invitrogen

Vector NTI™ *Express* Software

USER GUIDE

Publication Number MAN0006009

Revision A

For Research Use Only. Not for use in diagnostic procedures.

ThermoFisher SCIENTIFIC

The information in this guide is subject to change without notice.

DISCLAIMER

TO THE EXTENT ALLOWED BY LAW, LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) WILL NOT BE LIABLE FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING YOUR USE OF IT.

Corporate entity

Life Technologies | Carlsbad, CA 92008 USA | Toll Free in USA 1.800.955.6288

NOTICE TO PURCHASER: DISCLAIMER OF LICENSE

Purchase of this software product alone does not imply any license under any process, instrument or other apparatus, system, composition, reagent or kit rights under patent claims owned or otherwise controlled by Life Technologies Corporation, either expressly, or by estoppel.

TRADEMARKS

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

Contents

About This Guide

This user guide provides reference information for Vector NTI™ *Express* Software, a Java-based cross-platform application that can run on computers using the following operating systems:

- Microsoft® Windows® XP (SP3) and Windows® 7
- Mac OS® X
- Linux

For installation and licensing information, see the **Vector NTI™** ***Express*** **Release Notes and Installation/Licensing Guide**.

This guide uses conventions and terminology that assume a working knowledge of the your operating system, the Internet, and Internet browsers.

Note on screen captures

This guide includes screen captures for the Microsoft® Windows® version of the software. For Macintosh® and Linux users, the screens will look slightly different but the commands and functions will be the same.

Revision history

Revision	Date	Description
1.0	15 January 2012	New user guide
A.0	19 January, 2016	Updated Chapters on AlignX, Cone2Seq, and various screen shots.

1 Database Explorer

Overview

The Vector NTI™ *Express* local database is a database file and associated folder that by default are installed in the root directory of your local hard drive (e.g., C:\VntiExpress_Database). Database Explorer is the tool in Vector NTI™ *Express* for managing all the molecules, oligos, projects, analysis results, and other data in the local database.

Eight different types of objects are stored and organized in databases and subsets in the Vector NTI™ *Express* local database:

- DNA/RNA molecules—these contain a nucleic acid sequence as well as annotations, primers, restriction sites, analysis results, and other molecule data. Upon import from other sources, nucleic acid data are parsed and stored in an internal format. You can add molecules to the database by importing or creating basic or constructed molecules.
- Protein molecules—these contain an amino acid sequence as well as annotations and other features. Like DNA molecules, upon import from other sources, protein molecule data are parsed and stored in an internal format. You can add molecules to the database by importing or creating basic molecules.
- Projects—including Alignment, Contig Assembly, and Cloning projects.
- Restriction enzymes (RENs)—imported from the REBASE database. Data for restriction enzymes are parsed and stored in an internal format. You can add other RENs from the REBASE file included in the Vector NTI™ *Express* Software.
- Oligonucleotides—these can be created by the user or generated by primer design tools in Vector NTI™ *Express*. Several example oligos are included in the software for demonstration purposes.
- Gel markers—these can be created by the user. Commonly used gel markers are included with the Vector NTI™ *Express* installation.
- BLAST results—these are generated by BLAST searches run Vector NTI™ *Express*, and can be stored as results files in the database.
- Analysis results—such as PCR analysis or PFAM analysis of molecules, can be stored as results files in the database.

Archives

Parent-descendant relationships (to keep track of your constructs), user fields, comments, and keywords are kept for all molecules in the database.

All database molecules and other objects can be placed into "archives" – data files of special format – that can be transferred to another computer (Macintosh or Windows) and read by Vector NTI™ *Express* Software. Through archives, you can share molecules, constructs, or other objects with your colleagues, or simultaneously use them on several computers (for instance, at work and at home).

In Vector NTI™ *Express* Software archives:

- All DNA/RNA molecule information is written to, and read from, an archive file. This information includes molecule component fragments (if the molecule is constructed from other molecules) and parent-descendant connections between molecules.
- Vector NTI™ *Express* Software automatically checks the consistency of molecule archive information, adding necessary parents (including DNA parents of translated protein molecules) or disconnecting them if you neglected to transfer them to the archives.
- When the archive is loaded into a new database, Vector NTI™ *Express* Software checks the information consistency on any of database molecules and recalculates them if necessary.

Open the Database Explorer

The Database Explorer can be opened two ways:

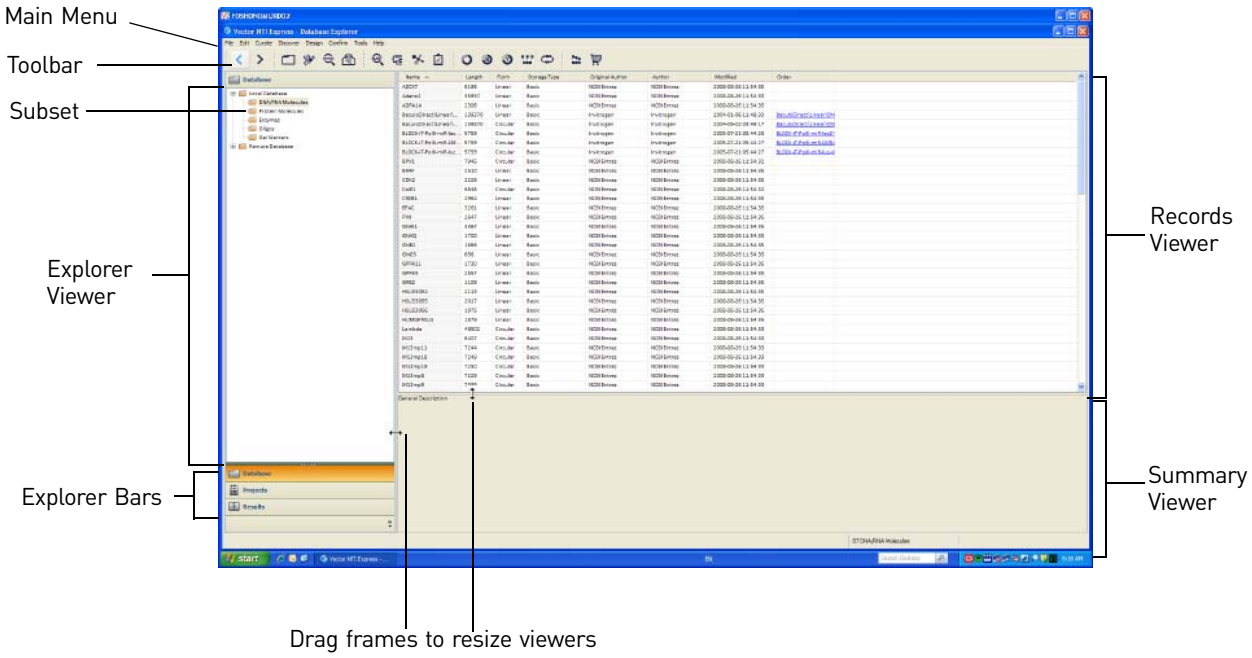
- In Windows, click **Start** ▸ **Programs** (or **All Programs**) ▸ **Life Technologies** ▸ **Vector NTI Express** ▸ **Vector NTI Express**. By default, Vector NTI™ *Express* opens with the Database Explorer screen.
- In the main toolbar, click the Database Explorer button:

Components of the Database Explorer

The Database Explorer screen consists of the following components:

- Three databases: Database, Projects, and Results.
- An Explorer Viewer for navigating among three types of databases. The viewer contents change when you click the Explorer buttons below the viewer.
- Explorer bars for opening and navigating the Database, Projects, and Results databases.
- A Records Viewer pane for viewing the data and records associated with the Database, Projects, and Results.
- A Summary Viewer for viewing a snapshot of a selected record.

Note: You can widen, narrow, lengthen, and shorten the size of viewers by clicking and dragging the viewer frames.



Types of databases

There are three types of databases in the Database Explorer:

- Database includes data associated with DNA and RNA records; protein records; and enzymes, oligos, and gel markers.
- Projects include data and results associated with alignment projects, contig assembly projects, and cloning projects.
- Results include BLAST and analysis results.

Subsets

A subset is a group of objects organized by a specified criteria, such as common features. For example, you might have one subset for each of your molecule families, and one for each taxonomic group. Subsets are shown as folders in the Database, Projects, and Results panes. You can search for records by name and create and store data in subsets.

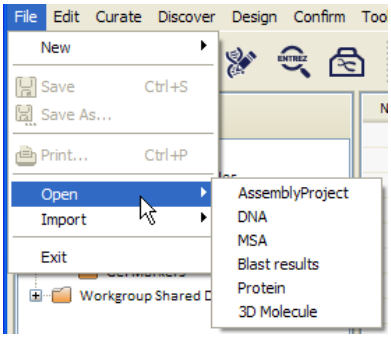
Open files

You can open assembly projects, molecules, and BLAST results. The following file formats are recognized by Vector NTI Express Software:

File	Format
DNA/RNA molecules	All nucleotide files (*.gb, *.gbwithparts, *.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq, *.embl, *.gcg, *.ma4, *.ddbj)
	Genbank (*.gb, *.gbwithparts)
	DDBJ (*.ddbj)
	Fasta (*.fasta, *.fas, *.mpfa, *.fna, *.fsa, *.seq)
	EMBL (*.embl)
	GCG (*.gcg)
	Vector NTI™ Archive (*ma4)
Protein molecules	All protein files (*.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fra, *.fsa, *.seq, *.gp, *.gpwithparts, *.swp, *.pa4, *.trembl)
	Fasta (*.fasta, *.fas, *.mpfa, *.fna, *.fsa, *.seq)
	GenPept (*.gp, *.gpwithparts)
	Swiss-Prot (*.swp)
	TrEMBL/EMBL (*.trembl,*.embl)
	Vector NTI™ Archive (*pa4)
3D molecules	*.pdb
Assembly projects	Contig Project (*.cepx, *.cep)
Alignment projects	Alignment Project (*.apr, *.aprx, *.msf)
BLAST results	Vector NTI™ Archive (*.ba6)

To open a file:

1. Click **File** ▸ **Open**.
2. In the drop-down list, select a subset.



3. In the Open File window, navigate to the file, select it, then click **Open**.

Database Explorer operations

In Database Explorer, you can:

- Create DNA/RNA and protein molecules (this page), gel markers, oligos, and enzymes (page 18), and projects (page 32).

- Import data (page 20) and export data (page 26) (not available in the demonstration version).
- Manage your data by organizing your data into convenient groups (subsets), and sorting, duplicating, and deleting data (page 22).
- Edit data (page 27).
- Search the database for text sequence, motifs, feature types, keywords, and other information (page 27).
- Open other applications (page 29).
- Copy a molecule (page 29) save a molecule (page 31), and print molecule data (page 32).

Create DNA/RNA and protein molecules

There are five different ways to create DNA/RNA and protein molecules:

- Import a molecule from DDBJ, EMBL, FASTA, GCG, GenBank, GenPept, SWISS-PROT, TrEMBL, and Vector NTI™ Archive formats, or from an ASCII file of flexible format. The sequence and feature map are converted from the file, and the new molecule becomes part of the Vector NTI™ *Express* Software database.
- Create molecules from nucleotide or amino acid sequences that you define. You can manually enter these sequences or you can paste from a clipboard and enter the sequence as a new molecule.
- Translate a coding region of an existing DNA or RNA molecule to create protein molecules.
- Construct new DNA/RNA molecules from user-defined compatible component fragments from other molecules.
- Design DNA/RNA molecules from components in a user-defined fragment list, using the built-in biological knowledge of Vector NTI™ *Express* Software to design the recombination process. All new molecules are integrated into the database and can be used in all future operations and analyses.

Create a DNA/RNA molecule

To create a DNA/RNA molecule:

1. In the main menu, click **File** ▸ **New**.
2. In the drop-down list, select **DNA** to open the New DNA Molecule window.
3. Enter a name, select the form, and enter a description, then click **OK**.

Create a Protein molecule

To create a protein molecule:

1. In the main menu, click **File ▸ New**.
2. In the drop-down list, select **Protein** in the New Protein Molecule window.
3. Enter a name and description, then click **OK**.

Create gel markers, oligos, and enzymes

Gel marker, oligonucleotide, and enzyme objects can be created "from scratch" using editors in Vector NTI™ *Express* Software or by importing them from FASTA files, Vector NTI™ Archive files, an oligo list, and a REBASE database.

Create a gel marker

To create a gel marker:

1. In the main menu, click **File ▸ New**.
2. In the drop-down list, select **Gel Marker** to open the Gel Marker window.
3. Enter a name. Click the browse button (**...**) to open the Choose Database Gel Marker window, then click **OK**.
4. Select a gel marker type.
5. Click the **Gel Marker** tab, enter a fragment length, then click **OK**.

Create an oligo

To create an oligo:

1. In the main menu, click **File ▸ New**.
2. In the drop-down list, select **Oligo** to open the Oligo window.
3. In the General tab, enter a name, then click the **Oligo** tab.
4. Enter the nucleotide sequence, then select the oligo type.

   To replace the oligo with its complementary sequence, check the **Reverse Complementary** box.

   Enter a description, then click the **User Fields** tab.
5. Click a cell in the table, click **Change Value** to open a data entry window, then enter a value.

   To remove a value, click **Remove Value**.

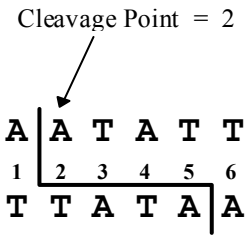
   After you have finshed entering values, click the **Comments** tab.
6. Enter comments about the oligo, then click the **Keywords** tab.
7. To add a keyword for the oligo, enter a new word or select an item in the list of existing keywords.

   To move the keyword into the keyword list, click **< Add**. To remove an item from the keyword list, select it, then click **> Remove**.
8. To save your data, click **OK**. To close the Oligo window without saving your data, click **Cancel**.

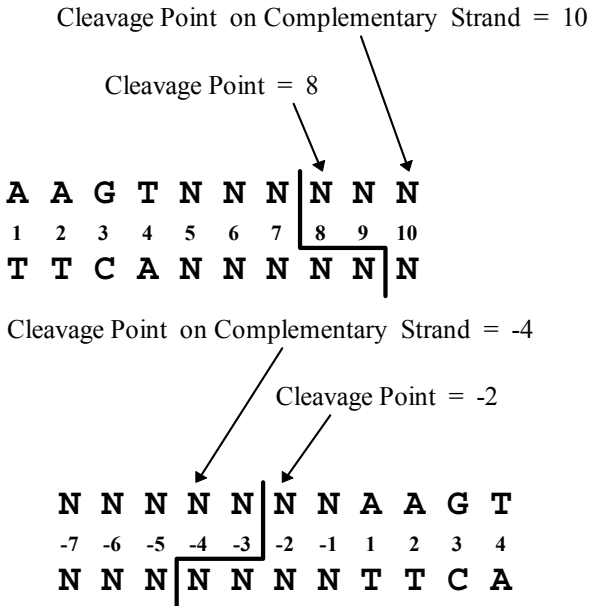
Create an enzyme

To create an enzyme:

1. In the main menu, click **File ▸ New**.
2. In the drop-down list, select **Enzyme** to open the Enzyme window.
3. In the General tab, enter a name, then click the **Enzyme/Methylase** tab.
4. Enter the recognition string of the enzyme.
5. In the Cleavage Point/Methylation Base field, enter the number of the nucleotide immediately after the direct strand cleavage point. The following example demonstrates how cleavage points of palindromic sites are defined.



In the Cleavage Point on Complementary Strand field, enter the number of the nucleotide immediately after the complementary strand cleavage point. The following example demonstrates how cleavage points are defined for non-palindromic sites on both direct and complementary strands.



6. Select the enzyme type, regular restriction enzyme or methylase, then click the **User Fields** tab.
7. Click a cell in the table, click **Change Value** to open a data entry window, then enter a value.

   To remove a value, click **Remove Value**.

   After you have finshed entering values, click the **Comments** tab.
8. Enter comments about the enzyme, then click the **Keywords** tab.

9. To add a keyword for the enzyme, enter a new word or select an item in the list of existing keywords.

   To move the keyword into the keyword list, click **< Add**. To remove an item from the keyword list, select it, then click **> Remove**.

10. To save your data, click **OK**. To close the Enzyme window without saving your data, click **Cancel**.

Import data

You can import DNA/RNA and protein molecules, enzymes, oligos, and gel markers, assembly projects, and BLAST results in the Vector NTI Express Software Local Database.

Import molecules

You can import molecules (including their feature tables) from DDBJ, EMBL, FASTA, GCG, GenBank, GenPept, SWISS-PROT, TrEMBL, and Vector NTI™ *Advance* Archive files. Two types of DNA/RNA molecules can exist in the Vector NTI™ *Express* Local Database:

- Circular DNA Molecules
- Linear DNA/RNA Molecules

Molecules are imported in the DNA/RNA Molecules subset and Protein Molecules subset within the Local Database. The DNA/RNA Molecules and Protein Molecules subsets contain descriptions of a molecule's features.

You can also import nucleotide or amino acid sequences from an ASCII file of flexible format, and Vector NTI™ *Express* Software will automatically create the new database molecule and assign the sequence to the molecule.

File formats

The following file formats can be imported into Vector NTI™ *Express*:

File	Format
DNA/RNA molecule	All nucleotide files (*.gb, *.gbwithparts, *.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq, *.embl, *.gcg, *.ma4, *.ddbj)
	Genbank (*.gb, *.gbwithparts)
	DDBJ (*.ddbj)
	Fasta (*.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq)
	EMBL (*.embl)
	GCG (*.gcg)
	Vector NTI™ Archive (*ma4)

File	Format
Protein molecule	All protein files (*fasta, *fas, *fa, *mpfa, *.fna, *fsa, *seq, *.gp, *.gpwithparts, *.swp, *.pa4, *.trembl)
	Fasta (*.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq)
	GenPept (*.gp, *.gpwithparts)
	Swiss-Prot (*.swp)
	TrEMBL/EMBL (*.trembl,*.embl)
	Vector NTI™ Archive (*pa4)
Enzyme	REBASE (*.rebase), Vector NTI Archive (*.ga4)
Oligo	Vector NTI Archive (*oa4), Oligo List (*.txt), Fasta (*.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq)
Gel marker	Vector NTI™ Archive (*.ga4)
Assembly Project	Contig Project (*.cepx, *.cep)
BLAST results	Vector NTI™ Archive (*.ba6)

Import files

There are two ways to import files: by dragging and dropping files in Windows Explorer or by using the main menu in the Vector NTI Express Software.

Drag and drop files into a subset

To drag and drop files into a subset:

Highlight the file names in Windows Explorer. Select the files, then drag and drop them onto a subset in the Local Database folder in the Explorer Viewer.

Import files via the main menu

To import objects via the main menu:

1. Click **File > Import**.
2. In the drop-down list, select the type of file you want to import.
3. In the Import into Database window, navigate to the file, select it, then click **Open**.

Manage data

This section describes how to:

- Manage Database, Projects, and Results databases (page 22)
- Manage Database, Projects, and Results subsets (page 22)

Manage data in the Database, Projects, and Results databases by creating subsets of subsets (subfolders of folders), grouping data within subsets, and organizing subsets. Subsets can be hierarchically organized down to six levels.

You can:

- Create DNA and protein molecules, gel markers, oligos, enzymes, and projects.
- Open an assembly project; DNA, protein, and 3D molecules; and BLAST results.
- Import DNA and protein molecules, gel markers, enzymes, oligos, BLAST results, and assembly projects.
- Edit molecules by inserting, deleting, and replacing sequence fragments and features. You can also modify the display format and general data of molecules.

