

## Tutorial for PinMol Mac version (contact us by email if you have questions–[icatrina@gmail.com](mailto:icatrina@gmail.com)):

### 1. Generate input file using mfold:

<http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form>

Paste/upload your transcript sequence (e.g. *nanos* mRNA) in FASTA format, and run an immediate (< 800 bases) OR batch job (800-9,000 bases).

2. **Input file:** Once the *mfold* job is completed, save the ss-count file as an **ASCII/ANSI** text file using the *Notepad*, *Atom* or *MS Word* software. When clicking the “ss-count” link (**red arrow**), a new window opens with the ss-count information.

Select the whole text (Win: CTRL+A; Mac: Command+A), copy/paste it into a text editor, and then save it (e.g. “nos\_sscount.txt”).

\*If you use *MS Word*, after the text is pasted, use: *File>Save As>Simple Text>Save>Other encoding>US ASCII*

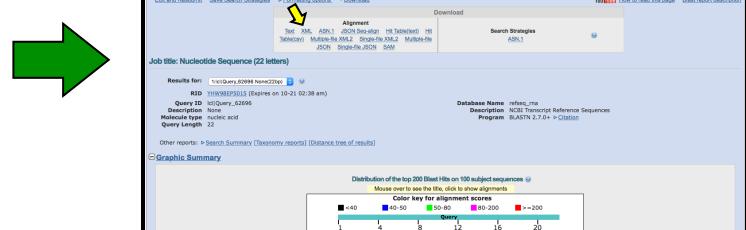
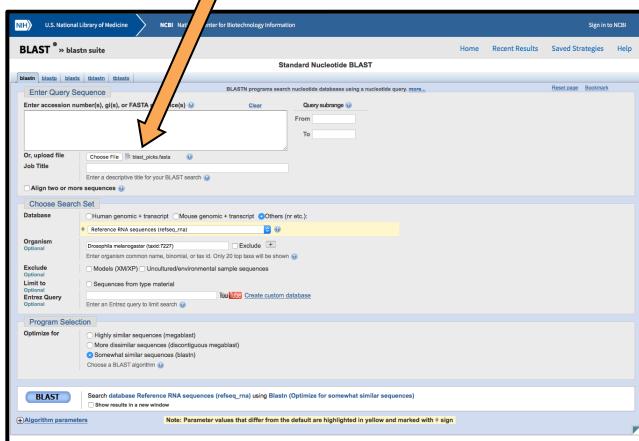
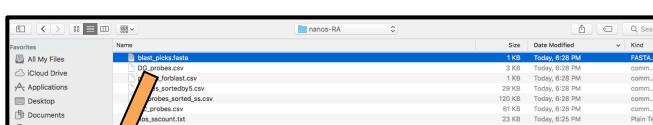
**3. Run the *PinMol* software from a terminal window and follow the instructions. Each step is described below.**

- A. Run software from command prompt.
  - B. Input the file path for the ss-count file saved at step 2

**Note:** Previous files will be overwritten so it is advised to create a new folder if you wish to keep those files.

- C. **Input the length of probe, between 18 and 26 bases (e.g. "22").**
  - D. **Input target region.** If the full length target OR a limited target region is considered, enter the nucleotide number for the first nucleotide ("1" OR the number for the desired start nucleotide) and last nucleotide (maximum possible – for *nanos-RA* mRNA "2,349" OR the desired number for the end nucleotide).  
**Note:** *PinMol* provides the total number of nucleotides of the RNA target.
  - E. **Input desired number of probes** (e.g. "50" or maximum possible if smaller than 50, which is listed in the output – for this *nanos-RA* example is "441", which is the default value if a larger number than the maximum possible is entered).
  - F. Select both "run BLAST" and "run tBLASTn" (indicated by "1")

4. **Run blast analysis using probe sequences.** “blast\_picks.fasta” file will be created in the folder containing your ss-count file and can be directly inputted into blastn suite with appropriate organism chosen prior to BLAST (e.g. *Drosophila melanogaster*).



