Dsembler User Manual

Aporva Gupta

July 12th, 2021

Contents

1. Launching Dsembler

Dsembler can be installed via GitHub. Users should have docker installed on their computer to run Dsembler locally. Docker can be downloaded from https://docs.docker.com/get-docker/. Ensure your BIOS settings are compatible with the docker application. Once docker is installed successfuly, run the following command in your terminal (Powershell for Windows).

```
$ ./docker_build.sh
```

A Dsembler image will be built locally using the Dockerfile present in the Dsembler source folder. Alternatively, pull the image from docker using the following command

```
$ docker pull image <>
```

To run the web application, run the following command in your terminal.

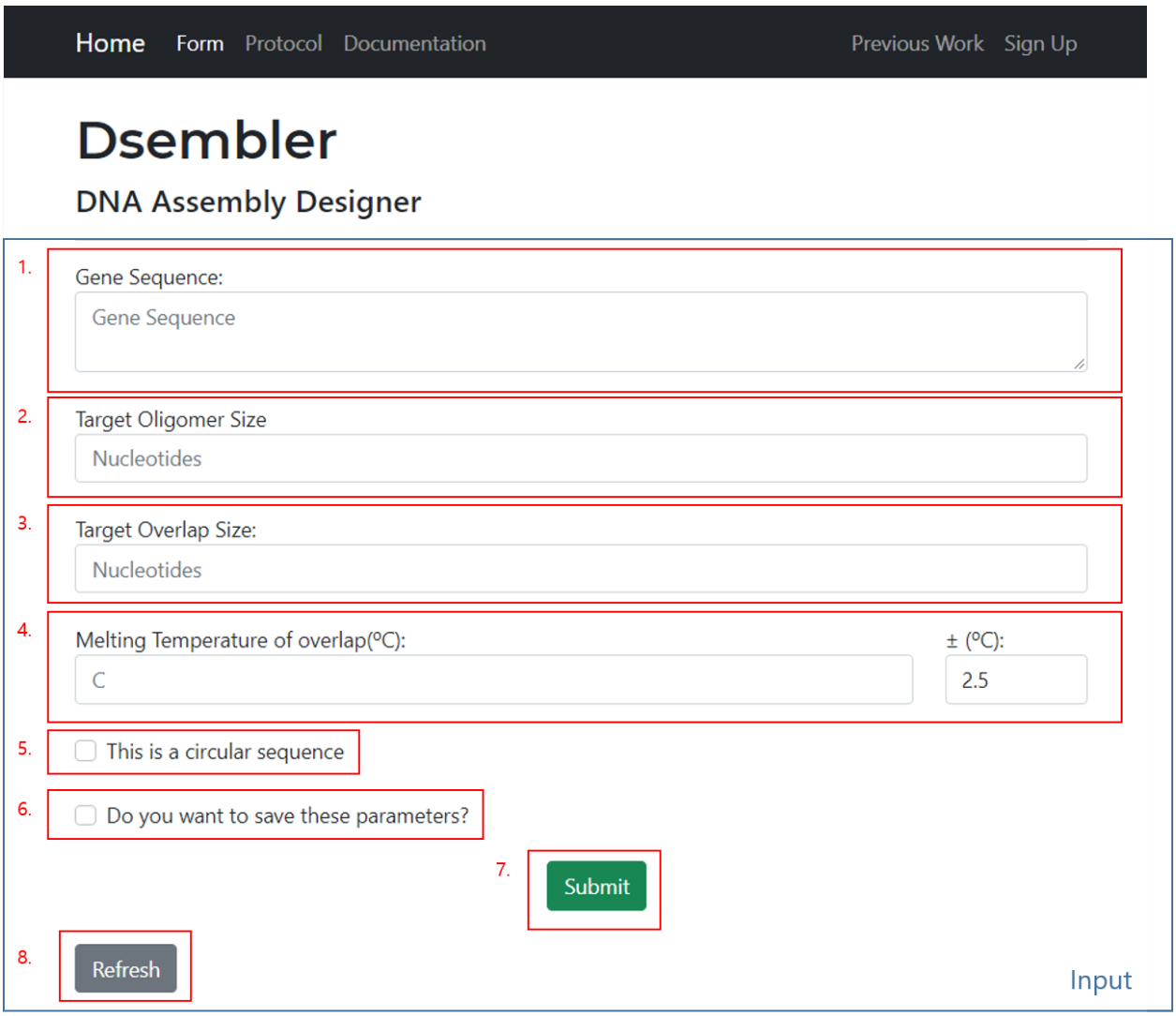
```
$ ./docker_run.sh
```

This will run the Dsembler application on your localhost via port 5000.

