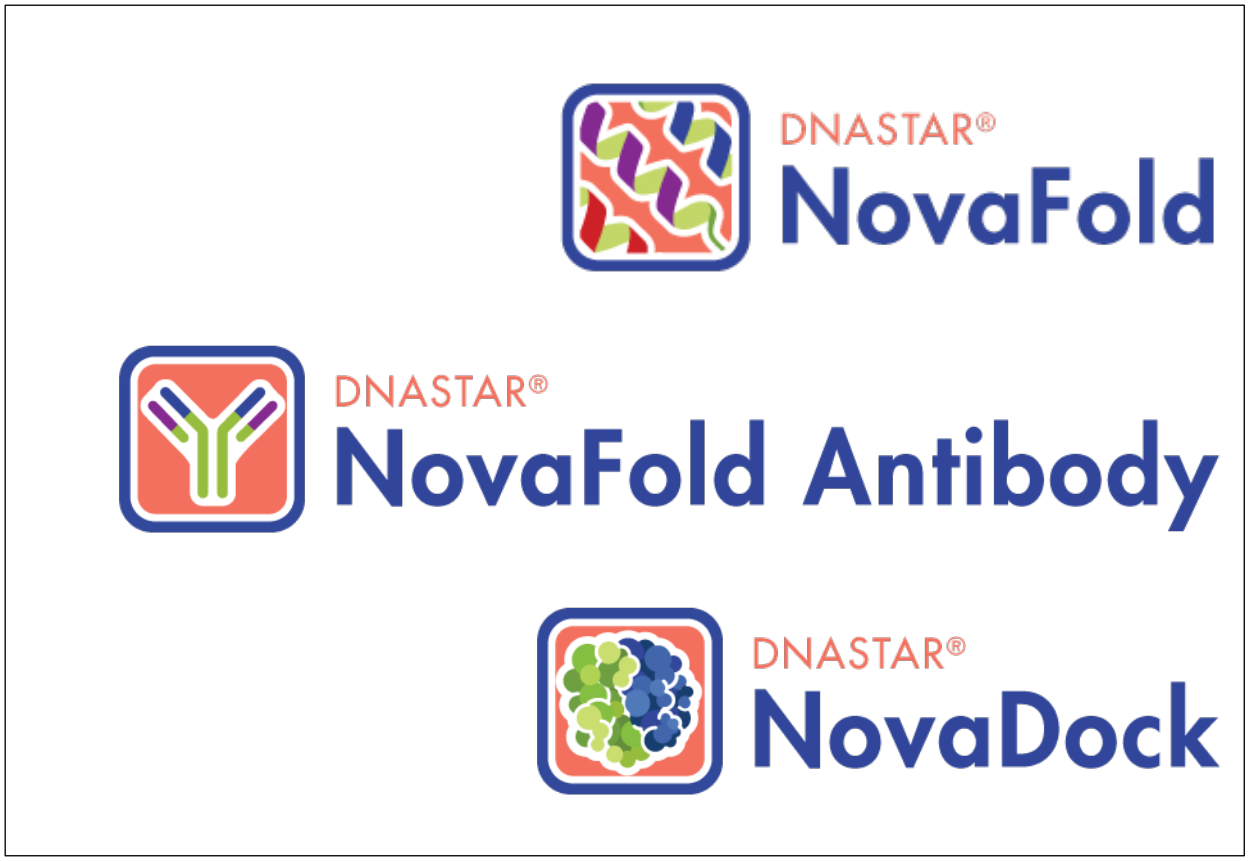
Installing and Running Local Versions of:



Version 15

DNASTAR, Inc. 2018

Contents

Before You Begin

We're here to help! If you have any difficulties with or questions about this application, please contact a DNASTAR support representative:

- E-mail: support@dnastar.com
- Phone (Madison, WI, USA): 608-258-7420
- In the USA and Canada, call toll free: 1-866-511-5090
- In the UK, call free on: 0-808-234-1643
- In Germany, call free on: 0-800-182-4747

This help document pertains to local versions of NovaFold 15, NovaFold Antibody 15 and NovaDock 15, and was last updated on October 23, 2018.

Overview

DNASTAR offers three protein folding, modeling and docking applications.