Manage Database, Projects, and Results databases

In the Database Explorer, you can create subsets in a database and search databases.

Create a Local Database subset

To create a subset:

1. In the Explorer Viewer, click **Local Database**, right-click a subset (for example, DNA/RNA Molecules), then select **Add subset** to open the Create Subset window.
2. Enter a subset name (for example, “Hexokinase”), description of the subset (for example, “A collection of hexokinase genes”), then click **OK**.

Search a Local Database subset

To search a Local Database subset:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a database, then select **Search** to open the Search window.
3. Enter text. To search for the exact text, check **Exact match only**. Click **OK** to begin the search.

Manage Database, Projects, and Results subsets

Manage Database, Projects, and Results subsets in the Explorer Viewer. You can:

- Add a subset (page 23)
- Search a local database or subset (page 23)
- Edit the properties of a local database subset (page 23)
- Rename a local database subset (page 23
- Clear a local database subset (page 23)
- Dismiss local database subsets (page 24)
- Delete the contents of a local database subset (page 24)
- View a summary of a local database subset (page 24)

To manage Local Database subsets:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a subset, then select a managerial operation.

Add a subset

To add a subset to a Local Database:

1. In the Explorer Viewer, click **Local Database**.
2. Right-click a database, then select **Add subset** to open the Create Subset window.
3. Enter a subset name and description, then click **OK**.

Search a local database or subset

To search a local database or subset in a local database:

1. In the Explorer Viewer, click **Local Database**.
2. Right-click a database, then select **Search** to open the Search window.
3. Enter text. To search for the exact text, check **Exact match only**. Click **OK** to begin the search.

Edit the properties of a local database subset

To edit the properties of a Local Database subset:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a database, then select **Edit properties** to open the Edit Subset window.
3. Enter a new name for, and/or description of, the subset, then click **OK**.

Rename a local database subset

To rename a Local Database subset:

1. In the Explorer Viewer, click **Local Database**, then select a database.
2. In the Records Viewer, right-click an object, then select **Rename** to open the Rename window.
3. Enter a new name, then click **OK**.

Clear a local database subset

Note: Clearing a Local Database subset of its contents does not permanently delete the contents from a local database.

To clear a subset of its contents:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a subset, select **Clear subset** to open the Subset Management window, then click **OK**.

Dismiss local database subsets

Dismissing a subset deletes the subset from a local database.

To dismiss Local Database subsets:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a subset, then select **Dismiss subsets** to open the Subset Managment window, then click **OK**.

Delete the contents of a local database subset

Note: Deleting the contents of a subset *permanently* deletes them from the Local Database.

To delete the contents of a Local Database subset:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a subset, select **Delete Subset Contents** to open the Vector NTI™ *Express* window, then click **OK**.

View a summary of a local database subset

To view a summary of a Local Database subset:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. Right-click a subset, then select **Subset Summary** to open the Subset Summary window.

Manage database objects

Manage database objects in the Records Viewer. You can:

To manage database objects:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click a database object, then select a managerial operation.

Group database objects

You can group database objects in subsets.

To group database objects:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, select several objects.

   Note: To select specific objects, press the **Ctrl** key and click objects. To select a block of objects, click the first object in the block, press the **Shift** key, then click the last object.
3. Add the objects to a subset:
   - Right-click the selected objects, then select **Add to subset** to open the Subset Management window. Select a subset, then click **OK**, or
   - Drag the selected objects into the subset in the Explorer Viewer.

   Note: The objects you grouped were not moved out of the parent subset.

Sort database objects

To sort database objects in ascending or descending order:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, click a column header.

Add an object to a subset

To add an object to a subset:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click an object, then select **Add to subset** to open the Subset Management window.
3. Select the subset you want to add the object to, then click **OK**.

Rename an object

To rename an object:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click an object, then select **Rename** to open the Rename window.
3. Enter a new name, then click **OK**.

Exclude an object from a subset

Note: Excluding an object from a subset does not permanently delete it from the database.

To exclude an object from a subset:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click an object, select **Exclude from subset** to open the Vector NTI™ *Express* window, then click **OK**.

Delete an object

To permanently delete an object:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click an object, select **Delete from Database**, then click **OK**.

Duplicate an object

To duplicate an object

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click an object or multiple objects, then select **Duplicate**.

Duplicates of objects are created in the database and included in the current subset. The duplicate of an object is named *Copy of*, for example, *Copy of EPAC*. Multiple duplicates are numbered, for example, *Copy#2 of EPAC*.

Note: Duplicates of objects are not related to the original objects. Therefore, changes that you make to the original object are not made in duplicates.

View object properties

To view object properties:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click an object, then select **Properties** to open the properties window.

Note: If one object is selected, all the named object fields with their values are displayed. Some object data (like sequence and comments) are not stored in named fields and are not displayed in the properties window.

Export data

To export DNA/RNA and protein molecules, enzymes, oligos, Contig Assembly projects, and BLAST results:

1. In the Explorer Viewer, click **Local Database**, then select a database.
2. In the Records Viewer, open the Export Data window:
   - Select an object, then click **File** ▸ **Export**, or
   - Right-click an object, then select **Export**.
3. Navigate to the export location.
4. (*Optional*) Rename the file and/or select a new file type, then click **OK**.

You can export data in the following file formats:

File	Format
DNA/RNA molecule	Genbank (*.gb, *.gbwithparts)
	DDBJ (*.ddbj)
	Fasta (*.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq)
	EMBL (*.embl)
	GCG (*.gcg)
	ASCII (*.txt)
Protein molecule	Fasta (*.fasta, *.fas, *.fa, *.mpfa, *.fna, *.fsa, *.seq)
	GenPept (*.gp, *.gpwithparts)
	SWISS-PROT (*.swp)
	TrEMBL (*.trembl)
	ASCII (*txt)
Enzyme	REBASE (*.rebase)
Oligo	CSV (*.csv)
BLAST results	Tab-delimited (*.txt)
Contig Assembly Project	Contig Project (*.cepx)

Show or hide columns

For all databases in the Local Database, you can show or hide columns in the Records Viewer.

To show or hide columns:

1. In the Explorer Viewer, click **Local Database**, then expand a database.
2. In the Records Viewer, right-click anywhere in the records table, then select **Column**.
3. In the drop-down list, check or uncheck the columns that you want shown or hidden.

Edit data

The operations described in this section apply to the Enzymes, Oligos, and Gel Markers databases in the Local Database. This section describes how to:

- Edit enzyme properties (page 27)
- Edit oligo properties (page 28)
- Edit gel marker properties (page 28)

Edit enzyme properties

To edit an enzyme:

1. In the Explorer Viewer, click **Local Database**, then select **Enzymes**.

2. In the Records Viewer, right-click an enzyme, then select **Edit**.
3. In the enzyme window, click the **Enzyme/Methylase** tab to edit the following properties:

Recognition String	Enter a nucleotide string that can be recognized as the enzyme.
Cleavage Point/Methylation Base	Enter the number of the nucleotide immediately after the direct-strand cleavage point.
Description	Enter or edit the enzyme's description.

4. Click the **User Fields** tab to edit the values of the following fields:
   - Commercial Sources
   - Freezer Position
   - Organism
   - WWW Source
5. Click the **Comments** tab to enter comments about the enzyme.
6. Click the **Keywords** tab to add or remove keywords that will improve database searches.

Edit oligo properties

To edit an oligo:

1. In the Explorer Viewer, click **Local Database**, then select **Oligos**.
2. In the Records Viewer, right-click an oligo, then select **Edit**.
3. In the oligo window, click the **Oligo** tab to edit the following properties:

Nucelotide Sequence	Enter or edit the oligo's nucleotide sequence. Valid characters: ATUCG
DNA/RNA Order	Select a DNA or RNA sequence, then click Order to open your browser and order your primer sequence from Life Technologies Corporation.
Reverse Complementary	Check the box to replace the oligo sequence with the complementary sequence.
Description	Enter or edit the oligo's description.

4. Click the **User Fields** tab to edit the values of the following fields:
   - Commercial Sources
   - Freezer Position
   - Organism
   - WWW Source
5. Click the **Comments** tab to enter comments about the enzyme.
6. Click the **Keywords** tab to add or remove keywords that will improve database searches.

Edit gel marker properties

To edit a gel marker:

1. In the Explorer Viewer, click **Local Database**, then select **Gel Markers**.
2. In the Records Viewer, right-click a gel marker, then select **Edit**.

3. In the General tab of the gel marker window, click the **...** button to select a name for the gel marker.
4. Click the **Gel Marker** tab to edit the following properties:

Fragment	Lists by length all the fragments making up the marker. Click **Add** or **Delete** to add a fragment to the marker, or remove a fragment from the marker.
Fragment (length in b.p.)	Enter the length of the fragment in base pair (bp), then click **Add**.
Description	Enter or edit the gel marker's description.

Open other applications

You can right-click on DNA/RNA and protein molecules in the database and open the following applications:

- Web Analyses (see page 73)
- Analysis Monitor (see Chapter 10 on page 133)
- BLAST Search (see Chapter 6 on page 103
- BioAnnotator (see Chapter 4 on page 93)
- AlignX (see Chapter 8 on page 117)

To open other applications:

1. In the Explorer Viewer, click **Local Database**, then select **DNA/RNA Molecules** or **Protein Molecules**.
2. In the Records Viewer, right-click a molecule, then select an application.

Copy, save, and print molecules

Copy a molecule by taking a screenshot with the Camera tool. Save the screenshot as text or as an image in an application such as Microsoft Word or Microsoft Excel. Print the text or image.

Copy a molecule

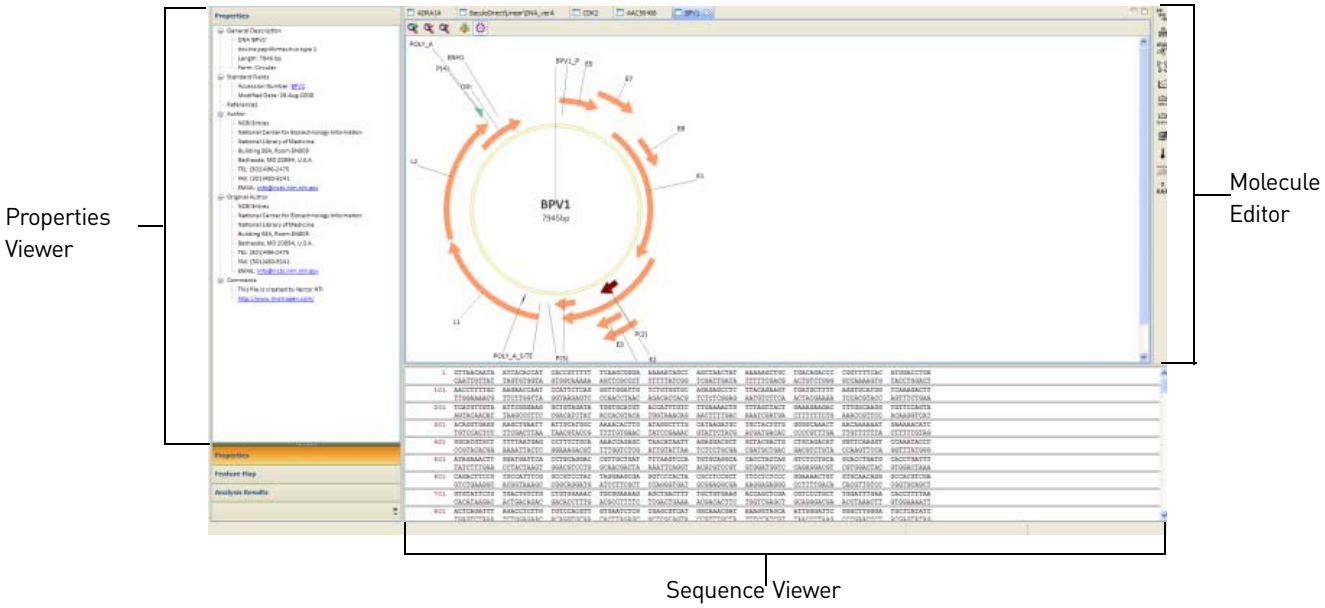
Use the Molecule Editor to copy molecules. In the Database Explorer, open the Molecule Editor by opening a molecule:

1. In the Explorer Viewer, click the **DNA/RNA Molecules** folder or the **Protein Molecules folder** in the Local Database folder to display the molecules in the Records Viewer.
2. In the Records Viewer, double-click a molecule to display a graphical map of a molecule and its sequence.

   The Explorer Viewer is replaced by the Properties Viewer. The Records Viewer is replaced by the Molecule Editor. The Summary Viewer is replaced by the Sequence Viewer.

Note: You can widen, narrow, lengthen, and shorten the size of the viewers and Molecule Editor by clicking and dragging the viewer frames.

For more information about the Molecule Editor, see Chapter 2 on page 39.



Copy molecule properties, feature map, and analysis results

To copy molecule properties, feature map, and analysis results:

1. In the Explorer Viewer, click **Local Database**, then select **DNA/RNA Molecules** or **Protein Molecules**.
2. In the Records Viewer, double-click a molecule to open it.
3. Right-click anywhere in the Properties, Feature Map, or Analysis Results viewers, then select **Camera** to take a screenshot of the data.

   The data is copied to the clipboard and can be pasted in other applications.

Copy molecule image

To copy an image of DNA/RNA molecules and protein molecules:

1. In the Explorer Viewer, click **Local Database**, then select **DNA/RNA Molecules** or **Protein Molecules**.
2. In the Records Viewer, double-click a molecule to open the Molecule Editor.
3. Right-click the graphical depiction, then select **Camera**.
4. Click anywhere in the graphic to take a screenshot. The image is copied to the clipboard and can be pasted in other applications.

Copy a sequence

To copy a sequence:

1. In the Explorer Viewer, click **Local Database**, then select **DNA/RNA Molecules** or **Protein Molecules**.
2. In the Records Viewer, double-click a molecule to open the Molecule Editor.
3. Select a sequence:
   - Click and drag across a sequence region in the molecule image, or
   - Click and drag across the text in the Sequence Viewer.

   Right-click, then select **Copy**

   The data is copied to the clipboard and can be pasted in other applications.

Save a molecule

Molecules you create can be saved to only to the Local Database. To save database molecules:

1. Click **File ▸ Save** in the main menu, or

   Click **File ▸ Save As** to save the molecule with a different name and/or to a different location:

   a. Enter a new name for the molecule.

   b. Select the destination folder for the molecule.

      Optionally save the object as a file. Select the file type:

      – For DNA/RNA molecules: Genbank, Genbank Old Format, EMBL, Fasta, DDBJ, or ASCII text

      – For protein molecules: ASCII text, GenPept, GenPept Old Format, Swiss-Prot, Fasta, or TrEMBL

2. Click **OK**.

Print molecule data

To print molecule data:

1. In the Explorer Viewer, click **Local Database**, then select **DNA/RNA Molecules** or **Protein Molecules**.
2. In the Records Viewer, double-click a molecule to open it.
3. Right-click in the Properties Viewer, Molecule Editor, or Sequence Viewer, then select **Camera** to take a screenshot.
4. Open an application such as Microsoft Word, Excel, Paint, or Notepad and paste the screenshot.
5. Print the data from the application.

Manage the Projects database

You can perform the following operations to manage the Projects database:

- Create a project (page 32)
- Open a project (page 32)
- Manage Projects subsets (page 32)
- Delete Projects subsets (page 33)

Create a project

To create a project:

1. In the main menu, click **File** ▸ **New**, then select the type of project you want to create.
   - Alignment
   - ContigExpress
   - TOPO Cloning
   - Gateway Cloning
   - Clone2Seq
   - Parts Assembler
2. In the project window, enter a project name and description, then click **New**.

Open a project

To open a project:

1. Click the **Projects** bar.
2. In the Projects Viewer, expand the Local Database.
3. Select the folder containing the type of project you want to open.
4. In the Records Viewer, double-click a project to open the project application (for example, Alignment) and view its properties.

Manage Projects subsets

To manage subsets in the Projects database:

1. Click the **Projects** bar.
2. In the Projects Viewer, expand the Local Database, then expand a project.
3. Select the subset you want to manage.
4. Right-click a subset, then select a managerial operation. For more information, see "Manage Database, Projects, and Results subsets" on page 22.

Delete Projects subsets

You can delete *subsets* of Alignment, Contig Assembly, or Cloning projects. You cannot delete the Alignment, Contig Assembly, and Cloning folders in the Local Database.

To *permanently* delete the **subset** of an Alignment, Contig Assembly, or Cloning project from the database:

1. Click the **Projects** bar.
2. In the Projects Viewer, expand the Local Database.
3. Select the folder containing the type of project you want to delete.
4. Right-click, then select **Dismiss subsets**.

Manage the Results database

You can view BLAST results and analysis results and manage the BLAST Results and Analysis Results databases.

View results

To view BLAST results and analysis results:

1. Click the **Results** bar.
2. In the Results Viewer, expand the Local Database, then select **BLAST Results** or **Analysis Results**.
3. In the Records Viewer, double-click a result to open the BLAST Viewer or Molecule Editor.

Manage Results subsets

To manage subsets in the Results databasse:

1. Click the **Results** bar.
2. In the Projects Viewer, expand the Local Database.
3. Right-click a subset, then select a managerial operation. For more information, see "Manage Database, Projects, and Results subsets" on page 22

Workgroup Shared Database

This section describes how to:

About workgroup shared databases

You can create special databases, repositories of DNA/RNA or protein molecules, enzymes, oligonucleotides, and gel markers and share them among several Vector NTI Express Software users on a network. Workgroup shared databases are not a replacement for local databases; each Vector NTI Express Software application still must have its own local database. The local database is used in all operations, for example, the construction, design, and creation of viewers. The local database is also the place for storing private and temporary data. The main purpose of the workgroup shared database is to store common data.

The only operations you can perform on workgroup shared databases are:

- Copying data to and from the local database. For instance, you can copy some of your molecules and enzymes from your local database into the workgroup shared database. In order to use them in the design process, your colleague must first copy them to his or her local database.
- Performing various database management operations such as creating and deleting subsets.
- Running database searches.

In addition to biological data, each database contains information about its creator and registered users. Only the creator and registered users can have access to database data. The database creator can also change database properties, remove registered users, and define the password required to become a register user.

Note: Vector NTI™ *Express* workgroup shared databases use portable data format (PDF) and file naming conventions to ensure that both Macintosh and Windows users of Vector NTI Express Software can access common databases. Workgroup shared databases can be located on a wide range of file servers.Vector NTI Express Software can use not only services native to each system (Microsoft Network or AppleTalk), but also various Unix (NFS or Samba) and NetWare services.

The workgroup shared database capability is a purchased addition to Vector NTI Express Software. When you purchase the workgroup shared database capability, you are issued a Vector NTI Express Software-workgroup shared database license that enables you to create workgroup shared databases. A workgroup shared database license is a special type of static license that enables you to create numerous workgroup shared databases. But the license also limits the number of users for each database you create. You do not need a workgroup shared database license to access workgroup shared databases.

Note: Workgroup shared databases can be accessed from Vector NTI Express Software using a license that is shared through a network server (Dynamic License).

