

## SUPPLEMENTARY MATERIAL - MFOLD WEB SERVER

A variety of output types and options are illustrated through the folding of a randomly generated DNA sequence.

The random sequence is

```
ACTTCGCAAG CACGCGTAGG GAAAGGCACC ATGTATCACG ATATTACATA CTAAGAGCGT CAACGTGAAT
ACCTGCTGGA TACTGTGTGG GCCGTGGTGA AAGTTTGATC CGCAAAGCAG CCCCTGTAAC TGTACTCGCG
GCAAGAGCAT CGCAGCAGTA TGTGCGTCTG AATGCGACAC GGAAGGCACG GCGGGACCCA
```

It is folded as DNA under default conditions on the regular mfold web server. Six foldings are generated. A text representation of the first folding is given in Figure 1. The sequence name is “RAN 130299481”, where the number refers to a “seed” for a random number generator.

The minimum folding free energy is -18.3 kcal/mole. The default 5% of this number is less than 1 kcal/mole, and so  $\delta\delta G$ , the free energy increment, is rounded up to 1 kcal/mole. Figure 2 is the default *energy dot plot* for this folding.

Structure 1

Folding bases 1 to 200 of RAN 130299481  
dG = -18.3 dH = -347.3 dS = -1060.8 Tm = 54.3

```
      10
.-A| T  AA
   CT CGC G
   GA GCG C
\ -^ T  CA
```

```
      20          30          40          50
.-GGAAAGGCACCATGTA      ATATTACATAC
                        TCACG          \
                        AGTGC          T
\ ----- AACTGCGAGAA
                        60
```

```
      70          80          90          100          110          120
ATA--      GATA--      .-GG      - AA      TGATCCGCAAA      C
      CCTGCTG      CTGTGT      GCCGTG GTG AGTT      GCAG \
      GGGCGGC      GGCACA      CGGCGC CAT TCAA      TGTC C
ACCCA      ACGGAA      \ --      T G-      ----- C
200          190          180          140          130
```

```
      150
AAGA      -----      G G
      GCAT      CGCA CA \
      CGTA      GCGT GT T
G---      AGTCT      - A
      170          160
```

Figure 1: The text output of folding 1 of a random DNA. It has 3 stems that touch the exterior loop, and a single multi-branch loop. This figure should be compared with other, easier to interpret drawings of this secondary structure. The | and ^ symbols point to the base pair that was selected for the double traceback to generate this folding.

## Fold of RAN 130299481 at 37 C.

$\delta G$  in Plot File = 1.0 kcal/mole

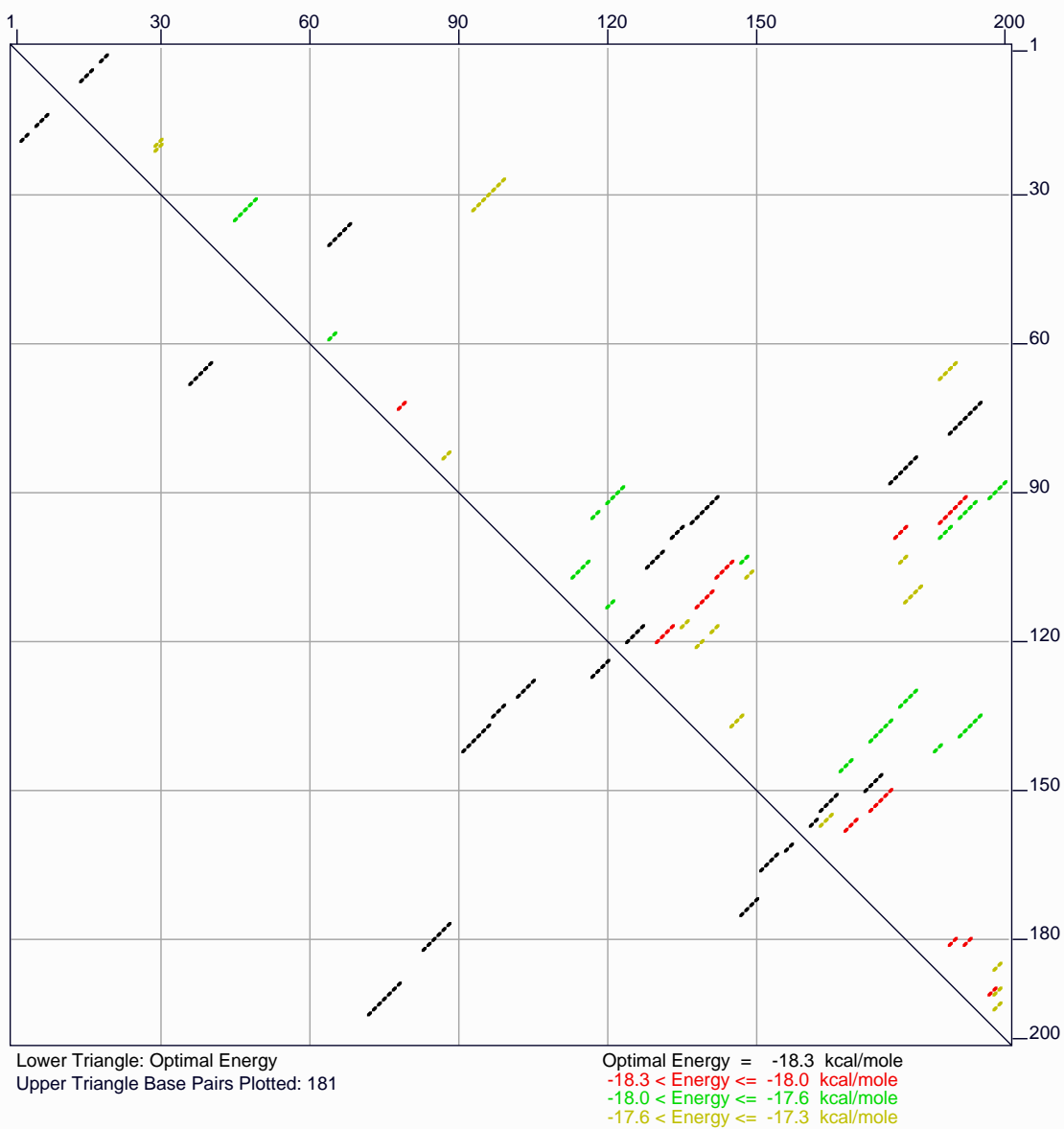


Figure 2: The default *energy dot plot* displays all close to optimal base pairs using 4 colors. Note that most of the variability in folding occurs in the sub-triangle with vertex (90,200)