- **NovaFold** – Protein structure prediction
- **NovaFold Antibody** – Antibody structure prediction
- **NovaDock** – Protein-protein docking

This help document discusses how to install and run the local versions of these applications through the terminal of a Linux computer. Since all three applications are installed at once, this guide will refer to the group of applications as the "Nova applications."

Note: Each of the three applications can also be run on the cloud from within DNASTAR's Protean 3D. For more information on the "NovaCloud" versions of the applications, consult the Protean 3D help.

Once installed, each Nova application is licensed separately. Note that you must have a NovaFold license if you wish to license NovaFold Antibody. After installing and licensing the product(s) and running a prediction, results are opened and viewed in DNASTAR's Protean 3D application.

NovaFold

NovaFold, DNASTAR's 3D protein structure prediction program, can use a sequence file to build a 3D protein structure model based on iterative assembly simulations, and to predict ligand binding sites and protein function. NovaFold runs locally using the terminal of a Linux computer.

NovaFold uses the I-TASSER algorithm developed by Professor Yang Zhang of the Department of Computational Medicine and Bioinformatics at the University of Michigan. This algorithm utilizes a combination of "threading" and "*ab initio* folding techniques" in predicting protein structure.

- Threading attempts to match portions of the query sequence to template sequences. The template sequences, and their experimentally solved structures, are part of the RCSB Protein Data Bank (PDB).
- *Ab initio* folding techniques use biophysical properties of the query sequence and simulations to determine the likely structure(s) of the protein.

Professor Zhang's algorithms have won the Protein Structure Prediction Center's five most recent Critical Assessment of Protein Structure Prediction (CASP) experiments. Each CASP experiment is a biennial, world-wide evaluation of structure prediction methods, with approximately 100 participating laboratories.

Note: In CASP, I-TASSER is referred to as "Zhang-Server."

- CASP11 Results - #1 Zhang-Server
- CASP10 Results - #1 Zhang-Server
- CASP9 Results - #1 Zhang-Server and QUARK (another Zhang service)
- CASP8 Results - #1 Zhang-Server
- CASP7 Results - #1 Zhang-Server

For references related to the I-TASSER algorithm, see the Research References page of the NovaFold online help.

Use Protean 3D to view NovaFold results, identify catalytic residues and view predicted enzyme and protein functions and predicted binding sites.

NovaFold Antibody

NovaFold Antibody is specifically designed to generate models of antibodies and antibody fragments (Fv, Fab, VH, sdAb) and, with the use of NovaDock, predict the structure of antibody-antigen complexes. Antibodies frequently have two chains of interest: a light and a heavy chain. When you submit a prediction using NovaFold Antibody, you can provide a light chain, a heavy chain, or both. If you submit two chains, the software will model the complex; if you submit one, it will instead model the chain.

The NovaFold Antibody algorithm utilizes a combination of homology modeling and *ab initio* loop prediction, resulting in highly accurate predictions. During modeling, NovaFold Antibody searches the input sequence(s) against thousands of non-redundant protein antibody structures from PDB and finds the best template matches for the chain or complex.

During modeling, NovaFold Antibody gives particular consideration to the complementary determining region (CDR) loops, hypervariable regions of an antibody that react dominantly with an antigen. The three CDR loops on the heavy chain are known as H1, H2 and H3, while the three on the light chain are L1, L2 and L3. H3 is commonly the most important region in antigen binding. Due to its increased length and flexibility, it is also the most difficult to model. As such, NovaFold Antibody limits H3 loop modeling to fifteen or fewer residues, a length commonly seen in antibody modeling problems. Finally, NovaFold Antibody performs energy minimization calculations to construct the final predicted structure model for the antibody chain or complex. This entire process takes approximately 5-15 minutes on a standard workstation computer.

After running a NovaFold Antibody prediction, the resulting *.antibody* structure file can be viewed and analyzed in DNASTAR's Protean 3D application. Protean 3D's graphical Structure View and Features area facilitate examination of antibody features, including the six annotated CDR loops for light and heavy chains.

NovaFold Antibody can also be used to create an input structure for NovaDock, DNASTAR's protein-protein docking application that can also be used to predict atomic-level antibody/antigen interactions. When setting up a NovaDock prediction, simply use Protean 3D to export the predicted structure in PDB format and specify the exported file as one of the two binding partners.

