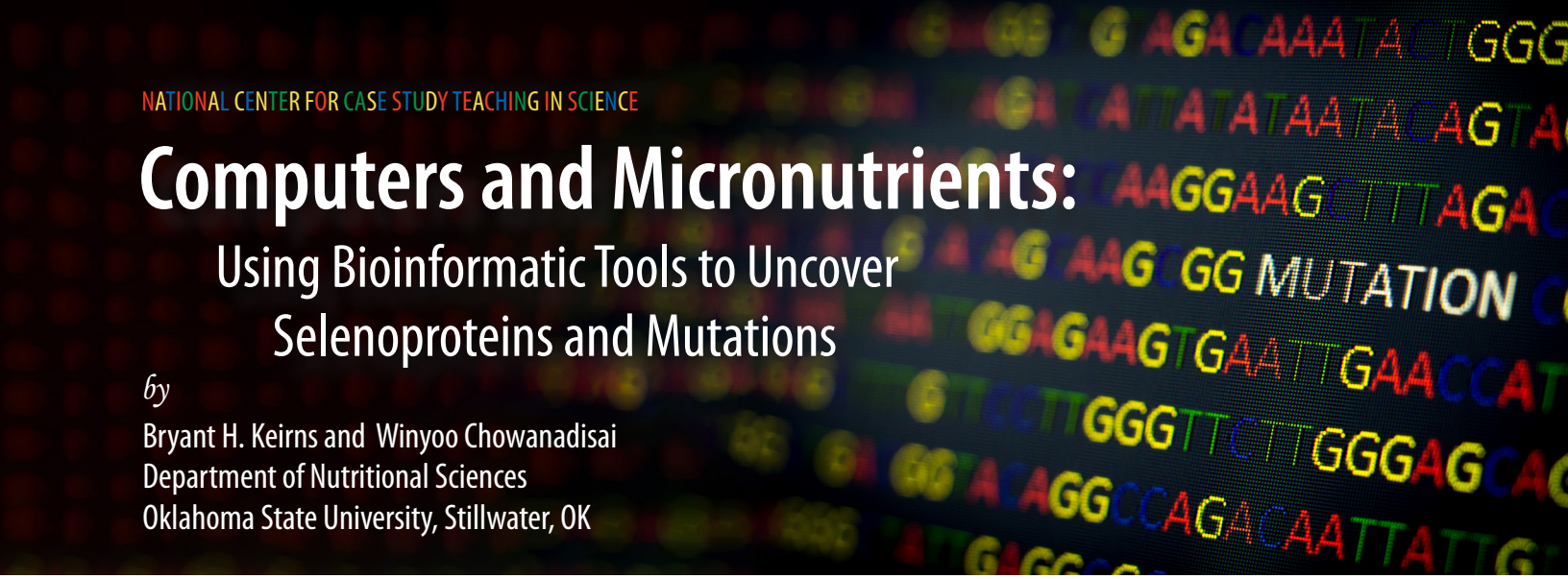

NATIONAL CENTER FOR CASE STUDY TEACHING IN SCIENCE

Computers and Micronutrients:

Using Bioinformatic Tools to Uncover Selenoproteins and Mutations

by

Bryant H. Keirns and Winyoo Chowanadisai
Department of Nutritional Sciences
Oklahoma State University, Stillwater, OK

Introduction

Maya, a 45-year-old Syrian woman, had suffered from muscle weakness, back discomfort, impaired mobility, and respiratory problems for as long as she could remember. Her health issues culminated last week when she was diagnosed with rigid spine muscular dystrophy (RSMD) by her doctor.

Maya's journey had taken a long and difficult road. Things were off to a terrible start in the first six months of her life when her single mother unexpectedly passed away. Luckily, her aunt and uncle took her in, but they quickly noticed that Maya seemed to have weak neck muscles and had difficulty keeping her head upright. During her childhood and teenage years, Maya continued to experience muscle weakness and pain. This physical ailment manifested as poor sports performance throughout her high school years, causing frustration and low self-esteem in addition to muscle pain. Through her 20s and 30s, Maya continued to face physical challenges related to muscle weakness, and also began to develop back pain and had trouble breathing.