## Energy Dot Plot at Magnification : 2

$\delta G$  in Plot File = 1.0 kcal/mole

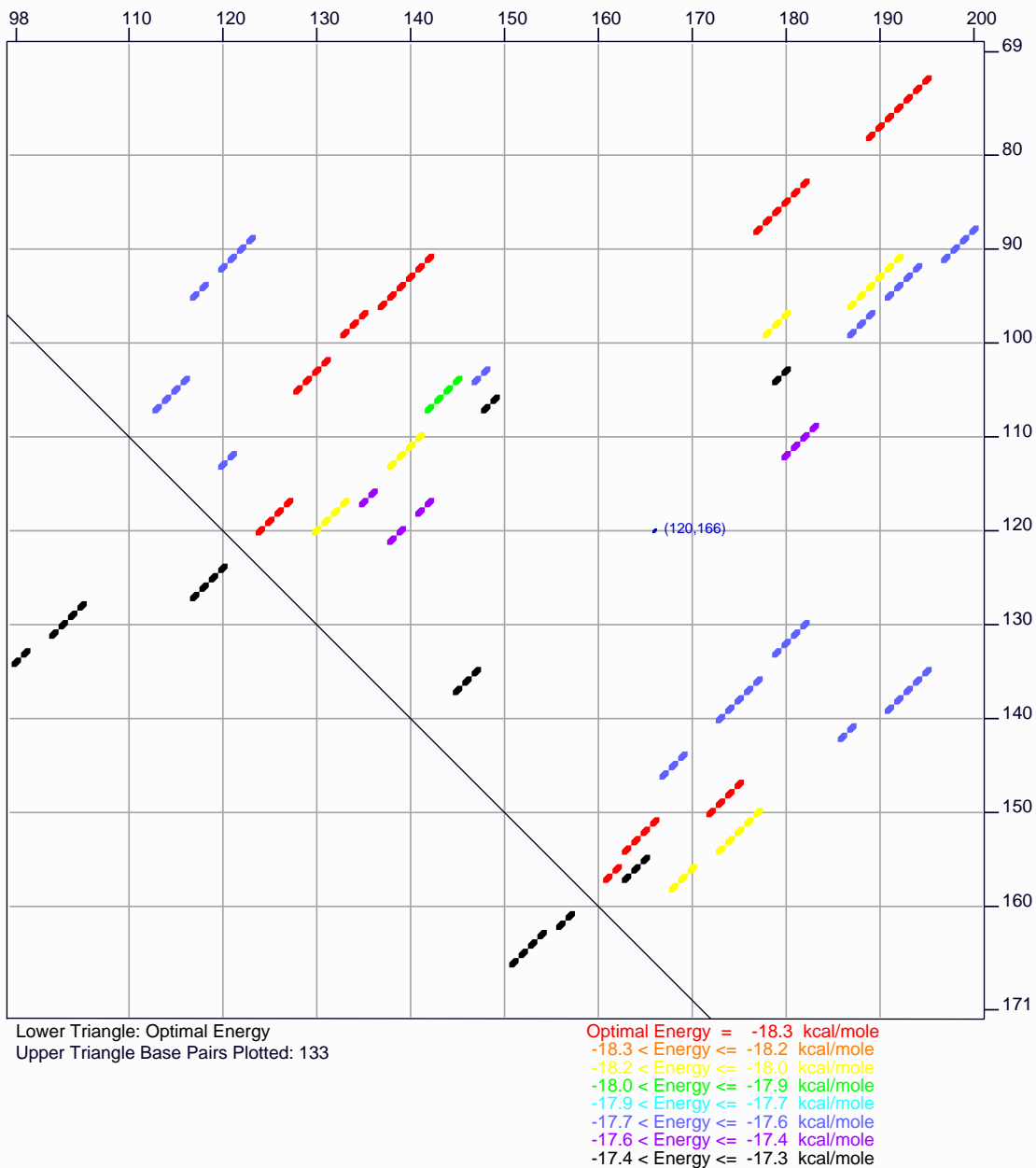


Figure 3: A magnification of 2 has been generated centered about (120,166). A total of 8 different colors have been chosen, although not all occur since  $\delta\delta G$  is so small. Base pairs in optimal foldings are colored in red.