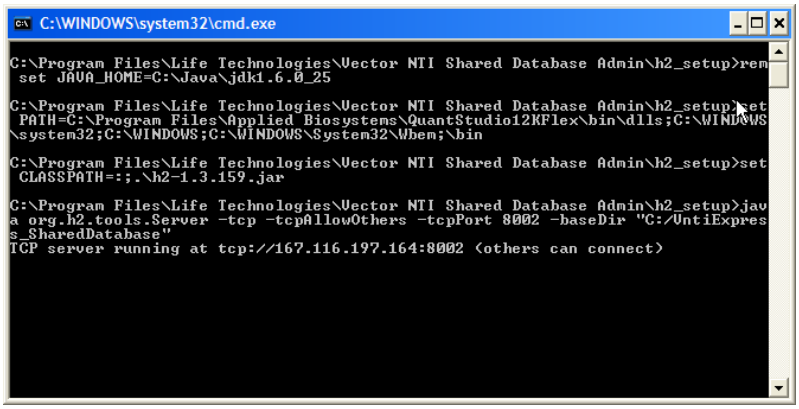
Connect to a workgroup shared database

You can connect to a workgroup shared database through only a hostname and IP address, not via the World Wide Web.

When the network directory for a new workgroup shared database is arranged, connect to a workgroup shared database. Before you can connect to a workgroup shared database, the workgroup shared database server must be started.

Start the workgroup shared database server

In *C:\Program Files\Life Technologies\Vector NTI Workgroup Shared Database Admin\h2_setup*, double-click **startSharedServer** and open the cmd.exe window:



Note: The cmd.exe window automatically opens after the installation of Shared Database Admin.

Connect to the workgroup share database server

1. Open Vector NTI™ *Express* Software, then:
   - In the main menu, click **Tools** ▸ **Connect to workgroup shared database**, or
   - Open the Database Explorer. In the Explorer Viewer, click **Workgroup Shared Database**.
2. In the Connect to Workgroup Shared Database window, log on:

- Host Name: **localhost**
- User Name: *Your user name*
- Password *Your user password*

Note: (Host Name field) If the Workgroup Shared Database Admin application and Vector NTI™ *Express* Software are not installed on the same computer, enter the IP address of the computer on which the Shared Database Admin application is installed. The IP address is shown in the cmd.exe window:

3. Click **Connect**.
4. Close the Connect to Workgroup Shared Database window. Your connection will not be broken.

Add users to a workgroup shared database

You must be an administrator to add users to a workgroup shared database.

To add users to a workgroup shared database:

1. In *C:\Program Files\Life Technologies\Vector NTI Workgroup Shared Database Admin\h2_setup*, double-click **startSharedServer** to start the server.
2. Double-click the Shared Database Administration icon on your desktop to open the Login window.
3. Log in:
   - Database Hostname/IP: **localhost**
   - Administrator username: **admin**
   - Password: **admin**

4. Click **Login** to open the Shared Database Administrator window.
5. Click **Add**, then enter information in the Add User window.

A user name should be one word without spaces.

Edit users in a workgroup shared database

To edit users in a workgroup shared database:

1. Open the Workgroup Shared Database Admin application and log in as an administrator.
2. In the Workgroup Shared Database Administrator window, double-click a user to open the Edit User window.
3. Edit the user information, then click **OK** to save your edits.

Delete users from a workgroup shared database

To delete users from a workgroup shared database:

1. Open the Workgroup Shared Database Admin application and log in as an administrator.
2. In the Workgroup Shared Database Administrator window, select a user, click **Delete**, then **OK**.

Upload data to a workgroup shared database

To upload data from the Local Database to a workgroup shared database:

1. Connect to a workgroup shared database.
2. In the Explorer Viewer, click **Local Database**, then select a subset.
3. In the Records Viewer, right-click an object, then select **Upload to workgroup shared database.**

Download data from a workgroup shared database

To download data from a workgroup shared database to the Local Database:

1. Connect to a workgroup shared database.
2. In the Explorer Viewer, click **Workgroup Shared Database**, then select a subset.
3. In the Records Viewer, right-click an object, then select **Upload to local database.**

Disconnect from a workgroup shared database

To disconnect from a workgroup shared database:

1. 1. Click **Tools** in the main menu, then **Connect to workgroup shared database**.
2. 2. In the Connect to Workgroup Shared Database window, click **Disconnect**.

Set preferences

You can set preferences for displaying molecules and sequences and register your email address with the National Center for Biotechnology Information (NCBI).

Set display preferences for molecules

To set display preferences for molecules:

1. In the main menu, click **Tools** ▸ **Preferences** to open the Preferences window.

2. In the Display Preference view, click the background color and highlight color buttons to open the Color palette.
3. Select background and highlight colors, then click **OK**.
4. To save your preferences, click **Apply**, or click **Restore Defaults** to keep the original settings.

Set display preferences for sequences

To set display preferences for sequences:

1. Expand Display Preference, then click **Sequence Preference**.
2. Click Sequence fill color and Sequence line color to open the Color palette and select colors.
3. Enter values for the sequence width and sequence line width parameters.
4. To change the title font, click the **Change** button, then select and define a font.
5. To change the title font color, click the button to open the Color palette and select a color.
6. Click **Apply** to save your preferences.

Register with the NCBI

To register your email address with the NCBI:

1. Click **NCBI Configuration**.
2. Enter your email address, then click Register to open your email application.
3. Click **OK** to close the Preferences window.

2 Molecule Editor

The Molecule Editor in Vector NTI™ *Express* displays a graphical representation of a DNA or protein molecule, the molecule sequence, information about the molecule, and the results of any analysis performed on the molecule.

Using the Molecule Editor, you can:

- Create and edit a molecule sequence
- Annotate the molecule
- Analyze the molecule
  - For DNA/RNA molecules, you can perform restriction analysis, design primers from the sequence, find ORFs, translate the sequence into amino acids, and other functions.
  - For protein molecules, you can scan for motifs, perform web searches on the sequence, perform back translation, and other functions.
  - Molecule sequences may also be analyzed using a variety of web tools and publicly available analysis engines
- Display the analysis results

Create or open a molecule

Create a new molecule

1. To create a new molecule, click on **File > New** and select **DNA** or **Protein** as the molecule type.

Note: For DNA or RNA molecules, specify **Linear** or **Circular**.

2. The blank molecule will be displayed in the Molecule Editor but will not yet be saved in the database.

Open an existing molecule

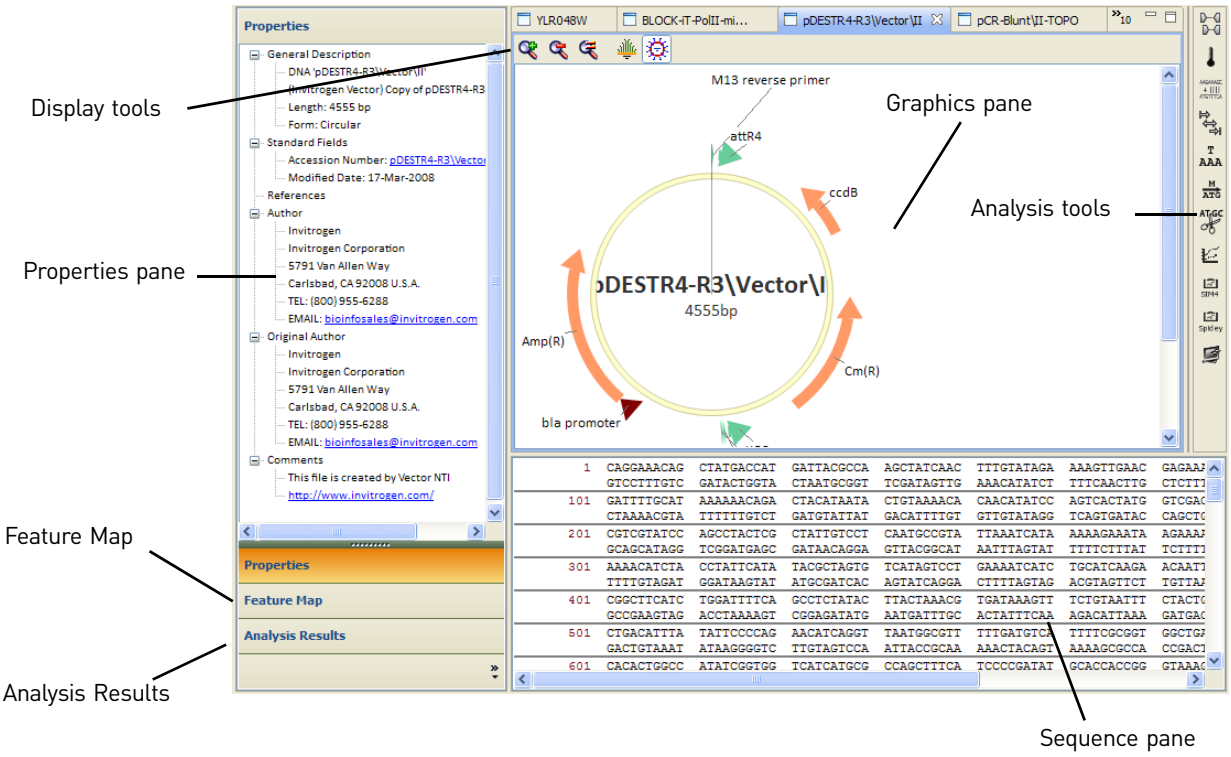
To open an existing molecule:

- For molecules in the database, double-click on the molecule name in the Database Explorer window
- For molecules that are not in the database, select **File > Open** and select a molecule of the appropriate file type (.gb, .fasta, etc.)

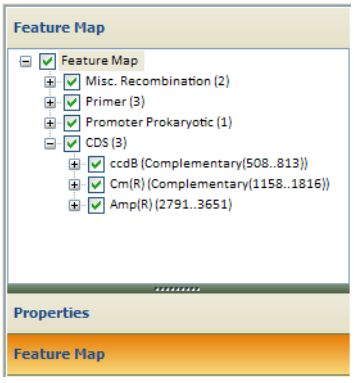
The molecule will be displayed in the Molecule Editor.

Molecule Editor window

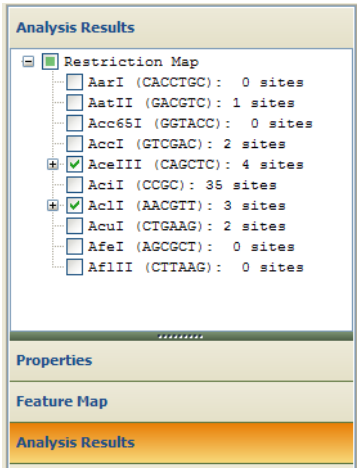
The Molecule Editor window consists of five main panes as well as various tools and analysis functions:



- The **Graphics pane** displays a graphical representation of the molecule, including features.
- The **Sequence pane** displays the molecule sequence, as well as the result
- The **Properties pane** displays information about the molecule.
- The **Feature Map** displays a list of the defined features in the molecule.



- The **Analysis Results** displays analyses that have been performed on the molecule.



The **Analysis tools** toolbar contains analysis functions specific to the Molecule Editor.

The **Display tools** toolbar contains tools to magnify the molecule, or display a DNA sequence as linear or circular.

Molecule display tools

Use the tools at the top of the window to zoom in and out on the molecule, fit the molecule within the window, or display DNA/RNA molecules as linear or circular.

Display Profile

Vector NTI™ *Express* Software uses Display Setup parameters to create the Molecule Display window contents. You can edit the Display Setup parameters before or after a display window is created, assign names to different settings and store them as a Display Profile, and change the default Setup settings.

The Display Profile is a way of viewing a molecule. You needn't bind a molecule with a particular profile. Instead, you can create a molecule short cut by opening a molecule with a desired profile.

The Vector NTI™ *Express* Software has pre-defined display profiles. You can edit or create your own display profile.

You can edit a display profile in two ways:

- From Preferences, at a system level, that can be reused across molecules
- In the Molecule Editor, for a specific molecule

Edit a display profile at a system level

To edit a display profile at a system level:

1. Go to **Tools** ▸ **Preferences**.
2. Select **Display Setup**.

Note: Select the Prompt for Display Setup on Opening Molecules check box to enable accessing the display setup before opening a molecule in the Molecule Editor.

3. Expand the Display Setup preference. Select from:
   - **DNA/RNA Molecules** to view the DNA/RNA Molecules Display Setup Configuration.
   - **Protein Molecules** to view the Protein Molecules Display Setup Configuration.
4. Click one or more check boxes to include the respective Setup Options in the Display Profile.

   Note: Click on each of the setup options below for more information in editing that setup option.

   a. For DNA/ RNA Molecules, you can choose from:
      - Feature Map
      - Sequence
      - Restriction Map
      - ORFs
      - Auto Load Analysis
      - Motifs Setup
      - Graphic Display
      - Picture type

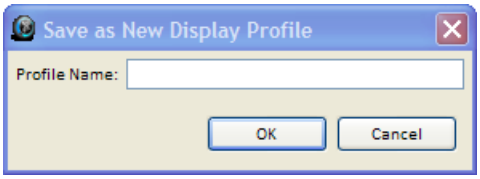
   b. For Protein Molecules, you can choose from:
      - Feature Map
      - Sequence
      - Auto Load Analysis
      - Motifs Setup
      - Graphic Display

   Note: Sequences and Graphic Display are mandatory setup items and cannot be deselected.
5. Once you have made the desired edits in each of the Setup Options, click **Save to file** to save the newly created display profile as an .xml file that can be accessed from anywhere.

   Click **Load from file...** to retrieve this display profile from the location you saved the .xml file.

   Note: We recommend that you avoid opening a Display Profile Settings file (*.ms5) saved from Vector NTI Advance® Software with Vector NTI™ *Express* Software.

6. Click **Save As** to save the newly created display profile.



   a. Enter a name in the Profile Name field in the Save as New Display Profile dialog box.

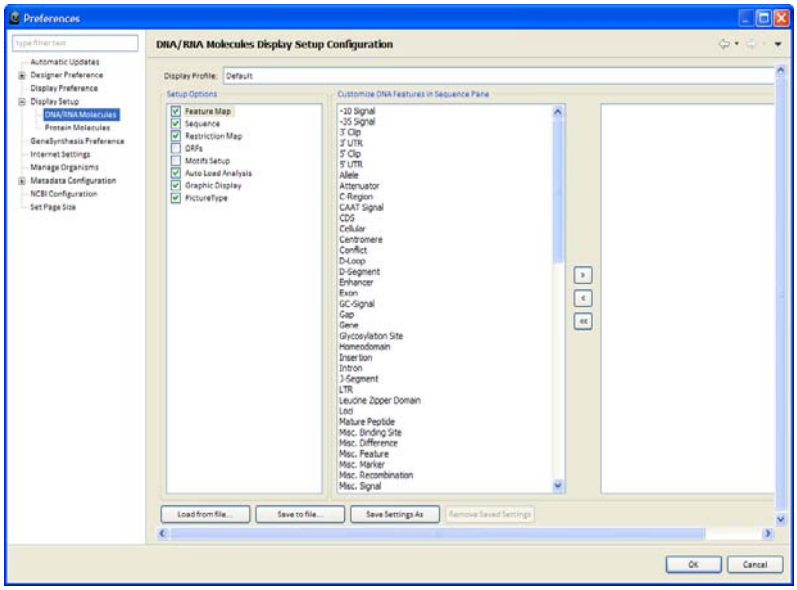
b. Click **OK**.

The newly created display profile will then appear in the Display Profile drop-down menu.

7. Click **OK** to submit the change or **Cancel** to exit the Preferences dialog box.

   Note: To remove a selected profile from the system, click **Remove Saved Settings**.

The Preferences dialog box for editing the display profile at the system level looks like this:



Edit a display profile for a specific molecule

To edit a display profile for a specific molecule:

1. Open a DNA/RNA molecule or Protein molecule in the Molecule Editor.
2. In the Molecule Editor toolbar, click  to open the DNA/RNA Molecules Display Setup dialog box.

   Note: If you are editing the display profile of a protein molecule, clicking on  will open the Protein Molecules Display Setup dialog box.

   Note: Clicking on the drop-down arrow on the above icon will allow you to choose from the existing system profiles:

   - Default
   - Green
   - Classic
   - Orange

3. You can also open the respective Display Setup dialog box by right-clicking on the molecule in the Molecule Editor and selecting Display Setup... from the drop-down menu.
4. Edit the Setup Options. Click on the options you want to include.

   Note: Click on each of the setup options for more information.

   a. For DNA/ RNA Molecules, you can choose from:
      - Feature Map
      - Sequence

- Restriction Map
- ORFs
- Auto Load Analysis
- Motifs Setup
- Graphic Display
- Picture type

b. For Protein Molecules, you can choose from:

- Feature Map
- Sequence
- Auto Load Analysis
- Motifs Setup
- Graphic Display

Note: Sequences and Graphic Display are mandatory setup items and cannot be deselected.

5. Click **Save to file** to save the newly created display profile as an .xml file that can be accessed from anywhere.

   Click **Load from file...** to retrieve this display profile from the location you saved the .xml file.

   Note: Display Profile Settings file (*.ms5) saved from Vector NTI Advance® cannot be loaded using the Vector NTI™ *Express* Software.

6. Click **Save As** to save the newly created display profile.

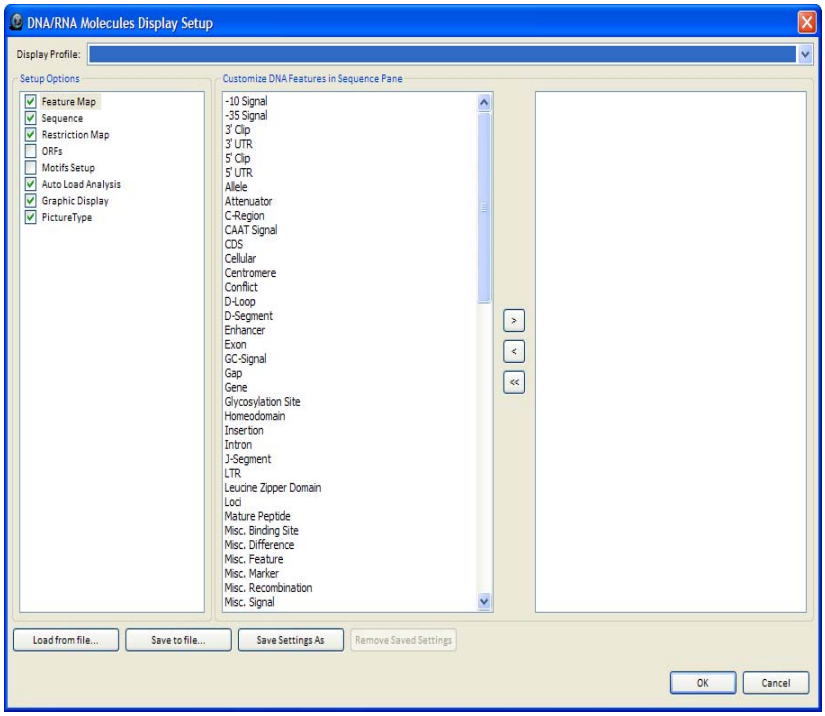
   a. Enter a name in the Profile Name field in the Save as New Display Profile dialog box.

   b. Click **OK**.

   The newly created display profile will then appear in the Display Profile drop-down menu.