Download XML file (yellow arrow) and enter **path/file name** into *PinMol*. The beginning (**green box**) and end (**blue box**) of the output data is presented below.

```

Student -- bash -- 135x24
Enter a file name: /Users/Student/Desktop/PinMol/nanos-RA/nos_sscount.txt
Enter the length of probe; a number between 18 and 26: 22
If a specific region within the target is needed, please enter the number of start base, or 1: 1
and the number of end base or max number of bases 2349: 2349
Maximum number of possible probes is: 441

How many probes do you want to save? Enter the maximum number of probes if smaller than 50, or a number between 2 and 50: 50
Do you want to use blast alignment information to determine cross homology? y/n: y

Please use the file blast_picks.fasta to perform blast with refseq-rna database, and desired organism.
For targets other than mRNAs make sure you use the Nucleotide collection (nr/nt) instead!
Enter path and file name for saved blast XML file: /Users/Student/Downloads/XR25BUF301R-Alignment.xml
1 MB sequence at base number 1806 is: CGACGUCUCCAUUCAUCAACUUUCGGAUUCGUUC

2 MB sequence at base number 1169 is: GCACGGUUGUACCGCUUUGUACCCUUGGUGC

3 MB sequence at base number 1170 is: GCACGUGUUGUACCGCUUUGUACCCUUCGUUC

4 MB sequence at base number 1805 is: CGACGUCCAUUCAUCAACUUUCGGAUUCGUUC

47 MB sequence at base number 1157 is: GCACGGUACACCUUUGUUGUUGUUGUACGUUC

48 MB sequence at base number 1160 is: CGAGCUUGUACACCUUUGUUGUUGUUGUUGUUC

49 MB sequence at base number 359 is: CGAGGGUGAAAAGCAGAAAAGUUAUCCCCUG

50 MB sequence at base number 165 is: CGACGUUUAUUCACUGAAGAUUUCGCCGUUC

Results for "/Users/Student/Desktop/PinMol/nanos-RA/nos_sscount.txt" using 22 as probe length,
for 50 probes, and blast choice = y, and for a target region between 1 and 2349 nucleotides:
1. Total number of possible probes = 2328
2. Number of probes that have a GC content between 30 and 56 = 1175
3. Number of probes that meet GC and energetic criteria = 441
4. Number of probes that have an ss-count fraction larger than 0.5 = 431

This information can be also be found in the file Final_molecular_beacons.csv

Check the structure for the selected probes using your favorite browser by opening the corresponding SVG files!

