

# A Method for Identification of Selenoprotein Genes in Archaeal Genomes

Mingfeng Li, Yanzhao Huang, and Yi Xiao\*

*Biomolecular Physics and Modeling Group, Department of Physics, Huazhong University of Science and Technology, Wuhan 430074, China.*

\*Corresponding author. E-mail: yxiao@mail.hust.edu.cn

DOI: 10.1016/S1672-0229(08)60034-0

**The genetic codon UGA has a dual function: serving as a terminator and encoding selenocysteine. However, most popular gene annotation programs only take it as a stop signal, resulting in misannotation or completely missing selenoprotein genes. We developed a computational method named Asec-Prediction that is specific for the prediction of archaeal selenoprotein genes. To evaluate its effectiveness, we first applied it to 14 archaeal genomes with previously known selenoprotein genes, and Asec-Prediction identified all reported selenoprotein genes without redundant results. When we applied it to 12 archaeal genomes that had not been researched for selenoprotein genes, Asec-Prediction detected a novel selenoprotein gene in *Methanosarcina acetivorans*. Further evidence was also collected to support that the predicted gene should be a real selenoprotein gene. The result shows that Asec-Prediction is effective for the prediction of archaeal selenoprotein genes.**

**Key words:** archaea, selenocysteine, selenoprotein, SECIS, SelB, gene prediction

## Introduction

Genetic code rules the translation from mRNA to protein. Sixty-one codons encode twenty amino acids and three remained codons (UGA, UAG and UAA) act as terminators. However, more and more works challenge this universal rule. Some reports declared that UGA codon has a dual function as it not only serves as a terminator but also encodes selenocysteine (Sec), the twenty-first amino acid in protein (1–5). The same phenomenon has also been found for UAG codon, which had been identified translating to pyrrolysine (Pyl) in several methanogenic archaeas (6, 7). However, most available gene prediction programs only annotated UGA codon as a terminator, which may lead to misannotation or completely missing selenoprotein genes in databases (8–11). Although several programs have already added the dual meaning of UGA codon in prediction, for example, the fully functional prediction for mammalian and insect genomes in Castellano's work (12), their applications did not extend to archaeal genomes.

Translating UGA codon to Sec is linked to two components, one is the Sec insertion sequence (SECIS) element, which is defined by characteristic nucleotide sequences and secondary structure base-

pairing patterns, and the other is the trans-acting factor SelB, which is a GTP-dependent elongation factor specific for Sec incorporation (13). The interaction mechanism of SelB with the Sec synthesis has been elucidated extensively in the literature (14, 15). The SECIS elements locate at different positions relative to the coding genes in three main domains of life. In bacteria, the SECIS elements are presented immediately downstream of UGA codon and are part of the coding regions (16, 17), whereas in archaea and eukaryotes the SECIS elements usually locate at the 3'-untranslated regions (3'-UTRs) (18–24). Gladyshev and co-workers have comprehensively studied their different features (17, 22–24). In archaea, they investigated 14 genomes by developing previous studies of Böck's group (20, 25, 26) and gave a consensus model of the SECIS elements. Using this model, they successfully predicted 15 archaeal selenoprotein genes (23). Currently, besides the 14 archaeal genomes, other 12 archaeal genomes have already been completely sequenced and deposited in the NCBI database, but their UGA codon misannotations have not been studied. In this study, we developed a computational method, named Asec-Prediction, to iden-

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

tify the potential misannotated selenoprotein genes in these genomes. Since Asec-Prediction properly adjusts previous prediction strategies, it achieved a quick and correct prediction. Moreover, Asec-Prediction successfully identified a novel selenoprotein gene in *Methanosarcina acetivorans*.

