



Submitting Sequences using Specific NCBI Submission Tools

Created: April 6, 2011; Updated: November 3, 2014.

Submission using BankIt

How do I create a submission to GenBank using BankIt?

1. **Review the Requirements for GenBank Submissions through BankIt**, and make sure you can provide the required information for your submission.
2. If you have never submitted to GenBank **scan the GenBank Sample Record to familiarize yourself with GenBank record field definitions**.
3. For examples of specific types of GenBank submissions, see the [GenBank Annotation Example](#) page.
4. **Login to MyNCBI:**
 - a. **Go to the BankIt home page**, and click on “Sign in to use BankIt” located in a yellow box on the right at the top of the page. You will go to the [MyNCBI login](#) page.
 - b. If you do not already have an NCBI account, **click on “Register for an NCBI account”** located below the login text boxes. The boxes marked with an asterisk (*) indicate the minimum amount of information we need to create an account for you. Enter the required information and click the “Create account” button.
5. **Login to BankIt and begin your submission.** The submission process has well marked steps where you will be prompted to provide contact information and your data.
6. The [BankIt home page](#) contains links to sequence annotation examples.

Submission using Sequin

How do I create a submission using Sequin?

1. **Download the Sequin program** from the Sequin Site:
 - a. Go to the [Sequin Home page](#), and click on the “Download Sequin” link located on the side-bar. You will go to the Sequin FTP site where you can download the correct file for your operating system.
 - b. If you are uncertain which FTP file to use, see the “[Select download type](#)” page to get downloading instructions for your specific operating system.
 - c. If you have trouble downloading or installing Sequin, see the [troubleshooting guide](#).
2. **Prepare a properly formatted FASTA file** of your sequence data:
 - a. **The FASTA format is raw sequence preceded by a definition line:**
The definition line begins with a > sign and is followed immediately by the name of your sequence (your own local identification code, or sequence ID) and a title that describes the sequence. Be sure to use a text editor when you create your FASTA file.

For more information on FASTA sequence formatting, see the [FASTA section](#) of the Sequin help document.

- b. Embed important information in the title portion of the definition line and Sequin will use this information to help construct your sequence record. For example:
 - **You can enter organism and strain or clone information** in the title portion of a nucleotide definition line using name=value pairs surrounded by square brackets: [organism=Drosophila melanogaster] [strain=Oregon R]
 - **You can enter gene and protein information** in the title portion of a protein definition line using name=value pairs surrounded by square brackets: [gene=eIF4E] [protein=eukaryotic initiation factor 4E-I]
3. **Launch Sequin.** During the Sequin submission process, Sequin will prompt you to provide the information we need to process your submission.
 - **Sequin has context-sensitive, on-screen help** that will open automatically when you start Sequin. Because it is context sensitive, the Help text will change and follow your steps as you progress through the program.
4. **Submit your completed Sequin file as ASN.1** rather than as a flat file:
 - Be sure to review and fix validation problems before saving your file.
 - To save the file as ASN.1, you can either click the “Done” button on the record viewer, or go to the “File” menu and select “Prepare Submission”, which will also save the file as ASN.1.
 - We cannot accept the flat file format since the flat file format is a display format only.
5. When you have completed the submission process, **you must email the.sqn submission files generated by the Sequin program to togb-sub@ncbi.nlm.nih.gov**, since Sequin does not automatically transmit the completed file for you at the end of the Sequin process (a dialog box will appear at the end of the Sequin submission process instructing you to email your submission files). Note: Do not encode the files before sending.
6. **When we receive a new Sequin submission**, an automatic reply will be generated and sent to the email address used to submit to GenBank. This automatic reply confirms that we received your submission, and states that you will be hearing from the GenBank submissions staff within two working days.

If I've created my submission file using Sequin, what file do I submit — a flat file or an ASN.1 file?