NovaDock

NovaDock is DNASTAR's protein-protein docking application, and uses an algorithm based on SwarmDock (Paul Bates' group, Biomolecular Modelling Laboratory, Francis Crick Institute, London UK). NovaDock does not use a library of templates, but instead makes docking predictions based on the "particle swarm optimization" algorithm.

NovaDock is used to predict atomic interactions between two binding partners, both of which must be proteins. One binding partner is the ligand ('L': usually the smaller partner or an antigen); the other is the receptor ('R': usually the larger partner or an antibody). Each binding partner can consist of one or multiple protein chains.

You can also direct NovaDock to use information regarding residues which are known or believed to be in the protein interface.

NovaDock requires a complete protein structure that uses only standard residues and has no missing residues and atoms. Before the docking simulation, NovaDock automatically replaces any missing atoms or replaces non-standard residues with the closest natural equivalent. If any missing data were replaced, the NovaDock Report (viewed within Protean 3D), will include a notation to that effect in its Details panel. NovaDock then uses "normal modes of motion" to explore protein flexibility during the docking simulation.

Technical Requirements

The technical requirements for the applications discussed in this document can be found on this page of our website.

Installing and Setting Up the Software

Before installing NovaFold, NovaFold Antibody, or NovaDock, you must first install Docker, a computing platform for distributed applications that is needed to run the programs.

On a Windows or Macintosh computer:

Download and install the DNASTAR License Server by following the "License server installation" instructions on approximately page 7 of the Lasergene Installation Guide PDF.

On the Linux computer where you will be using the Nova applications:

1) Log in to your DNASTAR account; you will automatically be taken to the My Licenses page. If you are already logged in, click the My Account link at the top of the DNASTAR header to go to the same page.

2) Use the button in the **Installers** column to download the installer archive file. Five downloadable archives are available, but you only need to download the installer package; the other four archives are downloaded automatically during the installation process. However, if desired, you may download and save all five archives to your Linux machine.

3) Unpack the installer archive file on your computer (e.g., using the `tar` command).

4) Install Docker using either of these methods, both of which require `sudo` / `root` privileges:

   - Follow the instructions on the Docker website.
   - Enter the following command in the terminal to run a script distributed in the installer archive file:

     ```
     $ get-docker
     ```

     Unless you add qualified users to the `docker` group, Docker requires `sudo` / `root` privileges. If you would like to add users, append their names to the `get-docker` command:

     ```
     $ get-docker $[USERNAME]
     ```

   If you are one of the users to whom access was granted, you must log out and log back in for changes to take effect.

5) Install the application by running the script located in the unpacked application image folder. Installation usually takes 5-20 minutes.

- **Simple install:** You will be prompted to accept the DNASTAR License Agreement ("yes"), and to enter your DNASTAR user name and password.

  ```
  $ install
  ```

- **Custom install:** Depending on the use of optional arguments (i.e., everything following '`install`,' below), you can run the installation without any further input:

  ```
  $ install [-h] [-D DOWNLOAD_DIR] [-s SUBSCRIBER] [-I
  INPUT_DIR] [-y]
  ```

  The following table describes each of the available arguments:

Optional argument	Description
-D DOWNLOAD_DIR, --download-dir DOWNLOAD_DIR	Specify a directory in which to save downloaded files.
-s SUBSCRIBER, --subscriber SUBSCRIBER	Authenticate with your DNASTAR subscriber number instead of user name and password.
-I INPUT_DIR, --input-dir INPUT_DIR	Specify the directory containing the local image and volume data files.
-y, --yes	Automatically respond to prompts with 'yes.'

6) Verify that the repository is showing the correct version of the application:

   ```
   $ docker images
   ```

7) (optional) If you need to resolve local servers by hostname, you can define the local DNS server by editing the `DOCKER_OPTS` variable to read `DOCKER_OPTS="--dns $[LOCAL_DNS_IP]"`. In Ubuntu, the variable is defined in `/etc/default/docker`.

Note: If you previously installed the Nova applications and are having issues with reinstallation, see Reinstalling the Software.

Once you have successfully installed Docker and the Nova applications, the next step is to prepare the input file(s).