2. Online Interface

Dsembler has a simple, easy to use user interface.

Input



1. **Gene Sequence:** Users can input the gene sequence as plain text. The Application will not run if non-DNA elements are used (spaces and newlines are removed by the program automatically).
2. **Oligomer size:** Input the target oligomer size. The oligomers must be at least 50bp
3. **Overlap size:** Input the target overlap size. Overlaps must be at least 20bp and must be at least 30bp smaller than the target oligomer size.
4. **Melting Temperature & Temperature range:** The target melting temperature is

inputted in ^{o}C. An acceptable range is required from the user, i.e ± n^{o}C. The default and minimum range is melting temp ± 2.5^{o}C. The recommended range is between 52 - 58^{o}C. **Note:** the melting temperature is measured using the Melting Temperature of overlaps are calculated based on the Nearest Neighbour Equation and the Sugimoto (1996) thermodynamic table.

5. **Sequence Orientation:** Check the box if the Sequence you are using is circular. If left unchecked, the sequence is processed as a linear sequence
6. **Save Data:** If you'd like to save your parameters, check this box. An input form for your user ID is provided if checked. **Note:** you should have signed up to access previous records (See Sign up).
7. **Submit:** Click here to initiate oligomer generation.
8. **Refresh:** Click this to refresh the form.

Output

Two buttons appear on the same page once the target parameters are submitted: - Download the excel sheet output
- Download the FASTA file

Excel File

The data is filled into the excel as the following columns.

