

# Dsembler User Manual

Aporva Gupta

July 12th, 2021

## 1. Launching Dsembler

Dsembler can be installed via GitHub.

```
# obtain dsembler package from GitHub  
$ git clone https://github.com/sblabkribb/dsembler.git
```

### Docker (recommended)

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 successfully, run the following command in your terminal (or Powershell in Windows).

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

```
# build the docker image  
$ docker build -t dsembler:latest .  
OR  
# pull image from dockerhub  
$ docker pull sblabkribb/dsembler:latest
```

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

Run docker container: Bash

```
$ docker run -d --rm -v $(pwd):/app --publish 5000:5000 --name dsembler dsembler
```

Run docker container: Powershell

```
> docker run -d --rm -v %cd%:/app --publish 5000:5000 --name dsembler dsembler
```

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

**Python** Dsembler can be run from the user's python environment as well. The user is responsible for installing the required packages (see requirements.txt) to run the program.  
Bash

```
$ export FLASK_APP=app  
$ flask run
```

Powershell

```
> $env:FLASK_APP = "app"  
> flask run
```

This would publish the application on your localhost:5000 port

## 2. Online Interface

Dsembler has a simple, easy to use user interface.

### 2.1 Input

The screenshot displays the 'Dsembler DNA Assembly Designer' web interface. At the top is a dark navigation bar with links: 'Home', 'Form', 'Manual', 'Documentation', 'Previous Work', and 'Sign Up'. Below this, the title 'Dsembler DNA Assembly Designer' is prominently displayed. The main content area is titled 'Input' and contains several form fields, each highlighted with a red box and a number:

- 1.** A text input field labeled 'Gene Sequence:' with the placeholder text 'Gene Sequence'.
- 2.** A text input field labeled 'Target Oligomer Size' with the placeholder text 'Nucleotides'.
- 3.** A text input field labeled 'Target Overlap Size:' with the placeholder text 'Nucleotides'.
- 4.** A section for 'Melting Temperature of overlap(°C):' containing a text input field with 'C' and a small box for '± (°C):' with the value '2.5'.
- 5.** A checkbox labeled 'This is a circular sequence'.
- 6.** A checkbox labeled 'Do you want to save these parameters?'.
- 7.** A green 'Submit' button.
- 8.** A grey 'Refresh' button.

- 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.

- Melting Temperature & Temperature range:** The target melting temperature is inputted in °C. An acceptable range is required from the user, i.e  $\pm n^{\circ}\text{C}$ . The default and minimum range is melting temp  $\pm 2.5^{\circ}\text{C}$ . The recommended range is between 52 - 58°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.
- Sequence Orientation:** Check the box if the Sequence you are using is circular. If left unchecked, the sequence is processed as a linear sequence
- 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).
- Submit:** Click here to initiate oligomer generation.
- Refresh:** Click this to refresh the form.