7. Click **OK** to submit the change or **Cancel** to exit the Display Setup dialog box.

The Display Setup dialog box for editing the display profile in the Molecule Editor looks like this:



Display Profile setup options

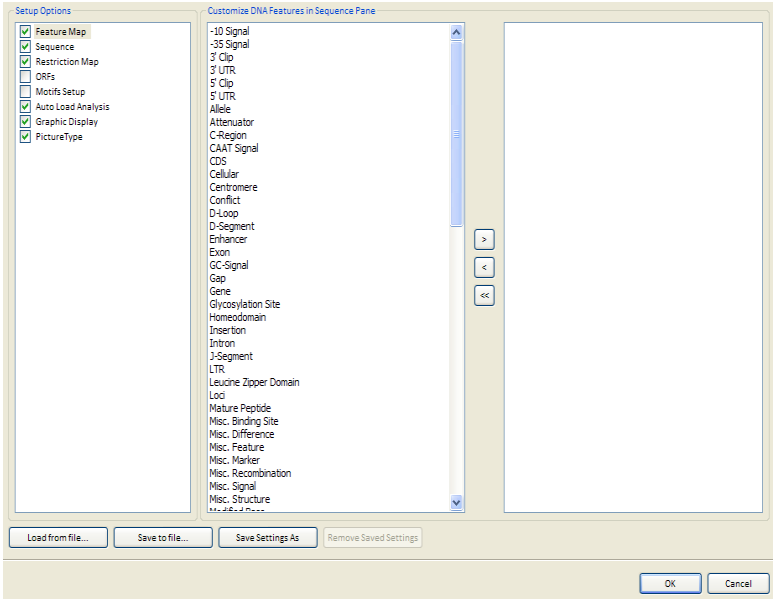
Feature Map

1. Check the **Feature Map** check box.
2. Customize the DNA features/ Protein features in the Sequence Pane by including one or more of the options available in the Customize DNA features in Sequence Pane (or Customize Protein Features in Sequence Pane) list.

   Note: The Feature Types available in the system are on the left side and the features you would like to be annotate in the sequence pane will be available on the right side.
3. Click [>] to add the DNA features and [<] to remove the DNA features.

   Note: Click [<<] to remove all the features in a single click.
4. Click **OK** to submit the Feature Map Setup Option edits or **Cancel** to exit the Preferences dialog box.

The Feature Map Setup dialog box looks like this:

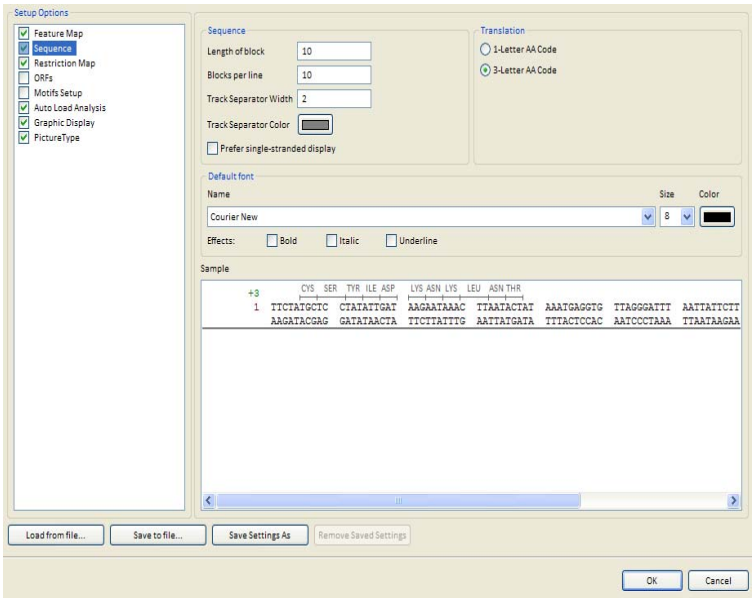


Sequence

1. Check the **Sequence** check box.
2. Enter the Sequence details:
   a. Length of block
   b. Blocks per line
   c. Track Separator Width
   d. Track Separator Color: Double-click the color button to change the default display.
   e. Check the Prefer single-stranded display check box if you want the single-stranded display.
3. Select the Translation Code you want to include. You can choose from:
   - **1-Letter AA Code**
   - **3-Letter AA Code**
4. Set the Default Font:
   - Select the Font name and size from the respective drop-down list.
   - Select the color from the color palette.
   - Select the respective check box for including the **Bold**, **Italic**, and **Underline** Effects.

   You can view the edits as you make them in the Sample pane in the lower half of the Sequence Setup Option dialog box.
5. Click **OK** to submit the Sequence Setup Option edits or **Cancel** to exit the Preferences dialog box.

The Sequence Setup dialog box looks like this:



Restriction Map

1. Check the **Restriction Map** check box.
2. Configure the Restriction Enzymes (RENs):
   a. Click **Configure List**.
   b. In the Select Restriction Enzymes dialog box, click the check box of one or more available enzymes.

      Note: Check the Select/ Deselect All check box to select or deselect, respectively, the available enzymes in one click.

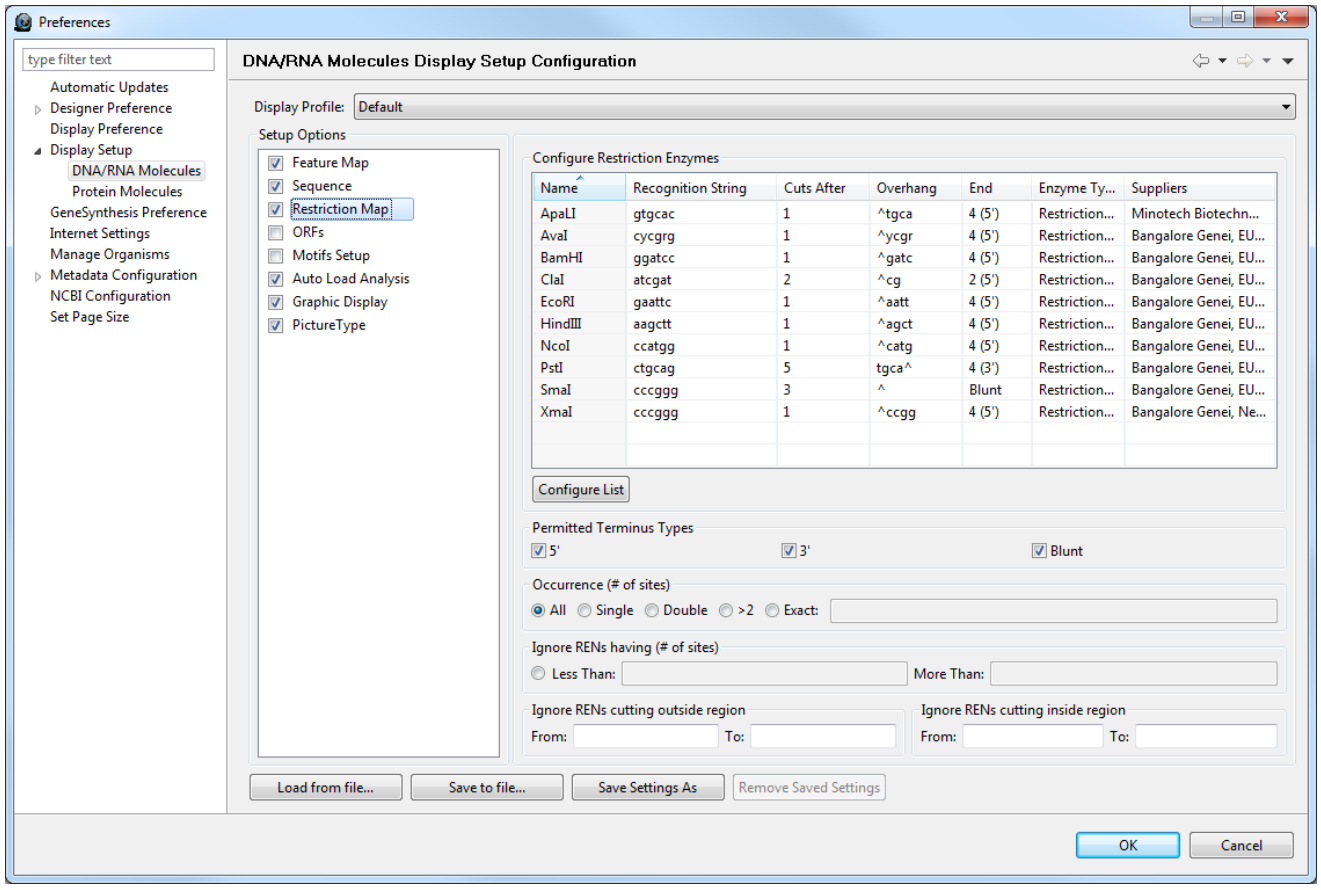
   c. Use the [>] button to add the available enzymes to the Select Enzymes pane. Use the [<] button to remove the selected enzymes to the Available Enzymes pane.

      Note: Click [<<] to remove all the selected enzymes in a single click.

   d. Click **OK** to confirm the edit or **Cancel** to exit the Select Restriction Enzymes dialog box.
3. Select the Permitted Terminus Types check boxes to filter the types of RENs displayed. You can select from:
   - 5′
   - 3′
   - Blunt
4. Enter a numerical value in the Ignore RENs having ... less than/ ... more than sites field.

   The feature removes from the restriction map RENs that do not fall within the specified cut site range. Such RENs are listed but grayed out in the Restriction Map folder in the Text Pane. They are not displayed in the Graphics and Sequence Panes.
5. Click **OK** to submit the Restriction Map Setup Option edits or **Cancel** to exit the Preferences dialog box.

The Restriction Map Setup dialog box looks like this:



ORFs

1. Check the **ORFs** check box.
2. Configure the ORF Parameters for specifying how the open reading frames (ORFs) are displayed:
   a. In **Start Codons** and **Stop Codons** fields, enter start and stop codons for the new ORFs.
   b. Click the **Default Start & Stop** button to set the start and stop codons to the following conventional values:
      - Start codons—ATG, GTG
      - Stop codons—TAA, TGA, TAG.
   c. Check the **Include Stop Codon in ORF** box if you want the stop codon to be considered as a part of the ORF. Otherwise, the stop codon is not considered as part of the ORF and is not included.
3. Select the ORF Type.

- **Complete**: The Complete check box is selected by default.
  a. In the Minimum Size field, specify the minimum size in base pairs. The default value is 150 and can be changed as per requirement.
  b. Check the Nested check box to look for nested ORFs. These are ORFs that occur within the main ORF.

- **Incomplete:**
  a. Check the **Incomplete** check box and in the **Minimum Size** field specify the minimum size in base pairs to display the incomplete ORF.
  b. Check the **Undefined Start/Stop** check box to search for ORFs with undefined starts, stops or both.
  c. Check the **Nested** check box to look for nested ORFs.

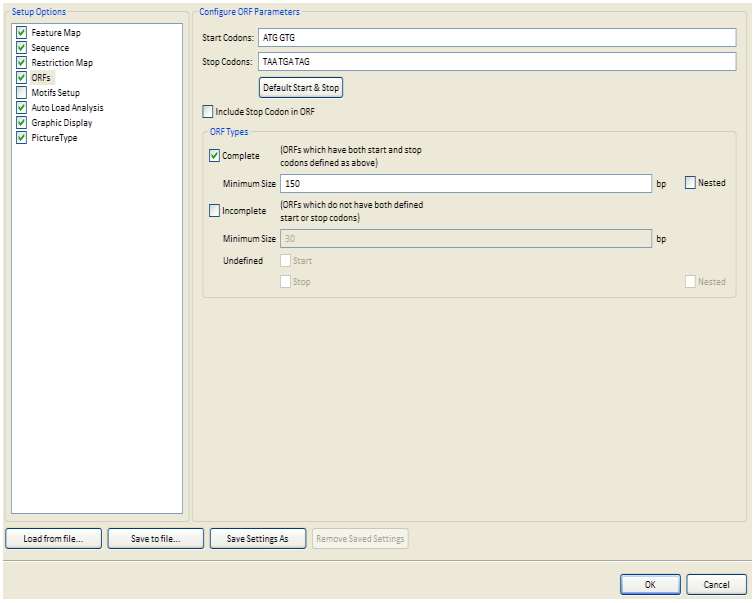
  Note: An ORF with an undefined start is one that has a stop codon at the end but no corresponding start codon in the same frame.

  An ORF with an undefined stop is one that has a start codon but no stop codon in the same frame.

  Incomplete ORFs are displayed as a dashed arrow in both the Graphics and the Sequence Pane. Complete ORFs are displayed as solid arrows.

4. Select the Permitted Terminus Types check boxes to filter the types of RENs displayed. You can select from:
   - 5′
   - 3′
   - Blunt
5. Click **OK** to submit the ORFs Setup Option edits or **Cancel** to exit the Preferences dialog box.

The ORFs Setup dialog box looks like this:



Auto Load Analysis

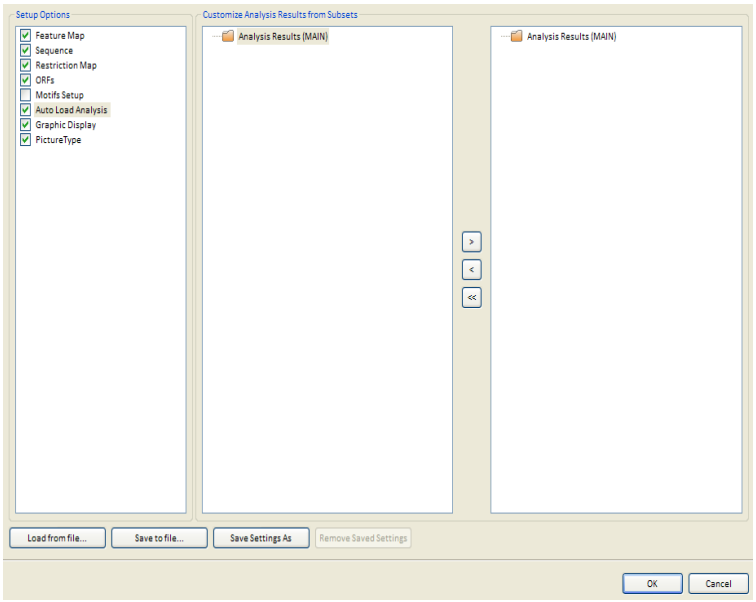
1. Check the **Auto Load Analysis** check box to specify the analyses to be shown in the Sequence and Graphics Panes.
2. From the Available Analyses Results Subsets drop-down box, choose the Vector NTI Database subset of saved analyses you want to display.

3. Click the > button to move the selected subset into the Load Analyses Results from Subset section. All analyses listed in this subset are shown in the Molecule Viewing window. Use the < button to remove the selected subset.

   Note: Click << to remove all the selected subsets in a single click.

4. Click **OK** to submit the Auto Load Analysis Setup Option edits or **Cancel** to exit the Preferences dialog box.

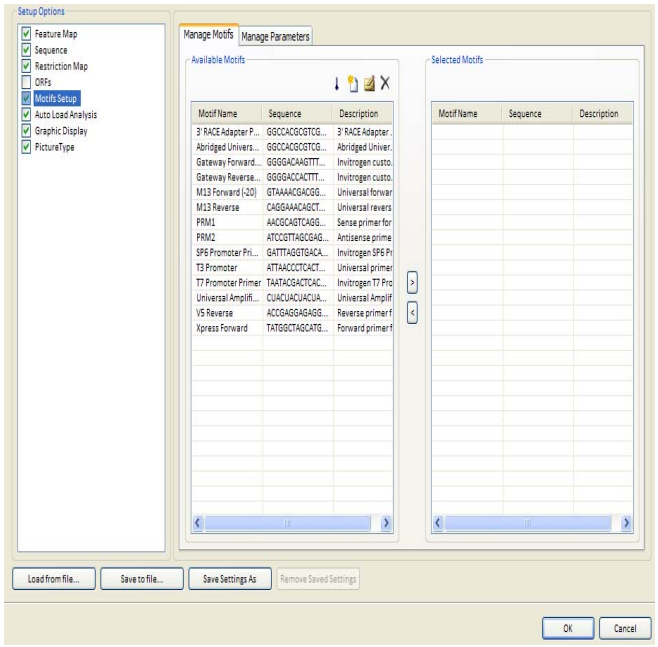
The Auto Load Analysis Setup dialog box looks like this:



Motifs Setup

1. Check the **Motif Setup** check box to display the motifs map.
   The motif maps display has two tabs: Manage Motifs and Manage Parameters
2. Click **Manage Motifs** to view the Manage motifs display.
   The Manage Motifs display includes the following panes:

- **Available Motifs**: includes motifs already chosen for a search
- **Selected Motifs**: includes motifs moved from the Available Motifs pane
- **Motifs Properties**: displays properties of the motif selected in the Available Motifs pane



Note: Except for General, all other tabs in the Motif Properties pane appear disabled when a motif is selected in the Available Motifs pane. Only when you use the Add and Edit tools, to add or edit motif respectively, the other tabs are enabled. For instructions on using the other features in the Motif Properties pane, refer to "Create an oligo" on page 24.

a. In the Manage Motifs display, move motifs from the Available Motifs pane to the Selected Motifs pane using the > button.

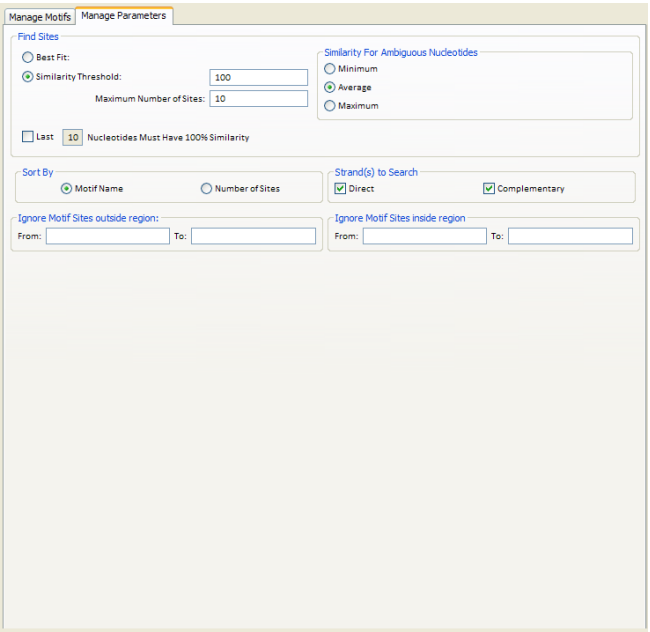
   You can move back a motif from the Selected Motifs pane to the Available Motifs pane using the < button.

b. In the Available Motifs pane, click
   - to add new motifs
   - to edit selected motifs
   - to initiate an analysis of a selected motif
   - to delete a selected motif

3. Click **Manage Parameters** to view the Manage Parameters display.

   The Manage Parameters display includes the following panes:

- **Find Sites**
- **Similarity For Ambiguous Nucleotides**
- **Sort By**
- **Strand(s) to Search**
- **Ignore Motif Sites outside region**
- **Ignore Motif Sites inside region**



a. In the Find Sites pane, specify the similarity between motif and molecule sequence for the motif's site to be accepted. Choose between the Best Fit and Similarity Threshold radio buttons.
   - **Best Fit** selects only those motif sites that received the best relative score.
   - **Similarity Threshold** displays all motif sites that have a similarity more than or equal to the similarity you specify in the Similarity Threshold box.

- To filter out all the motif sites that do not meet the specific requirement for 3′ end similarity, check the **Last** box and enter the number of nucleotides on 3′ end that should have 100% similarity.

Note: Any motifs that have more sites than specified in the Maximum Number of Sites box are displayed as grayed-out folders in the Text Pane; they are not displayed at all in graphics or Sequence Panes.

b. In the **Similarity For Ambiguous Nucleotides** pane, specify the acceptable similarity between ambiguous nucleotides (if any).