To save the file as ASN.1, you can either click the “Done” button on the record viewer, or go to the “File” menu and select “Prepare Submission”, which will also save the file as ASN.1. Be sure to review and fix validation problems before saving the file.

We cannot accept flat file format even if it is made in Sequin since the flat file format is a display format only.

Submission using tbl2asn

When is tbl2asn a good alternative to Sequin, and can you give me step-by-step instructions for using tbl2asn to create a submission to GenBank?

tbl2asn is a program that allows a user who has:

- Large batches of sequence
- A lot of annotation
- Complete Genomes
- Whole Genome Shotgun submissions

to create a Sequin (.sqn) file for submission without having to go through the step by step process of using the Sequin program.

All you need to do to use the tbl2asn program is:

- Place your data in appropriately formatted files
- Download and run the tbl2asn program
- Command the tbl2asn program to use the data files to generate a .sqn file, which you can submit by email to GenBank.

The main difference between Sequin and tbl2asn is that Sequin is a menu driven program with a graphical user interface, while tbl2asn is a command line program where the user interacts with the tbl2asn software by typing commands to perform specific tasks.

Below are step-by step instructions for creating a .sqn file using tbl2asn. Should you need further information about any part of this process, see the [tbl2asn home page](#) or the [tbl2asn documentation](#) available on the NCBI toolbox [FTP](#) site:

A. Place your data into appropriately formatted data files and place the data files together in a single directory:

1. **There are 6 types of data files that tbl2asn can use to construct a Sequin submission.** 3 are required, and 3 are optional:

Required files:

- [Template file](#) (see step “B” below) containing a text ASN.1 Submit-block object (use file suffix .sbt)
- [FASTA file for Nucleotide sequence data](#) (use file suffix .fsa)
- [Feature Table file](#) (use file suffix .tbl). [Required only if including annotation]

Optional files:

- [Quality Score file](#) (use file suffix .qvl)
- [Source Table file](#) (use file suffix .src) [useful when submitting multiple records with source qualifiers that have different values]
- Protein sequence file (use file suffix .pep) [These files are rarely needed]

2. **The prefixes (base names) of the different files** (with the exception of the .sbt file) **you are going to use together to construct a submission should be the same as that of the .fsa file** since tbl2asn will look for .tbl, .src, and .qvl files that have the same prefix as the .fsa file in order to make the Sequin file. For example:

- **template.sbt** (this is the only file whose prefix is different. Leave the prefix as is).
- chr01.fsa
- chr01.tbl
- chr01.qvl

3. **Save the files that you will be using in a single directory to construct a submission in the same directory.**

B. Create a Submission Template file:

(You can create this file by using the Online Submission Template page or by using Sequin.)

1. **Using the Online Submission Template page:**

- a. Go to the online [Create Submission Template page](#).
- b. Fill in the required (*) and optional textboxes in the “Contact Information”, “Sequence Authors” and “Reference Information” sections.
- c. Click on the "Create Template" button at the bottom of the page.
- d. SAVE the file as **template.sbt** to the same directory that contains the other files to be used for the submission.

2. **Using Sequin:**

- a. Load Sequin onto your computer.