## Energy Dot Plot at Magnification : 2

$\delta G$  in Plot File = 1.0 kcal/mole Filter: 4

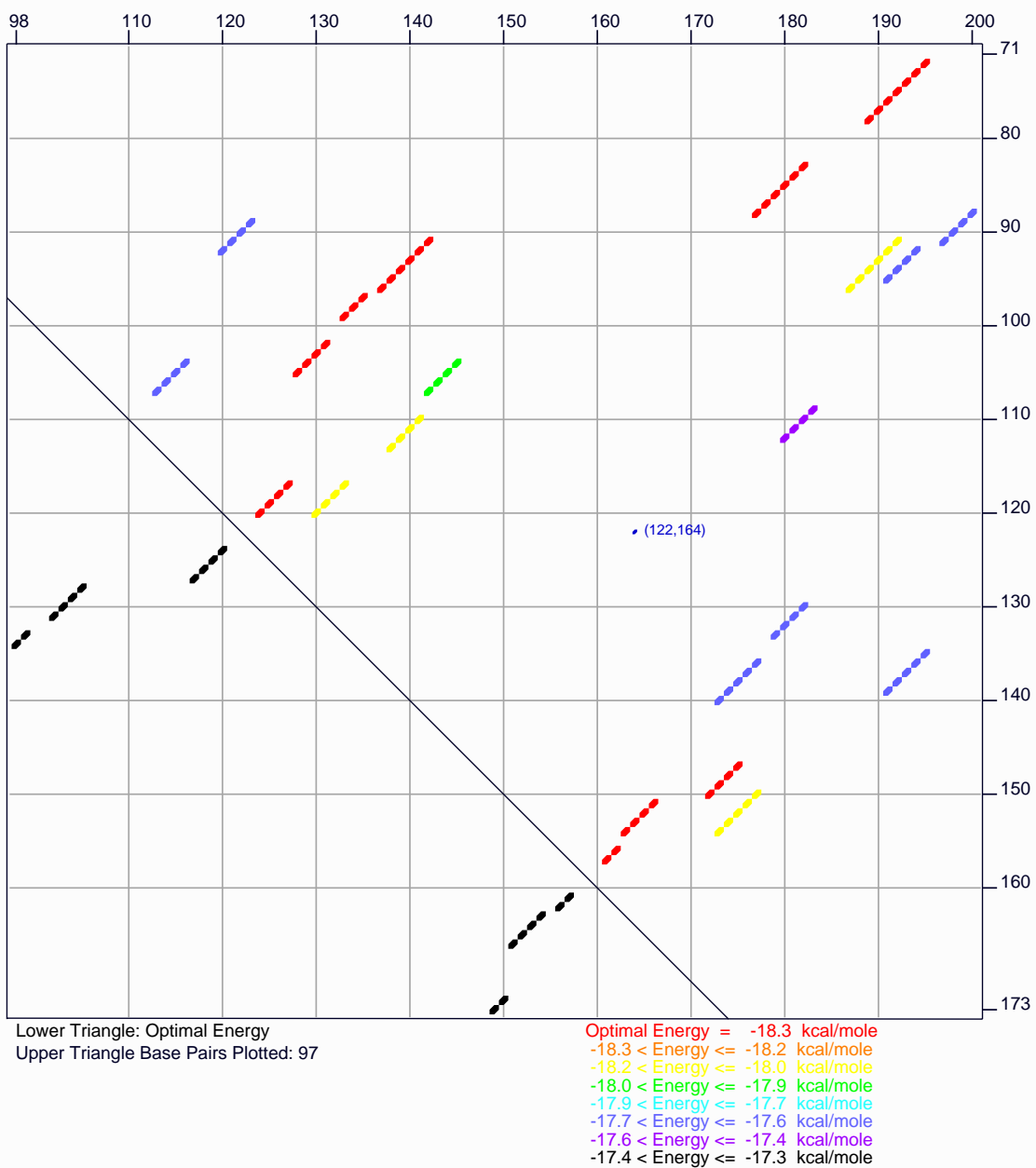
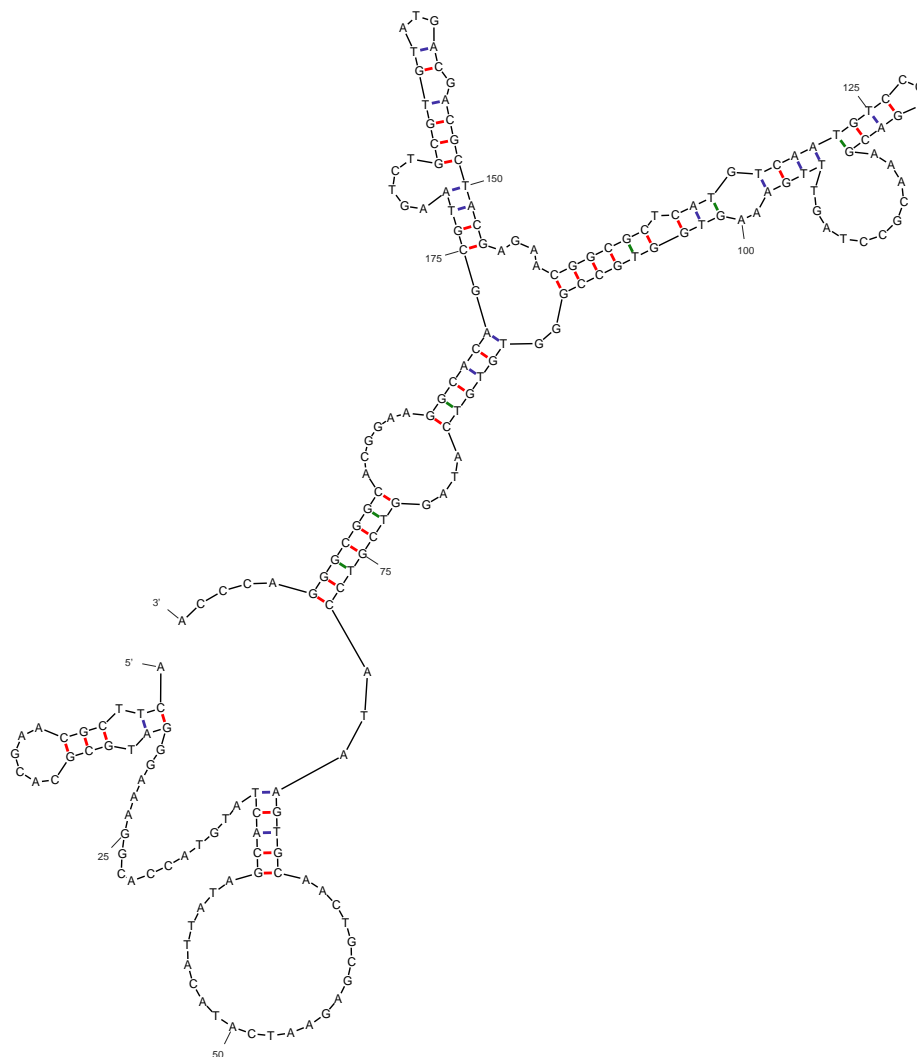
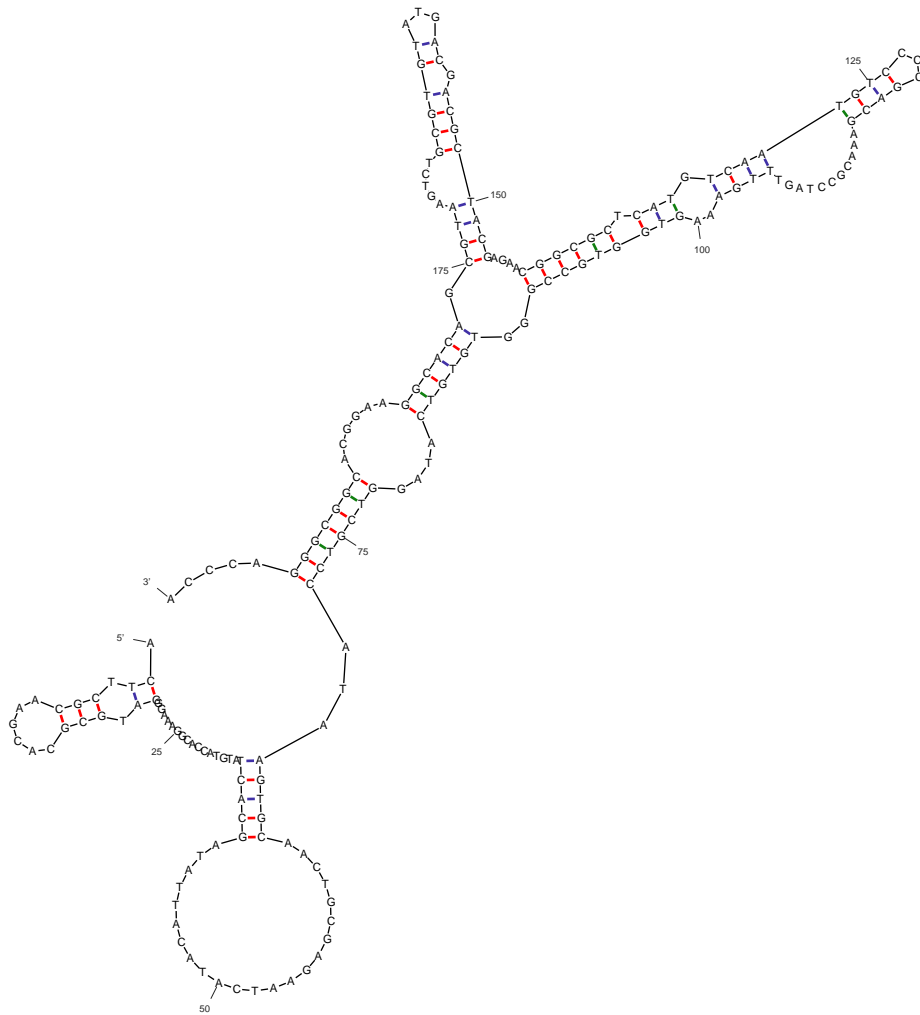


