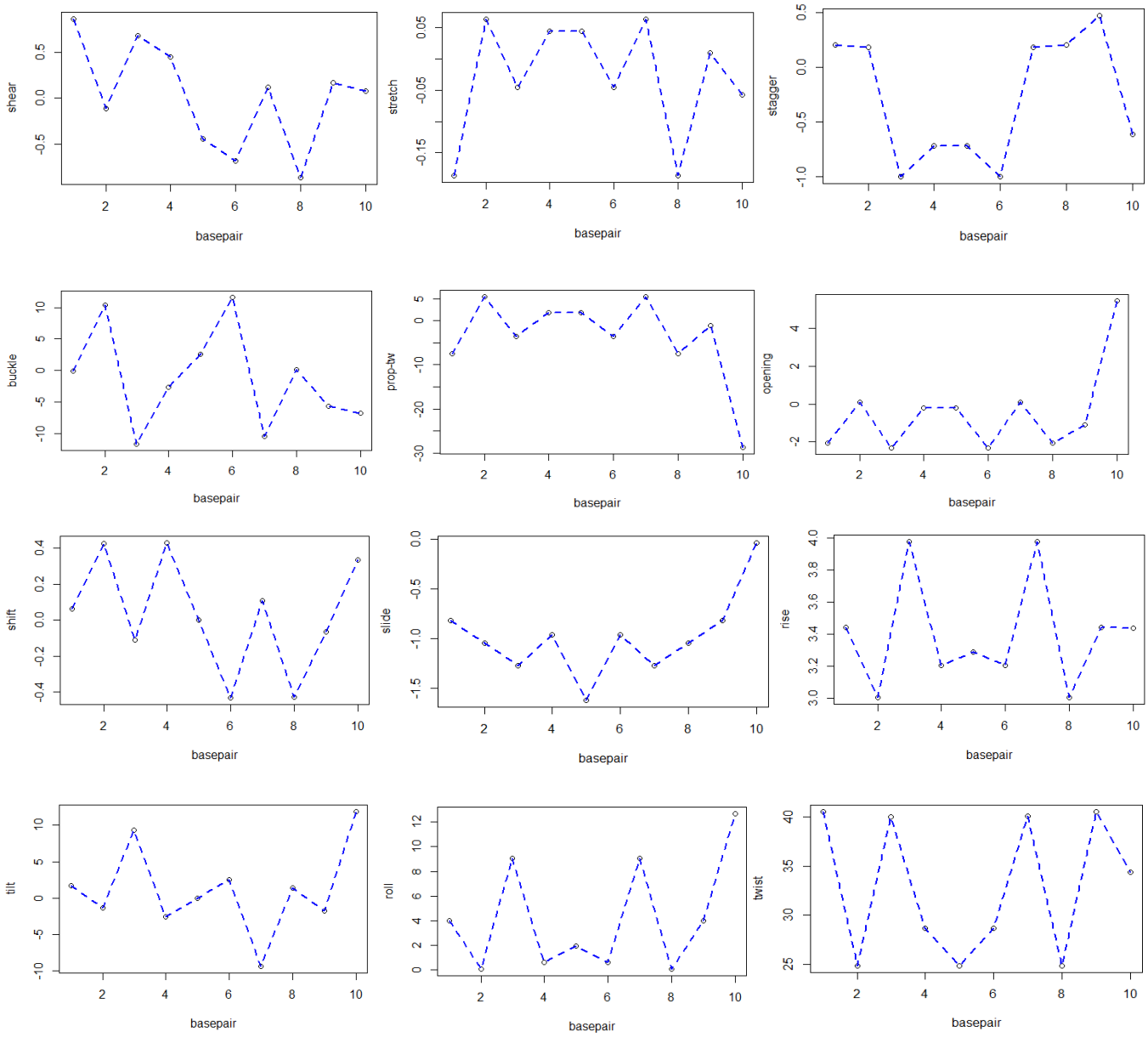
DNA modelling report.

Part 1.