Throughout all her suffering, Maya had longed to understand more about why she experienced weakness and pain. She finally decided to take matters into her own hands and do some of her own research. Through her research, Maya learned that modern genetic testing may hold the answer to her question. She consulted with a physician at a research hospital who recommended DNA sequencing. After several meetings, her physician told Maya he thought that Maya's clinical symptoms were very likely tied to a rare mutation in the *SEPN1* gene, which codes for the protein selenoprotein N. Maya learned that this protein is not fully characterized, but may have a role in protecting against oxidative stress and skeletal muscle development and contraction. In addition, Maya found out that mutations in *SEPN1* are associated with other types of myopathy besides RSMD. Maya left the appointment with a new understanding of her diagnosis.

Relevant Scientific Literature and Information

You may find the following resources helpful in answering the questions in this case study.

Wikipedia:

- Central dogma of molecular biology: <https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology>
- Selenoprotein: <https://en.wikipedia.org/wiki/Selenoprotein>
- Muscle contraction: <https://en.wikipedia.org/wiki/Muscle_contraction#Excitation-contraction_coupling>
- Selenium deficiency: <https://en.wikipedia.org/wiki/Selenium_deficiency>
- Selenium-rich foods: <https://en.wikipedia.org/wiki/Selenium_in_biology#Food_sources>
- Keshan disease: <https://en.wikipedia.org/wiki/Keshan_disease>
- Selenium toxicity: <https://en.wikipedia.org/wiki/Selenium_in_biology#Toxicity>

Original Report on Which This Case Study Is Based:

- Allamand, V., P. Richard, A. Lescure, C. Ledeuil, D. Desjardin, ..., and P. Guicheney. (2006). A single homozygous point mutation in a 3′ untranslated region motif of selenoprotein N mRNA causes SEPN1-related myopathy. *EMBO Reports* 7(4): 450–4. <https://doi.org/10.1038/sj.embor.7400648>

Role of SEPN1 in Muscle Contraction:

- Chernorudskiy, A., E. Varone, S.F. Colombo, S. Fumagalli, A. Cagnotto, ..., and E. Zito. (2020). Selenoprotein N is an endoplasmic reticulum calcium sensor that links luminal calcium levels to a redox activity. *Proceedings of the National Academy of Sciences* 117(35): 21288–98. <https://doi.org/10.1073/pnas.2003847117>

Review Article on Selenium:

- Prabhu, K.S. and X.G. Lei. (2016). Selenium. *Advances in Nutrition* 7(2): 415–7. <https://doi.org/10.3945/an.115.010785>

Comparison of mRNA for Normal Proteins and Selenoproteins (Figure 1):