Figure 4: This *energy dot plot* is essentially the same as in Figure 3, except that the magnification is about (122,164) and only helices of size 4 or greater are plotted (filter = 4).



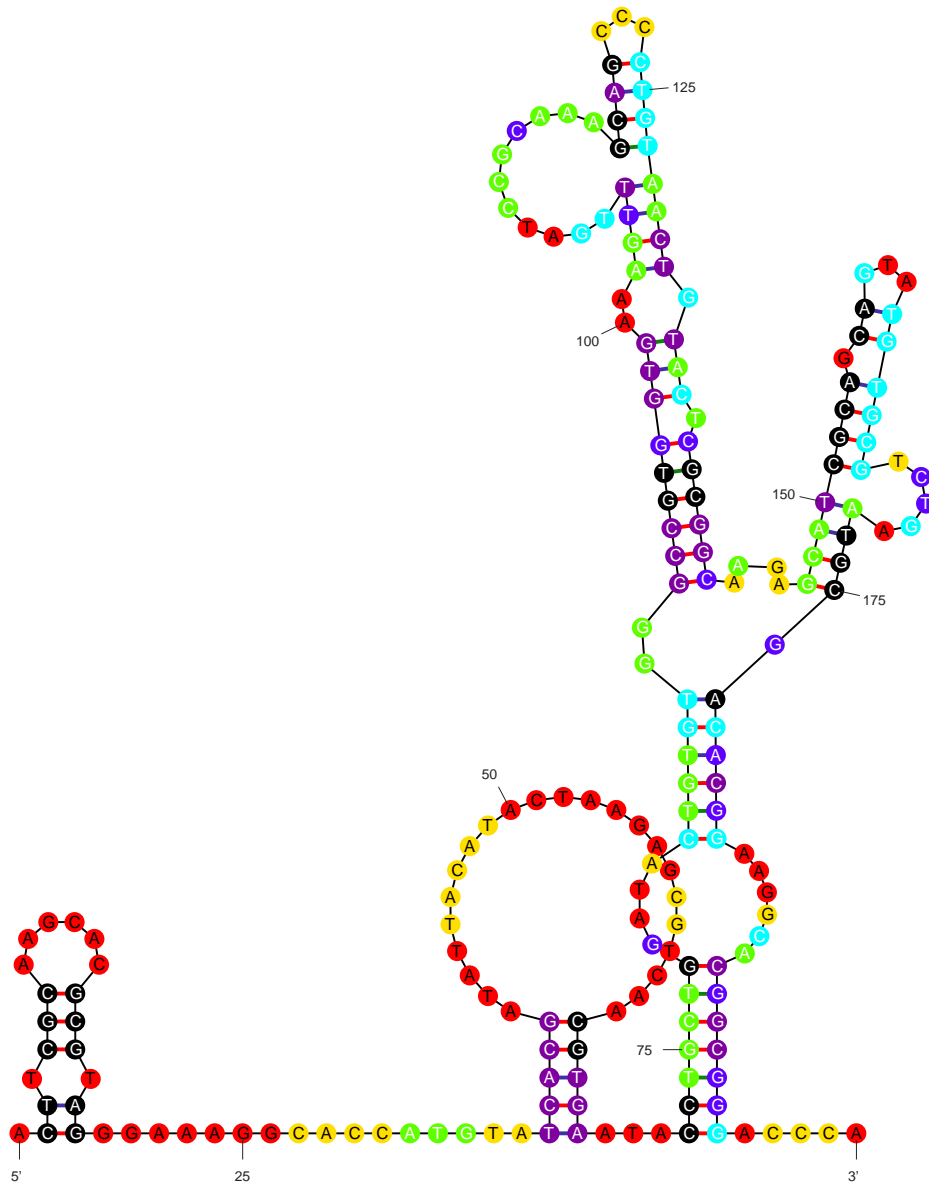
dG = -18.3 Random DNA

Figure 5: All folding plots are of the first predicted secondary structure and are drawn by sir\_graph. Here, the folding is displayed using default parameters, except that base pairs are displayed as line segments instead of by dots.



dG = -18.3 Random DNA

Figure 6: The folding is displayed using the simple option, which affects the layout of the multi-branch loops. Note the crowding of some single stranded bases in the exterior loop.

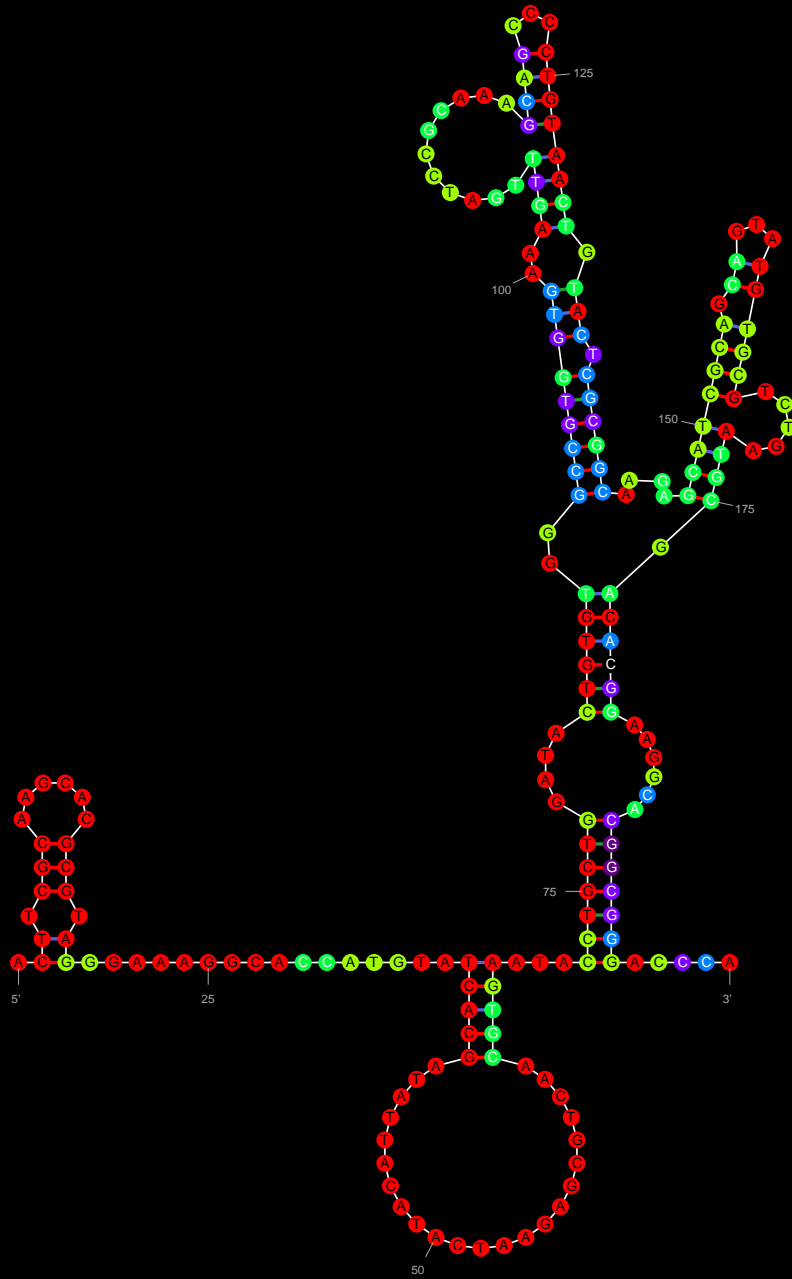


dG = -18.3 Random DNA

Figure 7:

Two new features are shown here. The is a “default” plot, but the exterior loop has been chosen to be “Flat”. Bases are labeled with the “Both” option using ss-count numbers. Note the undesirable overlap of two loops.





dG = -18.3 Random DNA

Figure 8: This is similar to Figure 7, expect that the background is black, p-num annotation is used, and the “Flat\_Alt” option has be chosen for the exterior loop. Note how this eliminates the loop overlap.