## Method

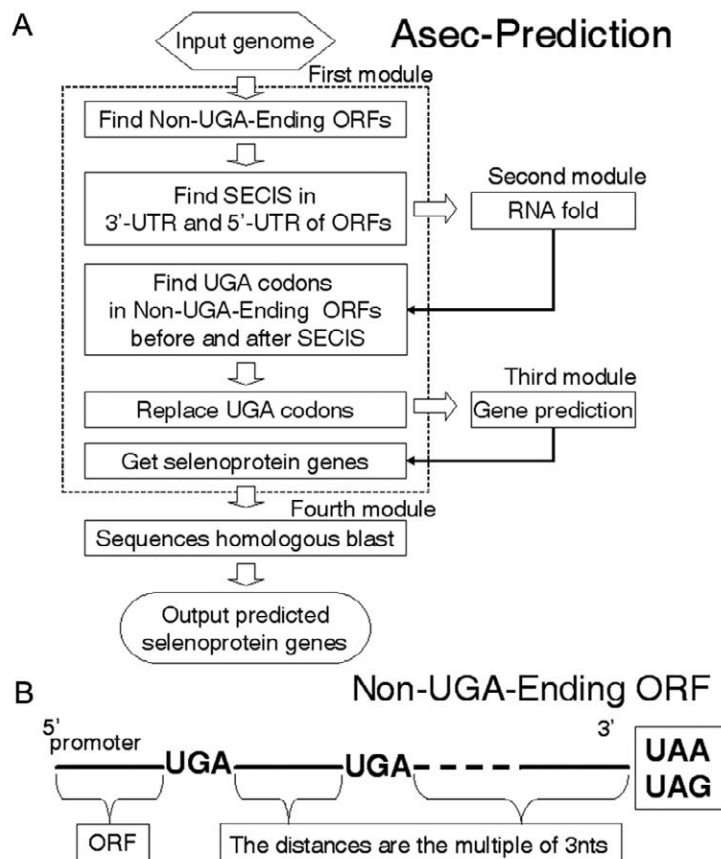
### Framework of Asec-Prediction

Asec-Prediction contains four modules (Figure 1A). The first module is composed of five functional parts. It mainly aims at detecting the locations of the non-UGA-ending open reading frames (ORFs), as well as the candidates of SECIS elements and selenoprotein genes. The non-UGA-ending ORF is obtained by extending the ORF terminated as UGA codon until meeting UAG/UAA terminator (Figure 1B). It must ensure that both the distance between any pair of UGA codons and the distance between the start

codon and the UAA/UAG terminator equal to the multiple of 3 nt. The second module predicts the RNA secondary structure and calculates the free energies for the putative SECIS elements by using Vienna RNAfold 1.4 (27). The third module predicts the coding genes in the nucleotide sequences with replaced UGA codons by using Glimmer 2.13 (11). The fourth module, integrating BLAST 2.2.14 (28), searches for the Cys-containing homologous sequences for the predicted selenoprotein genes in the NCBI non-redundant (NR) protein database. The Asec-Prediction program is written in ANSI C and implemented in Linux platform. The source codes of Asec-Prediction can be obtained freely from the corresponding author upon request.

### Flow of selenoprotein gene prediction

We introduce the procedure of Asec-Prediction in detail according to its flowchart (Figure 1A). The final output genes should be selenoprotein genes.



**Figure 1** The flowchart of Asec-Prediction (A) and the defined non-UGA-ending ORF (B). The modules are linked by arrows according to their executing orders. The function descriptions of each part are indicated via several key words in the rectangle boxes.

1. For a given archaeal genome, Asec-Prediction finds all non-UGA-ending ORFs.
2. Asec-Prediction finds the candidates of SECIS elements both in the regions of 500 nt upstream and 500 nt downstream of the non-UGA-ending ORFs. If the non-UGA-ending ORFs contain more than one UGA codon, the searching region in the downstream starts from the second UGA codon.
3. RNAfold 1.4 predicts the secondary structures and calculates the free energies of the putative SECIS elements. The cutoff of free energy is  $-16$  kcal/mol (23).
4. For the satisfied SECIS elements, Asec-Prediction calculates their positions relative to the non-UGA-ending ORFs. If they locate at the 3'-UTRs, Asec-Prediction finds all UGA codons contained in the non-UGA-ending ORFs. Otherwise, Asec-Prediction finds all UGA codons in-between the start codons and the SECIS elements except the last UGA codon that will act as terminator.
5. For these UGA codons in the nucleotide sequences, Asec-Prediction replaces them with other encoded codons. In order to avoid producing additional start or stop codons, Asec-Prediction chooses GGG codon or CGG codon to replace UGA codons in different situations. For the replaced nucleotide sequences, Asec-Prediction uses Glimmer 2.13 to predict genes.
6. Asec-Prediction extracts the predicted archaeal selenoprotein genes.
7. Asec-Prediction finds the homologous sequences in the NR protein database using BLAST 2.2.14.

