**PHYS 4xx Intro 3 - DNA composition and structure**

Composition and structure
- sugar, phosphate, base combination is called a *nucleotide*
- nucleotides polymerize to form DNA and RNA, through a linkage between a sugar from one nucleotide and a phosphate from another
     RNA: adenine, guanine, cytosine, uracil
     DNA: adenine, guanine, cytosine, thymine
- in double-stranded DNA, the bases lie in the interior of the helix, and hold it together through hydrogen bonding between base pairs: one purine and one pyrimidine: adanine/thymine and guanine/cytosine.
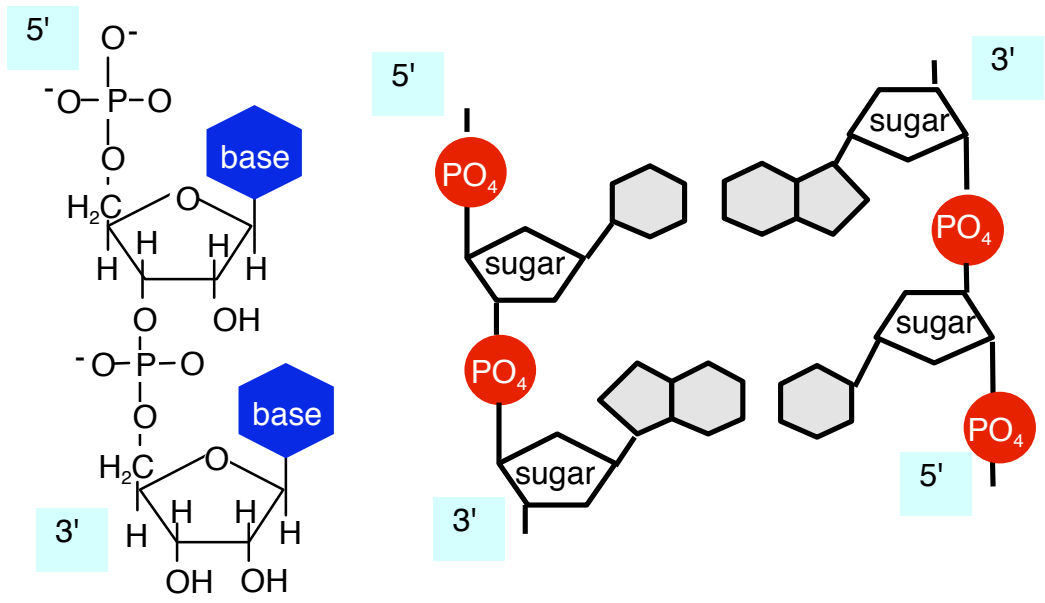
Species	genome size (10^6 bp)	genes	length (μm)
bacteriophage ϕX174	0.005386		
HIV type 2	0.009671	9	
mycoplasma genitalium	0.58	473	197
mycoplasma pneumonia	0.82	716	282
E. coli	4.7	4100	1600
Homo sapiens	3,000	~30,000	1.0 metres
amphibians	up to 80,000		27 km
plants	wide range, up to 900,000		

- typical protein has 300-400 AA, corresponding to ~1000 bp of DNA
- from above, DNA in the smallest cells is almost entirely coding
- human DNA is 99% non-coding

Microscopic geometry of DNA

So far, we have just considered the general structure of DNA as a helix with the bases facing into the interior of the helix. A more detailed picture would examine the orientation of the bases, which are small planar organic rings. One needs several variables to describe these orientations, both within a base pair, and between successive base pairs:

twist

twist, T

propeller twist
(between bases in a pair)