Select from **Minimum**, **Average**, and **Maximum** radio buttons. The buttons indicate the minimum, average, and maximum possible similarity that will be calculated respectively for any nucleotide pair. For instance, if you are calculating similarity between N and A, then the average similarity is 25%, the minimum similarity is 0%, and the maximum similarity is 100%. In case of R and A they are 50%, 0%, and 100%; in case of R and T—0%, 0%, and 0%. *For clarification of these IUB nucleotide symbols, refer Appendix A.*

	N-N	N-R	N-A	R-W	R-A	R-T
Maximum	100	100	100	100	100	0
Average	25	25	25	25	50	0
Minimum	0	0	0	0	0	0

c. In the **Sort By** pane, select the motifs sorting procedure to be used for the Motifs folder. Choose between the **Motif Name** and **Number of Sites** radio buttons. Motif Name sorts motifs alphabetically while Number of Sites sorts motifs by their number of sites on the molecule.

d. In the **Strand(s) to Search pane,** select or deselect the Direct and/ or Complementary check boxes to specify the strand(s) to be searched.

e. In the **Ignore Motif Sites outside region/ Ignore Motif Sites inside region** panes, specify additional limitations on the motif sites to be displayed.

- Specify the **Ignore Motifs Sites outside region** to display only those motifs located inside the region.
- Specify the **Ignore Motifs Sites inside region** to display only those motifs located outside the region.

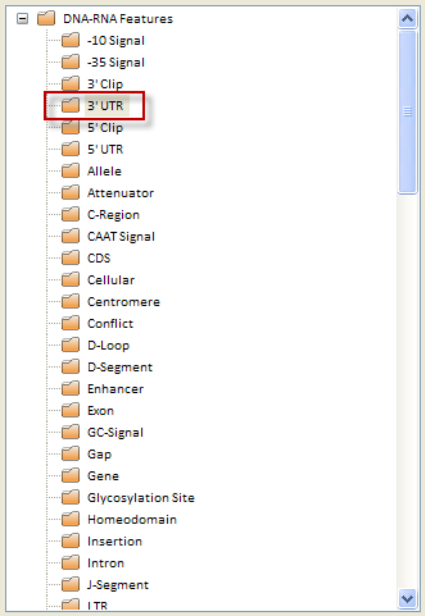
4. Click **OK** to submit the Motifs Setup Option edits or **Cancel** to exit the Preferences dialog box.

Graphic Display

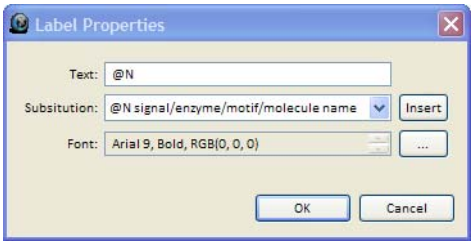
1. Check the **Graphic Display** check box to view the Graphics display editing features.

   You can edit the graphical styles for graphical map elements directly from the Display Setup dialog box (within Preferences) or in the Graphics Pane in the Graphic Editing Mode.

2. In the center pane, elect the DNA-RNA (or Protein) feature to be edited.

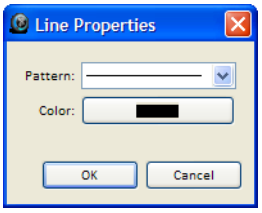


3. In the right hand side pane, select the Label from the Style drop-down menu. To add a new label, click **More...**



- In the Label Properties dialog box, enter the Text
- Select the Substitution from the drop-down menu and click **Insert**
- Click **...** to edit the font.
- Click **OK**

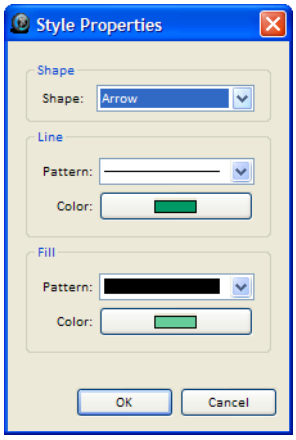
4. Select the Connector from the Style drop-down menu. To add a new connector, click **More...**



- In the Line Properties dialog box, select a pattern from the drop-down menu
- Select a color from the Color palette

- Click **OK**

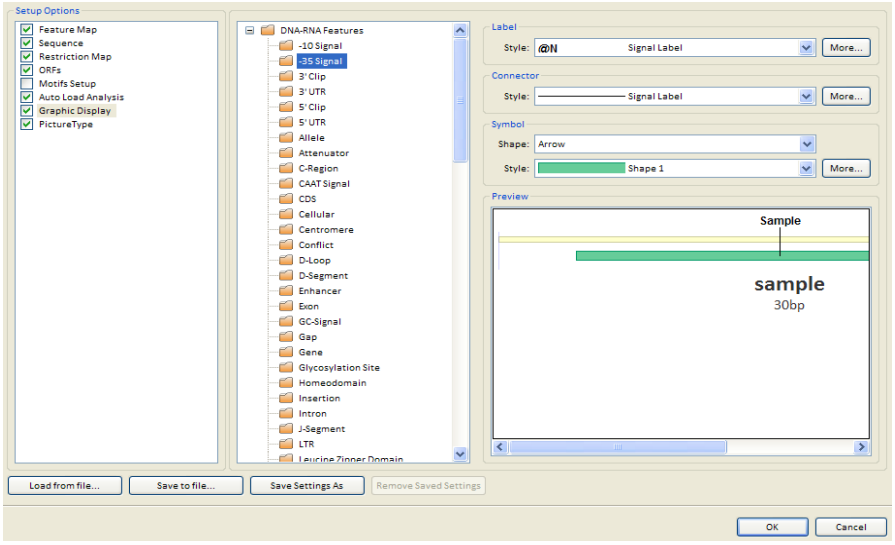
5. Select a Symbol from the Shape and Style drop-down menus. To add a new style, click **More...**



- In the Style Properties dialog box, set the Shape, Line, and Fill properties
- Click **OK**

6. Click **OK** to submit the Graphic Display Setup Option edits or **Cancel** to exit the Preferences dialog box.

The Graphic Display Setup dialog box looks like this:



Graphic Edit Mode: Formatting and Annotating Pictures

Vector NTI™ *Express* Software has two modes of operation in Graphics Panes.

- Molecule Editing mode (default) in which the Graphics Pane serves as a visual environment for viewing and *editing the content* of a molecule itself. In molecule editing, you are changing the nature of the molecule map and sequence. Selecting DNA molecule fragments for Construction and Design also take place in Molecule Editing mode.
- Graphic Editing mode where you can *edit the graphical representation* of an image such as format and arrange individual graphics symbols and labels. This mode is useful for preparing publication-quality figures of the molecule.

When in Graphic Editing mode, Vector NTI™ *Express* Software changes the way it selects objects in the Graphics Pane: the elementary units of selection are graphics objects like labels and symbols, not elements of the molecule content. In Graphic Editing mode, you can reposition and reformat labels and symbols, hide objects to prevent them from being displayed, and format multiple objects at once using centralized style sheets.

Note:

- If you intend to edit the molecule itself, it is recommended that you perform that first before you format molecule graphics.
- Graphic editing changes are not saved in database, but in a Molecule Document file.

To activate Graphic Editing Mode, in an active Graphics Pane, click **Graphic Edit** () on the toolbar.

All of the following operations are performed in Graphic Editing mode.

Formatting Objects on the Graphical Map

1. To format an object in the Graphics Pane, select the object by clicking on it or its label.

   Selected objects in Graphic Edit mode are displayed with selection boxes or handles at each corner. If this is not the case, check to make sure the **Edit Graphic** is depressed.

2. When the graphics objects to be formatted are selected, select **Properties** from the associated shortcut menu.

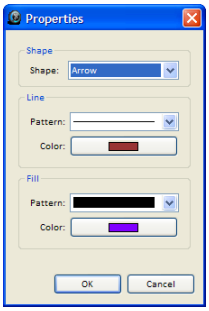
3. In the Properties dialog box, select **Line**, **Fill**, or **Text** attributes (if a label is selected). If a single symbol is selected, you may also change its **Shape**.

4. On the Properties tabs, select the pattern and color for Lines and Fill for selected graphics objects. On the Shape tab, click the down arrow for selections. Click **OK**.

Note:

- **Lines** associated with symbols are the lines that surround the main symbol. Lines associated with labels are the droplines connecting the labels to its corresponding symbol.
- **Fill** refers to the pattern and color with which the selected symbols are filled. Changing a symbol's filling does not affect the formatting of the line around that symbol.

  The Default **Shape** option forces Vector NTI to assign the default shape to the given symbol. You can edit the association between the element type and the default shape/style in the Graphics Display Setup dialog box accessible from the Display Setup dialog box. You cannot undo the Shape formatting.



If you have selected label(s), the Properties dialog box includes a Text field.

5. In the text box, enter ordinary text or choose from a variety of special codes that are converted into information about the symbol to which the label is attached. To insert a substitution code, choose the code from the Substitutions box and click **Insert**. The code is inserted at the current caret position.

   This substitution model allows assignment of a single text "style" to all labels of a certain kind. For example, if the text attribute for all restriction sites is "@N (@S)" then the actual label for a *Sma*I site 5354 bp will look like "*Sma*I (5354)" with the REN name and site position substituted for @N and @S. You might want to experiment with the different substitution models.

   Using the **Font** button on this tab, choose the font, font style, font size, effects, and display color attributes of labels you selected. *Attributes of currently selected font are displayed to the right of the Font button.*

   Note: The font size you choose is not necessarily equal to the font size on the screen or on the printed output. Displayed objects are stretched and shrunk as the picture magnification is changed.

Annotations

Annotations are elements of the graphical representation of the molecule only and are not stored in the database. To save an annotation, you need to save the display window as a Molecule Document. To associate text information with a molecule itself, not its temporary graphical representation, use the Comments or User Fields described in Chapter 1.

1. To insert a text annotation on a graphical map, click **Add Annotation** ( ) in the toolbar.
2. In the Annotation dialog box, enter the text of a new annotation label. Click **OK** to return to the Molecule Display window where the annotation is highlighted and selected with handles.
3. Click and drag it to the location you choose in the Graphics Pane.

   You can move annotations and change their font and formatting using the label formatting techniques described above.

Deleting Annotations

1. To delete an annotation, select it in the Graphics Pane, and select **Edit ▸ Delete Annotation**.
2. To delete an annotation from a Text Pane, click the annotation subfolder to select its line, right-click, and select **Delete Annotation** ( ) or select **Edit ▸ Delete Annotation**.

Note: Annotation subfolders are the only folders of the Text Pane that can be deleted.

Undo and Redo

You can undo/redo movement and formatting of graphics objects, redefinition of styles, etc. Undo/Redo also works in the Sequence Pane.

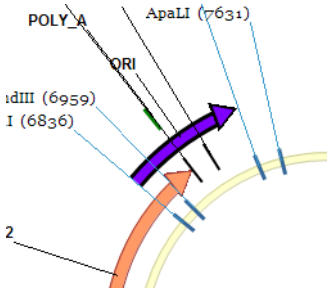
- To undo an action, press **Ctrl+Z** or right-click **Undo**.
- To redo an action, choose **Ctrl+Y** or right-click **Redo**.

Changing the Sizes And Shapes of Symbols

1. To change the size or shape of a symbol, click **Graphic Edit** ( ).

Note: You cannot modify a graphic unless the Graphics Pane is in Edit Picture mode.

2. Select the object to be changed with a mouse click. Black drag boxes will appear at its corners. Move the cursor to any drag box, changing it to a thin four-headed arrow ( ). Click and drag the black box, causing the symbol to change shape and/or size. Release the mouse when the symbol has the new shape you want.



Moving Objects on the Graphical Map

1. To move a graphics object around the Graphics Pane, select the symbol or label. When the cursor placed in the middle of the selected object turns into a four-headed arrow ( ), click and drag the object to the new location.

   As labels are dragged, their droplines follow, tracking the connection to the object to which they refer.

   Most symbols representing features, restriction sites, etc., do not move completely freely around the graphical map. They do move freely perpendicular to the "sequence" bar/ring, but Vector NTI™ *Express* Software prevents them from moving parallel to the sequence bar without your permission. This is to keep the objects in the places corresponding to their actual locations in the nucleotide or amino acid sequence.

2. To move a symbol parallel to the sequence bar of a graphical map, press and hold CTRL+SHIFT and drag the object.

   Note: This does not change the coordinates of the feature in the Feature map. To change the coordinates you need to edit the feature in the Molecule Editing mode described in Chapter 4.

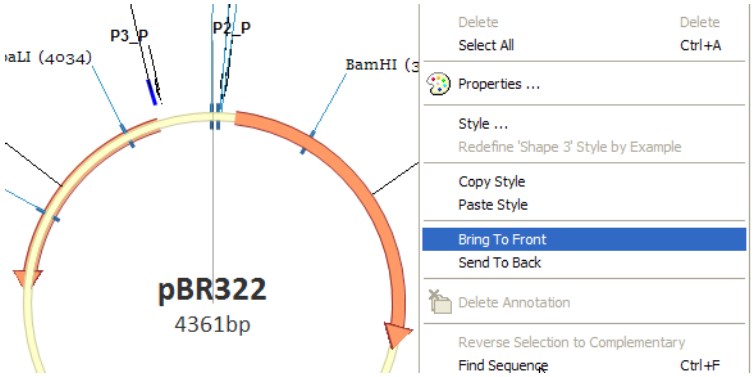
Changing the Display Order Of Overlapping Graphics Objects

To change the display order of overlapping graphics objects,

1. Select the overlapping graphic objects.
2. Right-click on the editor, and select **Bring To Front or Send To Back** or the corresponding commands from the shortcut menu.

   Note: To make these changes, the Graphics Pane must be in Graphic Edit mode.

   The following image demonstrates the results of bringing the DNA ring to the front of a circular molecule display:



Styles

A *style* is a named set of formatting attributes for graphics objects. Styles in Vector NTI™ *Express* Software work much the same as they do in most word processing programs. Styles provide a convenient and powerful way to organize the formatting of graphical maps.

Each style contains one or more of the following attributes: line, fill, text, and shape.

Although Vector NTI™ *Express* Software has standard styles for symbols and labels, you can also define your own. Once created, styles can be saved together with other parameters in the Display Setup Profile or in Vector NTI Setup files.

To learn what style a graphics object has

1. Turn on the Graphic Editing mode.
2. Select the object or **Style** from the shortcut menu.

   The style box in the Picture Element Style dialog box shows the style of the selected object. If more than one object is selected, the style box shows the style of the last object selected. The last selected object is displayed with small squares around it.

Assigning Styles to Graphics Objects

To assign an existing style to a graphics object,

1. Select the object(s) to which you want to assign a style and select **Style** from the shortcut menu.
2. In the Style box in the Picture Element Style dialog box, select an existing style from the pull-down list, or type a name, and click the **OK** button.

   If you select an existing style or type its name, that style is assigned to the selected graphics object(s), and the object(s) assume its format.

   If you type a name that is not in the style list, a new style with the format of the currently selected object(s) is created. If the selection includes more than one graphics object, a newly created style contains only those attributes that are the same for all selected objects.

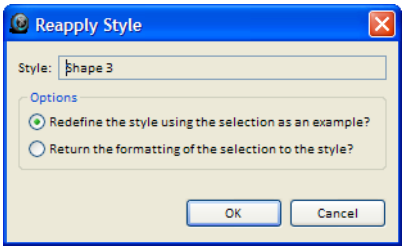
A style is the base format for a graphics object. You may make additional formatting changes to graphics objects using the methods described in preceding sections.

Redefining a Style's Format

To redefine the style of a graphic object, select the object,

1. Open the Properties dialog box, from the shortcut menu and make the desired changes for the object and click **OK**.
2. Select **Style** from the shortcut menu.
3. In the Style drop-down menu, the name of the current style is displayed. Click **OK** to apply the new format to an existing style.

   In the Reapply Style dialog box that opens, choose the redefine operation you wish to occur, and click **OK**:



To accomplish the same end, select **Redefine <name> Style by Example** or the corresponding command from the shortcut menu.

- If you choose the Redefine Style option, any objects (even unselected ones) that currently have that style will automatically change format to reflect the style's new settings.
- If you choose the Return Formatting to Style option, then the selected objects' format is restored to that Style's format. **Edit ▸ Undo** or the **Undo** button restores the former style.

Copy and Paste Styles

Once you have modified a symbol or label with a new format, you can easily reformat other objects with the same format.

1. Select the modified object, and select **Copy Style** from the shortcut menu associated with the modified object.
2. Move the cursor to another feature or label, and select **Paste Style** from the shortcut menu.

Saving Styles

The set of the styles and Standard Arrangement parameters appearing in the Molecule Display Window are a part of the Display Setup settings.

To save a display window's current styles,

1. Click **Display Setup** ( ) n the Window toolbar and choose **Display Setup** from the drop-down menu.
2. In the Display Setup dialog box, there are two options for saving styles:
   - **Save Settings As** saves all the settings in Display Setup. The settings are given a name that is then listed in the Display Setup drop-down menu. Properly saved Setup Profiles are easily retrieved and applied to display windows. Click this button and enter the name you want to assign to the current settings.

- **Save To File** saves current graphics display settings that you have defined. They are not as easily retrieved as when you use the Save Settings As option. Use this option as an example, when you are preparing graphics for publications and wish to transfer formatting instructions from one display window to another. In the standard File Save dialog box, choose a file name for current styles and Standard Arrangement information.

The Saving Styles are divided into two types:

- **Molecule Document Files**: In this saving style, you can export the file after editing. Go to **File ▸ Export** to export the edited file.You can only use the DNA/RNA Molecules Document or the Protein Document file format. In the absence of these two file formats, the Vector NTI™ *Express* Software will prompt you with the message that the Display Settings will not be saved.
- **Create Molecule Shortcut**: In this saving style, you can create a shortcut that contains the Display Settings configured in a non-graphical editing mode that can be saved. Changes done in the graphic edit mode will not be saved into the Molecule Shortcut file.

Loading Styles

To apply a saved Display Profile to a Molecule Display window,

1. Choose a profile entry name from the list in the Display Setup drop-down menu or in the Settings Profile drop-down menu in the **Display Setup** dialog box.
2. When you make the selection, Vector NTI™ *Express* Software loads the styles as well as other settings from the profile but does not apply them to the active window until you click the OK button.
3. Modify the settings further, if you like; click **OK** to accept the new setup or click **Cancel** to leave the current window's settings unchanged.

Change the Arrangement Setup

The Vector NTI™ *Express* Software automatically rearranges the symbols and labels on the graphics map when you change the set of restriction enzymes, edit the functional map of the molecule, etc. This is called the Standard Arrangement.

Go to Toolbar in the Molecule Editor and click  **Standard Arrangement** or by selecting Standard Arrangement Item from the Molecule Editor Shortcut menu.

Standard Arrangement repositions all the elements back to the original default position.

Picture type

1. Check the **PictureType** check box to view the type of the picture in the Preview pane.
2. Select from Linear and Circular Picture Type radio buttons.
3. Click **OK** to submit the PictureType Setup Option edits or **Cancel** to exit the Preferences dialog box.

Note: The Picture Type Display Profile setup option is applicable only to DNA/RNA molecules. Protein molecules can be displayed only in the Linear style.