Preparing the Input File(s)

Application	Directory contents
NovaFold	A FASTA-format protein sequence file containing a single sequence or chain. Maximum sequence length is 2,000 residues.
NovaFold Antibody	A FASTA-format protein sequence file containing either one or two sequences, representing the light and/or heavy chains of an antibody. If two sequences/chains are included, NovaFold Antibody models the complex. If one sequence/chain is included, NovaFold Antibody instead models the individual sequence or chain.
NovaDock	Two protein structure PDB files: one representing the ligand (L) and the other representing the receptor (R). You can obtain a protein structure file by downloading it from PDB or by predicting a protein sequence's 3D structure using NovaFold or NovaFold Antibody.

The next step is to initiate a prediction (i.e., run the application).

Initiating a Prediction

Predictions are initiated by typing a single command, followed by its parameters.

Note: Before initiating a prediction, you must install Docker and the Nova applications, and prepare the input file(s).

---

1) Launch the terminal.

2) Type the relevant application command and its parameters.

   `--lshost=ipaddress` (or `=hostname`)

   In this step, '`lshost`' must be in lower-case letters and preceded by two dashes (-).

3) (optional) Set up a persistent environment variable to define the license server address. Configuration varies by Linux distribution and command shell. Please consult your distribution's documentation to determine the preferred way to set environment variables.

   For example, if you are using a Bourne-style shell like `<code>bash</code>`, you can set the environment variable for your user account by adding the line:

   ```
   export DNASTAR_LSHOST=<license server IP address>
   ```

   to one of the following files:

$HOME/.profile

$HOME/.bash_profile

$HOME/.bashrc

Note that `DNASTAR_LSHOST` must be in capital letters and should not be preceded by any dashes. In addition, no spaces should be entered before or after the equal sign.

To test if the environment variable has been set, open a new shell and run the `<code>env</code>` command to confirm `DNASTAR_LSHOST` is listed.

4) (optional) Append any additional run parameters:

```
novafold --datadir value1 --seqname value2 --seqfile value3
```

```
novafold-antibody --datadir value1 –-seqname value2 --seqfile value3
```

```
novadock --output-dir value1 --name value2 --lfile value3 --rfile value4
```

…where:

`novafold`
`novafold-antibody`
`novadock`
	Command to run the associated application. A single Linux machine should only run one prediction and one application at a time since all resources will be used.

`value1`	Path to the workspace or output directory.
`value2`	Any desired name consisting of alphanumeric characters, hyphens and/or underscores. No spaces are allowed.
`value3`	Path to an input file, such as a sequence file or a ligand PDB file.
`value4`	Path to an input file, such as a receptor PDB file.

5) (optional) If you did not set up the environment variable in Step 3, above, append the IP address or hostname specification for your DNASTAR License Server:

`--LSHOST ipaddress` (or `hostname`)

6) (optional) Append any additional run parameters. See the example NovaFold script below. Also, see Scripting Commands for a list of available parameters for all three Nova applications.

7) Press **Enter** to begin the prediction. For NovaFold and NovaDock, the prediction may require several hours up to several days to finish, depending on the input. NovaFold Antibody predictions usually require only about 5-10 minutes.

To access the results for a completed prediction, see Accessing Prediction Results. To stop a prediction that is in progress, see Stopping a Prediction.

Example script:

```
novafold --datadir ~/Desktop/BioluminescentSeqs --seqname
firefly_luciferase --hours 100 --GO true --EC true --LBS true
```

The NovaFold script above denotes that the protein sequence `~/Desktop/BioluminescentSeqs/seq.fasta` should be used as the query, and each simulation should run for a maximum of 100 hours. Besides the structure prediction, the following optional predictions should be made: protein function GO terms, enzyme active sites, and ligand binding sites. Results are to be saved as `~/Desktop/BioluminescentSeqs/results.novafold`.

Accessing Prediction Results

After a prediction has finished successfully, result files will be added to the workspace folder (`datadir`) or output directory referenced in the command from Step 3 of Initiating a Prediction.

The *.novafold*, *.antibody* and *.novadock* results files can be opened on any Windows or Macintosh computer that has a suitable version of Protean 3D. To view *.novafold* files you must have Protean 3D version 12.2 or higher. Viewing *.antibody* and *.novadock* files requires version 14.0 or higher.