Cluster Number	Oligomer Number	Sequence (5' to 3')	Overlap length	Oligomer length	Melting Temperature of overlap	Overlap Score	Cause of errors	Repeat Sequences
Cluster 1	oligomer 1	ATGGCTTTGGCAAATGAAATCATTGGGTCATCTTTTCTTCACAAAACATCCTCAGTTTCTTCTCAGTTTCAGGGA	25	75	54.47	1	G	
Cluster 1	oligomer 2	CTTTTCGCAACTGGCACCACCCAAATAGGCCTGAAGAACTGTTTTCCCTGAAACTGAGAAGAAACTGAG	19	69	53.33	0.17	L	
Cluster 1	oligomer 3	TGGTGCCAGTTGCGAAAAGGCAAGTTGTAGCACAACTGAGGAAGGCAGTCAACTCCCCTGTCGCGGC	15	67	53.91	1	G	
Cluster 1	oligomer 4	CCTTGTACTTCACAGGTTTTTCATCAAGGGGGACTACCTTTACAAGATCCTCAGAAATAGCCGCGACAGGGGAG	26	74	53.99	1	G	
Cluster 1	oligomer 5	TGATGAAAAACCTGTGAAGTACAAGGTTCGGGCTGTGGTAACCATCCGTAATAAAAACAAAGAAGATATAAAAGAGAC	27	78	47.06	6.44	L	
Cluster 1	oligomer 6	CAACATTTTGACCAATCCTATCTGTCAAAGCATCCAAATGTTTCACTATAGTCTCTTTTATATCTTCTTTGTTTTTA	28	77	53.58	0		
Cluster 1	oligomer 7	TTGACAGATAGGATTGGTCAAAATGTTGTGTTACAACTGATATCTACAGAAATAGATCCTAAGACCAACGCACC	22	74	53.69	0		
Cluster 2	oligomer 1	CTTGAGATTAGTTTTTTTTGACCAATCTTTAAGTACAGCTTCGTTGGATTTCTTTGGTGCGTTGGTCTTAGGATCTA	29	77	51.67	1.83	L	
Cluster 2	oligomer 2	AAGATTGGTCAAAAAAAACTAATCTCAAGGCTGAGCGTGTGAACTATATAGCTGAGTTCCTTTTAACTTCTGACTTTG	28	78	55.21	0		
Cluster 2	oligomer 3	AGAAAAATTCTTGTTGGTGTTTGTTTGTAATTGTGATTGCGCCGGGTTCACCAAAGTCAGAAGTTAAAAGGAACTCAGC	29	79	53.81	0		
Cluster 2	oligomer 4	TACAAACAAACACCAACAAGAATTTTTCTTGGAAACTATTACCATTGAACAATTTGCAAATGATCCAATTCACTTTCC	26	78	53.63	0		
Cluster 2	oligomer 5	AATTCTCTTAGCTGGATGATCCTTTCTGGACTGGACCCATGAATTGCAAGGAAAGTGAATTGGATCATTTGCAAA	27	75	52.89	0.61	L	
Cluster 2	oligomer 6	AGAAAGGATCATCCAGCTAAGAGAATTTTCTTTAGTAACAAGCCTTATTTGCCAGGAGAAACACCCGC	18	68	55.38	1	G	
Cluster 2	oligomer 7	CCTTTCCCATCTCCTCTAATGTCCTTCAACTCTATCTCCCGCAATTTCTTAATGCCCGCGGGTGTTTCTCCTGGC	25	75	54.77	1	G	
Cluster 2	oligomer 8	AGGACATTAGAGGAGATGGGAAAGGTGAGAGAAAACTCTCAGATCGTGTGTATGACTTTGACGTTTACAACGATCTC	27	77	54.08	0		
Cluster 2	oligomer 9	TCCTCCCAATCTAGGTCTAGCATATTCTATTCCCTTGTCTGGGTTTCCGAGATCGTTGTAAACGTCAAAGTCATA	25	75	53.64	0		
Cluster 2	oligomer 10	ATATGCTAGACCTAGATTGGGAGGAGAGAAAATTCCGTATCCTCGTAGGTGTAGGACTGGTCGCGCTC	18	68	54.33	1	G	
Cluster 2	oligomer 11	CATTGGCAGTGGCTTCTCAACTCTTGACTCTGCAGTCATGTCAGTGTCACTTGGAGCGCGACCAGTCCTAC	21	71	53.28	0.22	L	
Cluster 3	oligomer 1	GTTGAGAAGCCACTGCCAATGTATGTTCCTAGAGATGAGCAATTTGAAGAATCCAAACAAACAACTTTCTCACTCG	25	76	54.53	0		
Cluster 3	oligomer 2	AATTGATGCTTTTAAACTAGGAATAAGATTGTGCAAGACTGCTTTCAGCCTCCCGAGTGAGAAAGTTGTTTGTTTGGA	29	78	50.04	3.46	L	
Cluster 3	oligomer 3	ATCTTATTCCTAGTTTAAAAGCATCAATTCTGAGTAATAAGCATGATTTCCATGGCTTTTCTGATATTGACAGTCTTTA	29	79	54.37	0		
Cluster 3	oligomer 4	AGTAACTCATCTTGGAGGCCTAGTTTAAGTAGCACCCCTTCAGAATAAAGACTGTCAATATCAGAAAAGCCATG	24	74	54.84	0		
Cluster 3	oligomer 5	ACTAGGCCTCCAAGATGAGTTACTAAAAAAACTTCCATTACCACGGGTTGTCTCAGAATCATCACAAGGACTTCTAAG	28	78	54.55	0		
Cluster 3	oligomer 6	AGCCAAGCAAATTTATCCTTGCTTAAAATCTTGGGAGTATTGTATCTTAGAAGTCCTTGTGATGATTCTGAGA	23	73	52.37	1.13	L	
Cluster 3	oligomer 7	AGCAAGGATAAATTTGCTTGGCTTAGGGATGATGAGTTTGCGCGACAGGCTATAGCAGGTGTAAATCCAGTGAATATTG	29	79	53.59	0		
Cluster 3	oligomer 8	TCCGGATCTAAATTCGACACCGGAGGGAAAACTTTTAGACGTTCAATATTCACTGGATTTACACCTGCTATA	22	72	54.49	0		
Cluster 3	oligomer 9	CGGTGTCGAATTTAGATCCGGATGTCTACGGTCCTTTAGAATCGAGTCTTAAGGAAGAACATATTTTGGGACAGATAAA	27	79	52.03	1.47	L	
Cluster 3	oligomer 10	ACAATAAAGAGCTTGTTCTCATCGAGTGCCTGTTGAACTGTCATCCCATTTATCTGTCCCAAAATATGTTCTTCC	27	75	54.54	0		