```

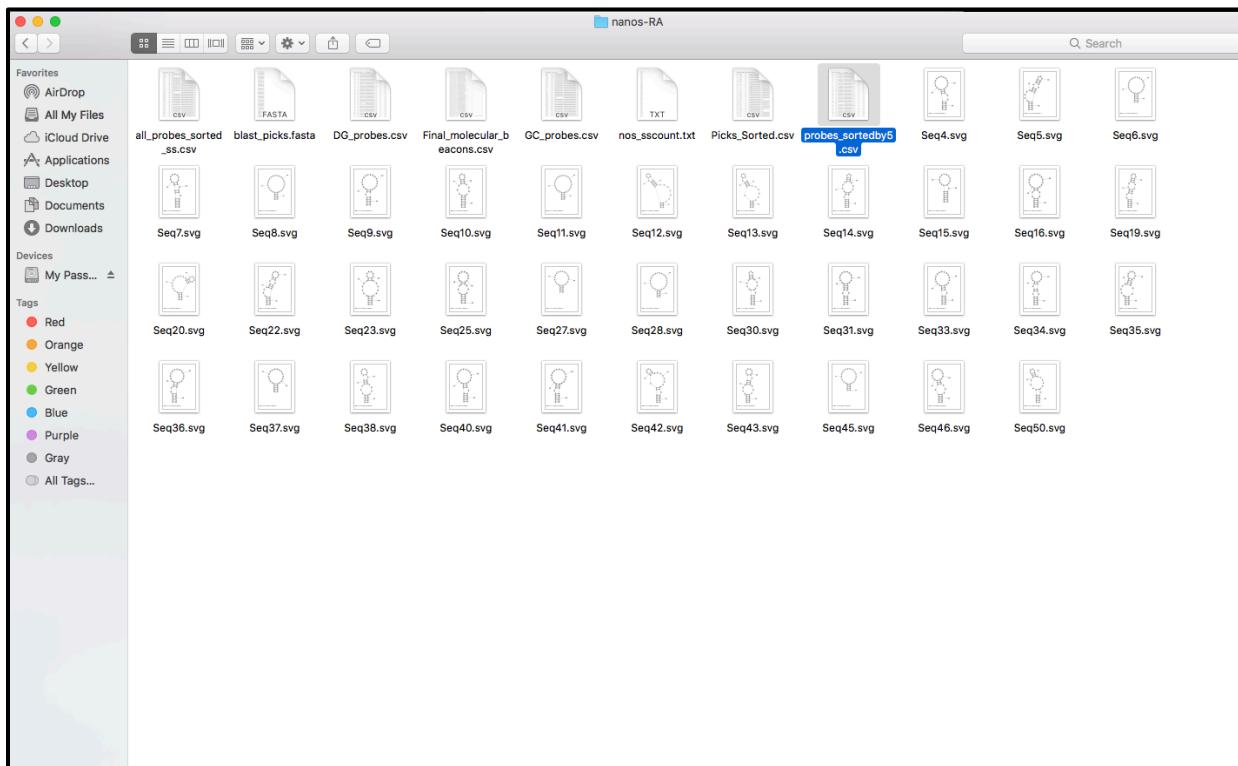
- 5. Analyze results.** The output data is saved in the folder containing the input file along with the following files, which are listed in alphabetical order. We used “probe” to refer to the region of the molecular beacon that is complementary to the target region.
- “**all\_probes\_sorted\_ss.csv**” – all possible probe sequences sorted in descending order by ss-count fraction – “1” for fully single stranded and “0” for fully double stranded, as predicted in the input file.
  - “**blast\_picks.fasta**” – probe sequences that should be used for BLAST analysis.
  - “**DG\_probes.csv**” – probes that meet both the GC (between 31 and 55%) and the energetic criteria,  $\Delta G_{\text{unimol}} > -2.5 \text{ kcal/mol}$  and  $\Delta G_{\text{bimol}} > -7.5 \text{ kcal/mol}$  – see point f below and our publication for more information.
  - “**GC\_probes.csv**” – probes that meet the GC content criteria (between 31 and 55%, also see publication for more information).
  - “**Final\_molecular\_beacons.csv**” – the output data, as seen in the terminal window.
  - “**probes\_sortedby5.csv**” – probes sorted by five criteria: in descending order by ss-count fraction,  $\Delta G_{\text{unimol}}$ ,  $\Delta G_{\text{bimol}}$  and probe’s GC percentage, and in ascending order by  $\Delta G_{\text{duplex}}$ . Where the ss-count fraction is described above in point a;  $\Delta G_{\text{unimol}}$  is the predicted free energy of uni- or intra-molecular folding of the probe sequence;  $\Delta G_{\text{bimol}}$  is the predicted free energy of bi- or inter-molecular folding of the probe sequence;  $\Delta G_{\text{duplex}}$  is the predicted free energy of hybridization of the probe sequence with a complementary RNA sequence (red highlighted portion of sequence indicates probe).

**g. “Picks\_Sorted.csv”** – probes sorted by increasing number of positive hits within genome where ‘positives’ are defined as number of nucleotide similarities between probe sequence and off-target transcripts (e.g. 16 nucleotides). Positive hits are based on the scoring matrix used by BLAST (red highlighted portion of sequence indicates probe sequence).

**h. “Seqi.svg”** – the drawing of the structure of each of the final molecular beacons.

Pick	Base Number	Positives	Probe Sequence	is-count fraction	
1	1806	16	UUCAUUCACAUACUUUGGAA	0.9845455	
2	1169	16	GUUGUAGCCUCUUCUUCUUC	0.5609091	
3	1169	16	GUUGUAGCCUCUUCUUCUUC	0.5609091	
4	1169	16	GUUGUAGCCUCUUCUUCUUC	0.5609091	
5	1805	16	UCAACAUACACCUUUCGAU	0.5845455	
6	1793	16	CUUUCGUAGUAGUAUAGAUU	0.54727273	
7	1793	16	CUUUCGUAGUAGUAUAGAUU	0.54727273	
8	1801	16	UCAACAUACACCUUUCGAU	0.53272727	
9	70	17	AAAAAAAUAGUAGUAGUAG	0.59272727	
10	11	17	UAGUAGUAGUAGUAGUAG	0.59272727	
11	1135	17	UAGUAGUAGUAGUAGUAG	0.5845455	
12	14	17	AAAAAAAUAGUAGUAGUAG	0.5845455	
13	16	18	UAGUAGUAGUAGUAGUAG	0.5845455	
14	187	17	UAGUAGUAGUAGUAGUAG	0.5845455	
15	1162	17	CGGUUCUACCUUOOUUUGU	0.56181818	
16	1164	17	AAGCUUUCUACCUUOOUUUGU	0.56181818	
17	1164	17	AAGCUUUCUACCUUOOUUUGU	0.56181818	
18	1168	17	GUUAUGGGGCAACUAAAAGGG	0.5518182	
19	32	17	UAAUGGCCACUACAAAGGGC	0.53272727	
20	32	17	UAAUGGCCACUACAAAGGGC	0.53272727	
21	38	17	GUUAUGGGGCAACUAAAAGGG	0.5518182	
22	38	17	GUUAUGGGGCAACUAAAAGGG	0.5518182	
23	1786	17	AAAAAAUAGUAGUAGUAGAC	0.5463636	
24	1216	17	UAGUAGUAGUAGUAGUAG	0.5463636	
25	1216	17	UAGUAGUAGUAGUAGUAG	0.5463636	
26	1215	17	UAGUAGUAGUAGUAGUAG	0.5463636	
27	42	17	CGGUUCUACCUUOOUUUGU	0.53272727	
28	42	17	CGGUUCUACCUUOOUUUGU	0.53272727	
29	48	17	CAGAGGCCCCAAAAAAUCUAG	0.53272727	
30	1223	17	GUUGUAGUAGUAGUAGUAG	0.5463636	
31	1161	17	GUUGUAGUAGUAGUAGUAG	0.5463636	
32	380	17	GUUGUAGUAGUAGUAGUAG	0.5463636	
33	380	17	GUUGUAGUAGUAGUAGUAG	0.5463636	
34	245	17	CGGUUCUACCUUOOUUUGU	0.53272727	
35	28	17	GGGGUAAAAGCAGAGAAGAAU	0.5518182	
36	31	17	GGGGUAAAAGCAGAGAAGAAU	0.5518182	
37	80	17	GGGGUAAAAGCAGAGAAGAAU	0.5518182	
38	40	17	GGGGUAAAAGCAGAGAAGAAU	0.5518182	
39	42	17	AUUCAUUCACUACUUUGGA	0.54	
40	42	17	ACCGAGAGGCAAAAGAACUUC	0.54	
41	39	17	UUCGUUUCUACUACUUUGGA	0.54727273	
42	1222	17	UUCGUUUCUACUACUUUGGA	0.54727273	
43	1220	17	UUCGUUUCUACUACUUUGGA	0.54727273	
44	50	17	GAGCGAGGCCCCAAAAAAAUUC	0.53272727	
45	6	1159	20	UUGUACACCUUOOUUUGUUGU	0.64145455
46	245	17	UUCGUUUCUACUACUUUGGA	0.54727273	
47	2	1155	21	ACACCUUOOUUUGUUGUUAU	0.67936364
48	3	1156	21	UACACCUCUUSGUUSGUUGUA	0.66183636
49	4	1157	21	GUACCUUSGUUSGUUGUA	0.65272727

**Note:** some molecular beacons were discarded (e.g. 1 through 3), because they were highly structured or do not fold into a hairpin shape.



- 6. Visualize molecular beacon results using a browser.** Open “Seqi.svg” files to view drawing of the final molecular beacons (red highlighted probe sequence and red circle indicate probe region in the final molecular beacon predicted secondary structure). Each “Seqi.svg” file corresponds to the “i” sequence listed for all sorted probes (see point 5g above).

	A B MB sequence at base number 70 is:	H	I	J	K	L	M	N	O	P	Q	R	S
1	MB sequence at base number 1306 is: CGACCUUCAUCAUCAUUUUGGAGUCGU												
2	MB sequence at base number 1189 is: GCACGUUUAACCGUUCUACACGUUCGU												
3	MB sequence at base number 1188 is: CGACGUUUAACCGUUCUACACGUUCGU												
4	MB sequence at base number 1305 is: CGACGUUCAUCAUCAUCCGUACGU												
5	MB sequence at base number 1793 is: GCACGUUUCGGGAUAGUAUAGAUUCCGU												
6	MB sequence at base number 1304 is: CGACGUUUCGGGAUAGUAUAGAUUCCGU												
7	MB sequence at base number 1301 is: CGGUCAUCAACJUUCGGAUAGAUUCCGU												
8	MB sequence at base number 1302 is: CGGUCAUCAACJUUCGGAUAGAUUCCGU												
9	MB sequence at base number 72 is: CGACGUUAAAAGUAGAUUCCGU												
10	MB sequence at base number 1135 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
11	MB sequence at base number 1303 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
12	MB sequence at base number 1187 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
13	MB sequence at base number 1188 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
14	MB sequence at base number 1189 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
15	MB sequence at base number 1184 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
16	MB sequence at base number 1185 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
17	MB sequence at base number 431 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
18	MB sequence at base number 430 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
19	MB sequence at base number 1186 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
20	MB sequence at base number 68 is: GCACGUUAAAAGUAGAUUCCGUAGGUCCG												
21	MB sequence at base number 1786 is: GCACGUUAAAAGUAGAUUCCGUAGGUCCG												
22	MB sequence at base number 1307 is: GCACGUUAAAAGUAGAUUCCGUAGGUCCG												
23	MB sequence at base number 1316 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
24	MB sequence at base number 1317 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
25	MB sequence at base number 77 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
26	MB sequence at base number 432 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
27	MB sequence at base number 1185 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
28	MB sequence at base number 1154 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
29	MB sequence at base number 1223 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
30	MB sequence at base number 1186 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
31	MB sequence at base number 980 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
32	MB sequence at base number 981 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
33	MB sequence at base number 982 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
34	MB sequence at base number 89 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
35	MB sequence at base number 1187 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
36	MB sequence at base number 1807 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
37	MB sequence at base number 81 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
38	MB sequence at base number 1188 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
39	MB sequence at base number 1222 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
40	MB sequence at base number 1189 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
41	MB sequence at base number 1221 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
42	MB sequence at base number 82 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
43	MB sequence at base number 1185 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
44	MB sequence at base number 164 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
45	MB sequence at base number 1186 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
46	MB sequence at base number 1156 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
47	MB sequence at base number 1157 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
48	MB sequence at base number 1187 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												
49	MB sequence at base number 399 is: CGACGUUAAAAGUAGAUUCCGUAGGUCCG												