The PictureType Setup dialog box looks like this:

Figure 1 Linear Picture Type

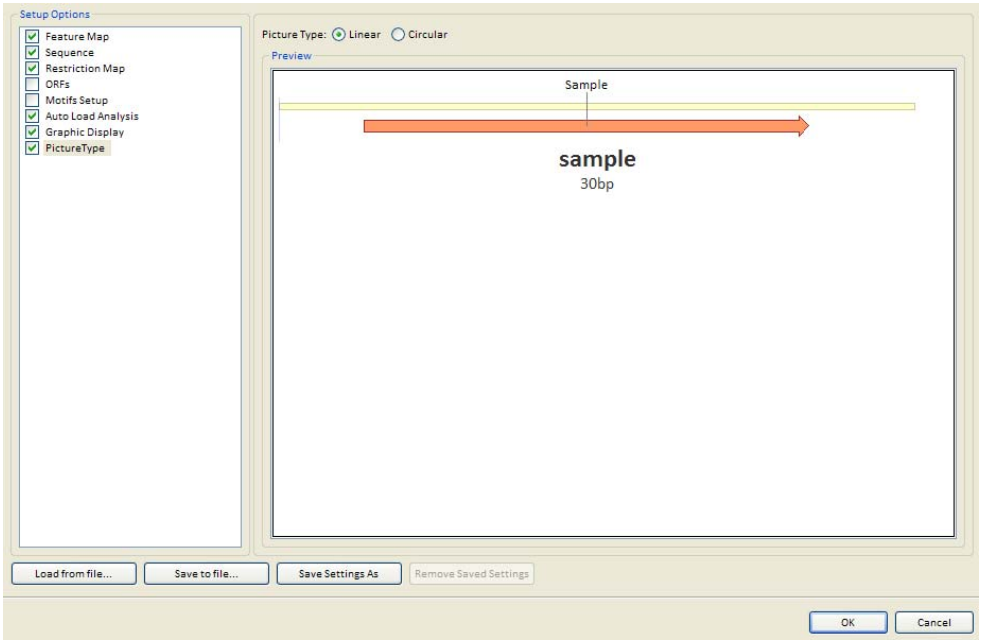
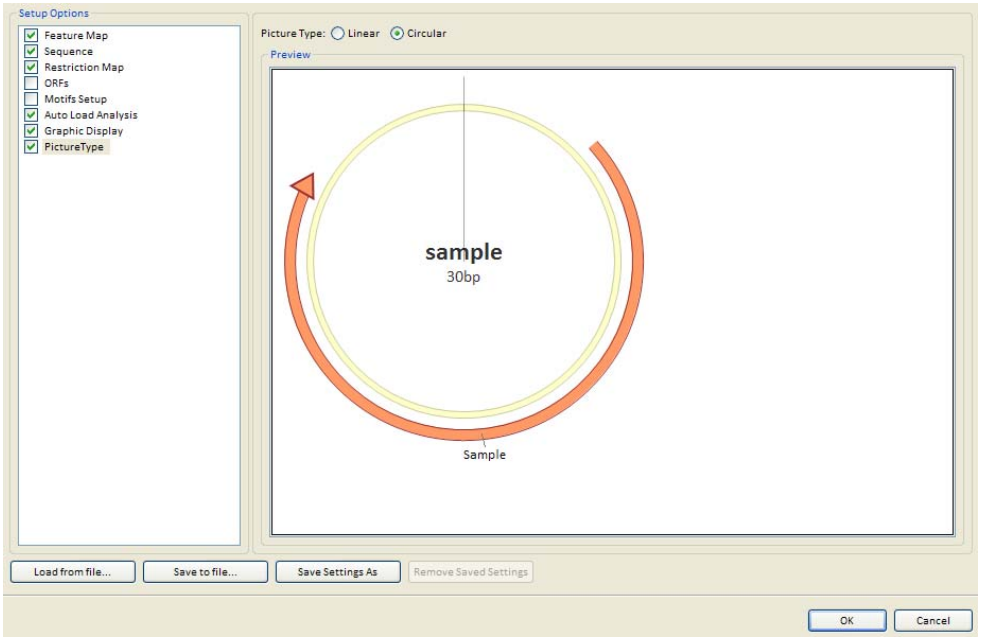


Figure 2 Circular Picture Type

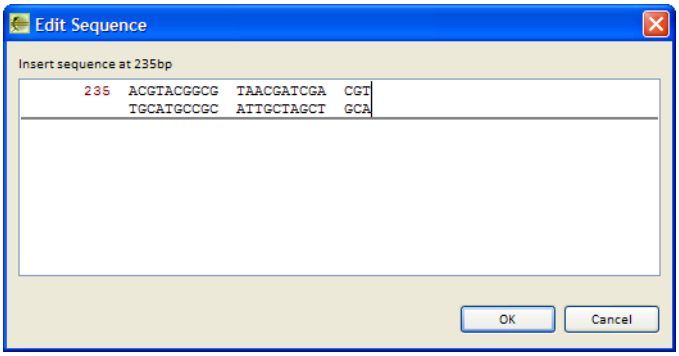


Enter or edit a sequence

The molecule sequence is displayed in the Sequence pane, and a graphical representation of the sequence is displayed in the Graphics pane.

Enter a sequence

1. To enter a new sequence, click on an insert point in the Sequence pane and begin typing.
2. The **Edit Sequence** dialog will open, displaying your typed sequence.



3. Continue typing and click on **OK** to insert the sequence.

Select a sequence

To select part or all of a sequence:

- Drag your cursor over the desired portion sequence in the Sequence pane or Graphics pane.
- Click on a feature in the Graphics pane or Feature Map that encompasses that sequence.
- Click **Ctrl+a** or right-click and select **Select all** to select the entire sequence.

Cut or copy a sequence

- To cut a sequence, select it in the Sequence or Graphics pane and click **Ctrl+x**, or right-click and select **Cut**, or select the command from the **Edit** menu. You will be prompted to remove the sequence.
- To copy a sequence, select it and click **Ctrl+c**, or right-click in the Sequence or Graphics pane and select **Copy**, or select the command from the **Edit** menu.

Paste a sequence

1. To paste a sequence from the clipboard, click on a particular insertion point in the Sequence or Graphics pane and click **Ctrl+v** or select **Edit > Paste**.
2. The **Edit Sequence** dialog will open, displaying your pasted sequence. Click on **OK** to complete the action.

Replace a sequence

To replace part or all of a sequence, select it in the Sequence Pane and type or paste as described above.

Delete a sequence

To delete all or part of a sequence, select it in the Sequence or Graphics pane and click the **Delete** key or right-click and select **Delete.**

Reverse a sequence

You can reverse all or part of a sequence:

1. Select all or part of the sequence in the Sequence or Graphics pane.
2. Right-click and select **Reverse Selection to Complementary**.

Molecule features

Molecule features are regions of a DNA or protein sequence that you define and annotate. The feature is then displayed in the Graphics pane and listed in the Feature Map.

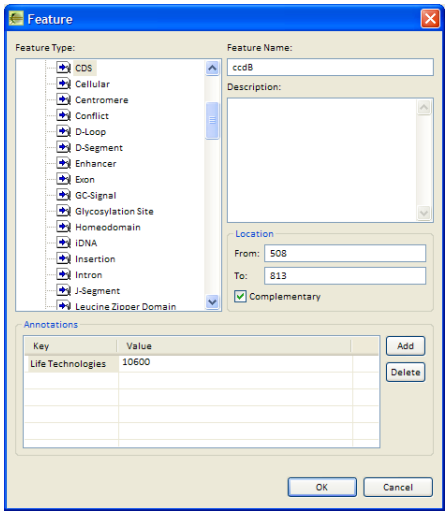
Typical DNA molecule features might include coding regions, specific ORFs, primer binding sites, etc. Protein molecule features might include structural features, binding sites, functional domains, etc.

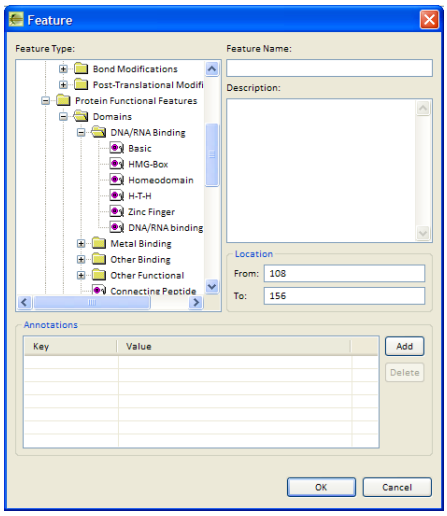
Create a molecule feature

To create a molecule feature:

1. Select a portion of the molecule sequence by dragging in the Graphics pane or in the Sequence pane.
2. Right-click and select **Create feature from selection**.
3. In the Feature dialog, select a category from the **Feature Type** list. Various miscellaneous categories are available under Misc.

Feature dialogs for DNA (left) and protein (right)





4. Enter a **Feature Name** and optional **Description**.
5. The **Location** range is auto-populated based on your selection. Enter a different range in the fields if desired.
6. For a DNA feature, select the **Complementary** checkbox if the feature is located on the reverse strand.
7. Click on **Add** or **Delete** to add or remove Annotation keys. When adding a key, a blank key is created in the table; type to replace the text.

8. When you are finished, click on **OK**. The feature will displayed in the Graphics pane and in the Feature Map.

Select a feature

To select a molecule feature, click on it in the Graphics pane or in the Feature Map.

Hide or display a feature

To hide a feature or group of similar features, deselect the appropriate checkbox in the Feature Map. To re-display it, re-select the checkbox.

Edit or delete a feature

To edit a feature, right-click on it in the Graphics pane and select **Edit feature**. The Feature dialog will open.

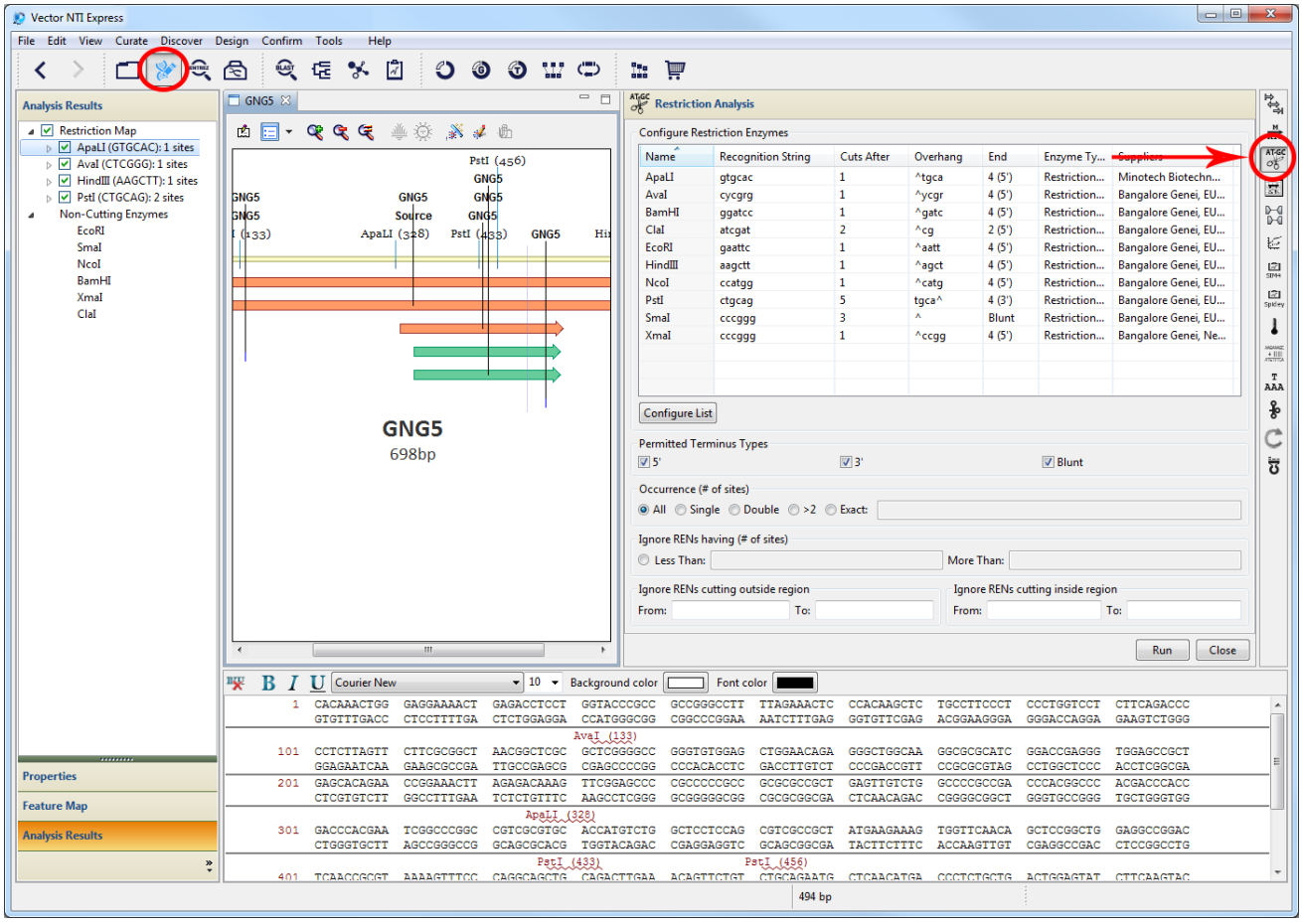
To deselect a feature, right-click on it in the Graphics pane and select **Delete feature**. You will be prompted to confirm the deletion.

Restriction Analysis

You can identify the restriction sites in a DNA/RNA sequence for hundreds of restriction enzymes.

To begin, with a DNA/RNA molecule open, click on the **Restriction Analysis** button on the Analysis toolbar.

The Restriction Analysis tool includes commands for selecting and configuring a list of restriction enzymes, and running the analysis.



Selecting enzymes for analysis

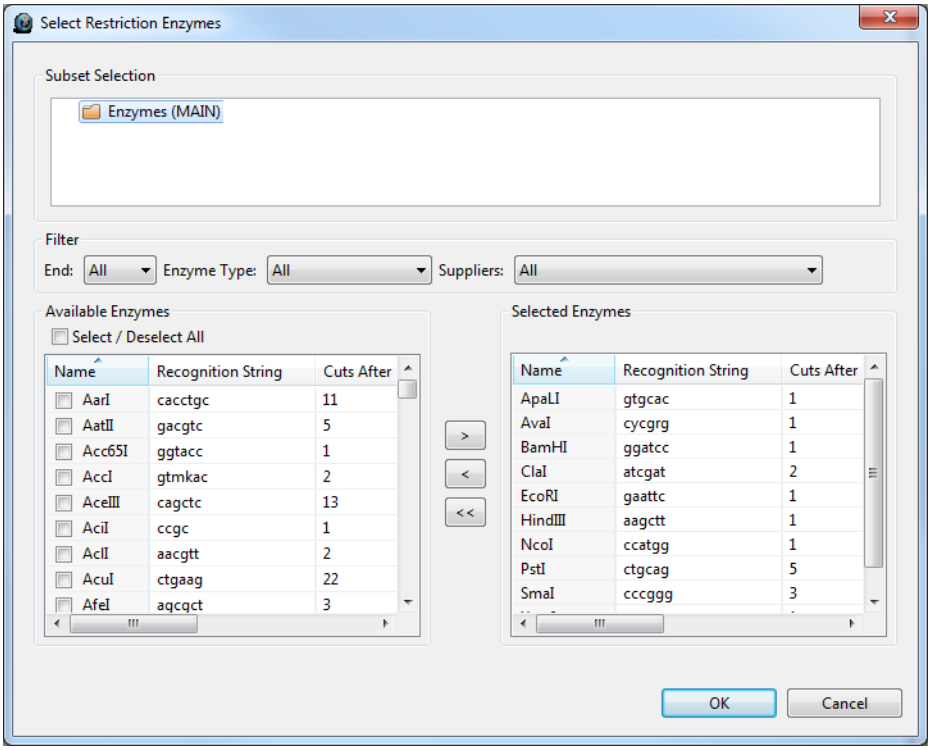
A default list of restriction enzymes to be used in the analysis will be selected.

- Click on enzymes in the list to select or deselect them.
- To clear the list, click on **Clear Selection**.
- To select all the enzymes in the list, click **Select All**.

Configuring the enzyme list

1. To configure the enzymes in the Restriction Analysis tool, click on **Configure List**.

2. Use the commands in the Select Restriction Enzymes dialog to add or remove the selected enzymes, then click **Apply**.



Filter the analysis results

To filter the analysis results by the terminus type of the restriction cut site, select or deselect the checkboxes under **Permitted Terminus Types**.

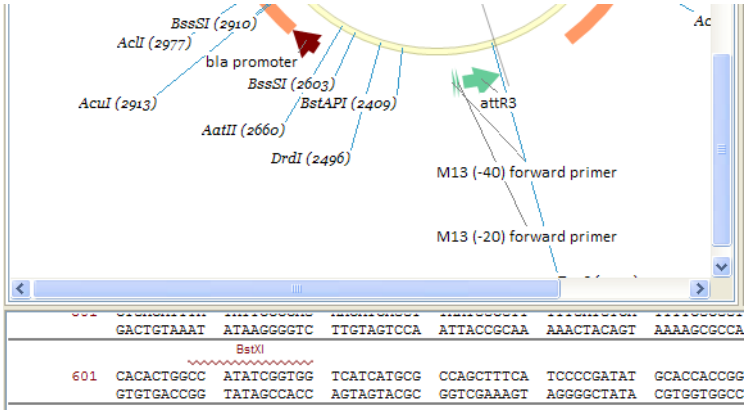
Note: **5′** and **3′** refer to the overhang of the resulting cut site, e.g., with **5′** deselected, restriction sites with 5′ overhangs will be removed from the analysis.

Ignore RENs Having Less Than/More Than … Sites hides restriction sites that do not fall within the range of the specified number of cut sites. Such RENs will be listed but deselected in the Restriction Map in the Analysis Results. They will not be displayed at all in the Graphics and Sequence panes.

Perform the analysis

When you have made your selections, click on **Run**. The sites will be displayed in the Graphics and Sequence panes and listed in the Analysis Results.

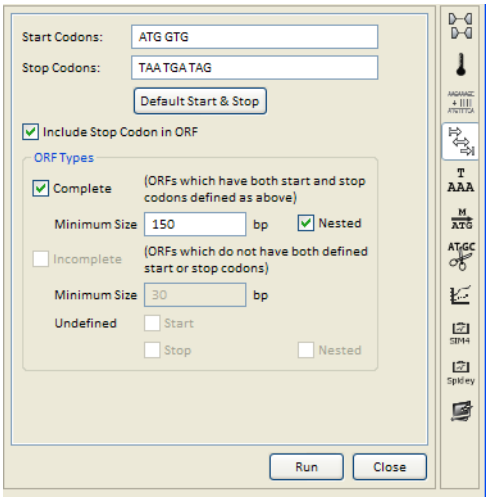
The results of Restriction Analysis are saved with the molecule.



ORF Finder

To identify ORFs in a DNA or RNA sequence, click on the **ORF Finder** button in the Analysis toolbar of the Molecule Editor.

The **ORF Finder** tool contains settings for identifying ORFs within the sequence.



Define start and stop codons

- In the **Start** and **Stop Codon** fields, enter start and stop codons for the new ORFs. Click the **Default Start & Stop** button to set the start and stop codons to the following conventional values: Start codons—ATG, GTG; Stop codons—TAA, TGA, TAG.
- Check the **Include Stop Codon in ORF** box if you want the stop codon to be considered as a part of the ORF. Otherwise, the stop codon is not considered as part of the ORF and is not included.