## Thermodynamics of Folding

---

$$\Delta G = \Delta H - T \Delta S$$

- $\Delta G^0 = -18.3$  kcal/mole at  $37^\circ$  C
  - $\Delta H^0 = -347.3$  kcal/mole
  - $\Delta S^0 = -1060.8$  cal/ ( $^\circ$ K·mol)
  - $T_m = 54.3^\circ$  C assuming a 2 state model!
- 

- linear DNA folding.
  - Ionic conditions:  $[\text{Na}^+] = 1.0$  M,  $[\text{Mg}^{++}] = 0.0$  M.
  - Standard errors are roughly  $\pm 5\%$ ,  $\pm 10\%$ ,  $\pm 11\%$  and  $2-4^\circ$  C for free energy, enthalpy, entropy and  $T_m$ , respectively.
- 

Figure 9: This Figure illustrates the presentation of thermodynamic details for the first folding. Because of the length of the output page, only the top part has been shown.

Structure 1  
 Random DNA  
 $dG = -18.30$   $dH = -347.3$   $dS = -1060.8$   $T_m = 54.3$

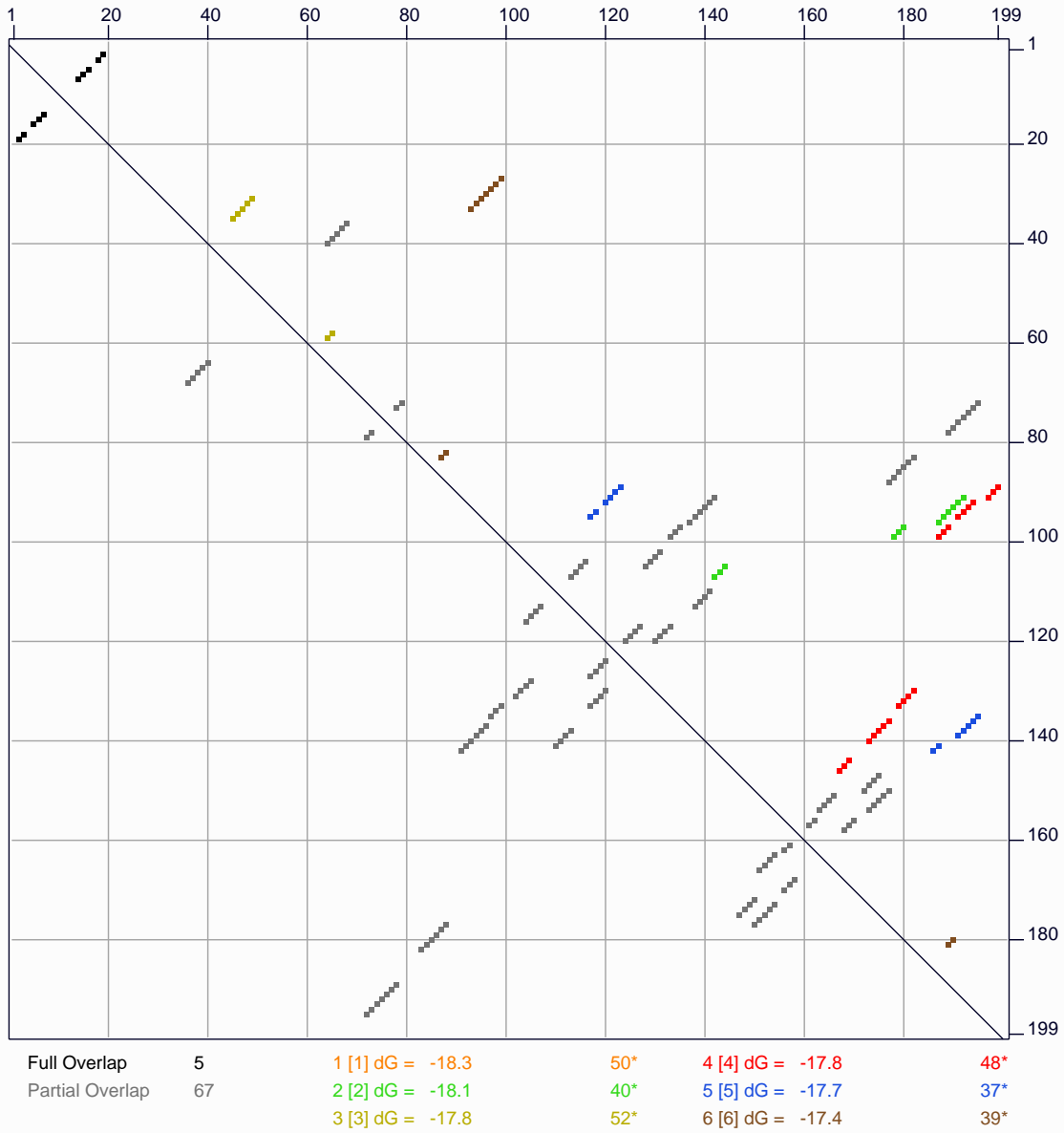
Structural element	$\Delta\Delta G$	Information
External loop	-3.85	25 ss bases & 3 closing helices.
Stack	-1.84	External closing pair is C <sup>72</sup> -G <sup>195</sup>
Stack	-0.32	External closing pair is C <sup>73</sup> -G <sup>194</sup>
Stack	-0.47	External closing pair is T <sup>74</sup> -G <sup>193</sup>
Stack	-2.24	External closing pair is G <sup>75</sup> -C <sup>192</sup>
Stack	-0.32	External closing pair is C <sup>76</sup> -G <sup>191</sup>
Stack	-0.47	External closing pair is T <sup>77</sup> -G <sup>190</sup>
<b>Helix</b>	-5.66	7 base pairs.
Interior loop	3.10	External closing pair is G <sup>78</sup> -C <sup>189</sup>
Stack	-0.32	External closing pair is C <sup>83</sup> -G <sup>182</sup>
Stack	-0.47	External closing pair is T <sup>84</sup> -G <sup>181</sup>
.....	.....	.....
Interior loop	1.13	External closing pair is T <sup>3</sup> -A <sup>18</sup>
Stack	-2.17	External closing pair is C <sup>5</sup> -G <sup>16</sup>
Stack	-2.24	External closing pair is G <sup>6</sup> -C <sup>15</sup>
<b>Helix</b>	-4.41	3 base pairs.
Hairpin loop	3.20	Closing pair is C <sup>7</sup> -G <sup>14</sup>

