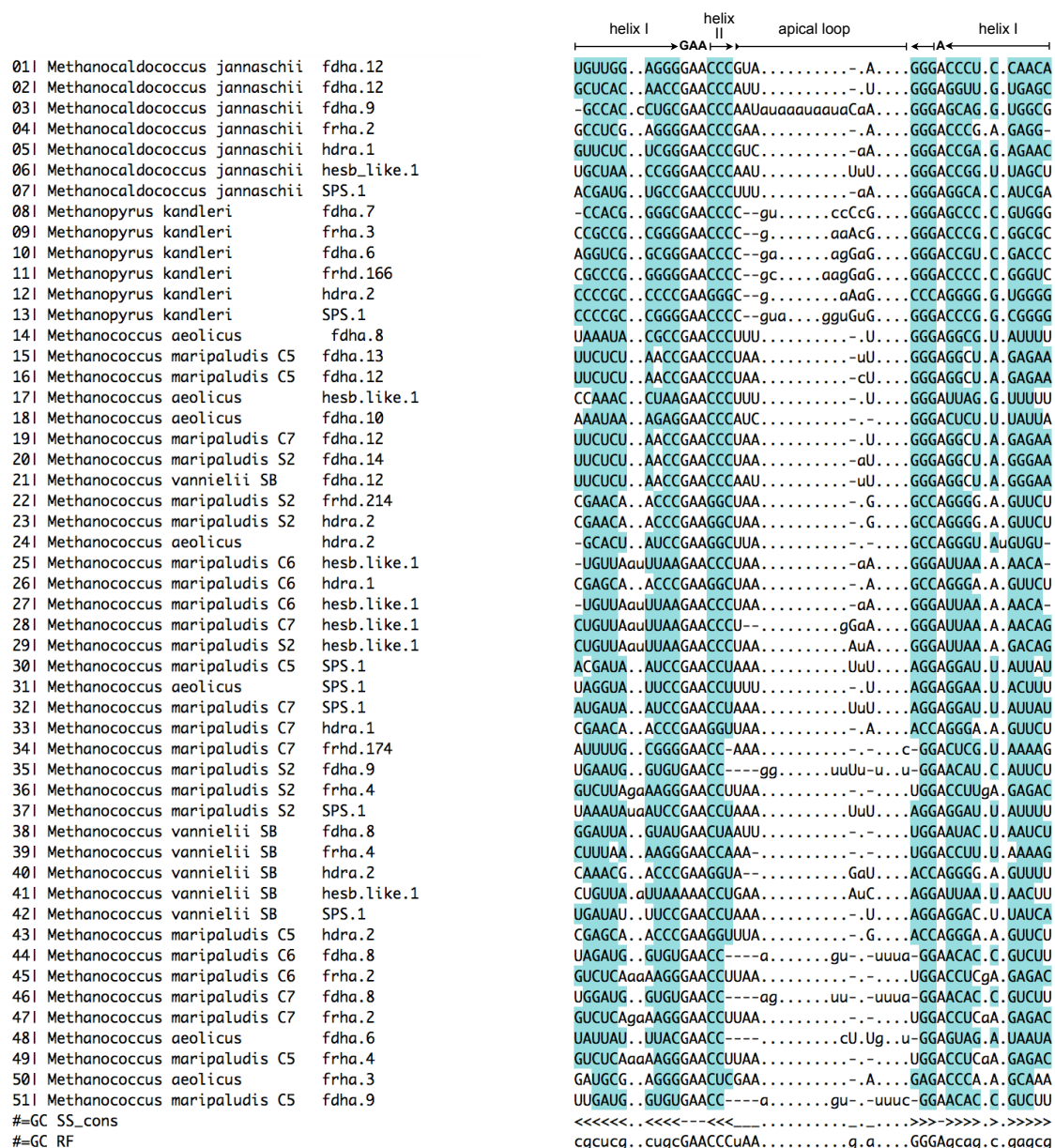
Sebastian (1) is a bioinformatic method for selenoprotein gene prediction in nucleotide sequences based on SECISearch3 results (Fig 5).

For each SECIS candidate, the sequence upstream is extracted and with this, Sebastian runs blastx against a protein database to make a gene prediction on the nucleotide sequence (36). The program parses the blastx results and consider all blast hits with an annotated selenoprotein. "Mostly two types of blast alignments are considered: those in which a Sec in a query protein is aligned with a UGA in the target sequence and those in which a cysteine in a query is aligned with a UGA in the target". It is important to remark that, in the output of this program, there are two types of Selenoprotein candidates: The known selenoproteins and the new ones homologues of known proteins. Thus, Sebastian is able to find not only annotated selenoproteins, but also new selenoproteins.

Finally, other types of blast alignments are also considered and kept to improve on sensitivity. Then, Sebastian joins all candidates that have the same query and seem to belong to the same gene. The next step in Sebastian program is to run Exonerate program (37) to refine and improve gene structures. The last candidates must pass another filter to finally complete the selenoprotein coding sequence prediction.