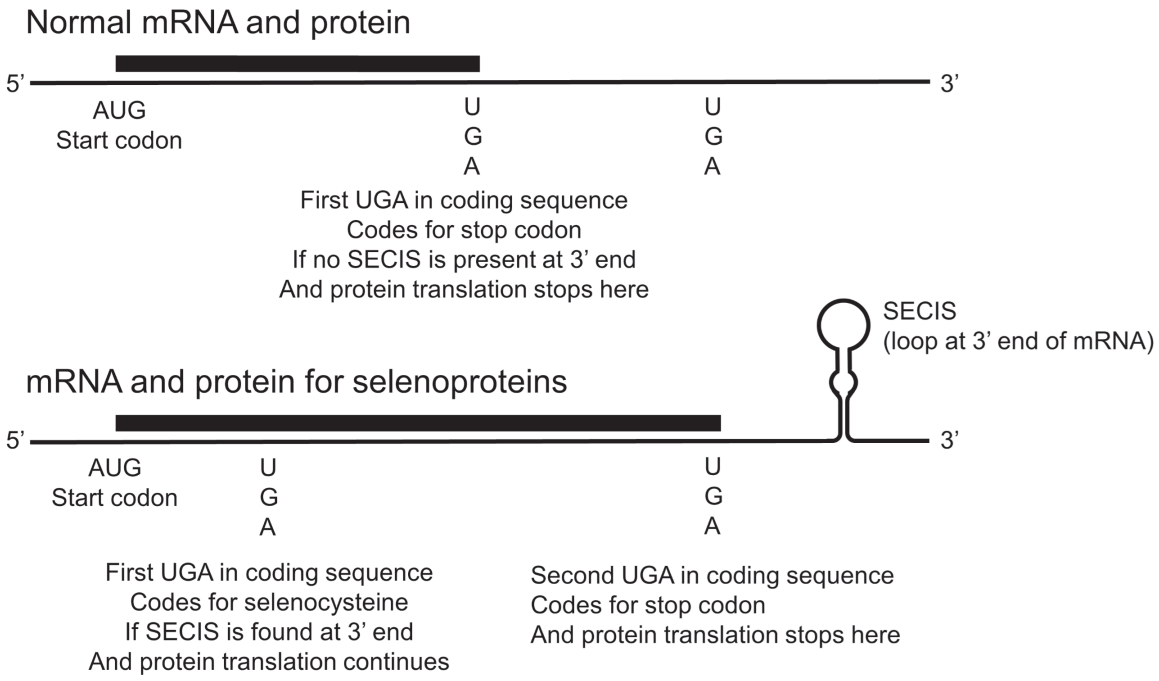

Figure 1. Comparison of mRNA (thin line from left 5′ to right 3′) and translated protein (thick line above mRNA) for both normal proteins and selenoproteins. For almost all proteins, protein translation starts at AUG. For normal proteins, protein translation stops at the first stop codon, which can be UGA (or UAG or UAA). For selenoproteins, if a SECIS loop is detected, the first UGA codon encountered encodes a selenocysteine, and protein translation continues afterwards until another stop codon is encountered. It should be noted that in mRNA, the adenine (a) is replaced with uracil (u). However, for the genetic sequences of the genes provided (including *SEPN1*), the sequences are provided from DNA. Thus, the output from the Seblastian program you will use later in this case study will show alignments for ATG instead of AUG for the start codon and TGA instead of UGA for the selenocysteine codon.

Preparation

Basic Steps for Creating Protein from the Genome

1. Define transcription.

2. What is translation? How does it differ from transcription?

3. What is a stop codon and how does it relate to translation?

4. What is the central dogma of molecular biology? How are transcription and translation related to the central dogma?

5. What would happen to a protein if it had a premature stop codon that occurred before the usual stop codon? Do you think the abrupt stop codon would result in a protein with normal function?

Role of Selenium in Human Health

1. What is selenium and why do some proteins need selenium?

2. Are proteins that require selenium, called selenoproteins, common among all the proteins in the human body?

3. What is the importance of selenium in the diet?

Part I – SEPN1 and Muscle Contraction

The pedigree below (Figure 2) is from the family data in the published case report (Allamand *et al.*, 2006). The arrow indicates the symbol representing Maya (Patient II:5). Use Figure 2 and the text from the published case report to answer the questions below.

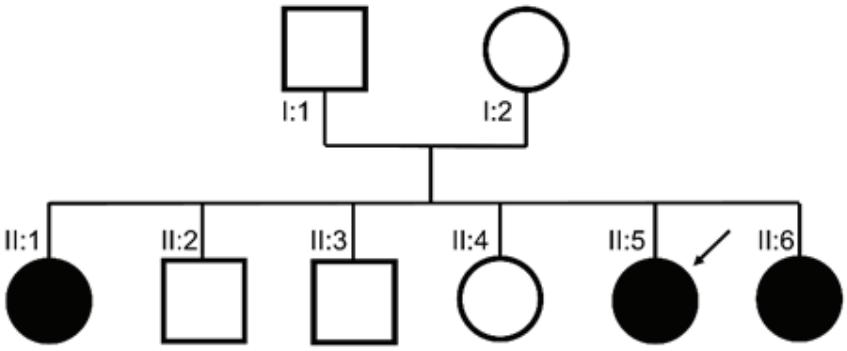

Figure 2. Pedigree of Maya (Patient II:5).