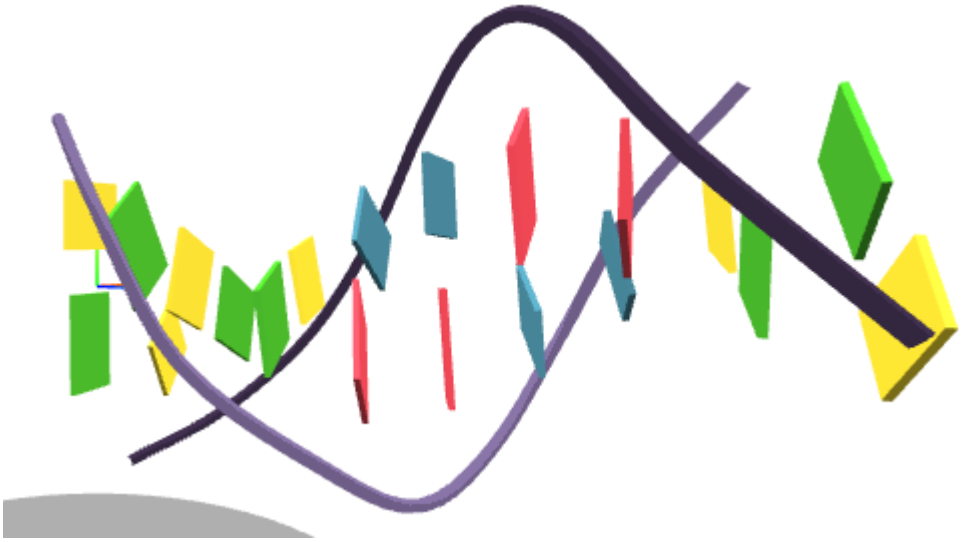
From the protein bank 1 DAU structure was chosen − analog of dickerson-drew DNA dodecamer with 6'-alpha-methyl carbocyclic thymidines, NMR, minimized average structure. It's not a standard DNA due to 'tt' in the sequence as well as not distorted by other molecules. The file with "simple" base-pair and step parameters:

```
  12 # base-pairs
   0 # ***local base-pair & step parameters***
#        Shear    Stretch   Stagger   Buckle   Prop-Tw   Opening    Shift     Slide     Rise      Tilt      Roll      Twist
C-G     -0.078    -0.057    -0.617     6.703   -28.705     5.498    0.000     0.000    0.000     0.000     0.000     0.000
G-C     -0.167     0.010     0.474     5.641    -1.098    -1.123   -0.336    -0.032    3.437   -11.921    12.691    34.385
C-G      0.862    -0.187     0.201    -0.118    -7.451    -2.077    0.065    -0.820    3.443     1.761     4.009    40.563
G-C     -0.116     0.064     0.180    10.393     5.457     0.092    0.428    -1.045    3.003    -1.376     0.080    24.809
A-t      0.679    -0.046    -1.003   -11.665    -3.588    -2.333   -0.109    -1.273    3.979     9.365     9.054    40.041
A-t      0.445     0.044    -0.717    -2.625     1.878    -0.194    0.432    -0.963    3.202    -2.505     0.620    28.676
t-A     -0.444     0.044    -0.718     2.630     1.901    -0.188    0.000    -1.622    3.286     0.010     1.966    24.841
t-A     -0.679    -0.046    -1.003    11.663    -3.600    -2.338   -0.432    -0.963    3.202     2.498     0.627    28.671
C-G      0.117     0.064     0.181   -10.397     5.477     0.085    0.109    -1.273    3.980    -9.368     9.061    40.046
G-C     -0.863    -0.187     0.202     0.125    -7.440    -2.073   -0.427    -1.045    3.002     1.370     0.070    24.803
C-G      0.167     0.010     0.474    -5.636    -1.097    -1.116   -0.065    -0.820    3.443    -1.748     4.006    40.562
G-C      0.079    -0.057    -0.617    -6.694   -28.714     5.484    0.335    -0.032    3.437    11.915    12.694    34.392
```

Plot of the values in R:



There are a lot of differences between experimental and model plots. Comparing the coordinate values of the experimental and model structures, twist, shift, roll and propeller coordinates differ most. The reason is also ‘tt’ addition in structure.
Constructed the cgDNA+ model for the sequence (CGCGAAttCG):