## Flexibility of Asec-Prediction

Asec-Prediction takes account of all possible cases of archaeal selenoprotein genes. For example, for some archaeal selenoprotein genes having more than one Sec residues, Asec-Prediction introduces the non-UGA-ending ORF so as to find those selenoprotein genes; for some archaeal selenoprotein genes observed to correlate with the SECIS elements locating at the 5'-UTRs, Asec-Prediction searches the SECIS elements both in the 3'-UTRs and the 5'-UTRs (20, 23). On the other hand, Asec-Prediction properly adjusts previous prediction strategy and thus achieves a higher efficiency. It firstly searches the non-UGA-ending ORFs and then the candidates of SECIS elements in the UTRs, reducing much more searching regions compared with the approach provided by Kryukov and Gladyshev that ordered them

in the reverse order (23). In addition, Asec-Prediction searches the candidates of SECIS elements only in the UTRs of the non-UGA-ending ORFs, which is different from searching thoroughly the whole genome sequence (23). This measure also reduces some searching regions. As we know, the consensus model of the SECIS elements is not an exclusive sequence motif (23). Numerous sequence motifs can satisfy the consensus model. The decreasing of searching regions will greatly decrease the numbers of the candidates. In Asec-Prediction, the most time-consuming part is to judge whether the candidates satisfy the SECIS elements or not. Undoubtedly, the two improvements of decreasing searching regions will accelerate the prediction of Asec-Prediction. As one exemplification, for *Methanococcus jannaschii* with  $\sim 1.6$  M nt, it only needs approximate 10 s to complete the prediction on a common personal computer.

The integrated programs in Asec-Prediction can be freely replaced with other optimal programs that can enable it achieving a higher efficiency. Presently, Asec-Prediction integrates RNAfold 1.4 and Glimmer 2.13. Sometimes its prediction precision will be limited by the deficiencies of these integrated programs. For example, RNAfold 1.4 cannot give the correct stem-loop structure for the SECIS element "CGCCCGGGGGGAACCCCGCAAGGAGGGGACCCCGGTC", while this element can be correctly predicted by RNAstructure 4.2 (29). This may be the result of slight variants in the implementation of the energy models between RNAfold and RNAstructure. Similarly, we found that Glimmer 2.13 sometimes cannot correctly predict the existed genes, such as HesB-like genes in *M. jannaschii* (23). In order to avoid these potential deficiencies, Asec-Prediction can flexibly replace RNAfold 1.4 and Glimmer 2.13 with higher versions like RNAfold 1.5 and Glimmer 3.01. Furthermore, other tools can also be freely integrated to Asec-Prediction if they enable it achieving better prediction. For example, Lambert *et al* reported that ERPIN is effective to detect SECIS elements (30). Thus, Asec-Prediction can be updated timely with much higher prediction accuracy.

## Evaluation

### Archaeal genome dataset

We downloaded 26 completely sequenced archaeal genomes from the NCBI database (updated on February 27, 2006) as our dataset (Tables 1 and 2). Among

**Table 1 Selenoprotein genes predicted by Asec-Prediction in fourteen researched archaeal genomes**

Species	Accession No.	Predicted selenoprotein genes	
		Previously reported	Asec-Prediction
<i>Aeropyrum pernix</i>	NC_000854	0	0
<i>Archaeoglobus fulgidus</i>	NC_000917	0	0
<i>Halobacterium sp.</i>	NC_002608	0	0
<i>Methanobacterium therm.</i>	NC_000916	0	0
<i>Methanococcus jannaschii</i>	NC_000909	8	8
<i>Methanopyrus kandleri</i>	NC_003551	7	7
<i>Pyrobaculum aerophilum</i>	NC_003364	0	0
<i>Pyrococcus abyssi</i>	NC_000868	0	0
<i>Pyrococcus furiosus</i>	NC_003413	0	0
<i>Pyrococcus horikoshii</i>	NC_000961	0	0
<i>Sulfolobus solfataricus</i>	NC_002754	0	0
<i>Sulfolobus tokodaii</i>	NC_003106	0	0
<i>Thermoplasma acidophilum</i>	NC_002578	0	0
<i>Thermoplasma volcanium</i>	NC_002689	0	0