Figure 1: Excel Output File Example

Cluster Number	Oligomer Number	Oligomer Sequence (5' to 3')	Overlap Length	Overlap Melting Temperature	Overlap Score	Score Faults	Repeats Sequences

Score of each oligomer's overlap:

- **H**: Overlap melting temperature is higher than the specified target temperature *(score = overlap_temp - temp_range_high)*
- **L**: Overlap melting temperature is higher than the specified target temperature *(score = temp_range_low - overlap_temp)*
- **T**: The overlap of each oligomer is not the smallest/largest within the possible oligomer range, and there is 'T' at the 3' end *(score = 1)*
- **G**: The last 5 bp of 3' end have 3 consecutive 'G', 'C', or a combination of both. *(score = 1)*
- **R**: There are repeats (10bp) within oligomer overlaps within each cluster *(score = 10)*

Overlap score is 0 if there are no possible areas of errors.

FASTA file

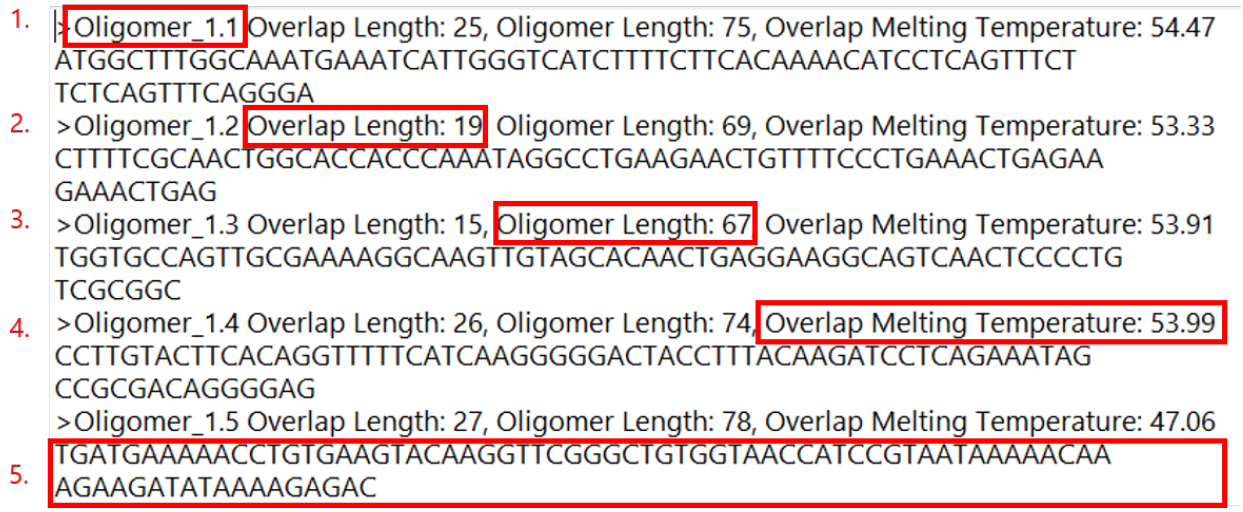


Figure 2: FASTA file output (Same example as Excel file; viewed on Notepad)

