

DNA mutation

DNA sequences can be represented as strings that only contain the letters A, C, G and T. One particular type of mutation that is observed in DNA, occurs when a particular subsequence of the DNA is completely inverted during the replication process. Usually such *reversals* occur in between what is called *inverted pairs*. For example, if the pattern TGAA is followed further downstream by the pattern AAGT, it is possible that the DNA segment that is enclosed in between these two patterns will be inverted and reattached, because binding regions at both ends are locally the same.



Mutation of an inverted pair.

Assignment

Write a function `mutation` that is passed a DNA sequence and a sequence pattern (both must be passed as string arguments to the function) and returns the mutated DNA sequence. The function should do this by finding the leftmost position of the pattern within the sequence, and next occurrence of the inverted pattern. The mutated DNA sequence is the result by reverting the segment in between the inverted pair. The function must return the original (non-mutated) DNA sequence in case the pattern does not occur in the original DNA sequence, or if it is not followed further downstream by the inverted pattern.

Example

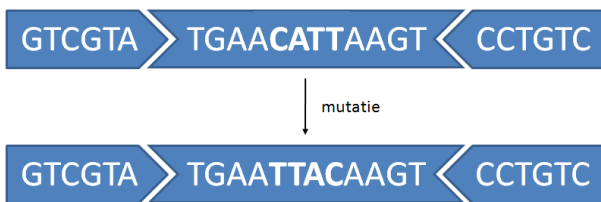
```
>>> mutation('GTCGTATGAACATTAAGTCCTGTC', 'TGAA')
'GTCGTATGAATTACAAGTCCTGTC'
>>> mutation('GTCGTATGAACATTAAGTCCTGTC', 'GTC')
'GTCCTGAATTACAAGTATGCTGTC'
>>> mutation('GTCGTATGAACATTAAGTCCTGTC', 'TCG')
'GTCGTATGAACATTAAGTCCTGTC'
```

The following table illustrates how the results in the above example were found, with inverted pairs being underlined.

SEQUENCE	PATTERN	RESULT
GTCGTAT <u>TGAACATTAAGT</u> CCTGTC	TGAA	GTCGTAT <u>TGAATTACAAGT</u> CCTGTC
<u>GTCGTATGAACATTAAGT</u> CCTGTC	GTC	<u>GTCCTGAATTACAAGTATGCT</u> GTC
<u>GTCG</u> TATGAACATTAAGTCCTGTC	TCG	GTCGTATGAACATTAAGTCCTGTC

DNA sequenties kunnen voorgesteld worden door strings die enkel bestaan uit de letters A, C, G en T. Een bepaalde vorm van mutaties in het DNA komt voor wanneer een deelsequentie van het DNA volledig wordt omgekeerd tijdens het replicatieproces. Normaalgezien doen dergelijke omkeringen (*reversals*) zich voor tussen wat men omgekeerde paren (*inverted pairs*) noemt. Wanneer het patroon TGAA bijvoorbeeld verderop wordt gevolgd door het omgekeerde patroon

AAGT, dan is het mogelijk dat het segment van het DNA dat tussen deze patronen zit ingesloten, wordt omgekeerd en terug wordt aangehecht, omdat de bindingen aan beide uiteinden lokaal dezelfde zijn.



Mutatie van een omgekeerd paar.

Opgave

Schrijf een functie `mutatie` die voor een gegeven DNA sequentie en een gegeven patroon (beide moeten als argument aan de functie doorgegeven worden) de gemuteerde DNA sequentie teruggeeft. De functie moet hiervoor eerst de meest linkse positie van het patroon binnen de sequentie bepalen, en ook het volgende voorkomen van het omgekeerde patroon. De gemuteerde DNA sequentie wordt dan bekomen door het segment tussen het omgekeerde paar om te keren. De functie moet de originele (niet-gemuteerde) DNA sequentie teruggeven indien het patroon niet voorkomt in de originele DNA sequentie, of indien het niet gevolgd wordt door een omgekeerd patroon.

Voorbeeld

```
>>> mutatie('GTCGTATGAACATTAAGTCCTGTC', 'TGAA')
'GTCGTATGAATTACAAGTCCTGTC'
>>> mutatie('GTCGTATGAACATTAAGTCCTGTC', 'GTC')
'GTCCTGAATTACAAGTATGCTGTC'
>>> mutatie('GTCGTATGAACATTAAGTCCTGTC', 'TCG')
'GTCGTATGAACATTAAGTCCTGTC'
```

De verklaring van de resultaten in bovenstaand voorbeeld vind je terug in onderstaand tabel, waarin de omgekeerde paren onderlijnd worden.

SEQUENTIE	PATROON	RESULTAAT
GTCGTAT <u>TGAACATTAAGT</u> CCTGTC	TGAA	GTCGTAT <u>TGAATTACAAGT</u> CCTGTC
<u>GTCGTATGAACATTAAGTC</u> CTGTC	GTC	<u>GTCCTGAATTACAAGTATGCT</u> GTC
<u>GTCGTATGAACATTAAGT</u> CCTGTC	TCG	GTCGTATGAACATTAAGTCCTGTC