**Table 2 Selenoprotein genes predicted by Asec-Prediction in twelve un-researched archaeal genomes**

Species	Accession No.	Selenoprotein genes by Asec-Prediction
<i>Haloarcula marismortui</i>	NC_006396	0
<i>Methanococcus maripaludis</i>	NC_005791	0
<i>Methanosarcina acetivorans</i>	NC_003552	1
<i>Methanosarcina barkeri str.</i>	NC_007355	0
<i>Methanosarcina mazei</i>	NC_003901	0
<i>Methanosphaera stadtmanae</i>	NC_007681	0
<i>Methanospirillum hungatei</i>	CP000254	0
<i>Nanoarchaeum equitans</i>	NC_005213	0
<i>Natronomonas pharaonis</i>	NC_007426	0
<i>Picrophilus torridus</i>	NC_005877	0
<i>Sulfolobus acidocaldarius</i>	NC_007181	0
<i>Thermococcus kodakaraensis</i>	NC_006624	0

them, 14 genomes had been researched previously for selenoprotein genes by Kryukov and Gladyshev (23), in which they reported 8 selenoprotein genes in *M. jannaschii* and 7 selenoprotein genes in *Methanopyrus kandleri*, while no selenoprotein genes were found in the remaining archaeal genomes. The other 12 genomes have not been researched for selenoprotein genes according to our knowledge.

### Prediction of previously known archaeal selenoprotein genes

We first applied Asec-Prediction to the 14 archaeal genomes with already known selenoprotein genes (Ta-

ble 1). Successfully, Asec-Prediction predicted the reported 15 selenoprotein genes in the two archaeal genomes (*M. jannaschii* and *M. kandleri*) without redundant results. As to the other genomes, Asec-Prediction did not find selenoprotein genes as well. Therefore, it shows that Asec-Prediction is effective for the prediction of archaeal selenoprotein genes.

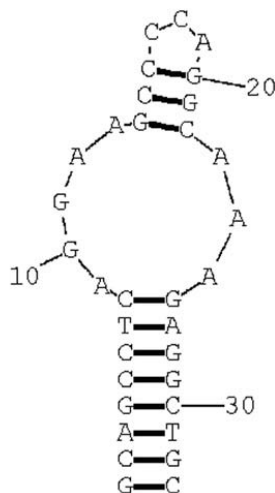
### Identification of a novel archaeal selenoprotein gene

#### *Recognition of SECIS element*

We then applied Asec-Prediction to the 12 archaeal genomes that have been completely sequenced but un-

researched for selenoprotein genes (Table 2). Among them, Asec-Prediction predicted a novel archaeal selenoprotein gene in *M. acetivorans*. Its ORF starts from 4,550,549 nt and ends at 4,550,942 nt in the nucleotide sequence. There is one TGA codon (4,550,732–4,550,734 nt) in this ORF translating to Sec. The corresponding SECIS element starts from 4,550,481 nt and ends at 4,550,513 nt with 36 nt upstream of the coding region. It is different from the general archaeal selenoprotein genes where the SECIS elements are downstream of the coding region. This phenomenon is just similar to that of the *fdhA* gene and seems to be the reason of its missing (20, 23). The secondary structure of this SECIS element displays a stem-loop structure (Figure 2). In detailed description, there are eight base pairs in the stem: six GC pairs and two AU pairs, which is different from ten base pairs in the consensus model (23). In the bugle, besides of containing GAA\_A pattern, it also contains two additional unpaired bases: AA pair and GA pair. The features of the up helix and the apical loop are unchangeable: three GC pairs locate at the up helix and three nucleotides in the apical loop. The minimum free energy of this SECIS element is  $-16.5$  kcal/mol, lower than the cutoff. Taken together, despite of some differences from the consensus model, this SECIS element still indicates a stem-loop structure and is reasonable to be considered as a satisfied SECIS element.