- b. Click the “Start New Submission” button on the Sequin startup page.
 - c. Enter the manuscript title if desired and click the “next page” button.
 - d. Enter your contact information and click the “next page” button.
 - e. Enter the author information and click the “next page” button.
 - f. Enter the affiliation information and click the “next page” button.
 - g. Click on the white submission tab to return to the submission window.
 - h. Click on “File” located at the upper left corner of the submission window to activate a drop-down menu and select “Export Submitter Info”.
 - i. Save the file as template.sbt.
- C. **Download the tbl2asn program appropriate to your operating system:**
1. Go to the tbl2asn [FTP](#) site.
 2. Click on the file appropriate for your operating system to download.
 3. **Uncompress** the tbl2asn file using the appropriate zip utility.
 4. Rename the file **tbl2asn**, and set permissions as required for your operating system.
- D. **Open a command line interpreter for Windows or Mac operating systems.** Because tbl2asn is a command line program, you can’t just click on the tbl2asn icon to open it like you would a standard graphical interface program like Sequin. You first have to open a [command line interpreter](#) for your operating system, and then once you are in the command line interpreter, you can command your computer to run the tbl2asn program:
- **In the Windows operating system (OS), the command line interface is called “Command Prompt”, which you can find by doing the following:**
 - a. Go to the “Start” menu.
 - b. Click on “All programs” to release a menu.
 - c. Click on “Accessories” to release a menu.
 - d. Click on “Command Prompt” to open the “Command Prompt” command line interpreter.
 - **In the Mac operating system (OS), the command line interface is called “Terminal”, which you can find by doing the following:**
 - a. Go to the “Applications” folder.
 - b. Double click on the “Utilities” folder to open it.
 - c. Double click on “Terminal” to open the “Terminal” command line interpreter.
 - **In the Linux operating system (OS),** open a shell and use `chmod +x` to allow the downloaded program to be executed. (File transfer with FTP does not retain UNIX file permissions.)
- E. **Move tbl2asn.exe to a directory that your computer will search automatically when you enter the **tbl2asn** command.** The benefit of moving tbl2asn.exe to an automatically searched directory is that you will only have to enter the command **tbl2asn** after the prompt without having to type out a lengthy path to tbl2asn.exe every time you want to use it.

Windows OS:

- a. Go to “Command Prompt” command line interpreter you opened in step D.
- b. Type **path** following the prompt, and hit the “Enter” button. The computer will list all the paths (the PATH directories) it automatically searches when you enter a command in the command line interface.
- c. For example:


```
C:\Documents and Settings\Owner>path
PATH=C:\WINDOWS\System32; C:\WINDOWS; WINDOWS\System32\Wbem
```

From the above response, you can see that the for the example computer, the WINDOWS directory is one of the directories that is automatically searched when the user enters a command (the PATH

directory). So, if you place `tbl2asn.exe` in this directory, the computer will find and run `tbl2asn.exe` if you type the command **tbl2asn** following the prompt.

- d. Move `tbl2asn` to the PATH directory mentioned in previous step.

Mac OS:

- a. Open the Applications folder.
- b. Create a new folder, and give it a recognizable name. For this example, we'll name the folder: `Command_line_apps`.
- c. Move the `tbl2asn` file you downloaded in step C into the `Command_line_apps` folder.
- d. Go to the "Terminal" command line interpreter you opened in step D, and enter the following command:

```
export PATH=/Applications/Command_line_apps:$PATH
```

At this point, you can start the `tbl2asn` program in the command line interpreter by using the command **tbl2asn**.

Note: you will have to repeat step d of these instructions for each new Mac "Terminal" session in order to use the command **tbl2asn**, since the command given in step d is not remembered from one "Terminal" session to the next.

F. **Change the default directory of your command line interface to the directory that contains your data files.**

The benefit of changing directories on your command line interface to the directory that contains your data files is that you can enter `tbl2asn` commands without having to type a lengthy path to the directory that houses the files for each command you use:

1. Following the prompt, type `cd` (change directories) followed by a space, and then by the path to the directory that contains your data files. Hit the "Enter" button.
2. The prompt should change to reflect the new directory (called `Sequence_Data` in the following examples):

- **Example of the `cd` command followed by the new prompt** (changed to reflect the new directory) **in a Windows OS command line interpreter.** In this example, the data files are housed in a file called "Sequence_Data", which in turn are housed in a directory called "My Documents":

```
C:\Documents and Settings\Owner\>cd C:\Documents and
+Settings\Owner\My Documents\Sequence Data
C:\Documents and Settings\Owner\My Documents
\Sequence_Data>
```

- **Example of the `cd` command followed by the new prompt** (changed to reflect the new directory) **in a Mac OS command line interpreter.** In this example, the data files are housed in a file called "Sequence_Data":

```
Apple1:~ Username$ cd Documents/Sequence_Data
Apple1:~/Documents/Sequence_Data Username$
```

Where Apple 1 = computer name in Mac OS; this name will be different for every computer as it reflects each individual computer's name.

