Transcribing DNA into RNA

Jan Emmanuel Samson

2024-07-28

Problem

An RNA string is a string formed from the alphabet containing 'A', 'C', 'G', and 'U'.

Given a DNA string t corresponding to a coding strand, its transcribed RNA string u is formed by replacing all occurrences of 'T' in t with 'U' in u.

- **Given**: A DNA string t of length at most 1000 nt.
- **Return**: The transcribed RNA string of t.

Sample Dataset

```
GATGGAACTTGACTACGTAAATT
```

Sample Output

```
GAUGGAACUUGACUACGUAAAUU
```

Intuition

The given string represents the coding DNA strand with $5' \rightarrow 3'$ directionality. Details of the transcription process can be demonstrated by first finding the DNA complement of the input strand, then finding the RNA complement of the resulting DNA complement. This is an unnecessary step as we only need to replace all occurrences of 'T' in the input strand with 'U'.

Iterate over each base of input string and check if the current base is a thymine ('T'). If so, replace the current base with uracil ('U'). Otherwise, continue scanning the string until the end is reached. To save memory, the input string is updated *in place*.

Solution

```python
def transcribe(seq: str, alphabet: str='ACGT') -> dict[str, int]:
    """Convert the input DNA string to an RNA string."""
    return seq.replace('T', 'U')
```

```python
seq = "GATGGAACTTGACTACGTAAATT"
result = transcribe(seq)
print(result)
```

```
GAUGGAACUUGACUACGUAAAUU
```