Application	File Name	Description
NovaFold	*results.novafold*	The main results file for NovaFold; opens as a report in Protean 3D (online help).
	results.zip	Can be decompressed to access the models from the structure prediction in PDB format.
	longstatus.txt	Has a detailed description of the run status, including logs and/or run progress information.
	snapshot.zip	Used for technical support issues only.
NovaFold Antibody	*results.antibody*	The main results file for NovaFold Antibody; opens as a 3D protein structure in Protean 3D.
	longstatus.txt	Has a detailed description of the run status, including logs and/or run progress information.
	model1.pdb	The predicted structure in PDB format.
	NovaFold-Antibody.log	A record of events that occurred during the modeling process.
	summary.txt	Formatted text file summarizing the project.
NovaDock	*results.novadock*	The main results file for NovaDock; opens as a report in Protean 3D (online help).
	novadock.log	Contains a brief description of the run and whether or not sequence repairs were performed.

Scripting Commands

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
Required command	`novafold`	Run a NovaFold structure prediction of the designated query sequence. Requires pre-installation of the ‘dnastar/novafold’ Docker image.		✓		
	`novafold-antibody`	Run a NovaFold antibody structure prediction. Requires pre-installation of the 'dnastar/novafold-antibody' Docker image.			✓	
	`novadock`	Run a NovaDock protein-protein complex prediction. Requires pre-installation of the 'dnastar/novadock' Docker image.				✓
Required arguments	`--datadir DATADIR`	The path to the location where output/results will be written.		✓	✓	
	`--seqname SEQNAME`	The unique name of the query sequence.		✓	✓	
	`--output-dir OUTPUT,` `-o OUTPUT`	The path to the result output directory.				✓
	`--name NAME`	The name for this complex prediction.				✓
	`--lfile LFILE`	Path to a ligand PDB file that includes all chains. The ligand is often, though not always, smaller than the receptor.				✓
	`--rfile RFILE`	Path to a receptor PDB file that includes all chains. The receptor is often, though not always, larger than the ligand.				✓

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
Optional arguments	`--force`	Performs these two functions: 1) Deletes any existing Docker container sharing the same `--seqname` or `--name` as the current prediction input. 2) After the current run has finished, deletes the Docker container for that run. The `--force` command has additional functionality related to reinstalling.		✓	✓	✓
	`-h, --help`	To display a list containing a subset of options and then to exit.		✓	✓	✓
	`--log [CRITICAL, ERROR, WARNING, INFO, DEBUG, NOTSET]`	To enable specific types of logging.	default: [INFO]	✓	✓	✓
	`--lshost LSHOST`	To define the network license server, where value is the IP address (or hostname) of the DNASTAR License Server. For capitalization rules, see Initiating a Prediction. **Note:** This parameter is not required if the environment variable has already been set.		✓	✓	✓
	`--recover`	To recover results from a stopped Docker container.		✓	✓	✓
	`--version`	To output the NovaFold version and exit.		✓	✓	✓
	`--supplemental-help`	To display a list containing additional NovaFold options and exit. An argument shown without brackets is required. The use of brackets signifies that the argument is optional.		✓		

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
	`--library LIBRARY`	The version of library volume for prediction.	default: [latest novafold-antibody-library]	✓	✓	
	`--seqfile SEQFILE`	To designate a local FASTA file as the query. `SEQFILE` represents the local path and filename of the query sequence, which must be in FASTA format. If you provide a path with no filename, NovaFold Local will look for a file named `seq.fasta` located in the specified directory.		✓	✓	

	`--lcontacts LCONTACTS`	To specify a file listing proposed contact residues for the ligand. Each line represents one contact residue using the format: `[CHAIN ID] [RESIDUE NUMBER]`
		• The `CHAIN ID` is case sensitive and is typically a single character.
		• The `RESIDUE NUMBER` is an integer and may include an insertion code.
		Examples:
		`L 15`
		`L 15A`
						✓

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND

	`--rcontacts RCONTACTS`	To specify a file listing proposed contact residues for the receptor. Using this argument causes NovaDock to focus docking around the specified contacts, reducing computation time.
		
		Each line represents one contact residue using the format: `[CHAIN ID][RESIDUE NUMBER]`
		
		• The `CHAIN ID` is case sensitive and is typically a single character.
		
		• The `RESIDUE NUMBER` is an integer and may include an insertion code.
		