ORF Types

- **Complete:** The **Complete** check box is selected by default and in the **Minimum Size** field the minimum size in base pairs is specified as **150**. This can be changed according to your requirement. Check the **Nested** check box to look for nested ORFs. These are ORFs that occur within the main ORF.

- **Incomplete:** Check the **Incomplete** check box and in the **Minimum Size** field specify the minimum size in base pairs for the incomplete ORF to be displayed. Check the **Undefined Start/Stop** check box to search for ORFs with undefined starts, stops or both. Check the **Nested** check box to look for nested ORFs.

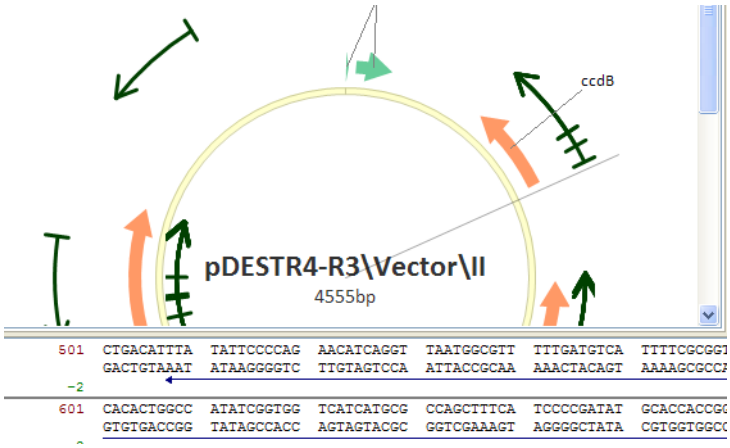
Note:

- An ORF with an undefined start is one that has a stop codon at the end but no corresponding start codon in the same frame
- An ORF with an undefined stop is one that has a start codon but no stop codon in the same frame.
- Incomplete ORFs are displayed as a dashed arrow in both the Graphics and the Sequence Pane. Complete ORFs are displayed as solid arrows.

Perform the analysis

Click on **Run** to perform the ORF analysis. The ORFs will be marked with directional arrows in the Graphics and Sequence panes and listed in the Analysis Results.

The results of ORF analysis are saved with the molecule

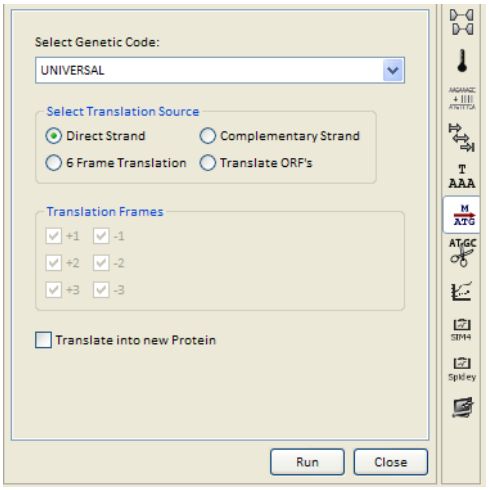


By default, ORFs are displayed for the direct and complementary strands. If single-stranded sequence is displayed, only the ORFs for that strand are shown.

Translation tool

To translate a RNA or RNA sequence into amino acids, click on the **Translate** button in the Analysis toolbar of the Molecule Editor. 

The Translation tool contains settings for translating the sequence.



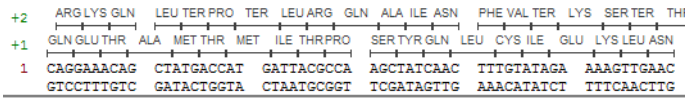
1. Select the genetic code for the desired organism to use as the basis for the translation from the Genetic Code dropdown list. For more information on genetic codes, visit **www.ncbi.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c**.
2. Select the Translation Source from among the following options:
   - **Direct Strand** – displays a translation of the direct strand sequence in the current frame
   - **Complementary Strand** – displays a translation of the complementary strand sequence in the current frame
   - **6 Frame Translation** – displays up to 6 translations of the sequence depending on the selections in the **Translation Frame** checkboxes
   - **Translate ORFs** – if you have used the **ORF Finder** to identify ORFs in the sequence, this selection will translate the ORFs
3. If you selected **6 Frame Translation,** select number of direct and complementary **Translation Frames** used to translate the sequence. Up to three direct strand (+1, +2, +3) and three complementary strand (-1, -2, -3) translations may be selected.
4. To create a new protein molecule from the translation, select the **Translate into new protein**. Otherwise the translation will be displayed within the DNA/RNA sequence.
5. Click on **Run** button to perform the translation.

Translation results

If you selected **Translate into new protein**, you will be prompted to name the new protein and the protein molecule will be created.

Otherwise, translations of the direct strand will appear above the DNA sequence and translations of the complementary strand will appear below the sequence.

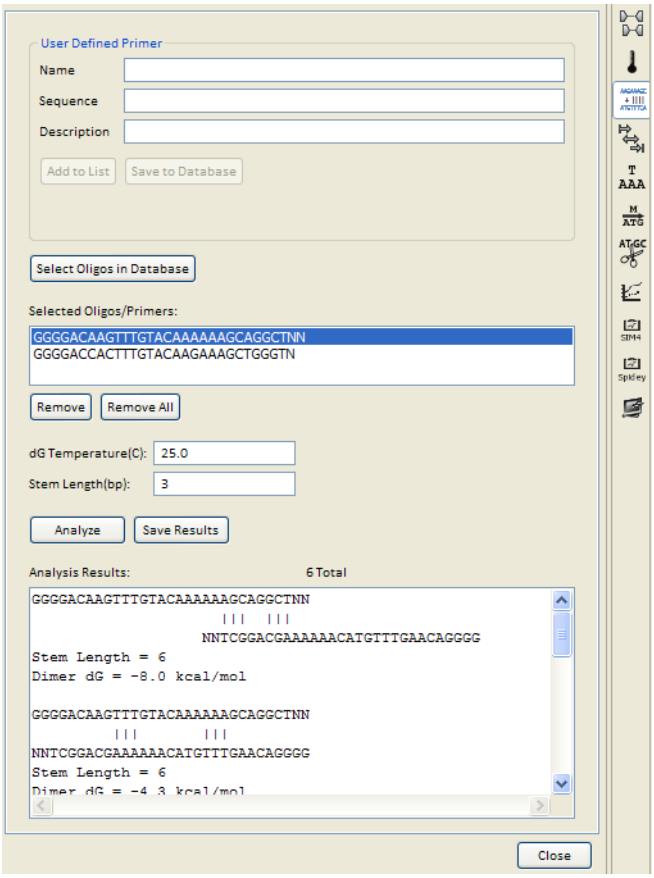
Translated sequence using 6 Frame Translation with +2 Frames selected



Oligo Duplex Analysis

Oligo Duplex Analysis enables the analysis of one or more oligos for potential cross-reactivity and dimerization.

To open the tool, click on the **Oligo Duplex Analysis** button in the Molecule Editor. 



Entering or selecting oligos

Using the tool, you can analyze oligo sequences that you enter in the fields at the top, or you can analyze saved oligos in the database.

- Enter an oligo name and sequence under **User Defined Primers** at top of the window and click on **Add to List** to add it to the list of oligo sequences below, and/or **Save to Database** to save it as an oligo in the database.
- Click on **Select Oligos in Database** to select one or more saved oligos in the database to analyze. Use **Ctrl+click** and **Shift+click** to select multiple oligos from the Oligos in Database dialog.

The selected oligos will appear in the **Selected Oligos/Primers** list.

Analysis parameters

Select from the following additional analysis parameters:

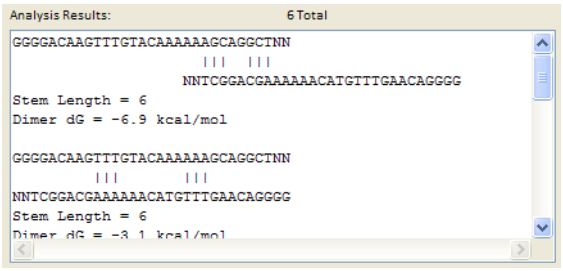
- **dG Temperature:** enter the temperature in degrees Celsius to be used for calculating free energy values.

- **Stem Length:** enter the minimum acceptable number of base pairs in a hairpin or dimer stem.

Run the analysis

Click on **Analyze** to run the analysis. Analysis results are displayed at the bottom of the window.

Click on **Save Results** to save the analysis as a separate text file.



Note: In the graphical depiction of duplexes, vertical lines indicate the primary interaction, based on the stem length set, and plus symbols indicate secondary interactions. The greater the dG value, the weaker the interaction; secondary interactions are not considered in the dG calculation.

Silent Mutation Analysis

In the Molecule Editor, you can search for "silent mutations" in a DNA/RNA sequence or selected region of a sequence that do not affect amino acid translation but result in the creation or disappearance of one or more restriction sites.

Note: You do not need to perform restriction analysis before running Silent Mutation Analysis.

1. Select a region of the sequence or make no selection to analyze the entire sequence.
2. Click on the Silent **Mutations Analysis** button to open the tool.

3. Click on **Add Enzyme from Database** to select from a list of restriction enzymes in the database. Use **Ctrl+click** and **Shift+click** commands to select multiple enzymes in the Enzyme in Database dialog.
4. Click on **OK** to add the selections to the **Available Enzymes** list.
5. Use the **>>**, **All>>**, **<<**, and **<<All** buttons to move enzymes from the **Available** to the **Use** list.
6. Click on **Run** to initiate the mutagenesis search. Vector NTI™ *Express* analyzes the sequence or selected region and attempts to generate suitable silent mutations. The reading frame for amino acids is defined by the start of the selected region so that the first three nucleotides of the selected region form the first codon.

The folder contains a list of mutation options that result in the appearing and/or disappearing of at least one restriction site. The options are sorted by the position of the first altered nucleotide. If you selected the complementary strand option, mutation coordinates on both complementary and direct strands are listed.

Note: The program is able to find both “single” (just one nucleotide altered) and “multiple” (several neighbor nucleotides altered) mutations for any elementary event (appearing and/or disappearing of at least one site) significantly widening the set of possible solutions compared to just “single”-mutation analysis.

Web analyses

A DNA or protein molecule sequence or part of a sequence can be analyzed using a variety of online databases, search engines, and analysis tools.

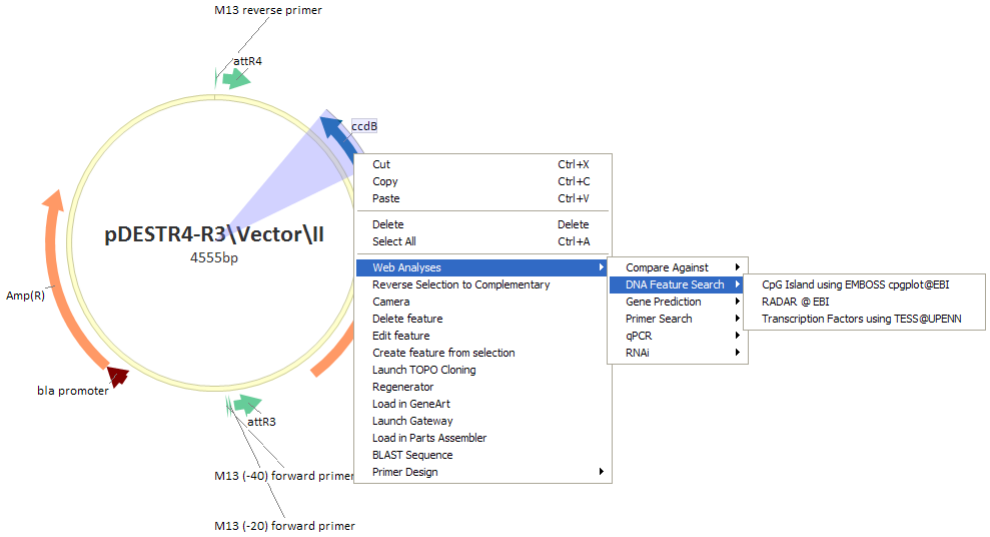
Note: These analyses require an active Internet connection.

1. In the Graphics or Sequence pane, select the region of the sequence to analyze, or select **Ctrl+A** to select the entire sequence.
2. Right-click and select **Web Analyses > [search type] > [search database]**.

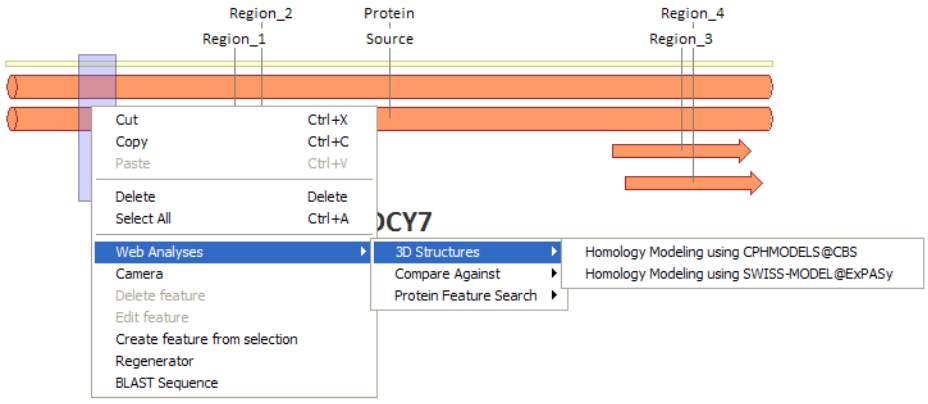
Note: You can also right-click on a molecule in the Database Explorer to analyze the entire sequence.

3. The sequence will be automatically transferred to the analysis engine on the web:

DNA Molecule Web Analyses.



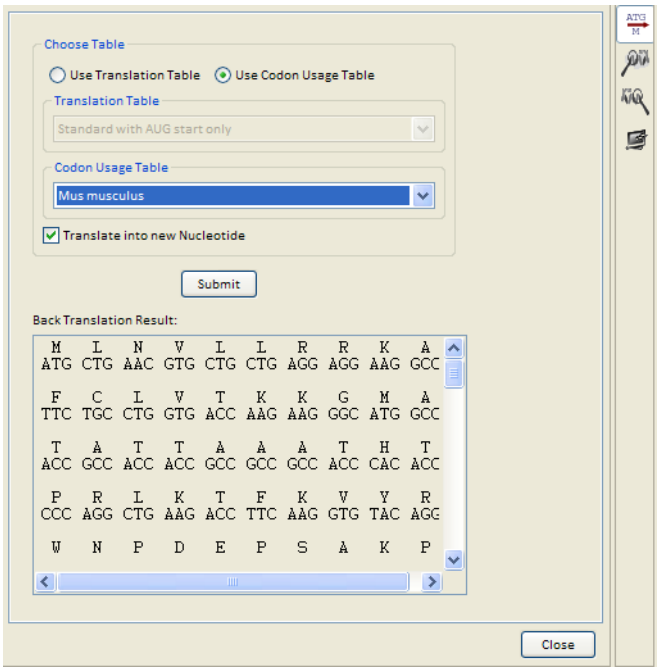
Protein Molecule Web Analyses



Back Translation

Back Translation allows you to obtain a DNA sequence from a protein sequence by reversing the translation process. You can select from a list of translation and codon usage table options

1. With a protein molecule open, click on the **Back Translation** button in the Analysis toolbar.



2. In the Back Translation Tool, select the translation table type: **Use Translation Table** or **Use Codon Usage Table.**
3. Select the specific translation table from the dropdown list.

Note: For more information about translation and codon usage tables, visit **www.ncbi.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c.**

4. Select the **Translate into new nucleotide** checkbox to create a DNA/RNA molecule from the resulting analysis.
5. Click on the **Submit** button.

If you selected the checkbox in step 3, you will be prompted to enter a name for the molecule and save it to the database.

The results will be displayed in the **Back Translation Result** field.

Protein Domain and Motif Finder Analyses

You can analyze a protein sequence in Vector NTI™ *Express* against databases of known protein domains, families, functional sites, and motif fingerprints.

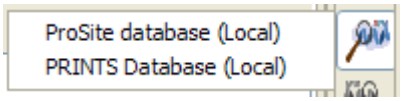
Protein Domain Analysis—PROSITE and PRINTS databases

PROSITE and PRINTS are databases of biologically significant protein patterns and profiles that can be used to identify the function of uncharacterized proteins. Comparing an unknown polypeptide sequence in Vector NTI™ *Express* to the motifs in these databases can help determine any known protein families and domains to which the sequence may belong.

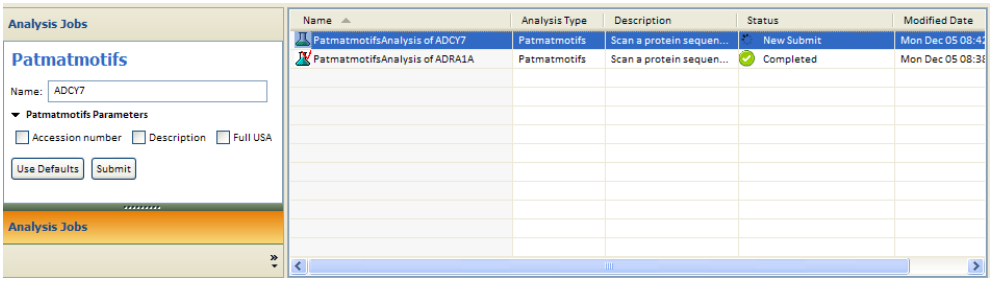
- For more information about the PROSITE database, visit **http://prosite.expasy.org/**.
- For more information about the PRINTS database, visit **www.bioinf.man.ac.uk/dbbrowser/PRINTS/index.php**.

Local versions of the PROSITE and PRINTS motif databases are provided as part of the Vector NTI™ *Express* installation.

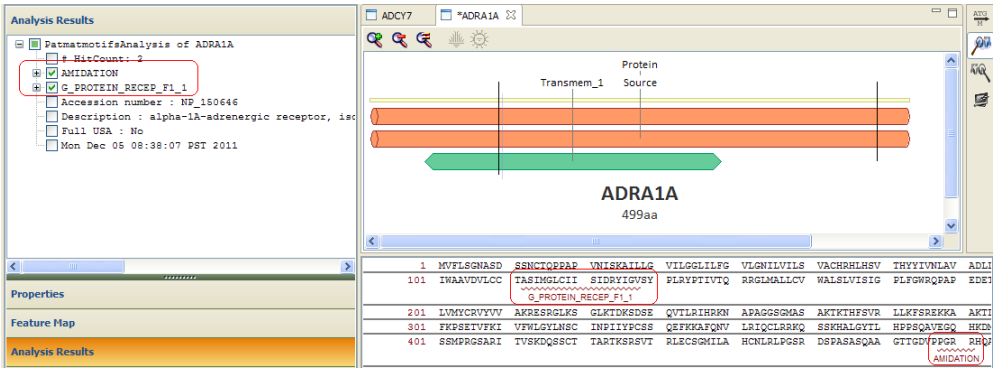
1. To begin the analysis, with a protein molecule open, click on the **Protein Domain** button in the Analysis toolbar of the Molecule Editor window.
2. Select the local version of the PROSITE or PRINTS database, installed with Vector NTI™ *Express*.



3. In the **Analysis Jobs** window, the name of the job will appear listed in the left-hand pane and the status will be "New submit..."