---

*mfold WWW server*  
 1996-2003, Michael Zuker, Rensselaer Polytechnic Institute

Figure 10: The bottom portion of the thermodynamic details page for the first folding.

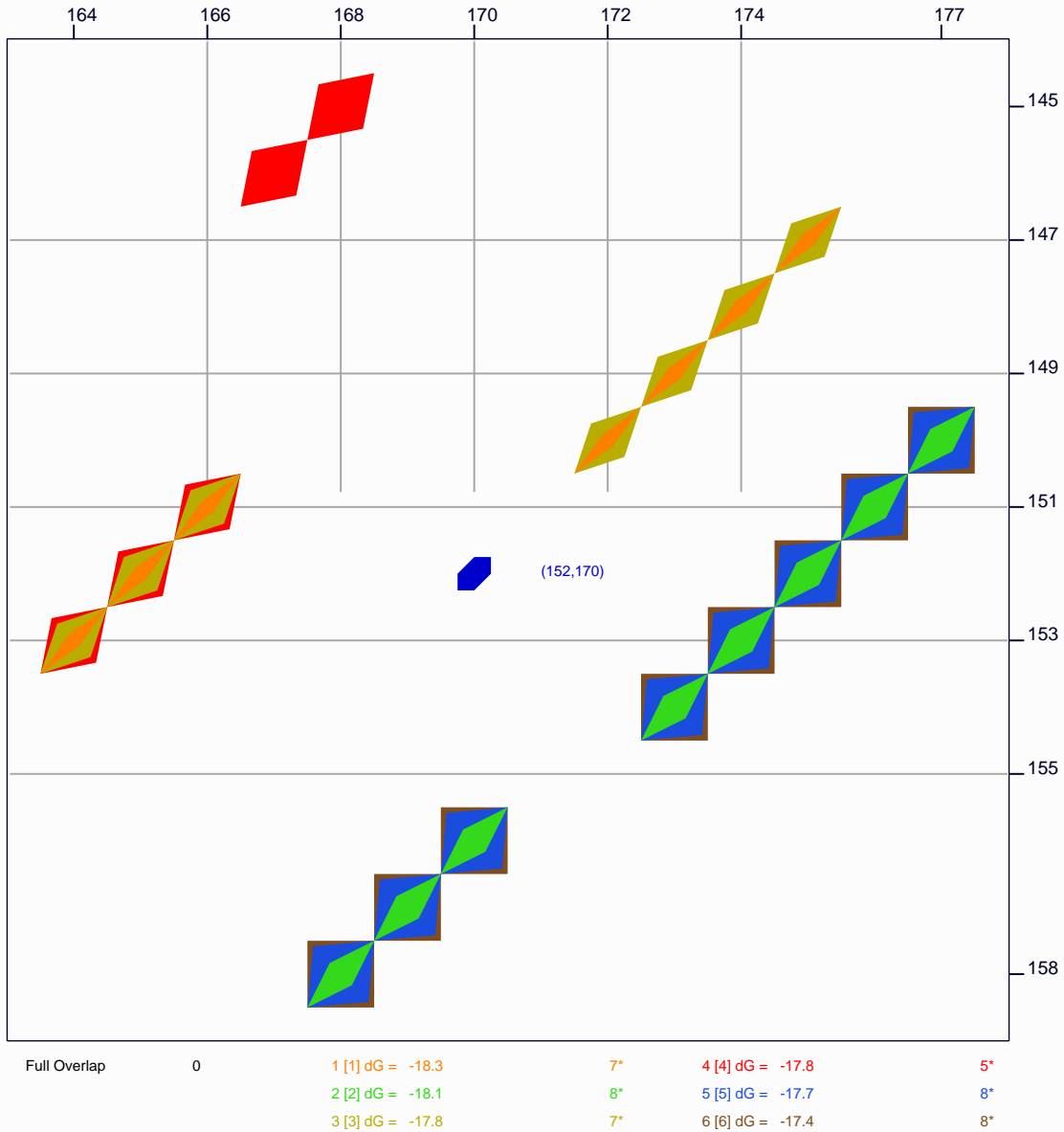
## Structure dot plot for RAN 130299481



\*Counts for each structure include overlap dots.

Figure 11: This is the default image of the structure dot plot that shows all 6 foldings superimposed in a single dot plot. The gray dots indicate base pairs that occur in some, but not all foldings.