		Examples:
		
		`R 15`
		`R 15A`
						✓

	`--lmodes LMODES`	Number of normal modes of motion to explore for the ligand. We suggest using the default to model large, concerted conformational changes and using a higher number (e.g. 20) to model more localized fluctuations, such as within antibody interfaces.	default: 5			✓
	`--rmodes RMODES`	Number of normal modes of motion to explore for the receptor. We suggest using the default to model large, concerted conformational changes and using a higher number (e.g. 20) to model more localized fluctuations, such as within antibody interfaces.	default: 5			✓

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
	`--ncopies NCOPIES`	Number of copies of each swarm point to run.	range: [1,50] default: 4			✓
Supplemental arguments	`--verbose –v`	To enable verbose output.		✓	✓	✓
	`--welcome`	To show a welcome message.		✓	✓	✓
	`--EC value`	To predict enzyme active sites.	allowed: [true, false] default: false	✓		
	`--enhancedSearch value`	To perform DNASTAR's experimental method for enhancing the structural diversity of the normal template set. For templates selected by protein threading, a proprietary process samples alternate structural conformations and replaces a subset of the templates with lower energy conformations. **Note**: This option typically adds 30-60 minutes to the prediction time but, in some cases, improves the accuracy of the prediction. We recommend running the prediction with and without this search option.	allowed: [true, false] default: false	✓		
	`--GO value`	To predict protein function GO terms.	allowed: [true, false] default: false	✓		
	`--homoflag value`	To use all templates or exclude homologs for benchmarking.	allowed: [real, benchmark] default: real	✓		
	`--hours value`	To set a maximum simulation runtime.	range: [1,200] default: 50	✓		
	`--idcut value`	To set a sequence identity cutoff for benchmarking.	range: [0,1] default: 0.3	✓		

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
	`--LBS value`	To predict ligand binding sites.	allowed: [true, false] default: false	✓		
	`--light value`	To enable fast mode (override 'hours' option to 5).	allowed: [true, false] default: false	✓		
	`--nmodel value`	To specify the maximum number of models to create.	range: [1,10] default: 5	✓		
	`--ntemp value`	To specify the maximum number of templates to be used from each threader.	range: [1,50] default: 20	✓		
	`--restraint1 value`	To provide a text file containing a collection of distance and/or contact restraints (*e.g.*, active sites, zinc fingers, disulfide bonds): • Pairwise distances between two atoms ("`i`" and "`j`") • Contact between two residues ("`i`" and "`j`") If both Distance and Contact are specified, they are described in different rows in the same restraint file. `Value` represents the path and filename of the text file containing the distance and/or contact information. A file located outside of the `datadir` data directory will be copied into the `datadir`. **IMPORTANT:** The text file may ***not*** contain any lower-case letters.	See the example below this table for a description of the text file and an example.	✓		

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
	`--restraint2 value`	To provide a text file containing a user-defined template structure and the alignment between that template and the query sequence. `Value` represents the path and filename of a text file containing the information below: • The pairwise FASTA-formatted sequence alignment between query and template. • The standard PDB format 3D structural coordinates of a single protein chain of the desired template. A file located outside of the `datadir` data directory will be copied into the `datadir`. The alignment file may only include one template sequence. When using `--restraint2`, note that: • The length of the aligned template residues must be ≥ 25% of the length of the query sequence. • In the coordinate section, the ATOM record indices need to be numbered sequentially, beginning at 1.	See the example below this table for a description of the text file and an example.	✓		

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
	`--restraint3 value`	To nominate a specific single-chained PDB structure as a template in the modeling prediction, along with other templates selected by NovaFold. `Value` represents the PDB and chain desired for the user template in the format `[PDB ID]:[CHAIN ID]`. • The `CHAIN ID` is case sensitive. • An underscore (_) may be used to designate the first listed chain in the PDB. • Downloading the designated file from the Protein Data Bank requires Internet connectivity. Example: `7tim:A.` Since 'A' is the first chain, `7tim:_` would work as an alternative expression.		✓		

	NovaFold:
	
	`--restraint4 value`
	
	NovaFold Antibody:
	
	`--add ADD`
		To nominate a local 3D structure (in PDB format) as a template in the modeling prediction, along with other templates selected by NovaFold or NovaFold Antibody. `Value` and `ADD` represents the path and filename of a standard PDB format text file. NovaFold requires a single protein chain; NovaFold Antibody can accept multiple chains.		✓	✓	

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND

	NovaFold:
	
	`--temp_excl value`
	
	NovaFold Antibody:
	
	`--exclude EXCLUDE`
		To exclude certain templates from the library (i.e., to prevent these templates from being considered) during structure prediction, where `value` and `EXCLUDE` represents the name of the file containing the list of structures to exclude.
		