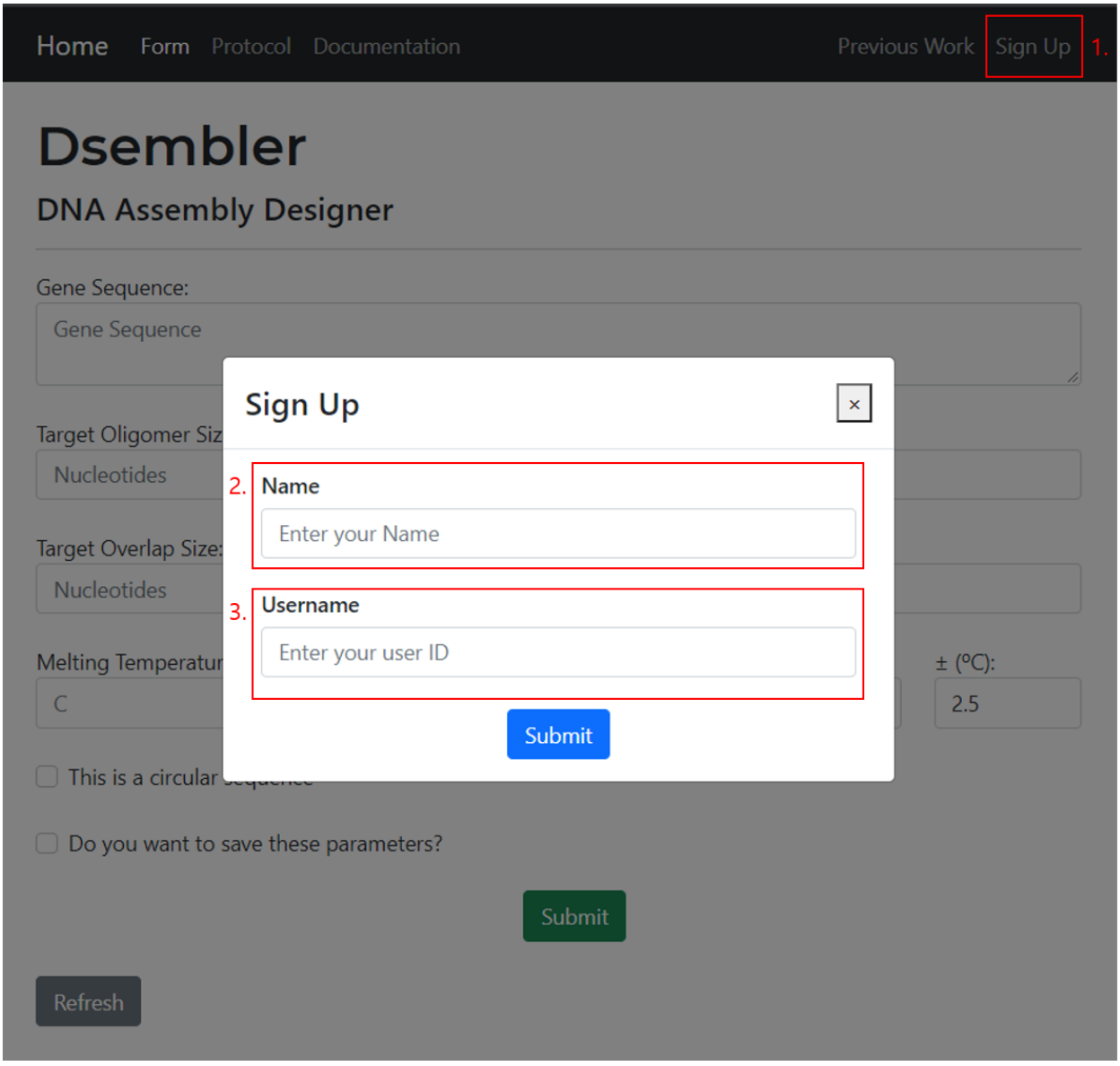
The oligomer FASTA file can be used for easy visualization of all oligomers. It can allow for amendments on other DNA visualising software, such as SnapGene. Similar to the Excel file (Fig.2), the Cluster and Oligomer number (1), the overlap length (2), The Oligomer length (3), the Overlap Melting Temperature(4), and the Oligomer sequence in a 5' to 3' orientation (5) is given for every oligomer.

Previous Work

A simple SQLite based storage is used to provide users to store and access any previous work.