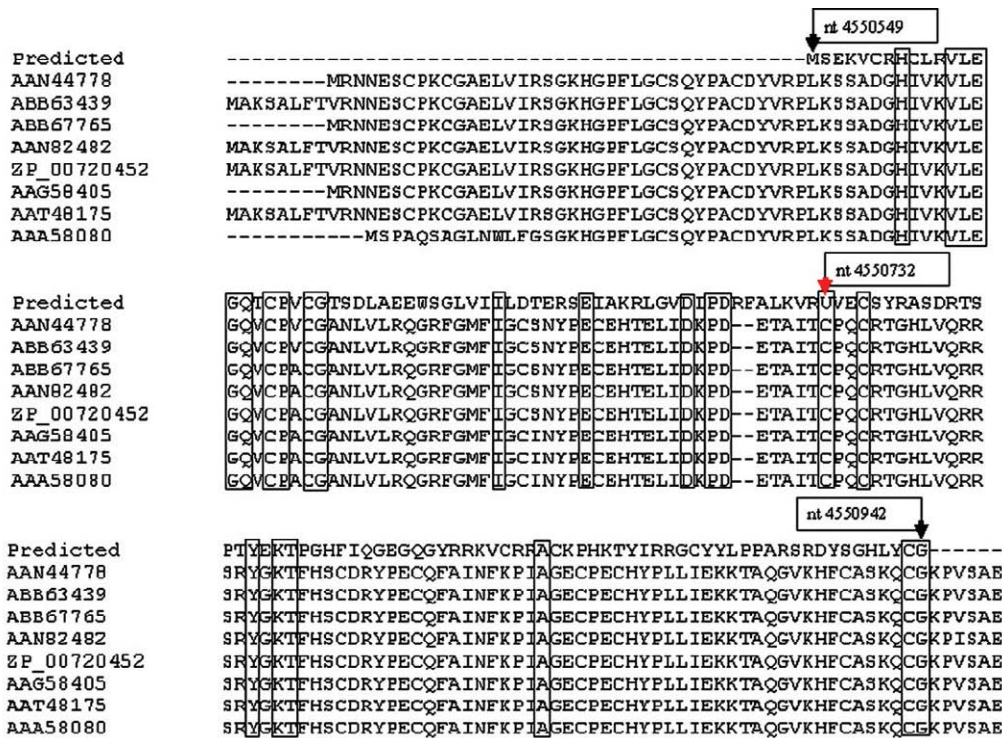
In the procedure of Sec/Cys-containing homology search in the NR protein database, Asec-Prediction



**Figure 2** The secondary structure of the SECIS element presented upstream of the novel selenoprotein gene in *M. acetivorans*. Numbers indicate the locations of some nucleotides in the SECIS element.

found eight Cys-containing homologous sequences of this predicted selenoprotein (Figure 3) (31). Multiple sequence alignment of them showed that the Cys residues and the Sec residue are identical. On both sides of the identical Sec/Cys, the alignment showed symmetrical patterns of conservation, such as the anterior “PD” pattern and the followed “C” pattern (Figure 3). As popular knowledge, a real selenoprotein shows symmetrical patterns of conservation around the Sec residue, as a result of similar strength of purifying selection at both sides of the recoded codon, together with different Sec- and Cys-containing orthologs across the phylogeny. Thus, the symmetrical conserved patterns seem to validate the translation from UGA codon to Sec. Furthermore, around the Sec residue the predicted selenoprotein gene contains a UXXC motif (Sec residue separated from Cys residue by two other residues) and the homologous sequences contain CXXC motifs. As reported in the literature (32–34), the [U/C]XXC motif always corresponds to redox function and presents in a variety of thiol-dependent redox enzymes. Besides, we noticed that these eight homologous proteins belong to two kinds of organisms. Five proteins belong to *Escherichia coli* (GenBank Accession No. AAA58080, AAG58405, AAN82482, AAT48175, and ZP\_00720452) and the other three belong to *Shigella* (GenBank Accession No. AAN44778, ABB63439, and ABB67765). Most of them have the same function annotations in the NCBI database, acting as topoisomerase. Therefore, it seems that the novel selenoprotein contains redox and enzymatic functions, similar with the real selenoprotein.

Before the SECIS element, there is another non-UGA-ending ORF, which starts from 4,549,509 nt and ends at 4,549,940 nt. Three TGA codons are included in the ORF. However, we could not find identical Sec/Cys in its homologous sequences since these homologous sequences all terminate at the Sec positions (data not shown). In the sequence alignment, the sharp decrease in sequence conservation most appears after the real stop codon, which is due to the reduced evolutionary constraint in UTR region. Therefore, these three TGA codons cannot recode as Sec residues. Although a similar exceptive example had been found, that is, glycine reductase selenoprotein A in bacteria has no Sec/Cys pair homologous sequences in the NCBI database (17, 24), in the present study we do not have other exact evidence to declare it as a selenoprotein gene.



