maxH - Using the Software

Sequence files must be "fasta" format, like this.

```
>seq1
MYYLKNTNFWMFGLFFFFYFFIMGAYFPFFPIWLHDINHISKSDTGIIFAAISLFSLLFQPLFGLLSDKLGLRKYLLWIITG
MLVMFAPFFIFIFGPLLQYNILVGSIVGGIYLGFCFNAGAPAVEAFIEKVSRRSNFEFGRARMFGCVGWALCASIVGIMFTI
NNQFVFWLGSGCALILAVLLFFAKTDAPSSATVANAVGANHSAFSLKLALELFRQPKLWFLSLYVIGVSCTYDVFDQQFANF
FTSFFATGEQGTRVFGYVTTMGELLNASIMFFAPLIINRIGGKNALLLAGTIMSVRIIGSSFATSALEVVILKTLHMFEVPF
LLVGCFKYITSQFEVRFSATIYLVCFCFFKQLAMIFMSVLAGNMYESIGFQGAYLVLGLVALGFTLISVFTLSGPGPLSLLR
RQVNEVA
>seq2
MTYYLKNTNFWMFGLFFFFYFFIMGAYFPFFPIWLHDINHISKSDTGIIFAAISLFSLLFQPLFGLLSDKLGLRKYLLWIIT
GMLVMFAPFFIFIFGPLLQYNILVGSIVGGIYLGFCFNAGAPAVEAFIEKVSRRSNFEFGRARMFGCVGWALCASIVGIMFT
INNQFVFWLGSGCALILAVLLFFAKTDAPSSATVANAVGANHSAFSLKLALELFRQPKLWFLSLYVIGVSCTYDVFDQQFAN
FFTSFFATGEQGTRVFGYVTTMGELLNASIMFFAPLIINRIGGKNALLLAGTIMSVRIIGSSFATSALEVVILKTLHMFEVP
FLLVGCFKYITSQFEVRFSATIYLVCFCFFKQLAMIFMSVLAGNMYESIGFQGAYLVLGLVALGFTLISVFTLSGPGPLSLL
RRQVNEVA
```

Sequences can be upper or lower case or mixed case. The number of characters per line does not matter. There can be "any" number of sequences in the file. The script was written to handle bacterial genomes efficiently so a few thousand sequences should be OK. The file must be saved as text with some name like "file_of_sequences" New_maxH_v3.3.pl should be capable of handling any number of such files at one time.

On UNIX:

1. Save the script as a text file with a name of your choice, perhaps, "New_maxH_v3.pl".
2. Type "perl New_maxH_v3.pl file_of_sequences"
3. Your results will be written to a file with .maxH appended to the name, "file_of_sequences.maxH" in the example here. This will first erase and then rewrite a file with that name in the same directory so be careful.

On Macintosh:

1. Open the script with (or paste it into) MacPerl and save it as a Droplet and quit MacPerl
2. Drag the icon of the sequence file, called for example "file_of_sequences"
3. Your results will be written to a file with .maxH appended to the name, "file_of_sequences.maxH" in the example here. This will first erase and then rewrite a file with that name in the same folder so be careful.
4. Quit from the droplet to run another file.

The output from this script, and from the runtime version mentioned above,

1. A listing for each protein giving the name, the maxH value, where in the sequence the value is, the length of the sequence and the probability of being integral (based on the E. coli analysis.)
2. An ordered list of all the maxH values which can be used to construct a histogram like those shown at the top of the page.