Code	Abbreviation	Amino acid
A	Ala	alanine
R	Arg	arginine
N	Asn	asparagine
D	Asp	aspartate
C	Cys	cysteine
Q	Gln	glutamine
E	Glu	glutamate
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
L	Leu	leucine
K	Lys	lysine

M	Met	methionine
F	Phe	phenylalanine
P	Pro	proline
S	Ser	serine
T	Thr	threonine
W	Trp	tryptophan
Y	Tyr	tyrosine
V	Val	valine

N.B. this is a subset of the FASTA amino acid code letter set (it does not include B,U,Z,X,*,-).

Amino acid mutation matrix:

,A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V
A,9867,2,9,10,3,8,17,21,2,6,4,2,6,2,22,35,32,0,2,18
R,1,9914,1,0,1,10,0,0,10,3,1,19,4,1,4,6,1,8,0,1
N,4,1,9822,36,0,4,6,6,21,3,1,13,0,1,2,20,9,1,4,1
D,6,0,42,9859,0,6,53,6,4,1,0,3,0,0,1,5,3,0,0,1
C,1,1,0,0,9973,0,0,0,1,1,0,0,0,0,1,5,1,0,3,2
Q,3,9,4,5,0,9876,27,1,23,1,3,6,4,0,6,2,2,0,0,1
E,10,0,7,56,0,35,9865,4,2,3,1,4,1,0,3,4,2,0,1,2
G,21,1,12,11,1,3,7,9935,1,0,1,2,1,1,3,21,3,0,0,5
H,1,8,18,3,1,20,1,0,9913,0,1,1,0,2,3,1,1,1,4,1
I,2,2,3,1,2,1,2,0,0,9871,9,2,12,7,0,1,7,0,1,33
L,3,1,3,0,0,6,1,1,4,22,9947,2,45,13,3,1,3,4,2,15
K,2,37,25,6,0,12,7,2,2,4,1,9924,20,0,3,8,11,0,1,1
M,1,1,0,0,0,2,0,0,0,5,8,4,9875,1,0,1,2,0,0,4
F,1,1,1,0,0,0,0,1,2,8,6,0,4,9944,0,2,1,3,28,0
P,13,5,2,1,1,8,3,2,5,1,2,2,1,1,9924,12,4,0,0,2
S,28,11,34,7,11,4,6,16,2,2,1,7,4,3,17,9840,38,5,2,2
T,22,2,13,4,1,3,2,2,1,11,2,8,6,1,5,32,9869,0,2,9
W,0,2,0,0,0,0,0,0,0,0,0,0,0,1,0,1,0,9976,1,0
Y,1,0,3,0,3,0,1,0,4,1,1,0,0,21,0,1,1,2,9947,1
V,13,2,1,1,3,2,2,3,3,57,11,1,17,1,3,2,10,0,2,9901
,10000

The evolutionary mutation model in this task is as follows. For each amino acid in the current sequence, its probability of mutating to any other amino acid is given by the appropriate entry in the mutation matrix. Note that we apply an independence assumption, namely that the mutation of any amino acid is a random event unaffected by any other amino acid in the sequence. The model is to be run for a fixed number of iterations, each corresponding to a "generation". The choice by "evolution" of a mutation in each amino acid is modelled by indexing into the mutation matrix using a random number. The initial current sequence is given as input. At each step, the current sequence has each of its amino acids "mutated" with a certain probability. (Note that this mutation process does not *necessarily* cause the amino acid to change at every round of mutation. In fact, looking at the mutation matrix, it is likely that amino acids will change only rarely.) This new, mutated sequence then becomes the current sequence for the next generation.

The basic steps in your program will be to read in the initial sequence and run the evolutionary model based on the mutation matrix for 500 generations. The original sequence, plus all the mutated sequences generated, one per iteration, should be saved in a single output file in FASTA format. The generation number should go in the description line preceding each sequence, with your initial sequence numbered 0, the first mutated sequence numbered 1 and the final mutated sequence numbered 500.

Inputs and Outputs:

s0 your input sequence file containing the initial sequence in FASTA format

s501 the output file containing the initial sequence plus all 500 mutated sequences in FASTA format

You may assume that the I/O comes from stdin and goes to stdout.
For example, the program may be executed as follows:

```
% evolve < s0 > s501
```

The only input and output routines you need are to read and write sequences in FASTA format. Design these well for part 1 and you can re-use them in part 2.

Your program should **report an error and halt** if the input is not in FASTA format or the sequence contains something other than the above 20 amino acid code letters.

Your program should **report an error and halt** if the input does not contain exactly one sequence in FASTA format.

FASTA Sequence Format Description [restricted version]:

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column. It is recommended that all lines of text be shorter than 80 characters in length. An example sequence in FASTA format is:

> description of the sequence
CAKKRNWCGKNEDCCCPMKCIYAWYNQQGSCQTTITGLFKKC

More than one sequence can be included in the same file:

> description of initial sequence
CAKKRNWCKKNEDCCCPMKCIYAWHNQQGSCQTTISGLFKKC
> description of another sequence
CAKKRNWCKKNEDCCCPMKCIYAWHNQQGSCQTTISGLFKKC
> description of yet another sequence
CAKKRNWCKKNEDCCCPMKCIYAWHNQQGSCQTTISGLFKKC