Questions

1. Are Maya's parents, brothers, or sisters also affected by the rare mutation in *SEPN1*?

2. Which symbols represent Maya's father and mother? Which symbols represent Maya's sisters? Which symbols represent Maya's brothers?

Maya's physician explained that a mutation in *SEPN1* may lead to issues with muscle contraction. Based on her musculoskeletal symptoms, Maya believed this was likely true for her. She learned that the protein for *SEPN1* is important for sensing low calcium concentrations within the sarcoplasmic reticulum of muscle cells so it can be replenished.

Questions

3. Using the Wikipedia page on muscle contraction (and excitation-contraction coupling), why is having enough calcium in the sarcoplasmic reticulum important for muscle contraction?

4. If having a mutation in *SEPN1* caused Maya to not have enough calcium within her sarcoplasmic reticulum, how would you expect that to affect her muscle contraction?

   Hints:

   - Review the role of calcium signaling in skeletal muscle contraction.
   - For more information, refer to the significance section from the paper "Selenoprotein N is an endoplasmic reticulum calcium sensor that links luminal calcium levels to a redox activity" (Chernorudskiy *et al.*, 2020).

Part II – Detecting Selenoproteins from mRNA Sequences

Selenoproteins have two distinguishing features in their mRNA sequence that change the interpretation of the TGA (UGA) sequence from a stop codon to a selenocysteine.

- There is a TGA (UGA) codon in the coding region of the mRNA sequence. Normally TGA encodes a stop codon.
- There is a SECIS (selenocysteine insertion sequence) in the 3′ end of the mRNA sequence (non-coding region) that forms a hairpin loop. This sequence and "stem- loop" formation recruit a variety of translation factors, which inform the ribosome to insert a tRNA bound to selenocysteine at the UGA codon.

Because of these two distinguishing features, it is possible to use bioinformatics programs to predict whether an mRNA sequence encodes a selenoprotein. Similar to Maya's case above, nutrigenomic approaches can be used to discover new selenoproteins. We can even predict what might be some symptoms of selenium deficiency by determining the functions of these selenoproteins, then predict what would happen if these selenoproteins were not fully functional due to insufficient selenium. One example of such a program is Seblastian. In this exercise, you will use Seblastian to investigate Maya's rare mutation and to identify other selenoproteins.

Begin by visiting <http://seblastian.crg.es>. Select the "Selenoprotein prediction" option, and then copy/paste each of the four mRNA sequences indicated below (one at a time) into the white box and hit the Submit button. To obtain the sequences, visit the provided links or obtain the sequences from your instructor.

- *Sequence 1:* Human Glutathione Peroxidase 1, <https://www.ncbi.nlm.nih.gov/nuccore/NM_000581.4>
- *Sequence 2:* Mouse Low Density Lipoprotein Receptor, <https://www.ncbi.nlm.nih.gov/nuccore/NM_010700.3>
- *Sequence 3:* Human Superoxide Dismutase 1 (SOD1), <https://www.ncbi.nlm.nih.gov/nuccore/NM_000454.4>
- *Sequence 4:* Human Iodothyronine Deiodinase 1, <https://www.ncbi.nlm.nih.gov/nuccore/NM_000792.6>

Note: If Seblastian generates an output for a given mRNA sequence, then the sequence codes for a selenoprotein. Conversely, if the mRNA does not code for a selenoprotein, the error message "Sorry, no SECIS elements could be predicted in your sequence" will appear. If a selenoprotein is predicted from the mRNA, then you will see a prediction of the protein sequence in the top portion of the screen (selenocysteine abbreviated by U and marked by *) and in the bottom half of the screen there is a prediction of the SECIS stem loop structure.