## 2.2 Output

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

### 2.2.1 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	ATGGCTTTGGCAATGAATCATTTGGTTCATCTTTCTCACAAACATCTCAGTTCTCTCAGTTTCAGGA	25	75	54.47	1 G		
Cluster 1	oligomer 2	CTTTTCCCACTGGCACCACCAATAGGCCTGAAGAAGCTTTTCCCTGAACTGAGAAGAACTGAG	19	69	53.33	0.17 L		
Cluster 1	oligomer 3	TGTTGCCAGTTGGCAAGGCAAGTTGTAGCACAACTGAGGAAGGCACTCACTCCCTGTCGCGGC	15	67	53.91	1 G		
Cluster 1	oligomer 4	CCTTGACTTCACAGGTTTTTTCATCAAGGGGGACTACTTTTACAAGATCTCAGAAATAGCCGCGACAGGGGAG	26	74	53.99	1 G		
Cluster 1	oligomer 5	TGATGAAAAACCTGTGAAGTACAAGGTTCCGGCTGTGGTAACCATCCGTAATAAAAAACAAGAGATATAAAGAGAC	27	78	47.06	6.44 L		
Cluster 1	oligomer 6	CAACATTTGACCAATCTATCTGTCAAGCATCCAAATGTTTACATATAGTCTCTTTTATATCTTTGTTTAA	28	77	53.58	0		
Cluster 1	oligomer 7	TTGACAGATAGATTGGTCAAAATGTTGTGTACACTGATATCTACAGAAATAGATCTTAAGACCAAGCACC	22	74	53.69	0		
Cluster 1	oligomer 8	CTTGAGATTAGTTTTTTTGAACCAATCTTAAGTACAGCTCTGTTGGAATTTCTTTGGTCTGGTCTAGGACTA	29	77	51.67	1.83 L		
Cluster 2	oligomer 2	AAGATTGGTCAAAAAAAGTAACTCTCAAGGCTGAGCGTGTGAATATAGCTAGTCTCTTTAACTCTTGACTTTG	28	78	55.21	0		
Cluster 2	oligomer 3	AGAAAAATCTTGTGTGTTGTTGTAATGTGATTGCGCGGTTCCAAAGTCAAGATGAGAATGAAAGCACTCAGC	29	79	53.81	0		
Cluster 2	oligomer 4	TACAAAAACAACCAACAAGATTTTCTTGGAACATATTACCAATGAACAATTTGCAATGATCCCAATCTTTCC	26	78	53.63	0		
Cluster 2	oligomer 5	AATTCTCTAGCTGGATGATCTTCTGGACTGGACCATGAATTGCAAGGAAGTGAATGGATCATTGCAAA	27	75	52.89	0.61 L		
Cluster 2	oligomer 6	AGAAAGGATCTCCAGTAAAGAGAATTTCTTAGTAACAAGCCTTATTGCTCCAGGAGAAACACCCGC	18	68	55.38	1 G		
Cluster 2	oligomer 7	CTTTCCCATCTCTTAATGCTCTCACTATCTCTCCGAATTTCTTAATGCGCGGGTGTCTCTGGC	25	75	54.77	1 G		
Cluster 2	oligomer 8	AGGACATAGAGGAGTGGGAAGGAGAGAAATCTCAGATCGTGTGTATGATGATCTGACGTTTACACCACTCTC	27	77	54.08	0		
Cluster 2	oligomer 9	TCTCCCAATCTAGGTCTAGCATATTCTATCCCTGTCTGGGTTCCGAGATCGTTGTAACGCTCAAGTCATA	25	75	53.64	0		
Cluster 2	oligomer 10	ATATGTAGACCTAGATTGGGAGGAGAGAAATCCGTATCTCTGAGGTGAGGATCTGTCGCGCTC	18	68	54.33	1 G		
Cluster 2	oligomer 11	CATTGGCAGTGGCTCTCAACTCTTGACTCTGAGTGTCACTGAGGCGGACCGCTCTAC	21	71	53.28	0.22 L		
Cluster 3	oligomer 1	GTGAGAAGCCATGCCAATGATGTCTCTAGAGATGAGCAATTTGAAGATCCAAACAACATTTCTCACTCG	25	76	54.53	0		
Cluster 3	oligomer 2	AATTGATGCTTTTAACTAGGAATAAGATTGTGCAAGACTGCTTTCAGCTCCGAGTGAGAAAGTGTGTTTGA	29	78	50.04	3.46 L		
Cluster 3	oligomer 3	ATCTTATCTCAGTTTAAAGCATCAATCTGAGTAATAAGCATGATTTCATGGCTTTCTGATATTGACAGCTTTA	29	79	54.37	0		
Cluster 3	oligomer 4	AGTATCATCTTGTGAGGCTAGTTTAAAGTAAAGCCCTCAGATAGAAAGACTGTCTAATATCAGAAAGCATG	24	74	54.84	0		
Cluster 3	oligomer 5	ACTAGGCTCCAAGATGAGTTACTAAAAAATCTCCATTACACAGGTTGTCTCAGATCATCAAGGACTCTTAAG	28	78	54.55	0		
Cluster 3	oligomer 6	AGCCAAGCAAAATCTCTGCTTAAATCTGGGAGTATTGTATCTAGAAGCTCTTGTGATGATTCTGAGA	23	73	52.37	1.13 L		
Cluster 3	oligomer 7	AGCAAGGATAAATTTCTGGCTTAGGGATGAGTTTGGCGCAGGCTATAGCAGGTTAAATCCAGTGAATATTG	29	79	53.59	0		
Cluster 3	oligomer 8	TCCGATCTAAATTCGACACCGGAGGAGAACTTTAGACGTTCAATATTCAGTGGATTACACCTGCATA	22	72	54.49	0		
Cluster 3	oligomer 9	CGGTGTGAAATTTAGATCGGATGTCTACGGTCTTTAGAAATGAGTCTTAAGGAAGAACATATTTGGGACAGATAA	27	79	52.03	1.47 L		
Cluster 3	oligomer 10	ACAATAAGAGGCTGTGTTCTCATCGAGTGCTGTGAACGTGATCCCATTTATCTGTCCCAAAATATGTTCTTC	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.

