

ECES T580 Lab 6

Bhautik (Brian) Amin

Contents

- [Lab 6.1.1](#)
- [Lab 6.2.1](#)

Lab 6.1.1

```
%Import data
[headers, seq] = fastaread('data.fasta');
```

Calculate Entropy and then Rseq. Calculate height for b x l table and then plot Rseq vs position (Assume $e(n) = 0$)

```
H = [];
R = [];
% find how many positions are in the sequence
temp = seq(1);
positions = length(temp{1}); % This assumes all sequences given are the same length
no_of_sequences = length(seq);
p_a = 0; p_t = 0; p_g = 0; p_c = 0;

% Make a bxl table
height_table = eye(4,20);
for i=1:positions % For each position of the sequence
    for j=1:no_of_sequences % For each sequence within the array of sequences
        temp = seq(j);
        seq_at_pos = upper(temp{1}(i)); % Pull out the char in the position
        switch seq_at_pos % Run through switch statement and increment found NT
            case 'A'
                p_a = p_a + 1;
            case 'G'
                p_g = p_g + 1;
            case 'T'
                p_t = p_t + 1;
            case 'C'
                p_c = p_c + 1;
        end
    end

    % Calculate the probability of each NT
    p_a = p_a/positions;
    p_g = p_g/positions;
    p_c = p_c/positions;
    p_t = p_t/positions;
    % calculate H for this position
    H_pos = -p_a*log2(p_a) - p_t*log2(p_t) - p_g*log2(p_g) - p_c*log2(p_c);
    H = [H, H_pos];
    R = [R, 2-H_pos]; % Calculate Rseq for this position
    % Construct the height vs position table for this specific position
    height_table(1,i) = p_a*(2-H_pos); % A
    height_table(2,i) = p_c*(2-H_pos); % C
    height_table(3,i) = p_g*(2-H_pos); % G
    height_table(4,i) = p_t*(2-H_pos); % T
    % Loop again for the next position
end
```

Now that we have our Rseq, positions, and B x L table we can plot:

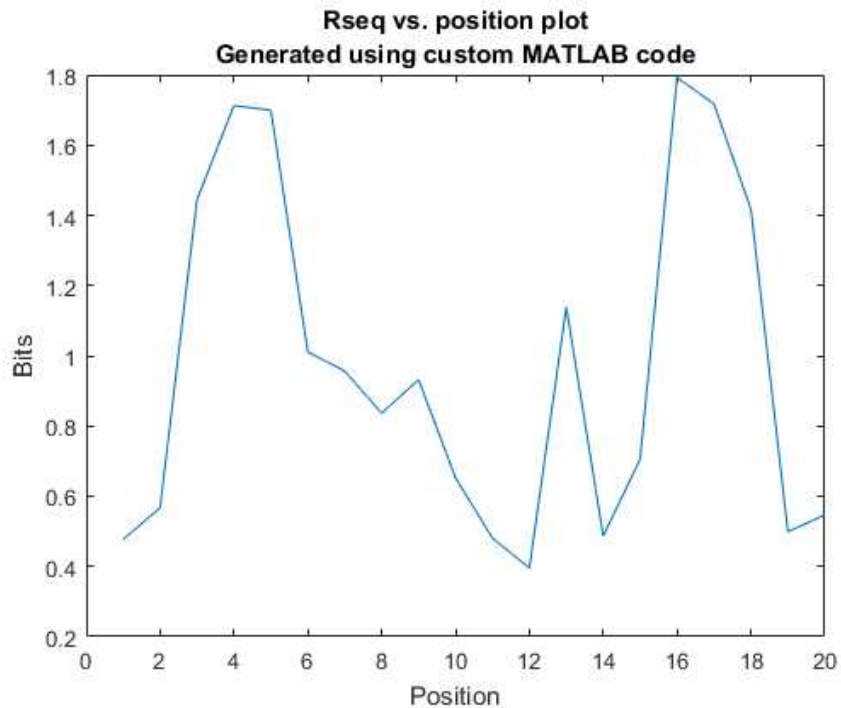
Plot R vs Position using custom MATLAB code above

figure(1)

```

plot(1:positions, R)
title('Rseq vs. position plot','Generated using custom MATLAB code')
xlabel('Position')
ylabel('Bits')