G. **Run `tbl2asn` within a command line interpreter and access all current `tbl2asn` commands.**

A list and definitions for the most commonly used `tbl2asn` commands is available on the [tbl2asn home page](#). You can access a complete list of all `tbl2asn` commands by doing the following (these instructions will work for all operating systems):

1. Open the command line interpreter for your operating system.

2. Type **tbl2asn** followed by a space and a hyphen after the command line interpreter prompt:
Prompt>tbl2asn -
3. Hit the “enter” key to display all the tbl2asn commands.
4. **The following TBL2ASN commands will be used in the example below to create a sequin file** using the minimum required files: a FASTA file, a feature table, and a template file.
 - -p specifies the path for the table and sequence files [required]
 - -t specifies the template file (including the path) [required]
 - -j allows the addition of source qualifiers that will be the same for each submission
Example: -j "[organism=Saccharomyces cerevisiae] [strain=S288C]"
 - -V is a verification function when combined with the following:
 - v performs a validation [optional but strongly recommended]
 - b generates GenBank flatfiles with a .gbf suffix
 (Sample command line: -V vb). Note: The flatfile format is for viewing only. It cannot be submitted.

H. Use **tbl2asn** commands to create a **Sequin (.sqn) file using required data files**. **NOTE:** If you intend to submit multiple sequences in a single submission, see Step H6 of this question before beginning. The following example will show how to use the above **tbl2asn** commands to create a Sequin (.sqn) file using the minimum required files —a template file (.sbt), a FASTA file (.fna) and a feature table file (.tbl). The following instructions will work for all operating systems.

In order to use the instructions as written below, you must place **tbl2asn.exe** in a path directory (see step E), and you must change the command line interpreter’s default directory to one that houses all the files you intend to use in a single submission (see step F).

1. **Type **tbl2asn** after the prompt** (do not hit “Enter” yet).
This command tells the computer to run **tbl2asn.exe**
Example:
Prompt>tbl2asn
2. **Type a space after **tbl2asn**, then type **-t**, another space, and the name of your template file** (do not hit “Enter” yet).
The **-t** command followed by the name of your template file tells the computer the name of the template file to use in creation of your .sqn file when using **tbl2asn**.
Example:
Prompt>tbl2asn -t template.sbt
3. **Type a space following the name of your template file, then type **-p**, another space, then type a period (dot)** (do not hit “Enter” yet).
-p alone tells the computer where to look for the table and sequence files. **-p** followed by a space and then a dot tells the computer to look for the table and sequence files in the current directory.
Example:
Prompt>tbl2asn -t template.sbt -p .
4. **Type a space following the dot, then type **-j**, then type another space, and then provide source modifier information inside of quotation marks and brackets**. For example:
"[organism=Saccharomyces cerevisiae] [strain=S288C]" (do not hit “Enter” yet)
-j tells the computer to add the source information that follows to each submission. If

there is annotation and the genetic code is not the standard code, then include the [correct code](#) in the [fsa definition line](#), or with the `-j` in the command line, to avoid errors.

Example:

```
Prompt>tbl2asn -t template.sbt -p . -j
"[organism=Saccharomyces
cerevisiae] [strain=S288C]"
```

NOTE: The method stated in step 4 is good if you have source information that is common to all the files in the directory. If you have additional source information that is specific to particular submissions, omit the `-j` command, and:

- include the source information in the definition line of each [FASTA](#) (.fsa) sequence file.

OR

- create a [tab-delimited source table](#) (file suffix .src) for each .fsa file, and place it in the directory where the other files specific to a particular submission are housed.

