

Biopython – Cheat Sheet

Getting started

```
import Bio
```

```
from Bio.Seq import Seq
dna = Seq("ACGTTGCA")
print dna
```

(alternative)

```
from Bio.Alphabet import IUPAC
dna = Seq("AGTACACTGGT", IUPAC.unambiguous_dna)
```

Reverse complement, transcribing & translating

```
dna.reverse_complement()
rna = dna.transcribe()
rna.translate()
```

(alternative)

```
from Bio.Seq import reverse_complement, transcribe, translate
reverse_complement("GCTGTTATGGGTCGTTGGAAGGGTGGTCGTGCT")
```

Calculating GC-content

```
from Bio.SeqUtils import GC
GC(dna)
```

Calculating molecular weight (DNA only)

```
from Bio.SeqUtils import molecular_weight
molecular_weight("ACCCGT")
```

Loading sequences from a FASTA file

```
from Bio import SeqIO
for record in SeqIO.parse("ls_orchid.fasta", "fasta"):
    print record.seq, len(seq)
```

Plotting a histogram of seq lengths with pylab (needs to be installed separately)

```
import pylab
pylab.hist(sizes, bins=20)
pylab.title("%i orchid sequences\nLengths %i to %i" \
            % (len(sizes), min(sizes), max(sizes)))
pylab.xlabel("Sequence length (bp)")
pylab.ylabel("Count")
pylab.show()
```