```



Here is the Height vs Position table

```
height_table
```

```
height_table =
```

Columns 1 through 7

0.1193	0.3755	0.0478	0.0028	0.0001	0.0506	0.7197
0.0477	0.0879	1.3111	0.0777	0.0039	0.0001	0.0478
0.0239	0.0581	0.0074	0.0004	1.6145	0.1998	0.0094
0.2625	0.0439	0.0778	1.6315	0.0810	0.7610	0.1795

Columns 8 through 14

0.0733	0.7031	0.1552	0.2939	0.1308	0.9296	0.1659
0.0857	0.0048	0.0328	0.0732	0.1217	0.0744	0.0989
0.0422	0.1421	0.0376	0.0254	0.0010	0.0002	0.0000
0.6351	0.0820	0.4272	0.0878	0.1421	0.1343	0.2219

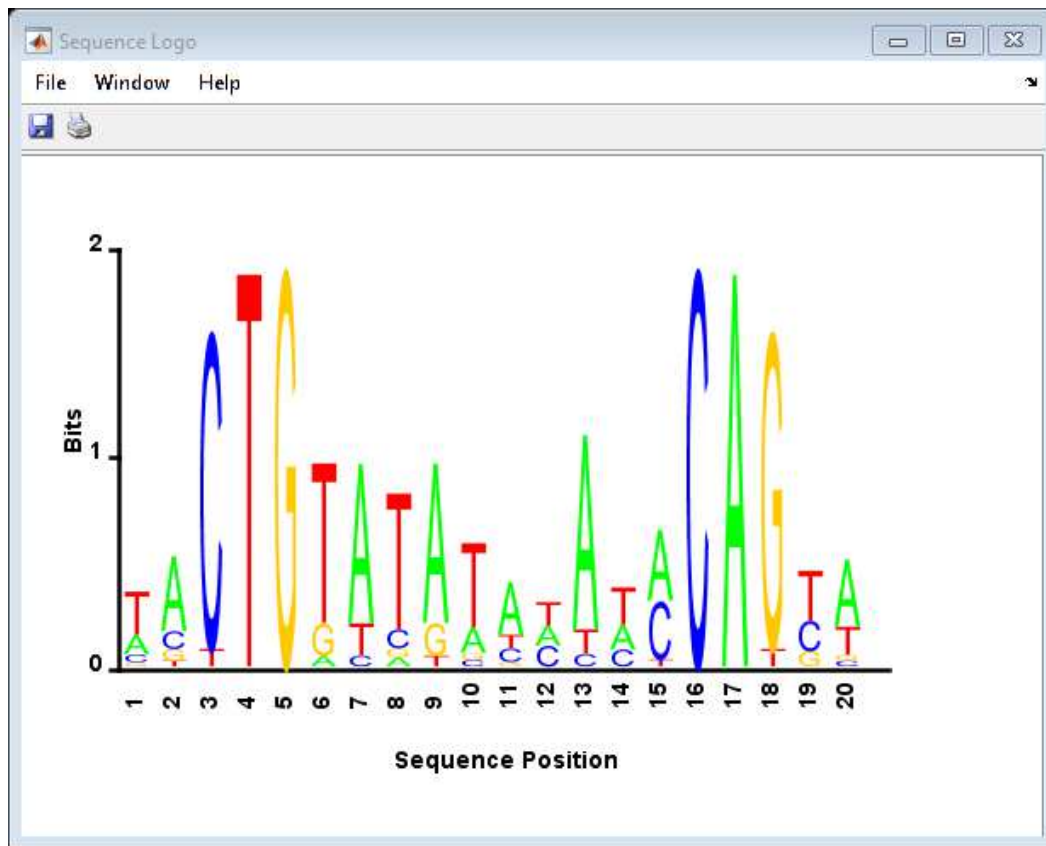
Columns 15 through 20

0.3647	0.0463	1.6337	0.0674	0.0012	0.3279
0.2893	1.7394	0.0833	0.0034	0.1499	0.0355
0.0000	0.0000	0.0000	1.2753	0.0974	0.0327
0.0514	0.0065	0.0003	0.0709	0.2509	0.1503