		In both NovaFold and NovaFold Antibody, templates can be excluded solely by name. In addition, NovaFold lets you specify a sequence identity cutoff value, such that all templates with an identity at that threshold or higher are excluded. By default, if no sequence identity cutoff is specified, a value of 100% is used.
		
		The tab delimited file listing templates to exclude must have the following format: `[PDB ID][CHAIN ID]`.
		
		Example:
		
		`1wor:A`
		
		A percent sequence identity can be specified at the end, if desired. For example, `1wor:A 70` would specify a 70% sequence identity cutoff. If no number is specified, the percent sequence identity is assumed to be 100%.
		
		An asterisk (`*`) may be used to designate any chain in the PDB file. For example, `1wor:*`
			default sequence identity cutoff: 100%	✓	✓	

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND

	`--include INCLUDE`	To define one or more chains from a PDB structure as the template(s) to use in the modeling prediction. `INCLUDE` represents a file listing the only templates to use. Each line represents one template using the format `[PDB ID][CHAIN ID][HEAVY/LIGHT]`
		• The `PDB ID` is four characters, and starts with a number, followed by three letters or numbers.
		• The `CHAIN ID` is case sensitive and is typically a single character.
		Examples:
		`1IGT:A Heavy`
		`1IGT:A Light`
					✓	

	`--models MODELS`	The H3 loop is the generally the hardest region of the antibody structure to predict. NovaFold Antibody offers a template-based approach that uses a machine learning model to choose the best templates for the H3 loop. The `--models` option lets you specify how many results models to output, each using a unique H3 loop template.	range: [1-10] default: 1		✓	
	`--max-abinitio-h3 MAX_ABINITIO_H3`	To specify a cutoff for switching from *ab initio* "Distance Guided" prediction of the H3 loop to a template-based prediction. Example: The default setting of '3' means loops of length 3 or shorter would use the *ab initio* methodology, while loops longer than 3 residues would be built with the template-based approach.	range [3-15] default: 3		✓	

Type	Name	Definition	Arguments and Defaults	NF	NFA	ND
	`--min-coverage`	The default behavior of NovaFold Antibody (`--min-coverage` = 0.0) is to select a template framework based on its statistical significance to the query sequence; coverage criteria is not considered. The threader picks the template with the highest (log-likelihood score)/(background log-likelihood) for matching a sequence. If a single domain matches, only that domain will be proposed as a model. The result is that the most significant template may not always cover the entire query. To ensure the selected template exceeds a particular fractional coverage, specify a higher number for the `--min-coverage`. This compels NovaFold Antibody to locate a template with both variable and constant domains.	range: [0.0-1.0] default: 0.0		✓	
	`--no-orient-refine`	By default, NovaFold Antibody optimizes the rigid-body orientation between the light and heavy antibody chains to remove atomic clashes if they were introduced during the modeling process. To skip the optimization step, use the argument `--no-orient-refine`.			✓	

Notes regarding the template-based restraint commands (`--restraint2`, `--restraint3/INCLUDE` and `--restraint4/ADD`):

- Only one template-based restraint parameter can be used in the `novafold` command string.
- The `add` option can be used multiple times to introduce multiple templates to the modeling process with the `novafold-antibody` command. The `include` option can be used in combination with the `add` option.

During a folding prediction, the threader and user template alignments are each ranked. Therefore, user provided templates may not necessarily appear in the NovaFold Report's 'top ten' template list.

Example text file for `--restraint1`:

```
DIST   12   HG21  50   HB1   8.1
DIST   14   HA    57   1HE   6.2
DIST   21   HB2   43   HD11  4.0
DIST   124  CA    84   CA   17.4
DIST   36   UNK   120  CA   17.4
CONTACT   33    6
CONTACT   60    29
CONTACT   37    345
CONTACT   109   42
```

Column requirements in the `--restraint1` text file:

Distance rows contain the following columns from left to right:

`"DIST"` (without quotes)
`Res_No.i`
`Atom_type_i`
`Res_No_j`
`Atom_type_j`
`Distance in Angstroms`

Contact rows contain the following columns from left to right (see definitions below):

`"CONTACT"` (without quotes)
`Res_No.i`
`Res_No.j`

In both cases, `UNK` can be used in a row to represent an unknown atom.