## RESULTS

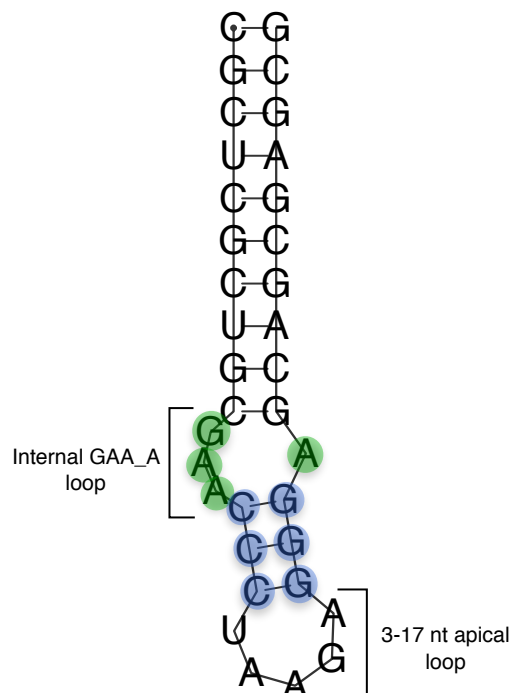
An archaeal SECIS alignment was built using the Infernal program cmbuild. 51 aSECIS elements, from our set of 59 archaeal selenoprotein genes identified by Selenoprofiles, were finally included. Nine iterations were needed to achieve this finally number of aSECIS elements included into the model. Thus, eight aSECIS element were not found with Infernal, and they could not be incorporated to the aSECIS alignment.



**Fig 6 |** Alignment of aSECIS elements using RALEE. An alignment of 51 aSECIS elements from five archaeal species were aligned using the Infernal program cmalgn. It is based on both primary sequences and secondary of aSECIS elements. Nucleotide base pair is highlighted in blue.

The final aSECIS alignment is shown in the Fig 6. It contains aSECIS elements from 50 selenoprotein genes: six in the *M. jannaschii* genome, six in the *M. kandleri* genome, seven in the *M. aeolicus* genome, in the strain Nankai, six in *M. vannielii* genome and 25 in *M. maripaludis*, for which the genomes of 4 different strains are available (C5, C6, C7, S2). However, the total number of aSECIS element in the alignment is 51. The explanation of this difference between the number of alignment aSECIS elements (51) and the number of selenoprotein genes from which aSECIS were extracted (50) is that, the *M. jannaschii fdha* gene contains two SECIS elements in its 3' UTR, both with good SECIS characteristics.

The aSECIS consensus nucleotide sequences and secondary structure are presented in the Fig 7. The conserved region GAA\_A remains in the internal loop, and is followed by three C-G base pairs. Our current model developed from this project has a variable apical loop length of 3-17 nucleotides.



**Fig 7 |** Reference sequence and secondary structure of the archaeal SECIS element model.

Conserved GAA\_A region in the internal loop are highlighted in green and the conserved C-G base-pairs in yellow. Other important structural information is properly annotated.



## CONCLUSIONS

---

We have developed a new archaeal SECIS model based on the alignment of 51 aSECIS elements. This archaeal SECIS model is an improvement of the existing one, based exclusively in thirteen aSECIS elements.

Interestingly, the conserved aSECIS regions and the archaeal SECIS element consensus of our model are similar to the previously reported (26,31), but the number of aSECIS elements included in our model is considerably higher. Therefore, the fact that the new model has a very similar consensus aSECIS element, based on both sequence and secondary structure, and almost equal conserved regions, compared to the previously reported models, gives important validity to our model. In parallel, the inclusion of 38 new aSECIS elements provides greater robustness and accuracy of this model.

This model can be used for more accurate and exhaustive identification of SECIS element in archaeal genomes. Thus, the identification and acceptance of new aSECIS candidates will allow the incorporation of these new aSECIS elements to the existing model. This will contribute to further improve the robustness of the model.

The application of this model in a positive set of known archaeal selenoprotein genes seems to indicate that this is a good prediction model of aSECIS element. Even so, in the immediate future, with the application of this model to other archaeal databases, we need to evaluate it more exhaustively and try to amplify it with the most number of selenoprotein-containing genomes.

Currently, we are working in the development of a new version of Sebastian program that runs on Archaea species. This archaeal version (aSebastian) will be able to predict archaeal selenoprotein genes based on the improved aSECIS model as first step.