Sign up

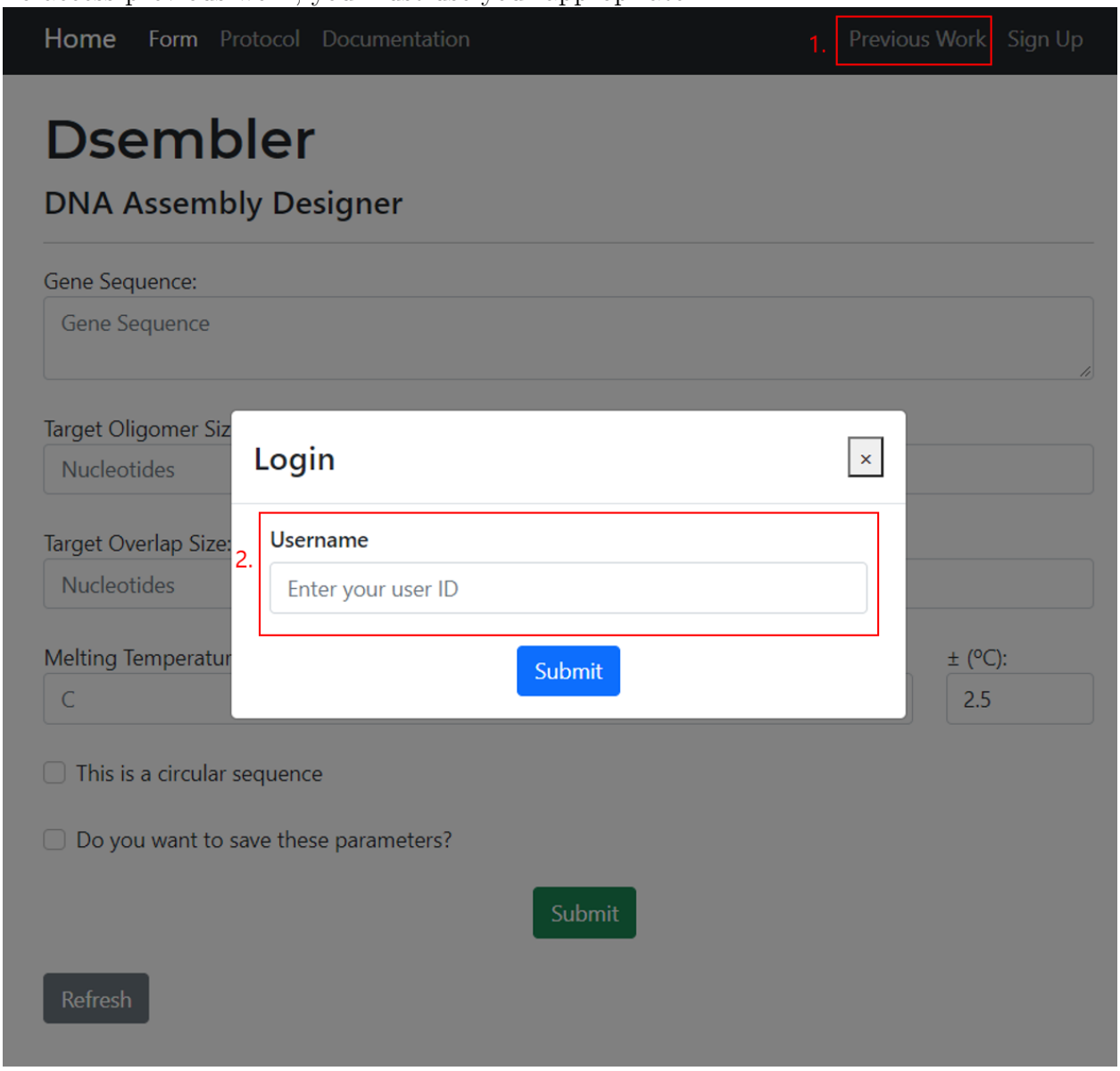
First, you must register into the system by signing up.



