**Goal:**

1)Input size-genome level

2)make a visualizer

3)change input type from di(crystallized image) to tri/tetra(md simulated image)

**Understanding of the paper:**

**Test\_Seq.txt has 100 sequences of length 1001**

* Why promoter prediction is needed? Pg1

For genome annotation-to understand transcription regulatory networks

* Why has it been tough?? Pg1

Genomic sequences are highly adaptive and diversified to enable survival in diverse conditions.

Variability in the length of 5’ un-translated regions and presence of multiple TSS

High gene density, very less inter-genic spaces or having overlapping coding regions

* What is meant by parameter signals?

Structural/energetic properties like bendability, curvature etc.  
 STRUCTURAL: DNA shape is cumulation of certain parameters. So instead of taking their

Cumulative effect, we consider different parameters. (28)

ENERGETIC: hydrogen bond, stacking energy and solvation energy.

**UNDERSTANDING THE getParameterDetails file:**

**FLOW:**

getParameterDetails-> iterateSequences-> calculateParameters-> calculateMovingAverages-> normalizeMovingAverages->return

then iterateSequences-> combineStructEnergyParams->return

then iterateSequences -> transformStructEnerMap

**WHAT STORES WHAT: UNDERSTANDING THE FUNCTIONS**

**1)readSequenceFile**

Reads the sequence file, stores sequences of size 1001 nucleotides as array elements. Array “content” is an array of 100 such sequences.

We check if content[i] is not none and size is >1000

Content is changes to a dict called sequence\_map with index as the key and value: sequence of 1001 nucletides

**2)getParameterDetails()**

Takes as input the sequence\_map. Then runs the function iterateSequences which calles calculateParameters.

**getParameterDetails-> iterateSequences-> calculateParameters->**

Calculate parameters assigns 31 (‘a’,b,c,d,e,……,’ae’) parameter values to each nucleotide pair in the sequence. On passing one sequence of 1001 nucleotides, we get a param\_map for each letter of 1000 values each. (**1000 nicleotide pairs in 1001 nucleotides)**

The corresponding param\_map is now sent as parameter to calculateMovingAverages. We take a moving window size of 25. Similary normalizedAverage is calculated.

combineStructEnergyParams: takes in as input the parameter to define energyStructparams(understand better)