Column definitions for a `--restraint1` text file:

Given two residues that contact one another ('Residue i' and 'Residue j') or two atoms at a distance from one another ('Atom i' and 'Atom j'):

- `Res_No_i` – Residue sequence number for Residue i.
- `Atom_type_i` – Atom name for contacting atom of Residue i.
- `Res_No_j` – Residue sequence number for Residue j.
- `Atom_type_j` – Atom name for contacting atom of Residue j.

Example text file for `--restraint2`:

The following is a `-restraint2` file for mammoth myoglobin (query) against whale myoglobin (target structure). "ATOM" rows 6-1211 have been omitted for space. The format for ATOM records is described on this PDB web page.

```
>query
MGLSDGEWELVLKTWGKVEADIPGHGLEVFVRLFTGHPETLEKFDKFKHLKTEGEMKASE
DLKKQGVTVLTALGGILKKKGHHQAEIQPLAQSHATKHKIPIKYLEFISDAIIHVLQSKH
PAEFGAD---------------------------
>1MBN:A
-VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE
DLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRH
PGDFGADAQGAMNKALELFRKDIAAKYKELGYQG

ATOM      1  N   VAL A   1      -2.900  17.600  15.500  1.00  0.00           N
ATOM      2  CA  VAL A   1      -3.600  16.400  15.300  1.00  0.00           C
ATOM      3  C   VAL A   1      -3.000  15.300  16.200  1.00  0.00           C
ATOM      4  O   VAL A   1      -3.700  14.700  17.000  1.00  0.00           O
ATOM      5  CB  VAL A   1      -3.500  16.000  13.800  1.00  0.00           C
…
ATOM   1212  NE2 GLN A 152      -1.600  24.200  -1.500  1.00  0.00           N
ATOM   1213  N   GLY A 153       1.500  24.700  -6.400  1.00  0.00           N
ATOM   1214  CA  GLY A 153       1.100  24.000  -7.600  1.00  0.00           C
ATOM   1215  C   GLY A 153       0.300  22.700  -7.500  1.00  0.00           C
ATOM   1216  O   GLY A 153      -0.900  22.800  -7.100  1.00  0.00           O
TER    1217      GLY A 153
```

Troubleshooting

Stopping a Prediction

Once you have initiated a prediction, the process will normally stop running only after the prediction is finished. If you need to stop a prediction before it is complete:

- If you are in the terminal with the running process, press **Ctrl+C** to stop the prediction.
- If you are in a different terminal, instead enter the command:

  ```
  docker stop seqname
  ```

  … where `seqname` is the same one used when the run was initiated.

  Once you terminate a prediction, you may run a new prediction, if desired.

Uninstalling the Software

You can remove a Nova application using the following command.

```
# /opt/dnastar/<product name>/uninstall
```

```
$ sudo/opt/dnastar/<product name>/uninstall
```

Note that removal requires root access or elevated (sudo) privileges. If there are multiple components installed in the '/opt/dnastar' directory, you will need to repeat this command for each component.

Reinstalling the Software

If you have installed the Nova applications successfully in the past, subsequent reinstallation may result in one or more error messages. Simple steps for resolving these errors are described below:

Error: failed to remove one or more images

This issue occurs when one or more Docker containers reference the image being removed. To address this issue:

1) Find out which containers reference the image by running: `docker ps –a`

2) Remove the Docker containers one by one using this command: `docker rm value`

…where `value` is the either the "Container ID" or the "Name" in the table resulting from Step 1.

Two Docker commands were used to resolve the second issue, above. The following table contains a list of these and other useful Docker commands.

Script	Definition
`docker images`	To view a list of installed Docker images.
`docker rmi value`	To remove a particular Docker image, where `value` is the "Image ID."
`docker ps`	To view a list of Docker containers that are currently running.
`docker ps -a`	To view a complete list of Docker containers.
`docker rm value`	To remove a particular Docker container, where `value` is the "Container ID" or the "Name."