slide, S

roll, R
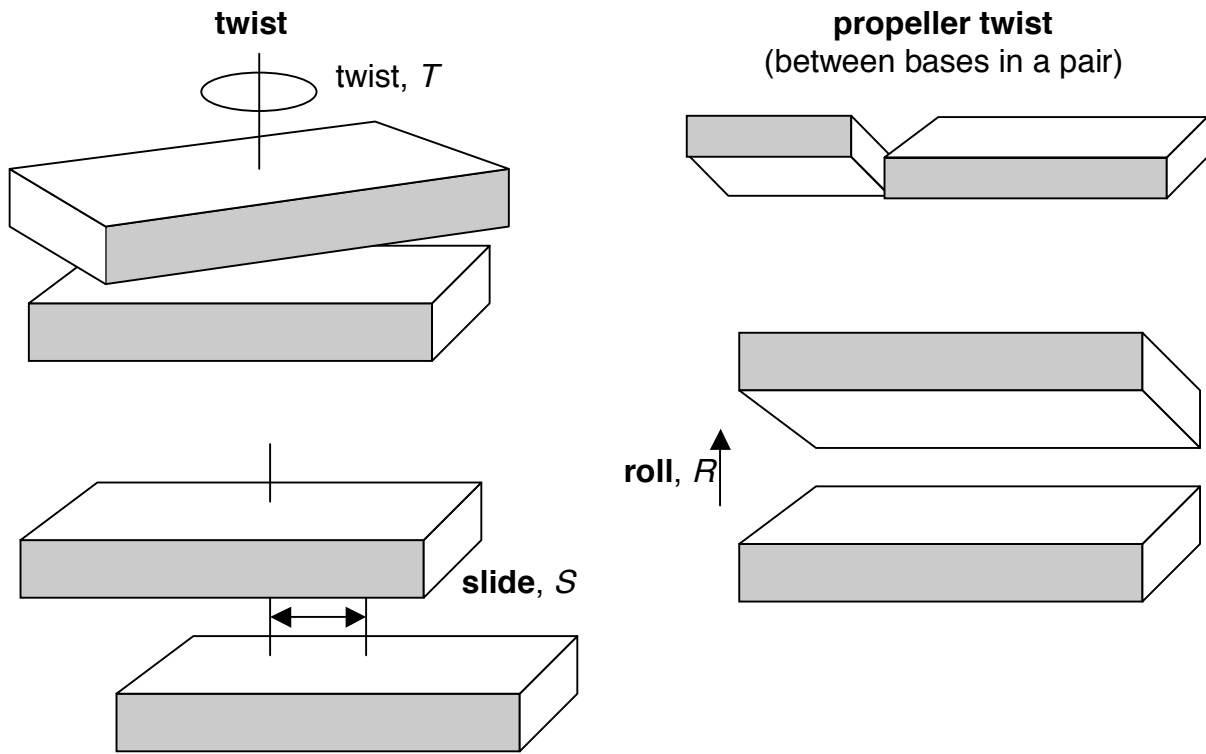
The average value of twist is approximately 32°, which corresponds to 360 / 32 = 11.3 base pairs per complete turn. But there is variation, with 36° for   -AAAAAA-
-TTTTTT-

Examples (*Calledine and Drew, Fig. 3.12*): scatter plot of data from X-ray studies
AA/TT steps are rigid