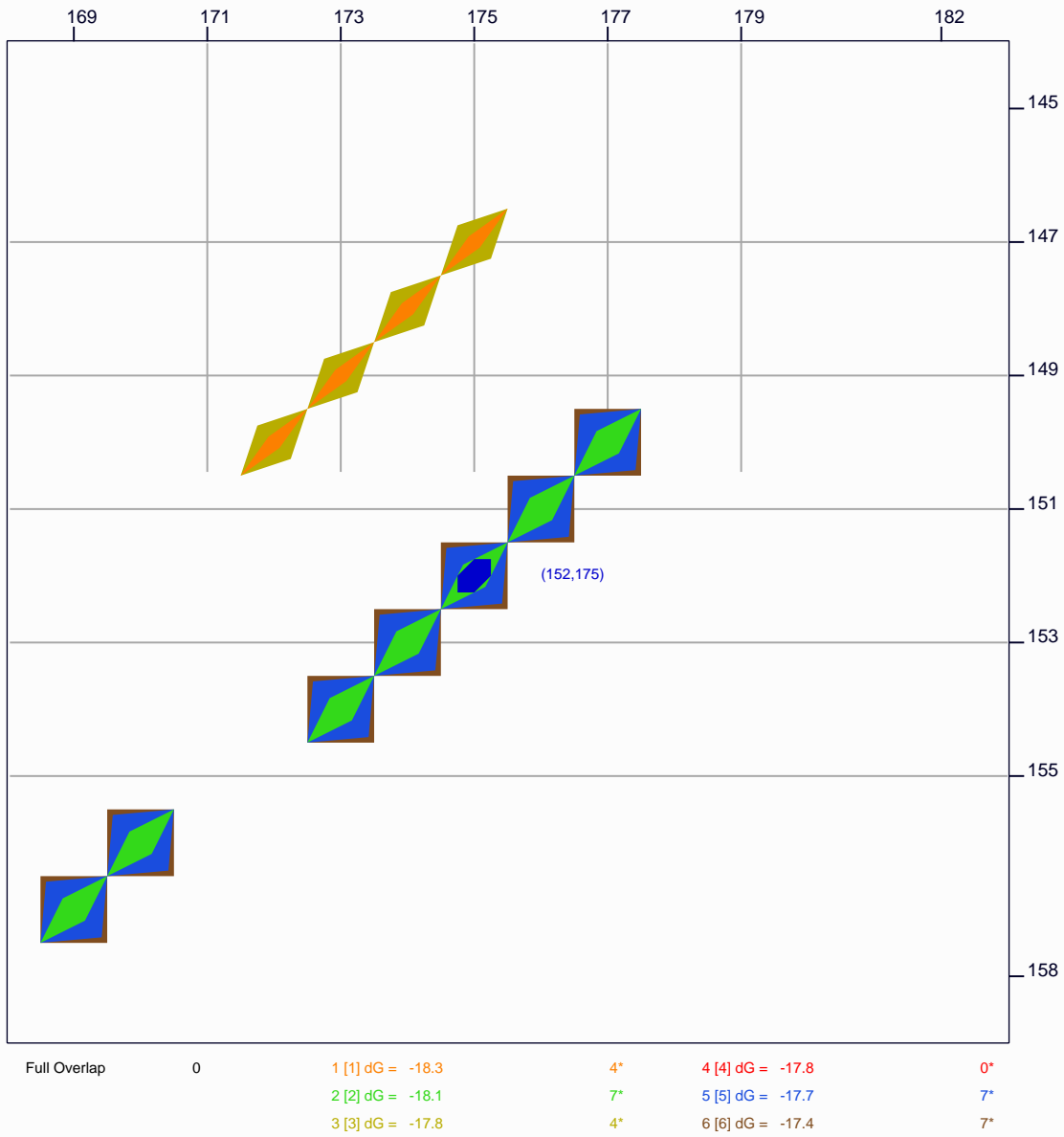
### Structure dot plot for Random DNA



\*Counts for each structure include overlap dots.

Figure 12: The default structure dot plot has been magnified by a factor of 20 about (152,170), shown in blue. The multicolor option has been chosen. By eye, one can see that an orange-yellow base pair is in foldings 1 and 3. A yellow-orange-red base pair is in foldings 1,3 and 4. The green-blue-brown base pairs are in foldings 2, 5 and 6. The chosen base pair was not in a folding because it is too difficult to hit on a base pair at the resolution in the default dot plot for this case.

### Structure dot plot for Random DNA



\*Counts for each structure include overlap dots.

Figure 13: The dot plot in Figure 12 has been redrawn with the same magnification, but the 152,175 base pair has been deliberately chosen. This can easily be done as high resolution. Note the blue “chosen dot” on top of the base pair.

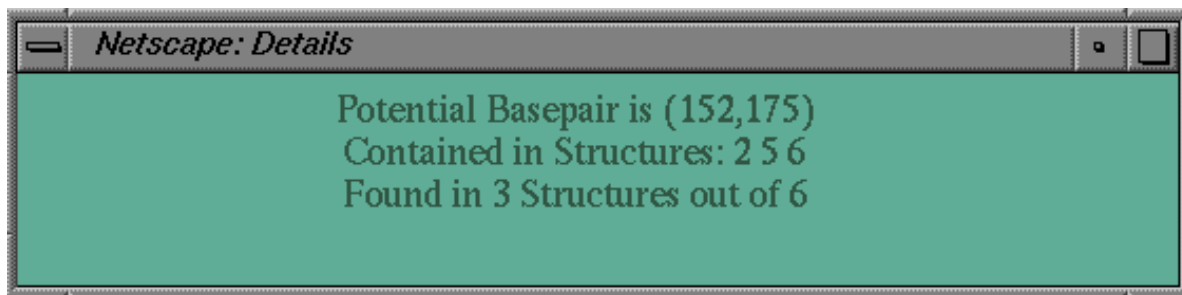


Figure 14: Choosing the (152,175) base pair in Figure 13 causes the Details window to display a message that this base pair occurs in structures 2, 5 and 6, or 3 out of 6. In this case, we could determine this fact visually. It becomes harder when there are more foldings, especially when there are more than 15 foldings and unique colors can no longer be used.