Questions

1. The Seblastian bioinformatic tool has been developed by researchers to predict whether mRNA (DNA) sequences of genes will ultimately be translated into selenoproteins. Using the four sequences above and the bioinformatic tool, which genes (proteins) are selenoproteins?

2. After identifying which genes code for selenoproteins, use Wikipedia to determine the functions of these two selenoproteins.

3. One hope for a bioinformatic tool such as this is that it could be used to predict clinical symptoms associated with selenium deficiency. Based upon the functions of these selenoproteins, what are some symptoms that might be associated with selenium deficiency?

4. What would you predict would happen if a patient were to have a mutation in the SECIS, or the "stem-loop" portion of a selenoprotein? If the UGA codon no longer codes for selenocysteine, then what will happen to the protein? Will the protein still contain selenocysteine as an amino acid and will it still be full length? Do you think the bioinformatic tool would still predict that the mRNA (DNA) sequence would code for a selenoprotein?

Part III – Mutations Impair a Critical Motif in Selenoproteins

Recall that *SEPN1* is the gene mutated in Maya's genome. The reference sequence for the mRNA sequence for *SEPN1* is unmutated and found in healthy individuals; it codes for the protein "selenoprotein N." Enter the following reference sequence into the Seblastian website.

- *Sequence 5:* Human Selenoprotein N1 (SEPN1, Reference Sequence), <https://www.ncbi.nlm.nih.gov/nuccore/NM_206926.2>

Questions

1. Selenoproteins contain a TGA (UGA) codon in the coding region and an SECIS (stem loop structure) in the non-coding 3′ region of the mRNA sequence. Do the figures generated for human selenoprotein N1 (SEPN1) match the criteria for selenoproteins in general? Using the alignment for *SEPN1* as a guide, draw a rough sketch of the *SEPN1* mRNA and point out the two criteria for selenoproteins. Figure 1 (see page 2 of this handout) containing the general diagram for selenoprotein mRNA structure may be helpful here.

2. Maya is homozygous for a four-base-pair deletion mutation and is missing nucleotides from 2825–2828. Now insert the mutated sequence of *SEPN1* (Sequence 6) by following the instructions below.
   - *Sequence 6:* Human Selenoprotein N1 (Mutated SEPN1 Sequence from Maya, Missing Nucleotides 2825–2828 (aatc). You will need to create this sequence by modifying Sequence 5 above. Starting with Sequence 5, go to the line beginning with 2821: "gctgaatccg"; then delete "aatc" so that the line now reads: "gctg cg". Now copy and paste the entire new, slightly modified, mutated sequence into Sebastian. Does the software still recognize it as a selenoprotein?

Part IV – Dietary Selenium Deficiency vs. Mutations in Specific Selenoprotein Genes

Questions

1. What are symptoms of selenium deficiency and examples of potential causes? (Wikipedia can provide answers to this question.)

2. What are symptoms of Keshan disease, a disease linked to a combination of selenium deficiency and viral infection? (Wikipedia can provide answers to this question.)

3. Would increasing dietary selenium intake, through either food or dietary supplements, prevent Keshan disease? Explain why or why not. What foods are rich in selenium that could improve selenium status?

Part V – Would Dietary Selenium Supplements Help Maya?

Questions

1. The selenium status of an individual likely affects their ability to generate selenocysteine and ultimately selenoproteins. In contrast, a mutation restricted to *SEPN1* would specifically affect that selenoprotein, even with sufficient selenium status of the affected individual. In Maya's case, would increasing selenium intake in her diet, through either food or dietary supplements, improve her symptoms? Explain why or why not. Are there any cases where selenium supplements may improve her health?

2. Maya remains convinced that dietary supplements are worth a try, considering the severity of her symptoms and how they impact her quality of life. Are there any risks of selenium toxicity? What are some symptoms of selenium toxicity? How would you explain to your patient, Maya, the potential hazards of selenium supplements in individuals who already consume sufficient selenium? (Wikipedia can provide some answers to this question.)