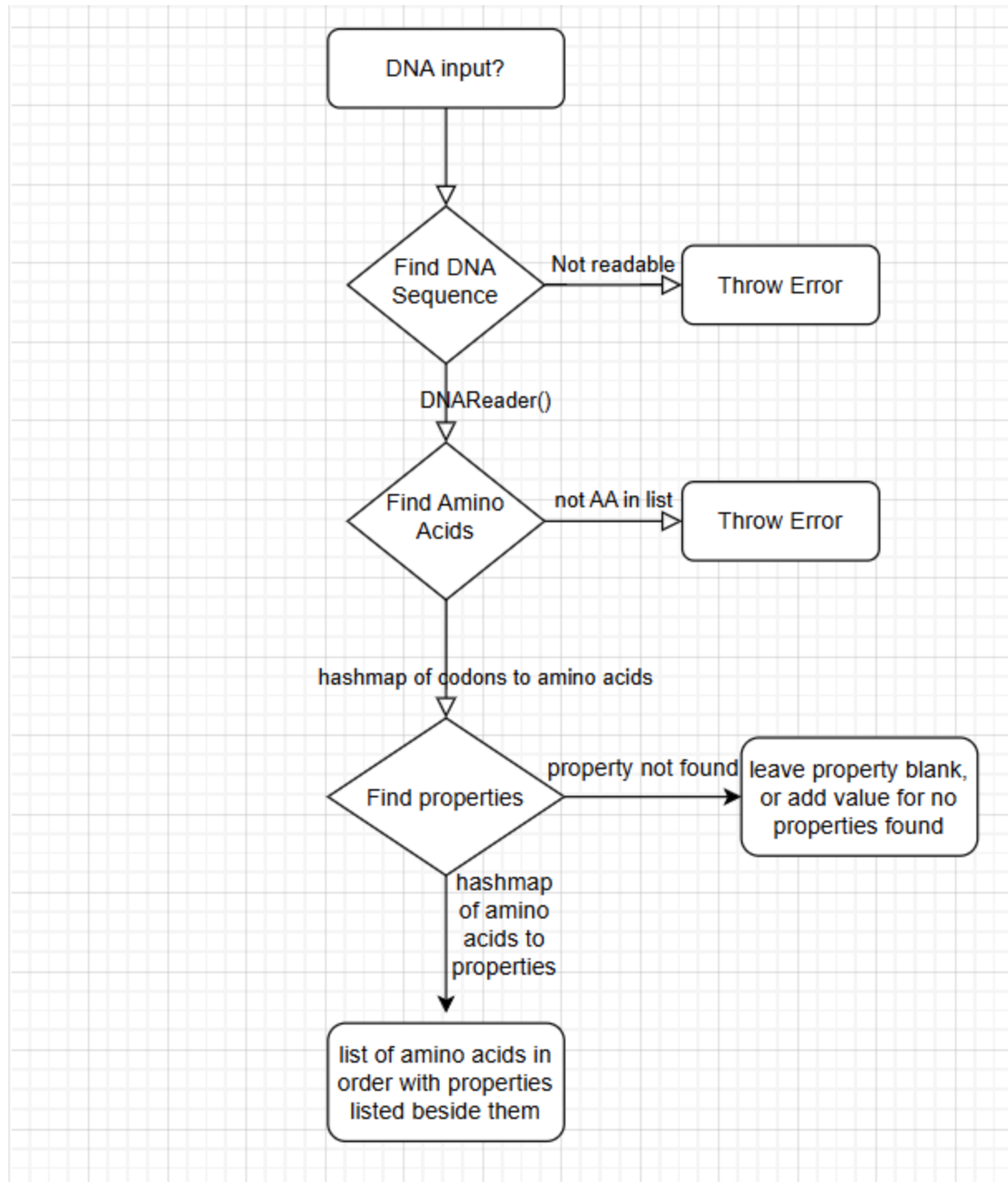My project will consist of code that will analyse a DNA sequence, identify start and stop points, and create a list of the amino acids. Then the code will analyse the amino acid sequence, identifying key properties about the amino acids and quantifying them. Some properties that could be counted is how many of the amino acids are polar, nonpolar, acids, or bases. The exact numbers of certain amino acids could also be counted. The properties of an amino acid could also be added to the sequence to show the order of amino acids and properties. This is useful because this could make DNA analysis much faster, because rather than having to go through and identify properties by hand the program will compile it all in easy counts. The combination of sequence and properties could be used to simplify folding calculations, as everything would be together and easily readable rather than scattered. In the future there could even be a protein folding model made that take in the list of amino acid sequences and attached properties to model how it would fold.

This project will be done alone, not in a group. The sole contributor is Tanya Klochko.

Flowchart:

```
          ┌─────────────┐
          │  DNA input? │
          └──────┬──────┘
                 │
                 ▼
             ╱Find DNA╲      Not readable      ┌──────────────┐
            ╱ Sequence ╲ ─────────────────────▶│  Throw Error │
            ╲          ╱                        └──────────────┘
             ╲        ╱
              DNAReader()
                 │
                 ▼
            ╱Find Amino╲      not AA in list    ┌──────────────┐
           ╱   Acids    ╲ ────────────────────▶│  Throw Error │
           ╲            ╱                       └──────────────┘
            ╲          ╱
       hashmap of codons to amino acids
                 │
                 ▼
                                property not found  ┌──────────────────────┐
        ╱Find properties╲ ──────────────────────────▶│  leave property blank,│
        ╲               ╱                            │   or add value for no │
         ╲             ╱                             │   properties found    │
          ╲           ╱                              └──────────────────────┘
           hashmap
           of amino
           acids to
           properties
                 │
                 ▼
        ┌──────────────────────┐
        │ list of amino acids in│
        │ order with properties │
        │  listed beside them   │
        └──────────────────────┘
```

This will mainly be done in R.

In the case that the project can't be completed, a scaled down version could be to simply identify amino acid sequences and compile their quantities. This can still help

with figuring out possible shapes for the resulting protein, and also make translating the DNA sequence to a possibly understandable form easier.

I do not plan on using AI in this project, I will write and debug all the code myself. This is because I feel like I learn more when not using AI, and also I want to challenge myself to use new and advanced techniques. In addition, I know multiple professional software engineers I can turn to if I have questions on how to structure a function or accomplish something. They can at the least help me with cleaning the code up.