+20 °

roll

-AA-
-TT-

0 °

-10 °

-3 Å   slide   +3 Å
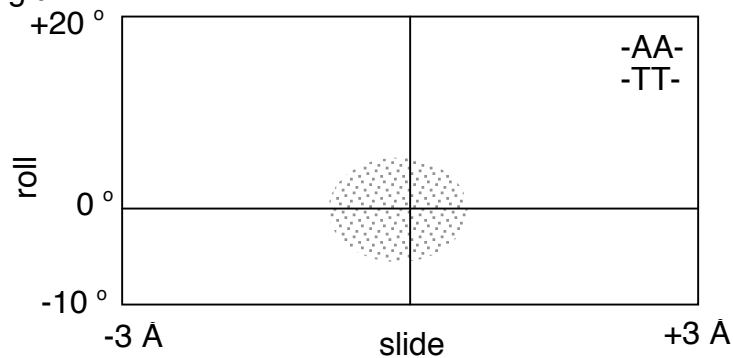
GG/CC steps are bimodal
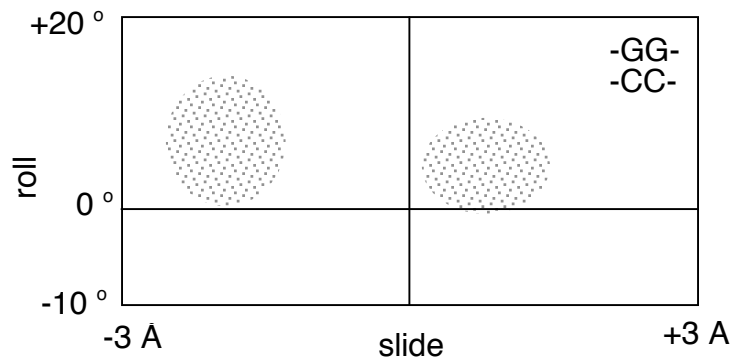
Length
- 0.34 nm per base pair:
  - at 3×10^9 bp, human DNA has a contour length of 1.0_2 m
- random coil has a much smaller dimension, $\xi_p = 53$ nm for helix
- in eukaryotic cells, DNA is wound around histones to aid in packaging

Supercoiling
- when subject to a torsional stress, DNA may form supercoils like a telephone cord (*demo*)
- geometry of coil is described by *twist* and *writhe*

Protein synthesis
Genetic information is stored in the DNA sequences using the GCAT alphabet. Both the sequence for a gene, and it's complement (ie, CGTA) is stored, although other information is also encoded to indicate which is the correct direction for transcription.

The sequence on the DNA master blueprint corresponding to a specific protein is transcribed onto a string of messenger RNA or mRNA, written in the GCAU alphabet. It is the mRNA sequence that is called the sequence for the protein.

The mRNA is then read to create the amino acid sequence. There are 20 commonly used amino acids. Each one of these can be attached to a short string of RNA called tRNA (t for transfer), about 70-90 nucleotides long. At one end of the tRNA is an attachment site for an amino acid, further along the string is a series of three nucleotides which are the **anticodon** for the AA, complementary to the **codon** carried by the mRNA.

Folded, tRNA looks like an L, but somewhat unfolded, it appears like a cloverleaf:

3'
attachment site for
amino acid

anticodon
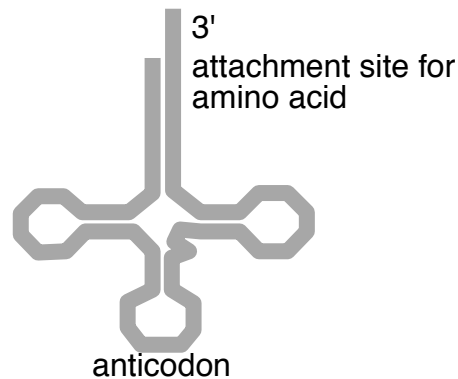
At 3 nucleotides per codon, there are 4^3 = 64 unique codes for amino acids - far more than 20. Thus:
- some AAs can be carried by more than one tRNA
- a given tRNA may be able to fit into several different slots on the mRNA, if it is primarily sensitive to the first two nucleotides in the three-letter codon.

```
Ala    GCA  GCC  GCG  GCU
Arg    AGA  AGG  CGA  CGC  CGG  CGU
Asp    GAC  GAU
Asn    AAC  AAU
Cys    UGC  UGU
Glu    GAA  GAG
Gln    CAA  CAG
Gly    GGA  GGC  GGG  GGU
His    CAC  CAU
Ile    AUA  AUC  AUU
Leu    UUA  UUG  CUA  CUC  CUG  CUU
Lys    AAA  AAG
Met    AUG
Phe    UUC  UUU
Pro    CCA  CCC  CCG  CCU
Ser    AGC  AGU  UCA  UCC  UCG  UCU
Thr    ACA  ACC  ACG  ACU
Trp    UGG
Tyr    UAC  UAU
Val    GUA  GUC  GUG  GUU

start  AUG (met)
stop   UAA  UAG  UGA
```