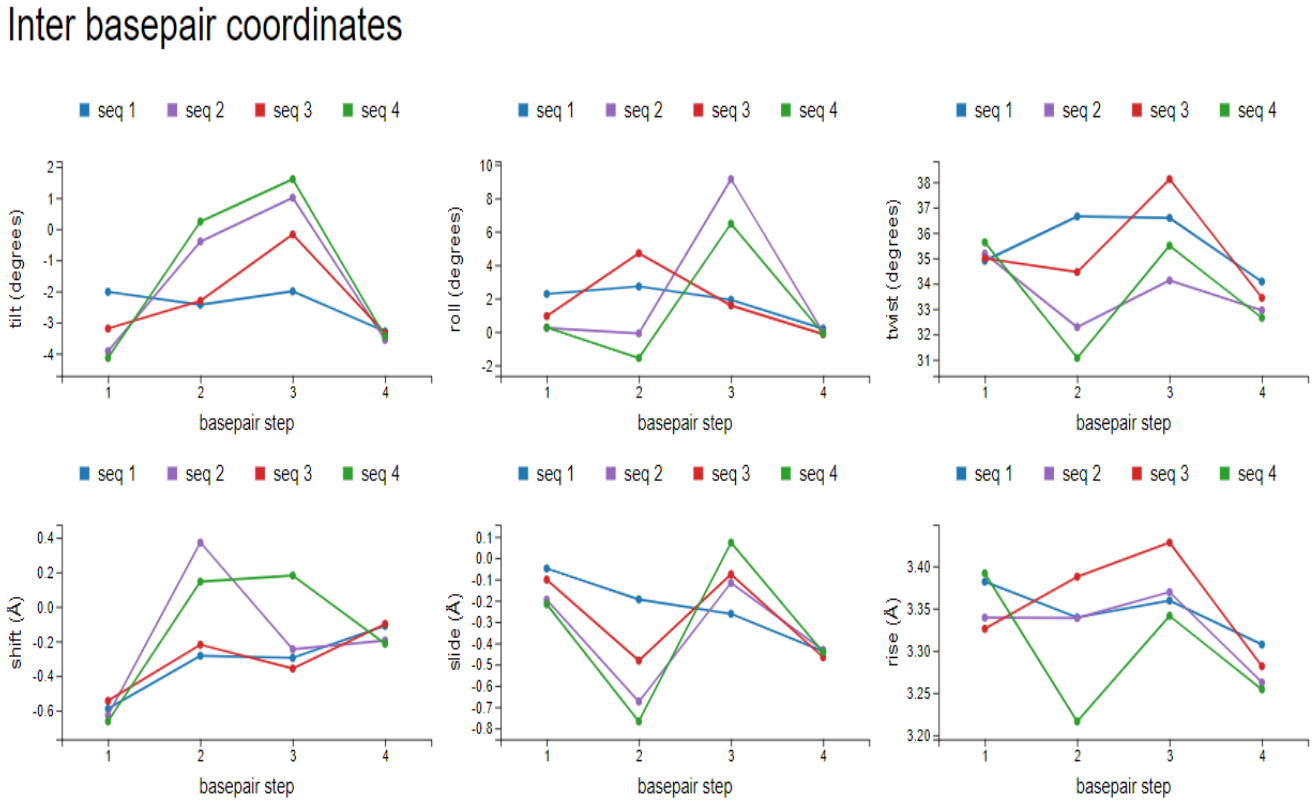


Both structures look similar, so the cgDNA+ model prediction is appropriate.

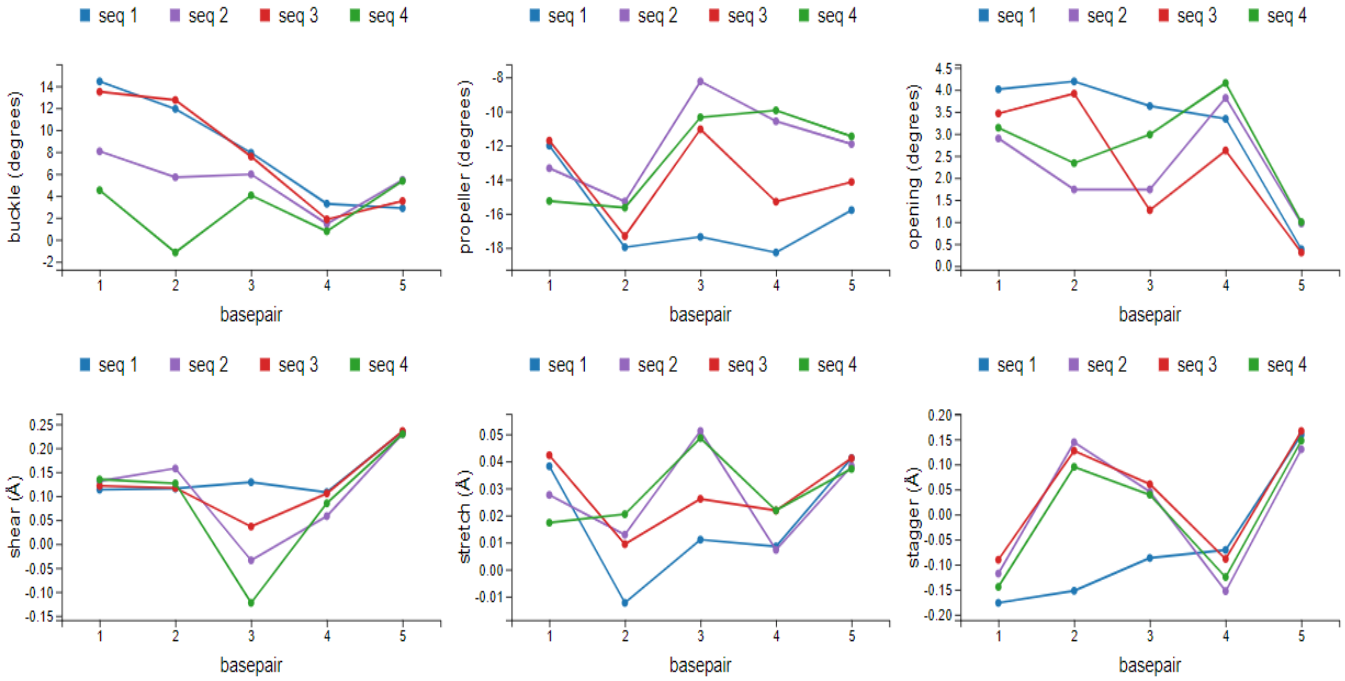
Part 2.