**Figure 3** Multiple sequence alignment of the predicted novel selenoprotein. The novel gene starts from 4,550,549 nt and ends at 4,550,942 nt. The arrow for 4,550,732 nt indicates the locations of the predicted Sec (U) and the corresponding Cys in homologies. Rectangle boxes in the alignment show the identical residues. The alignment was generated using ClustalW and edited manually.

**Recognition of SelB**

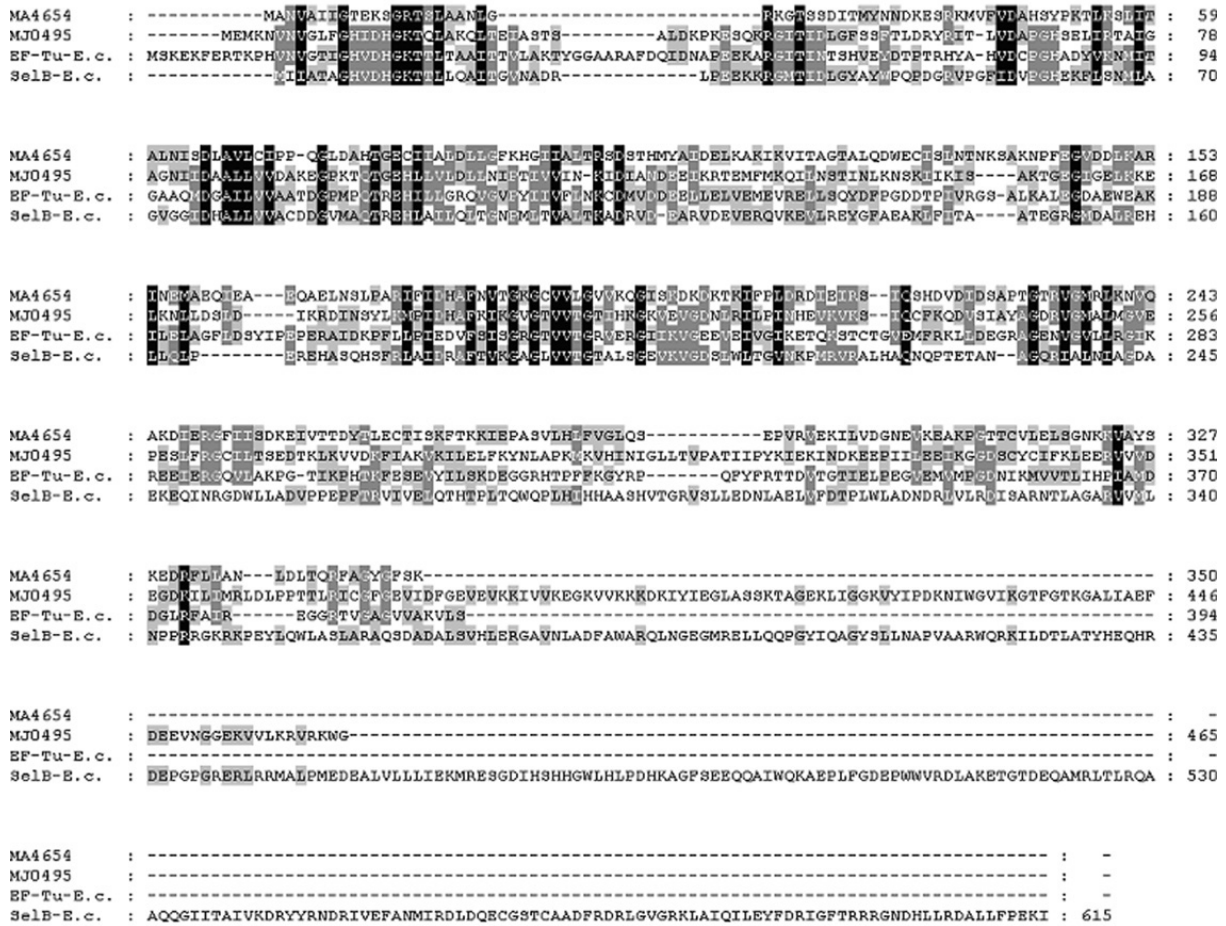
The decoding of UGA codon to Sec is always accompanied by the interaction with the translation elongation factor SelB, especially for *E. coli* (35, 36). For archaea and animal, the mechanisms of Sec incorporation are more complex and unclear. Böck and colleagues previously identified ORF MJ0495 in *M. jannaschii* to share some functions with bacterial SelB and acts as the Sec-specific translation factor SelB (37, 38). In *M. acetivorans*, we noticed that ORF MA4654 is annotated as a SelB gene in the NCBI database. However, some researchers doubt the correctness of this annotation (Methanosarcina Sequencing Project; <http://www.broad.mit.edu/>). We aligned MA4654 with MJ0495 as well as with the elongation factors SelB and EF-Tu in *E. coli*. The alignment shows that MA4654 shares significant homologies with them (Figure 4). Nevertheless, unlike MJ0495, MA4654 does not contain C-terminal extension. Böck and colleagues demonstrated that the C-terminal extension of MJ0495 did not contribute to its binding to the SECIS element as SelB does in *E. coli*. They speculated that another protein might exist and contributed to the SECIS element binding in