5. **Type a space following the source information in quotation marks and brackets, then type `-V`, another space, and then `vb`** (do not hit “Enter” yet).

`-V` is a verification command when used in conjunction with `v` (strongly suggested), which will tell the computer to run a validation step to insure that there are no errors in your submission.

This validation step will generate a report (with suffix .val) for each .fsa file and place it in the same directory that houses the data files and tables used in the submission.

If you add a `b` command (optional) following the `v` command, the computer will generate a GenBank flat file (.gbf) of your submission and deposit it in the same directory that houses the data files and tables used in the submission. Note that .gbf files are not suitable for submission. They are only to view the file in GenBank flatfile format.

Example:

```
Prompt>tbl2asn -t template.sbt -p . -j
"[organism=Saccharomyces
cerevisiae] [strain=S288C]" -V vb
```

6. Optional Step

If your submission contains multiple records, put them in a single .fsa file, and add the following to the command line:

Type a space following your last command, then type `-a` then type another space and then type `s`

Example:

```
Prompt>tbl2asn -t template.sbt -p . -j "[organism=Saccharomyces
cerevisiae] [strain=S288C]" -V vb -a s
```

The `-a` command used in conjunction with the `s` command instructs `tbl2asn` to read multiple FASTA components in one file as a set of unrelated sequences. This creates a single file of multiple submissions.

Remember: Each `.tbl` and `.qvl` file will need to contain the information for the sequences in the corresponding `.fsa` file.

Note: You will have to try to achieve a balance between the number of files you submit and the number of records per file submitted. In general, limit the number of records (sequences) per file to a few thousand when there is no annotation, and a few hundred when there is annotation.

7. **Hit the “Enter” key on your keyboard.**

The response will be as follows:

[`tbl2asn` and the current `tbl2asn` version number] **Flatfile** followed by the prefix used for the tables and files used for this specific submission.

Followed by

[`tbl2asn` and the current `tbl2asn` version number] **Validating** followed by the prefix used for the tables and files used for this specific submission.]

Followed by

the command line interface prompt.

Example:

```
[tbl2asn 15.2] Flatfile chr01
[tbl2asn 15.2] Validating chr01
Prompt>
```

8. **Find the Sequin file (.sqn), the validation file (.val) and the GenBank Flatfile (.gbf) generated by the `tbl2asn` program for each `.fsa` file:**

Once `tbl2asn` generates the `.sqn`, `.val` and `.gbf` files, it automatically places them in the same directory that houses the data files and tables used in the submission.

- I. **Open the `.val` file using a text editor to see the errors. Open the `.sqn` file in Sequin to correct any errors that are mentioned.** Or you can make the appropriate changes in the `.tbl` file and remake the `.sqn` file.
 - Taxonomy-related errors about missing lineages can be ignored
 - If there is annotation and the genetic code is not the standard code, then include the [correct code](#) in the [.fsa definition line](#), or with the `-j` in the command line, to avoid errors
- J. **Optional:** open the Genbank Flatfile with a text editor (if you chose to include the generation of a GenBank Flatfile in your commands) and review. The `.gbf` files are only for display, not for submission. Do not make any changes to these files; the changes need to be in the `.sqn` files.
- K. **After you have checked your files and corrected any errors that the validation step found, send the files to Genbank** either by email to gb-sub@ncbi.nlm.nih.gov, or by using [SequinMacroSend](#) for regular submissions, and [GenomesMacroSend](#) for genome (complete or incomplete whole genome shotgun) submissions.

When submitting to GenBank using `tbl2asn`, do I always have to provide a project ID?

The answer to this question depends upon what you are submitting.

A project ID is only required if you are submitting:

- A bacterial or eukaryotic (non-organelle) genome
- Genome-scale studies including:
 - Targeted loci studies
 - Metagenomic studies
 - Multi-isolate studies

For information on how to register your project with BioProject and get a Project ID, see the “[Submitting to BioProject](#)” section of the [BioProject help manual](#).