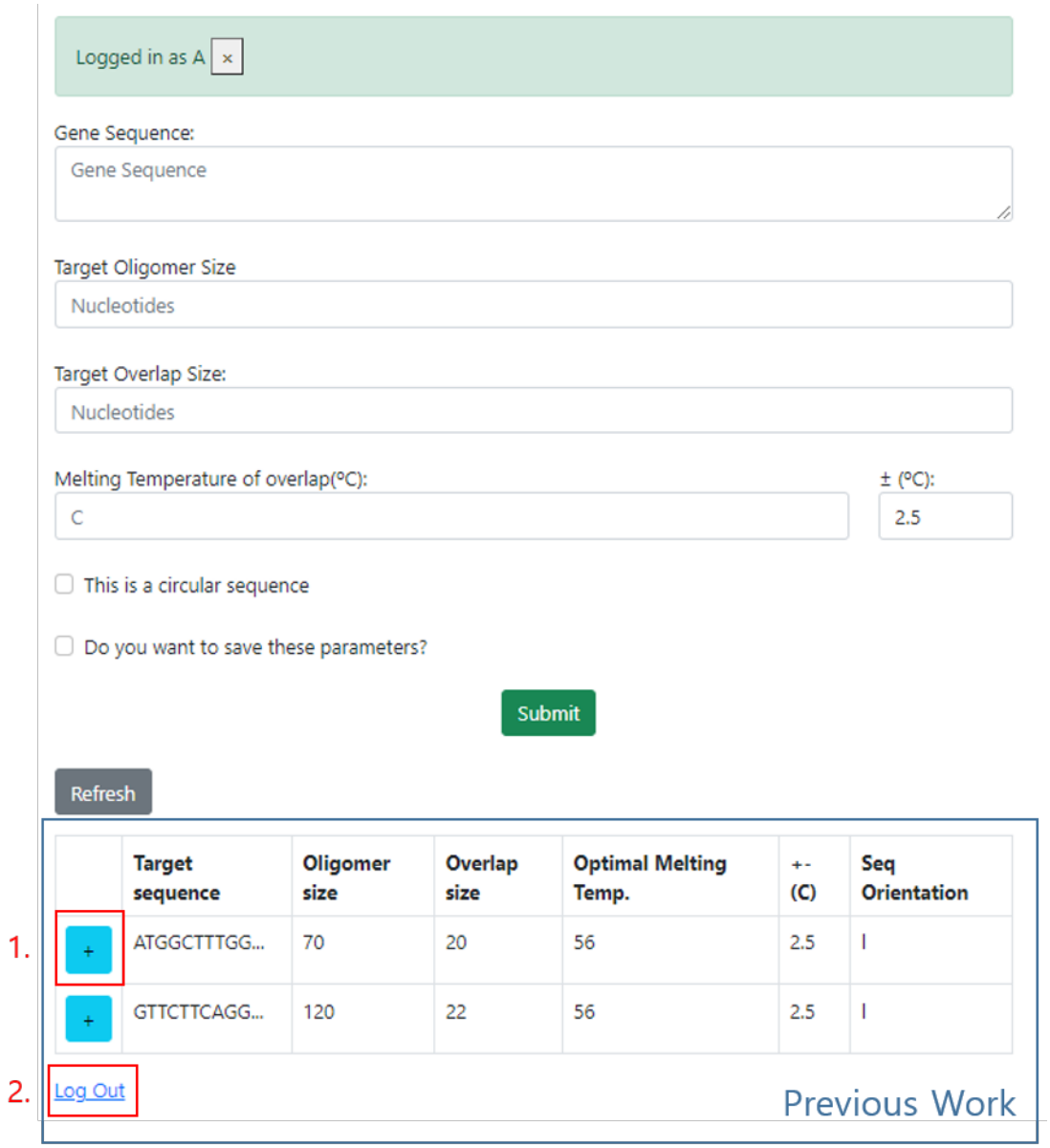
As the webpage does not cater to having password based logins, only the user's name and preferred username is registered. Only unique ID's are registered per container. Success and Failure messages are given appropriately.

Login

To access previous work, you must use your appropriate ID.



Once Logged in, you can find your previous queries, and can access the oligomers for the same.



The oligomers for a specific query can be accessed by clicking on the blue '+' button, which will generate the oligomers in a similar manner as shown in Output.

3. Contact

To report bugs and provide suggestions for the Dsembler User Interface and functionality, please contact the maintainer, Aporva Gupta (aporva03@kribb.re.kr)