# Genome sizes

Another interesting set of numbers with a huge range are the sizes of genomes.

```
In [1]:  genome_sizes = {
         'viroid': 3e2,      # (one tweet)
         'RNA_virus': 1e4,   # typical RNA viruses and phages (one page)
         'DNA_virus': 1e5,   # animal DNA viruses (10 pages)
         'bacteria': 5e6,    # bacterial genome (a serious book)
         'human': 3e9        # haploid human genome (one tenth of all of English wik
         }
```

As double stranded DNA, each base pair has a diameter of about 2nm and a height of 0.34nm. We can now readily estimate how long the total genome of these entities is:

```
In [2]:  for k,v in genome_sizes.items():
             print(f"{k} genome length: {v*0.00034:1.3e}um")
```

```
viroid genome length: 1.020e-01um
RNA_virus genome length: 3.400e+00um
DNA_virus genome length: 3.400e+01um
bacteria genome length: 1.700e+03um
human genome length: 1.020e+06um
```

The typical diploid human cell therefore contains about 2m of DNA, the typical bacterium about 2mm. These lengths are much longer than the linear dimensions of the cells.

Cartoons you often see of DNA loosly floating around in cells are therefore very misleading. The genome of a bacterium goes about 1000 times back and forth across the cell.
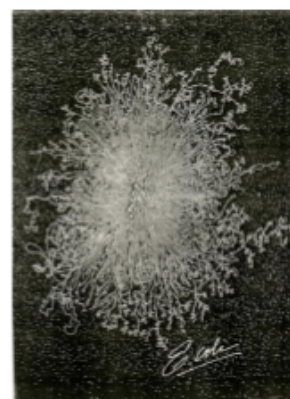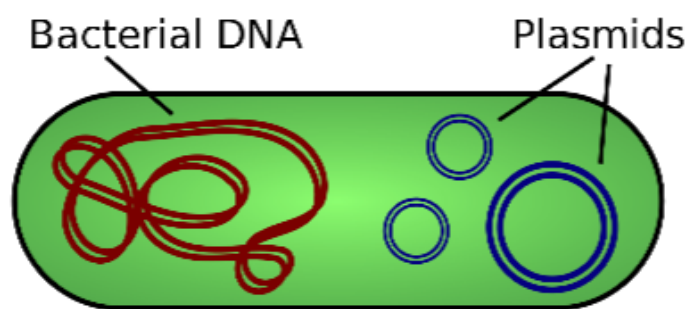


Left: a typical cartoon suggesting the length of a bacterial chromosome is similar to the length of the cell. Image source: kissclipart. Right: The DNA from an actual E.~coli cell that was lysed (the blob in the middle). A bacterial chromosome of length $5 \times 10^6$bps is about 1.5mm long -- roughly 1000-fold the length of the cell. Clearly, the DNA has to be pretty tightly packed inside the cell. Image source: R. Kavanoff

# How fast is a DNA polymerase?

To estimate this, we need

- the division time of the cell
- the genome size
- the number of origins of replication

Let's assume a single origin and typical bacterium with genome size $L = 5 \times 10^6$ and a division time of $\tau = 2500s \approx 40min$. At the origin, replication forks run off in both directions leaving $L/2$ for each fork. The polymerisation rate should naively be

$$\kappa = \frac{L}{2\tau} = \frac{5 \times 10^6}{5000s} = 1000/s$$

This estimate is more or less correct, but about a factor of 3 too high. The underlying reason is that rapidly dividing cells start replicating their daughters DNA before dividing. But overall, this estimate gave as a pretty idea of the rate.

## Dig deeper:

- If the human genome had one origin of replication, how long would it take? How long would it take to replicate the largest chromosome if there was one origin per chromosome?
- What is your guess for the number of replication origins in the human genome?

In [ ]:

In [ ]: