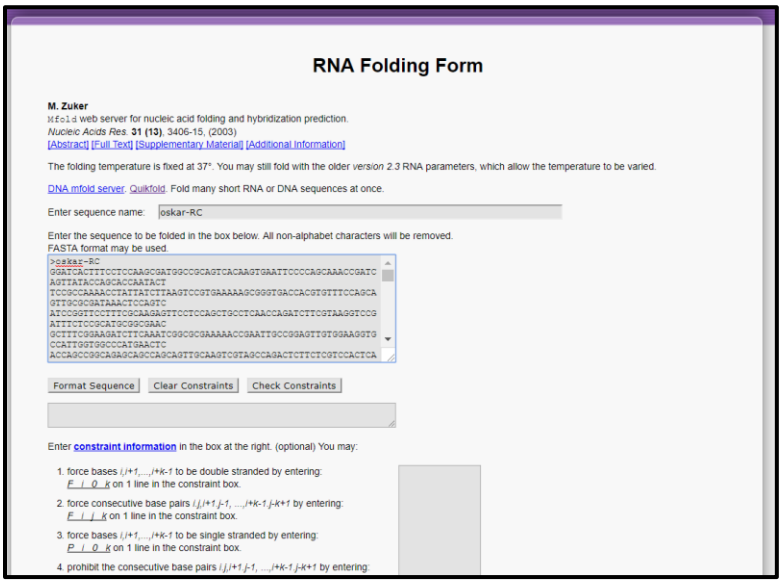
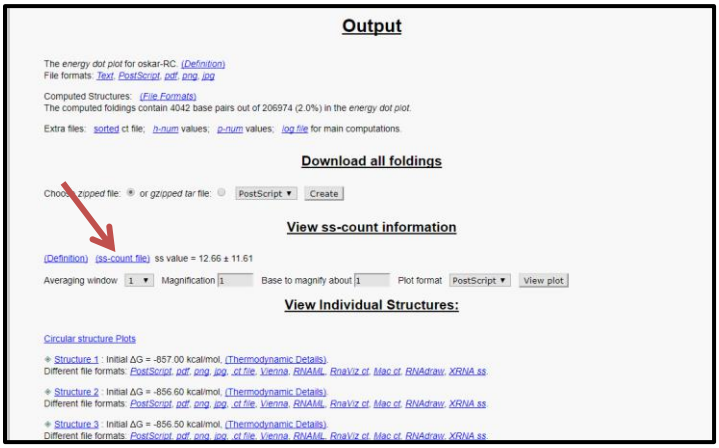
Tutorial for *PinMol* Mac version (contact us by email if you have questions– icatrina@gmail.com):

1. **Generate input file using mfold:**
   http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form
   Paste/upload your transcript sequence in FASTA format, and run an immediate (<800 bases) OR batch job (800-9,000 bases).



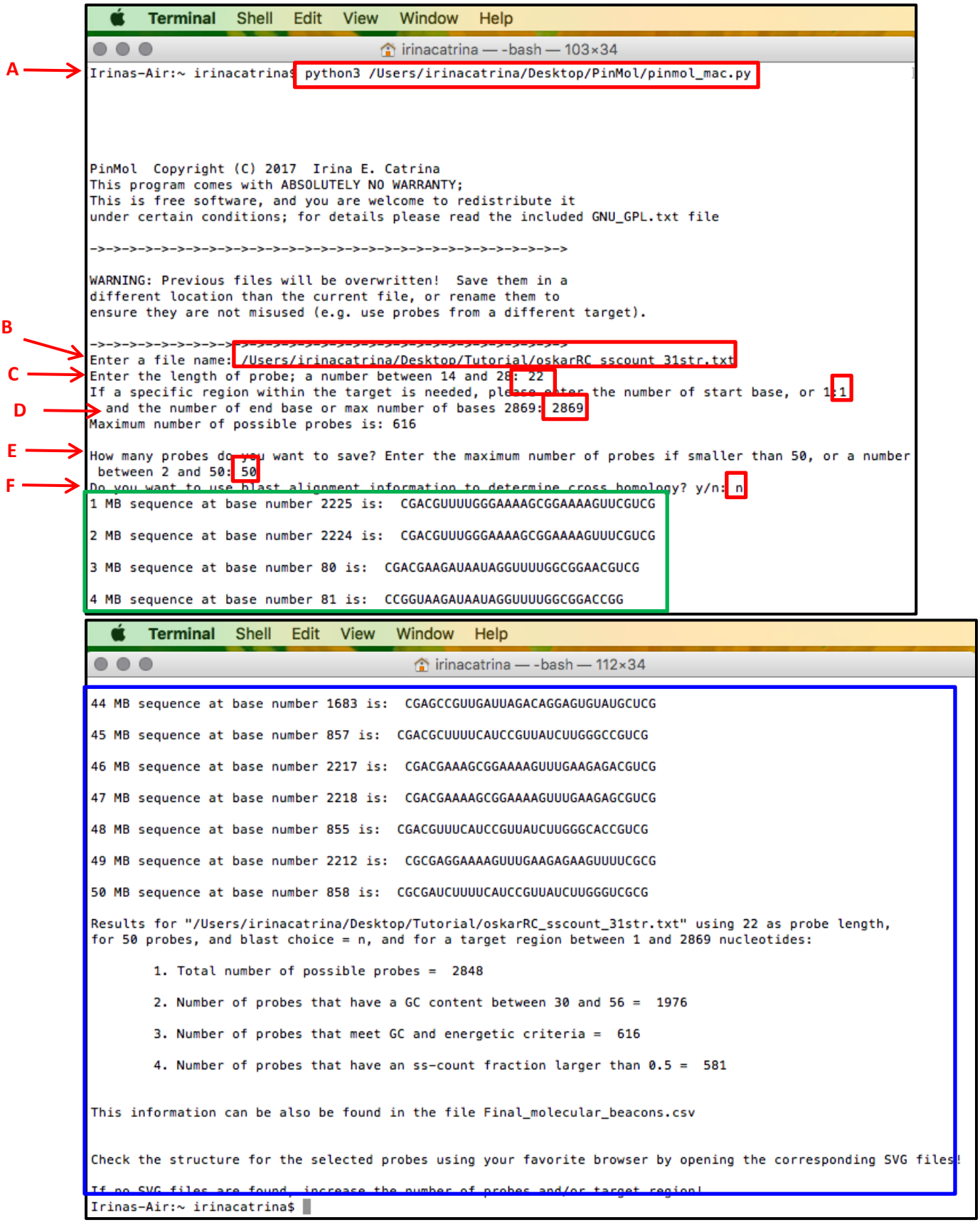
2. **Input file:** Once the *mfold* job is completed, save the ss-count file as an **ASCII/ANSI** text file using the *Notepad*, *Atom* or *MS Word* software. When clicking the "ss-count " link (red arrow), a new window opens with the ss-count information.
   Select the whole text (Win: CTRL+A; Mac: Command+A), copy/paste it into a text editor, and then save it (*e.g.* "oskarRC_sscount_31str.txt").



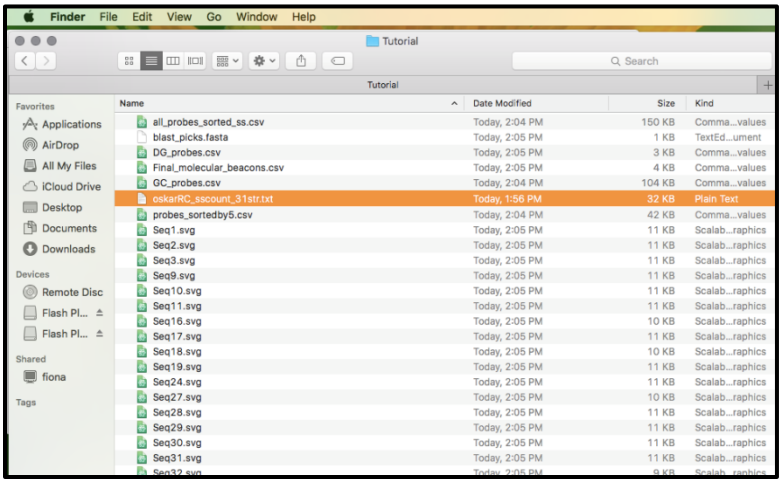
If you use *MS Word*, after the text is pasted, use:
File>Save As>Simple Text>Save>Other encoding>US ASCII

3. **Run the *PinMol* software from a terminal window and follow the instructions.** Each step is described below with the entered text/values highlighted with red boxes.
    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").**
    Note: this tutorial shows 14 and 28 as probe length limits, which were used to identify the optimum values of 18 and 26 (see our publication for more details).
    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 *oskar-RC* mRNA "2,869" 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 *oskar-RC* example is "616", which is the default value if a larger number than the maximum possible is entered).
    F. **Select whether a user-provided BLAST analysis should be taken into consideration (*e.g.* "n").**

    The beginning (green box) and end (blue box) of the output data is presented below.