archaea (38). Unfortunately, they did not find this protein. Therefore, we think MA4654 may take the role like MJ0495 and acts as the Sec-specific translation factor SelB in *M. acetivorans*.

In eukaryotes, the SelB sequences do not contain the SECIS-binding function (39). However, another protein, SECIS-binding protein 2 (SBP2), was found to contain RNA-binding motif (40, 41). We tried to find the similar functional protein in *M. acetivorans* through searching its homologous sequences. We downloaded all currently published SBP2 and aligned them with the present annotated proteins in *M. acetivorans*. Unfortunately, no prominent homologous proteins were found. A possible reason for this deficiency may be that the SBP2 dataset is not large enough. Another reason may be that the SECIS-binding protein in archaea is different from that in eukaryotes.

**Conclusion**

Asec-Prediction takes proper prediction strategy that avoids some redundancies, hence achieves quick and



**Figure 4** Multiple sequence alignment of ORF MA4654 in *M. acetivorans* with ORF MJ0495 in *M. jannaschii*, SelB sequence (SelB-E.c.) and EF-Tu sequence (EF-Tu-E.c.) in *E. coli*. Residues are shaded in black for more conservation and in grey for less conservation. The alignment was generated with ClustalW and edited with GeneDoc.

correct prediction. Asec-Prediction correctly predicted 15 previously reported archaeal selenoprotein genes in 14 researched archaeal genomes without any redundant results. In 12 un-researched archaeal genomes, Asec-Prediction identified a novel selenoprotein gene in *M. acetivorans* with further supporting evidence: (1) the SECIS element shows a stem-loop structure and locates at 5'-UTR; (2) the Sec/Cys pair is identical in the alignment; (3) the [U/C]XXC motif suggests that it contains redox and enzymatic functions like the real selenoprotein; (4) the ORF MA4654 in *M. acetivorans* acts as Sec-specific translation factor SelB like the ORF MJ0495 in *M. jannaschii*. All of them suggest that the predicted gene should be a real selenoprotein gene. The result shows that Asec-Prediction is effective for the prediction of archaeal selenoprotein genes.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant No. 30525037 and the National 863 Program under Grant No. 2004AA231060.

## Authors' contributions

ML collected the datasets, conducted data analyses, and prepared the manuscript. YH and YX supervised the project and co-wrote the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors have declared that no competing interests exist.