### 2.2.2 FASTA file

```

1. >Oligomer_1.1 Overlap Length: 25, Oligomer Length: 75, Overlap Melting Temperature: 54.47
   ATGGCTTTGGCAAATGAAATCATTTGGGTCATCTTTCTTCACAAAACATCCTCAGTTTCT
   TCTCAGTTTCAGGGA
2. >Oligomer_1.2 Overlap Length: 19, Oligomer Length: 69, Overlap Melting Temperature: 53.33
   CTTTCGCAACTGGCACCACCCAAATAGGCCTGAAGAACTGTTTCCCTGAAACTGAGAA
   GAAACTGAG
3. >Oligomer_1.3 Overlap Length: 15, Oligomer Length: 67, Overlap Melting Temperature: 53.91
   TGGTGCCAGTTGCGAAAAGGCAAGTTGTAGCACAACCTGAGGAAGGCAGTCAACTCCCCTG
   TCGCGGC
4. >Oligomer_1.4 Overlap Length: 26, Oligomer Length: 74, Overlap Melting Temperature: 53.99
   CCTTGTAATTCACAGGTTTTTCATCAAGGGGGACTACCTTTACAAGATCCTCAGAAATAG
   CCGCGACAGGGGAG
   >Oligomer_1.5 Overlap Length: 27, Oligomer Length: 78, Overlap Melting Temperature: 47.06
   TGATGAAAAACCTGTGAAGTACAAGGTTTCGGGCTGTGGTAACCATCCGTAATAAAAAACAA
   AGAAGATATAAAAGAGAC

```

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. The first sequence is the user's inputted sequence with the parameters as the sequence description. 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.

## 2.3 Previous Work

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

### 2.3.1 Sign up

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

The screenshot displays the Dsembler DNA Assembly Designer web application. The main interface is dark grey with a navigation bar at the top containing links: Home, Form, Manual, Documentation, Previous Work, and a highlighted Sign Up button (labeled 1.). The main heading is "Dsembler DNA Assembly Designer". Below this, there are input fields for "Gene Sequence:", "Target Oligomer Size:", "Target Overlap Size:", and "Melting Temperature:". The "Sign Up" modal is open, showing two input fields: "Name" (labeled 2.) and "Username" (labeled 3.), both with placeholder text "Enter your Name" and "Enter your user ID" respectively. A blue "Submit" button is at the bottom of the modal. The background form has a green "Submit" button and a "Refresh" button.

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.

### 2.3.2 Login

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

The screenshot displays the Dsembler DNA Assembly Designer web application. At the top, a dark navigation bar contains links for Home, Form, Manual, and Documentation. On the right side of this bar, a red box highlights the 'Previous Work' link, preceded by a red number '1.'. Below the navigation bar, the main header area features the title 'Dsembler' and the subtitle 'DNA Assembly Designer'. The main content area contains several input fields for configuring a DNA assembly: 'Gene Sequence' (with a placeholder 'Gene Sequence'), 'Target Oligomer Size' (with a placeholder 'Nucleotides'), 'Target Overlap Size' (with a placeholder 'Nucleotides'), and 'Melting Temperature' (with a placeholder 'C'). There are also checkboxes for 'This is a circular' and 'Do you want to save these parameters?'. A large green 'Submit' button is positioned at the bottom center of the main form. A modal window titled 'Login' is overlaid on the form. Inside the modal, a red box highlights the 'Username' section, which includes a text input field with the placeholder 'Enter your user ID' and a blue 'Submit' button. A red number '2.' is placed to the left of the 'Username' label. The modal also has a close button (an 'x' in a square) in the top right corner. At the bottom left of the main form, there is a 'Refresh' button.

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