Terminal Shell Edit View Window Help
irinacatrina — -bash — 103×34
A
Irinas-Air:~ irinacatrina$ python3 /Users/irinacatrina/Desktop/PinMol/pinmol_mac.py
PinMol Copyright (C) 2017 Irina E. Catrina
This program comes with ABSOLUTELY NO WARRANTY;
This is free software, and you are welcome to redistribute it
under certain conditions; for details please read the included GNU_GPL.txt file
WARNING: Previous files will be overwritten! Save them in a
different location than the current file, or rename them to
ensure they are not misused (e.g. use probes from a different target).
B
Enter a file name: /Users/irinacatrina/Desktop/Tutorial/oskarRC_sscount_31str.txt
C
Enter the length of probe; a number between 14 and 28: 22
If a specific region within the target is needed, please enter the number of start base, or 1: 1
D
and the number of end base or max number of bases 2869: 2869
Maximum number of possible probes is: 616
E
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
F
Do you want to use blast alignment information to determine cross homology? y/n: n
1 MB sequence at base number 2225 is: CGACGUUUUGGGAAAAGCGGAAAAGUUCGUCG
2 MB sequence at base number 2224 is: CGACGUUUGGGAAAAGCGGAAAAGUUUCGUCG
3 MB sequence at base number 80 is: CGACGAAGAUAAUAGGUUUUGGCGGAACGUCG
4 MB sequence at base number 81 is: CCGGUAAGAUAAUAGGUUUUGGCGGACCGG
Terminal Shell Edit View Window Help
irinacatrina — -bash — 112×34
44 MB sequence at base number 1683 is: CGAGCCGUUGAUUAGACAGGAGUGUAUGCUCG
45 MB sequence at base number 857 is: CGACGCUUUUCAUCCGUUAUCUUGGGCCGUCG
46 MB sequence at base number 2217 is: CGACGAAAGCGGAAAAGUUUGAAGAGACGUCG
47 MB sequence at base number 2218 is: CGACGAAAAGCGGAAAAGUUUGAAGAGCGUCG
48 MB sequence at base number 855 is: CGACGUUUCAUCCGUUAUCUUGGGCACCGUCG
49 MB sequence at base number 2212 is: CGCGAGGAAAAGUUUGAAGAGAAGUUUUCGCG
50 MB sequence at base number 858 is: CGCGAUCUUUUCAUCCGUUAUCUUGGGUCGCG
Results for "/Users/irinacatrina/Desktop/Tutorial/oskarRC_sscount_31str.txt" using 22 as probe length,
for 50 probes, and blast choice = n, and for a target region between 1 and 2869 nucleotides:
1. Total number of possible probes = 2848
2. Number of probes that have a GC content between 30 and 56 = 1976
3. Number of probes that meet GC and energetic criteria = 616
4. Number of probes that have an ss-count fraction larger than 0.5 = 581
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!
If no SVG files are found, increase the number of probes and/or target region!
Irinas-Air:~ irinacatrina$

4. **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 (the input file is highlighted in orange in the screenshot shown below). We used "probe" to refer to the region of the molecular beacon that is complementary to the target region.
    a. **"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.
    b. **"blast_picks.fasta"** – probe sequences that should be used for BLAST analysis (see tutorial file with BLAST analysis).
    c. **"DG_probes.csv"** – probes that meet both the GC (between 31 and 55%) and the energetic criteria, ΔGunimol > -2.5 kcal/mol and ΔGbimol > -7.5 kcal/mol – see point **f** below and our publication for more information.
    d. **"GC_probes.csv"** – probes that meet the GC content criteria (between 31 and 55%, also see publication for more information).
    e. **"Final_molecular_beacons.csv"** – the output data, as seen in the terminal window.
    f. **"probes_sortedby5.csv"** – probes sorted by five criteria: in descending order by ss-count fraction, ΔGunimol, ΔGbimol and probe's GC percentage, and in ascending order by ΔGduplex. Where the ss-count fraction is described above in point **a**; ΔGunimol is the predicted free energy of uni- or intra-molecular folding of the probe sequence; ΔGbimol is the predicted free energy of bi- or inter-molecular folding of the probe sequence; ΔGduplex is the predicted free energy of hybridization of the probe sequence with a complementary RNA sequence.
    g. **"Seqi.svg"** – the drawing of the structure of each of the final molecular beacons.
    Note: some molecular beacons were discarded (*e.g.* 4 through 8), because they were highly structured or do not fold into a hairpin shape.



5. **Visualize molecular beacon results using a browser.**
   Open “Seqi.svg” files to view drawing of the final molecular beacons (red circle indicates probe region). Each “Seqi.svg” file corresponds to the “i” sequence listed for all sorted probes (see point 4f above).