Seq1 − AAAAA, seq2 – AACAA, seq3 – AAGAA, seq4 – AATAA.

The 2D coordinate plots:



Intra basepair coordinates

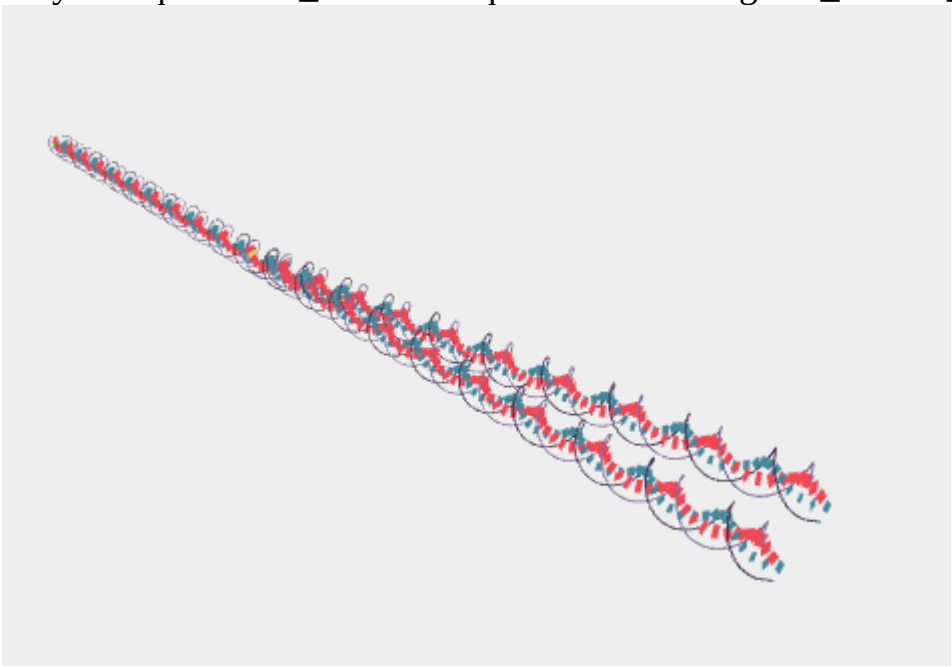


All sequences are not so far from one another, however some coordinates are affected much. The most affected coordinate, from my point of view, is roll. Both A to G and A to T mutations have an impact on the shape of DNA.

When an A to G mutation occurs, the purine-pyrimidine base pairing in DNA is disrupted, as guanine (G) is a purine and adenine (A) is a pyrimidine. This can cause changes in the local DNA structure, such as altered base stacking, helix geometry, and minor groove width.

Similarly, when an A to T mutation occurs, the base pairing between adenine (A) and thymine (T) is disrupted, leading to changes in local DNA structure. However, since both A and T are pyrimidines, the impact of an A to T mutation may be less pronounced than that of an A to G mutation.

Poly A sequence: A_205 and sequence with change: A_102CA_102



This change disrupts the continuous string of adenine nucleotides in the poly(A) sequence and the shape of new DNA is different.

Generated random sequences:
1 tttagtgtca
2 gacccaaatt
3 tggcctcgtc
4 gtggttaatg
I presume that the 2 last seq are the closest in shape.