Lab 6.2.1

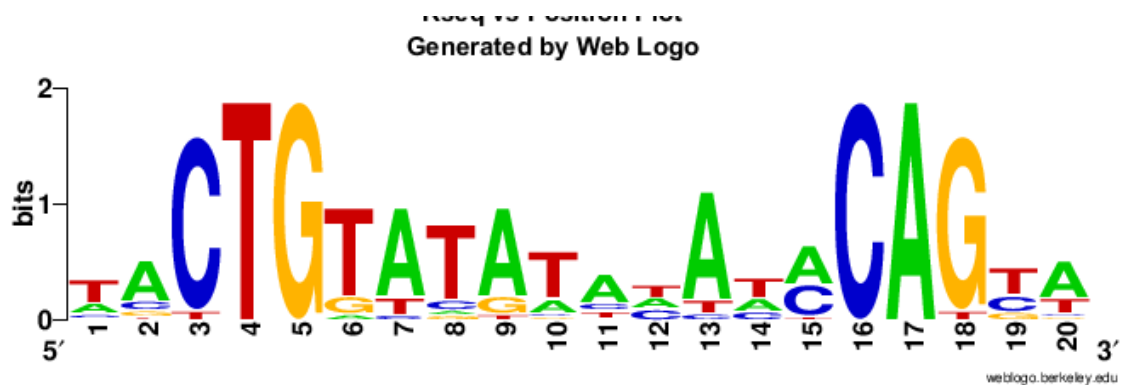
Now we can try to plot using the built in MATLAB function seqlogo

```
[W, handle] = seqlogo(seq);
```



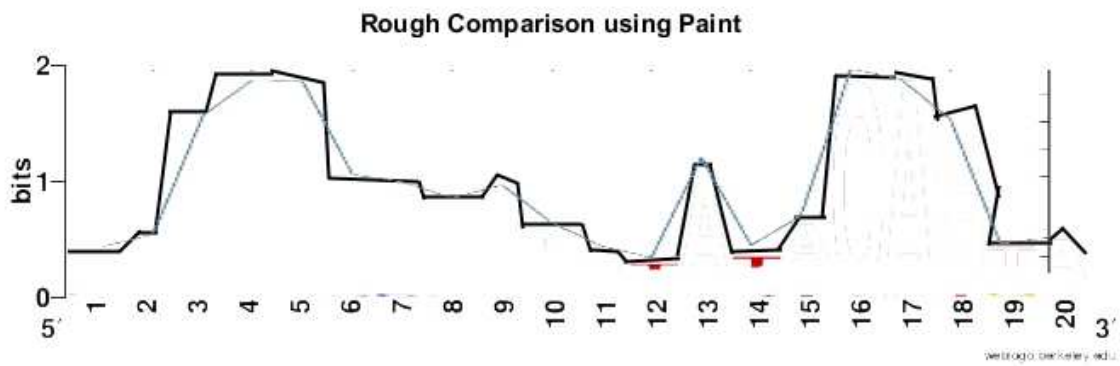
Show plot generated by Weblogo as well

```
figure(3)
imshow('web_logo.png')
title({'Rseq vs Position Plot', 'Generated by Web Logo'})
```



As we can see from all three plots, the general trend is the same. Where there are large features at positions 3,4,5, 13, 16, 17, 18 and so forth. These are consistent with all three plots. We can use a program such as paint to super impose the custom code generated plot and the web logo plot to see how they compare visually:

```
figure(4)
imshow('rough_comparison.jpg')
title('Rough Comparison using Paint')
```



Where in blue is the MATLAB generated plot and the black line is a hand traced (Not accurate) sketch over the web logo image. We can see that they both nearly line up