## References

1. Chambers, I., *et al.* 1986. The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the “termination” codon, TGA. *EMBO J.* 5: 1221-1227.
2. Low, S.C. and Berry, M.J. 1996. Knowing when not to stop: selenocysteine incorporation in eukaryotes. *Trends Biochem. Sci.* 21: 203-208.
3. Atkins, J.F. and Gesteland, R.F. 2000. The twenty-first amino acid. *Nature* 407: 463-465.
4. Ibba, M. and Söll, D. 2004. Aminoacyl-tRNAs: setting the limits of the genetic code. *Genes Dev.* 18: 731-738.
5. Berry, M.J. 2005. Knowing when not to stop. *Nat. Struct. Mol. Biol.* 12: 389-390.
6. Srinivasan, G., *et al.* 2002. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* 296: 1459-1462.
7. Hao, B., *et al.* 2002. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science* 296: 1462-1466.
8. Borodovsky, M. and McIninch, J. 1993. GeneMark: parallel gene recognition for both DNA strands. *Comput. Chem.* 17: 123-133.
9. Burge, C. and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268: 78-94.
10. Lukashin, A.V. and Borodovsky, M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* 26: 1107-1115.
11. Delcher, A.L., *et al.* 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* 27: 4636-4641.
12. Castellano, S., *et al.* 2001. *In silico* identification of novel selenoproteins in the *Drosophila melanogaster* genome. *EMBO Rep.* 2: 697-702.
13. Hüttenhofer, A., *et al.* 1996. Solution structure of mRNA hairpins promoting selenocysteine incorporation in *Escherichia coli* and their base-specific interaction with special elongation factor SELB. *RNA* 2: 354-366.
14. Fourmy, D., *et al.* 2002. Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB. *J. Mol. Biol.* 324: 137-150.
15. Driscoll, D.M. and Copeland, P.R. 2003. Mechanism and regulation of selenoprotein synthesis. *Annu. Rev. Nutr.* 23: 17-40.
16. Liu, Z., *et al.* 1998. The nature of the minimal ‘selenocysteine insertion sequence’ (SECIS) in *Escherichia coli*. *Nucleic Acids Res.* 26: 896-902.
17. Zhang, Y. and Gladyshev, V.N. 2005. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. *Bioinformatics* 21: 2580-2589.
18. Berry, M.J., *et al.* 1991. Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region. *Nature* 353: 273-276.
19. Walczak, R., *et al.* 1996. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA* 2: 367-379.
20. Wilting, R., *et al.* 1997. Selenoprotein synthesis in archaea: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion. *J. Mol. Biol.* 266: 637-641.
21. Fagegaltier, D., *et al.* 2000. Structural analysis of new local features in SECIS RNA hairpins. *Nucleic Acids Res.* 28: 2679-2689.
22. Kryukov, G.V., *et al.* 2003. Characterization of mammalian selenoproteomes. *Science* 300: 1439-1443.
23. Kryukov, G.V. and Gladyshev, V.N. 2004. The prokaryotic selenoproteome. *EMBO Rep.* 5: 538-543.
24. Zhang, Y., *et al.* 2005. The microbial selenoproteome of the Sargasso Sea. *Genome Biol.* 6: R37.
25. Böck, A. 2000. Biosynthesis of selenoproteins—an overview. *Biofactors* 11: 77-78.
26. Rother, M., *et al.* 2001. Selenoprotein synthesis in archaea. *Biofactors* 14: 75-83.
27. Hofacker, I.L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Res.* 31: 3429-3431.
28. Altschul, S.F., *et al.* 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
29. Hofacker, I.L., 1994. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* 125: 167-188.
30. Lambert, A., *et al.* 2002. A survey of metazoan selenocysteine insertion sequences. *Biochimie* 84: 953-959.
31. Gladyshev, V.N. and Kryukov, G.V. 2001. Evolution of selenocysteine-containing proteins: significance of identification and functional characterization of selenoproteins. *Biofactors* 14: 87-92.
32. Chivers, P.T., *et al.* 1996. The CXXC motif: imperatives for the formation of native disulfide bonds in the cell. *EMBO J.* 15: 2659-2667.
33. Fomenko, D.E. and Gladyshev, V.N. 2003. Identity and functions of CxxC-derived motifs. *Biochemistry* 42: 11214-11225.
34. Moutevelis, E. and Warwicker, J. 2004. Prediction of pKa and redox properties in the thioredoxin superfamily. *Protein Sci.* 13: 2744-2752.
35. Baron, C. and Böck, A. 1991. The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of *Escherichia coli* is the determinant for binding to elongation factors SELB or Tu. *J. Mol. Biol.* 266: 20375-20379.
36. Kawashima, T., *et al.* 1996. The structure of the *Escherichia coli* EF-Tu.EF-Ts complex at 2.5Å resolution.



- lution. *Nature* 379: 511-518.
37. Kromayer, M., *et al.* 1996. Domain structure of the prokaryotic selenocysteine-specific elongation factor SelB. *J. Mol. Biol.* 262: 413-420.
38. Rother, M., *et al.* 2000. Identification and characterization of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*. *J. Mol. Biol.* 299: 351-358.
39. Fagegaltier, D., *et al.* 2000. Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation. *EMBO J.* 19: 4796-4805.
40. Copeland, P.R., *et al.* 2000. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J.* 19: 306-314.
41. Fletcher, J.E., *et al.* 2001. The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA* 7: 1442-1453.