## REFERENCES

---

1. Mariotti M, Lobanov a. V., Guigo R, Gladyshev VN. SECISearch3 and Seblastian: new tools for prediction of SECIS elements and selenoproteins. *Nucleic Acids Res.* 2013 Jun 19;41(15):e149–e149.
2. Böck a, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, et al. Selenocysteine: the 21st amino acid. *Mol Microbiol.* 1991;5(3):515–20.
3. Whanger PD. Selenocompounds in plants and animals and their biological significance. *J Am Coll Nutr.* 2002;21(3):223–32.
4. Mariotti M. Computational genomics of selenoproteins.[Doctoral Thesis] Universitat Pompeu Fabra. Barcelona 2013.
5. Zhong L, Arnér ES, Holmgren a. Structure and mechanism of mammalian thioredoxin reductase: the active site is a redox-active selenolthiol/selenenylsulfide formed from the conserved cysteine-selenocysteine sequence. *Proc Natl Acad Sci U S A.* 2000;97(11):5854–9.
6. Roman M, Jitaru P, Barbante C. Selenium biochemistry and its role for human health. *Metallomics.* 2014;6(1):25–54.
7. Krol A. Evolutionarily different RNA motifs and RNA–protein complexes to achieve selenoprotein synthesis. *Biochimie.* 2002 Aug;84(8):765–74.
8. Papp LV, Lu J, Holmgren A, Khanna KK. From Selenium to Selenoproteins: Synthesis, Identity, and Their Role in Human Health. *Antioxid Redox Signal.* 2007 Jul
9. Ip C, Ganther HE. Comparison of selenium and sulfur analogs in cancer prevention. *Carcinogenesis.* 1992;13(7):1167–70.
10. Whanger PD. Selenium and its relationship to cancer: an update. *Br J Nutr.* 2004;91(1):11–28.
11. Rayman MP. Selenium and human health. *The Lancet.* 2012. p. 1256–68.
12. Donovan J, Copeland PR. Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins. 2009;2(Form 2):1–18.
13. Hoffmann PR, Berry MJ. Selenoprotein synthesis: a unique translational mechanism used by a diverse family of proteins. *Thyroid.* 2005;15(8):769–75.
14. Hatfield DL, Carlson B a., Xu XM, Mix H, Gladyshev VN. Selenocysteine Incorporation Machinery and the Role of Selenoproteins in Development and Health. *Prog Nucleic Acid Res Mol Biol.* 2006;81:97–142.
15. Donovan J, Copeland PR. Threading the Needle : Getting Selenocysteine Into Proteins. 2010;12(7).
16. Srinivasan G, James CM, Krzycki J a. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science.* 2002;296(5572):1459–62.
17. Gaston M a, Zhang L, Green-Church KB, Krzycki J a. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. *Nature.* 2011;471(7340):647–50.
18. Commans S, Böck A. Selenocysteine inserting tRNAs: An overview. *FEMS Microbiol Rev.* 1999;23(3):335–51.
19. Carlson BA, Xu X-M, Kryukov G V, Rao M, Berry MJ, Gladyshev VN, et al. Identification and characterization of phosphoseryl-tRNA[Ser]<sup>Sec</sup> kinase. *Proc Natl Acad Sci U S A.* 2004;101(35):12848–53.
20. Palioura S, Sherrer RL, Steitz TA, Söll D, Simonovic M. The human SepSecS-tRNA<sup>Sec</sup> complex reveals the mechanism of selenocysteine formation. *Science.* 2009;325(5938):321–5.
21. Xu XM, Carlson BA, Mix H, Zhang Y, Saira K, Glass RS, et al. Biosynthesis of selenocysteine on its tRNA in eukaryotes. *PLoS Biol.* 2007;5(1):0096–105.

22. Lescure A, Fagegaltier D, Carbon P, Krol A. Protein factors mediating selenoprotein synthesis. *Curr Protein Pept Sci.* 2002;3(1):143–51.
23. Stock T, Rother M. Selenoproteins in Archaea and Gram-positive bacteria. *Biochimica et Biophysica Acta - General Subjects.* 2009. p. 1520–32.
24. Rother M, Wilting R, Commans S, Böck A. Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*. *J Mol Biol.* 2000;299(2):351–8.
25. Rother M, Resch A, Wilting R, Böck A. Selenoprotein synthesis in archaea. *Biofactors.* 2001;14(1-4):75–83.
26. Wilting R, Schorling S, Persson BC, Böck a. Selenoprotein synthesis in archaea: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion. *J Mol Biol.* 1997;266(4):637–41.
27. Martin, G.W., Harney, J.W. and Berry M. Selenocysteine incorporation in eukaryotes: insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element. *RNA.* 1996;2:171–82.
28. Fletcher JE, PR, Driscoll DM, Krol A. The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA.* 2001;7(10):1442–53.
29. Latrèche L, Jean-Jean O, Driscoll DM, Chavatte L. Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine. *Nucleic Acids Res.* 2009;37(17):5868–80.
30. Rother M, Resch A, Gardner WL, Whitman WB, Böck A. Heterologous expression of archaeal selenoprotein genes directed by the SECIS element located in the 3' non-translated region. *Mol Microbiol.* 2001;40(4):900–8.
31. Kryukov G V, Gladyshev VN. The prokaryotic selenoproteome. *EMBO Rep.* 2004 May;5(5):538–43.
32. Nawrocki EP, Kolbe DL, Eddy SR. Infernal 1.0: Inference of RNA alignments. *Bioinformatics.* 2009;25(10):1335–7.
33. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013;29(22):2933–5.
34. Griffiths-Jones S. RALEE - RNA alignment editor in Emacs. *Bioinformatics.* 2005;21(2):257–9.
35. Mariotti M, Guigó R. Selenoprofiles: Profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *Bioinformatics.* 2010;26(21):2656–63.
36. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research.* 1997. p. 3389–402.
37. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005;6:31.