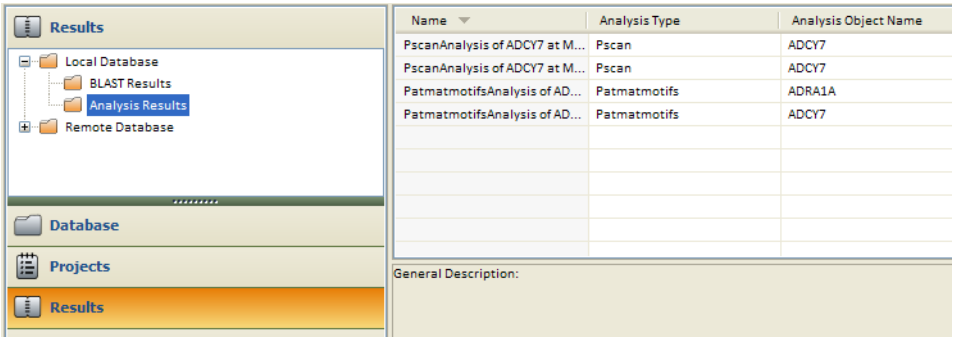


4. Select the parameters that will be included with the results, then click on **Submit**.
5. When the analysis is complete, it will be listed in the left-hand pane as "Complete."
6. Double-click on the complete job to return to the molecule with the results displayed in the Sequence pane and the Analysis Results.



- Click on a feature in the Analysis Results pane to show expanded details for that result.
- Click on a checkbox in the Analysis Results pane to show or hide that feature in the Sequence pane.
- Right-click in the Analysis Results pane and select **Save All Analysis Results to Database** to save these to Analysis Results in your local database.

- Analysis Results saved to the database will listed in the Database Explorer, under **Results** in the **Analysis Results** folder.



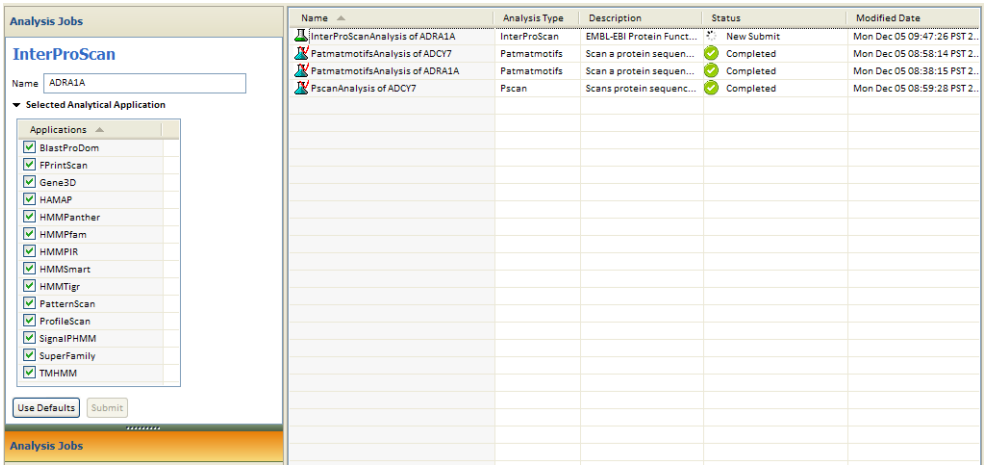
Motif Finder—InterProScan sequence search

Motif Finder uses a local version of the InterProScan sequence search engine to scan a protein sequence for protein domains and functional sites. This search tool integrates analysis engines from multiple protein signature databases. Using this tool, you can analyze a sequence using all InterPro analysis engines or a subset.

- For more information about InterProScan and the databases/analytical tools it includes, visit **www.ebi.ac.uk/Tools/pfa/iprscan/**.

A local version of the InterProScan search engine is provided as part of the Vector NTI™ *Express* installation.

1. To begin the analysis, with a protein molecule open, click on the **Motif Finder** button in the Analysis toolbar of the Molecule Editor window.
2. In the **Analysis Jobs** window, the name of the job will appear listed in the left-hand pane and the status will be "New submit..."

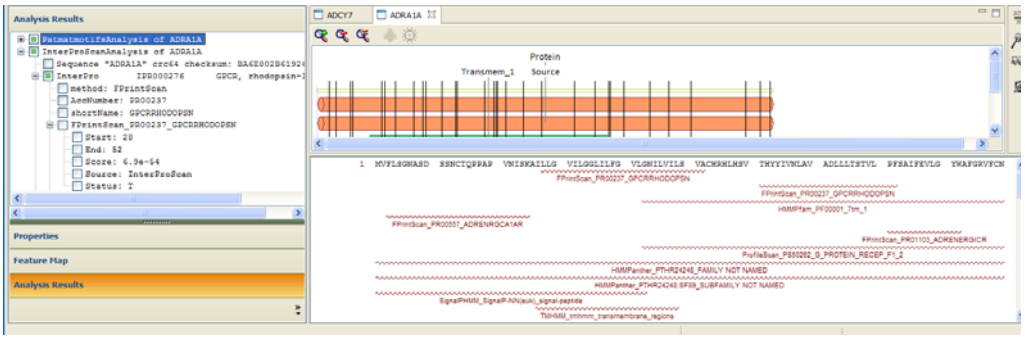


3. Select the analytical application(s) to use from the available list of InterProScan tools, then click on **Submit**.

Note: The more applications you select, the longer the analysis will take.

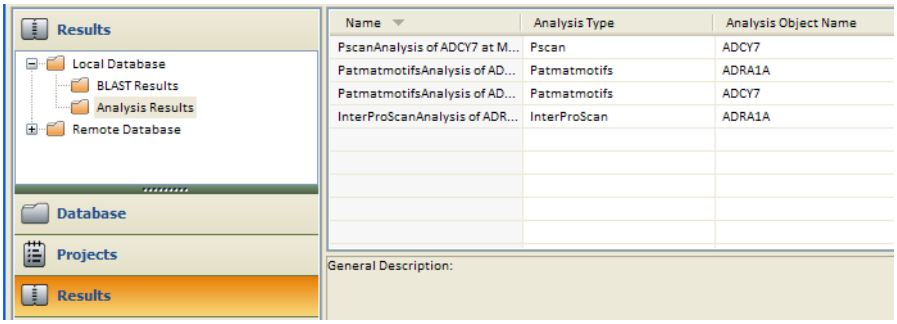
4. When the analysis is complete, it will be listed in the left-hand pane as "Complete."

5. Double-click on the complete job to return to the molecule with the results displayed in the Sequence pane and the Analysis Results.



- Click on a feature in the Analysis Results pane to show expanded details for that result.
- Click on a checkbox in the Analysis Results pane to show or hide that feature in the Sequence pane.
- Right-click in the Analysis Results pane and select **Save All Analysis Results to Database** to save these to Analysis Results in your local database.

Analysis Results saved to the database will listed in the Database Explorer, under **Results** in the **Analysis Results** folder.



3 Primer Design

This chapter describes the functions for designing primers and probes in Vector NTI™ *Express*, including settings for designing PCR primers, sequencing primers, and hybridization probes. Vector NTI™ *Express* can design primers for an entire DNA molecule sequence or part of a sequence selected in the Molecule Editor window. After selecting the target sequence, the maximum and minimum product length and parameters are determined, and the software evaluates, rates and sorts several design options. You can further fine-tune the oligos and annealing parameters if you wish, save the primers or probes as separate molecules in the database or to the Oligo List, order custom oligos from Thermo Fisher Scientific, or use the primers in recombinant cloning strategies.

The following table summarizes the various primer/probe design options in Vector NTI™ *Express*:

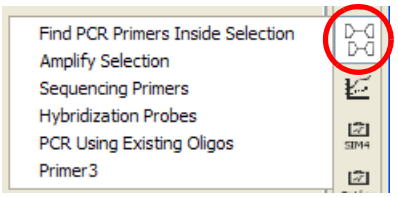
Design Tool	Purpose
Find PCR Primers Inside Selection	Specify limits for PCR primer search such as length of target sequence, output options, attach restriction sites, etc.
Amplify Selection	Similar to **Find PCR Primers** except that primer hybridization domains upstream and downstream from the target sequence can be specified. Primers will be generated anywhere within the designated upstream and downstream domains.
Sequencing primers	Set parameters for sequencing and primer regions and primer; analyze primers.
Hybridization Probes	Set parameters for target region, output options; analyze probes.
PCR Using Existing Oligos	Similar to **Find PCR Primers**, but, for the selected amplification region, allows you to search for suitable PCR primers from among those selected from the Vector NTI™ *Express* oligo list.

Open the primer/probe design tools

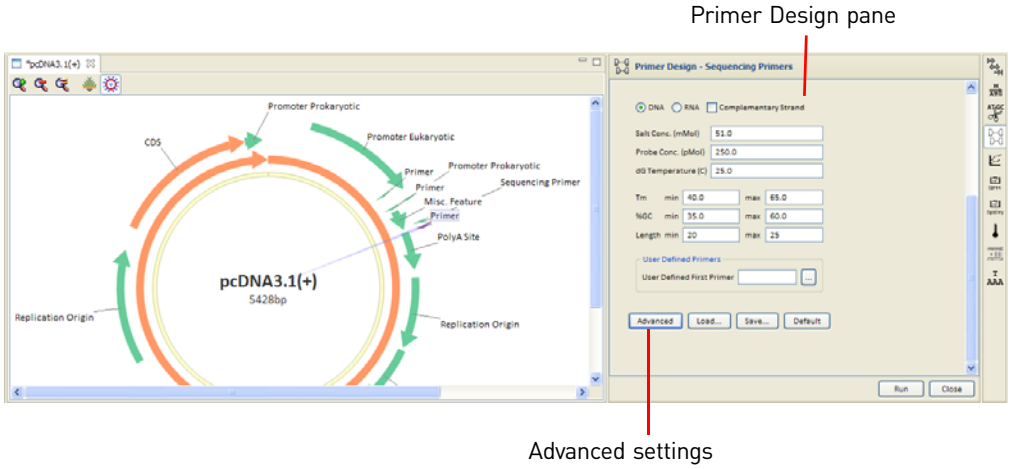
The design tools are located on the Molecule Editor toolbar.

1. With a molecule loaded in the Molecule Editor, you can design oligos for the entire molecule sequence or a selected region:
   - To design primers/probes for the entire molecule, make no sequence selection before you open the design tool.
   - To design primers/probes for a specific region of the sequence, select the region (e.g., by dragging or clicking on a feature in the Graphics or Sequence pane) and then open the tool.

2. Click on **Primer Design** in the toolbar and select from the dropdown list of design tools:



3. Each primer tool opens in a separate pane in the Molecule Editor window, with the basic settings accessible in the pane. For advanced settings, click on the **Advanced** button in the pane to open a separate dialog box.



Save and load settings

The **Save** and **Load** buttons at the bottom of each pane (or at the bottom of each tab in the **Advanced** settings dialog) allow you to save your primer/probe design settings to a file and load the settings file for subsequent analyses. Design settings are saved as ***.pcr** files.

This is useful for saving frequently used settings.

Run the design tool

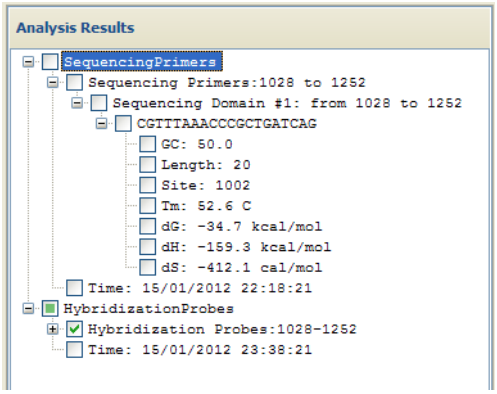
After you have selected the desired settings (see the following pages for more details about individual settings), click on **Run** in the primer design tool, or click on **OK** in the Advanced settings dialog of the tool.

Primer/probe design results

The primer or probe designs will be added as features in the molecule, and listed in the **Feature Map** and displayed in the **Graphics pane** of the Molecule Editor (see "Molecule features" on page 64).

Primer and probe designs are sorted in descending order in the Feature Map according to their rating values calculated based on the importance factors assigned in the Qualities tab (see page 91). The molecule region of each design is listed in parentheses.

The results are also listed under **Analysis Results** in the Molecule Editor window, along with specific information about each oligo sequence.

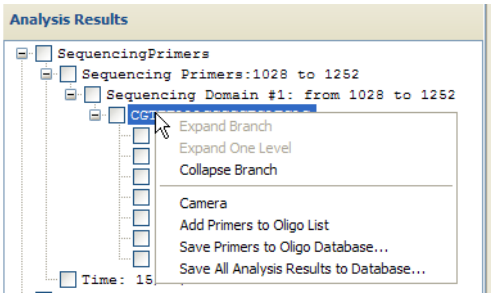


Primer and probe designs are also be listed in the **Ordering** dialog, under Primers.

Save primer/probe designs

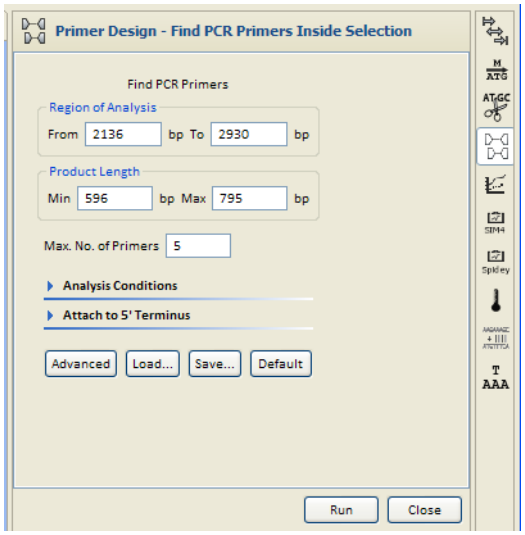
There are a number of ways to save primer/probe designs:

- To save them as Analysis Results in the database, right-click on a top branch in the Molecule Editor Analysis Results pane and select **Save All Analysis Results to Database**.
- To save them as oligos in the Oligos Database, right-click on a primer/probe sequence in the Molecule Editor Analysis Results pane and select **Save Primers to Oligo Database**.
- To add them to the Oligos List, right-click on a primer/probe sequence in the Molecule Editor Analysis Results pane and select **Add Primers to Oligo List**.



Find PCR Primers Inside Selection settings

Select **Find PCR Primers Inside Selection** to design primers within a selected region or the entire molecule for PCR amplification. With the molecule open in the Molecule Editor, select the region to be sequenced, then select **Primer Design > Find PCR Primers Inside Selection** from the Molecule Editor toolbar.:



The following are the unique settings for **Find PCR Primers**. These are also available under the Primers tab if you click on the **Advanced** button:

Find PCR Primers setting	Description
Region of Analysis	The start and end coordinates of the region to amplify. You can enter new coordinates, or select a region of the sequence before opening the tool to pre-populate the region coordinates.
Product Length	Enter the maximum and minimum lengths of the molecule target region. Note: Unless you specify differently here, the minimum amplicon length may be less than the target sequence you selected.
Maximum Number of Primers	Enter the number of sense-antisense primer pairs to be found. The actual result may contain fewer than this number if there are not enough possible primers.

Analysis Conditions settings

Click on the **Analysis Conditions** dropdown to access these settings:

Analysis Conditions setting	Description
DNA/RNA radio button	Select the type of target nucleotide sequence.
Salt Concentration	Enter the PCR reaction salt concentration in mMol, if known.
Probe Concentration	Enter the value of probe concentration in pMol, if known.
dG Temperature	Enter the temperature in degrees Celsius to be used for calculating free energy values.

Analysis Conditions setting	Description
Tm	Enter limits in degrees Celsius for primer melting temperature (T_m) (temperature at which 50% of primer is a duplex) and the difference between T_m for sense and antisense primers.
%GC	Enter the limits of G/C percentage in the primer and the difference between GC percentages for sense and antisense primers.
Length	Enter primer length limits. Note: Nucleotide sequences such as RENs attached to a primer's 5' end are included when calculating primer length.

Note: The calculation for Tm is dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Make sure to adjust these parameters according to your reaction conditions when performing your PCR analysis to ensure that you obtain accurate Tm values.

Attach to 5' Terminus settings

Click on the **Attach to 5' Terminus** dropdown to access these settings

Attach to 5' Terminus setting	Description
Attach to 5' Terminus of Sense Primer and/or Antisense Primer	Enter a short (=/<18 bp) nucleotide sequence (if any) to be attached to the 5' end of either primer. To choose from recognition sites of database RENs, click the **Browse** button next to each field. **Note:** This sequence, while considered in primer parameters, does not affect the calculation of complementarity between primer and molecule. A sequence can be attached to the primer whether or not the primers are user-defined or designed by the software.

Advanced settings: Primers tab

Click on the **Advanced** button below the main settings to open the Advanced settings dialog. Under the Primers tab, the settings described above are listed, in addition to the following:

Advanced Primers setting	Description
User-Defined Primers	Enter user-defined primer sequences or a primer from the oligo database. The search engine checks the compatibility of the primers according to primer parameters.

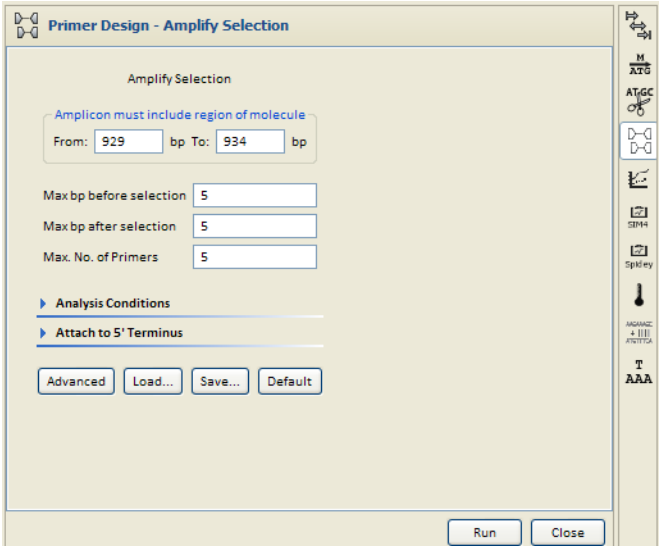
Additional advanced settings are described in "Shared Advanced settings" on page 88

Amplify Selection settings

Select the **Amplify Selection** primer design tool to amplify an entire selected region of a molecule. With the molecule open in the Molecule Editor, select the region to be sequenced, then select **Primer Design > Amplify Selection** from the Molecule Editor toolbar.

These settings are similar to the Find PCR Primers settings except that you can specify primer hybridization regions upstream and downstream of the target sequence.

Amplify Selection picks primers to amplify the entire selection. If suitable primers cannot be found inside the selected region, the search will expand within the specified upstream and downstream flanking regions.



The following are the unique settings for **Amplify Selection**. These are also available under the Primers tab if you click on the **Advanced** button:

Amplify Selection setting	Description
Amplicon Must Include Region of Molecule	Set the 5' and 3' positions for region of the molecule that must be included in the final amplified product.
Max bp before selection	Provides additional upstream region where the primer may be made.
Max bp after selection	Provides additional downstream region where the primer may be made.
Maximum Number of Primers	Enter the number of sense-antisense primer pairs to be found. The actual result may contain fewer than this number if there are not enough possible primers.

Analysis Conditions settings

For information about the **Analysis Conditions** settings, see page 82.

Attach to 5' Terminus settings

For information about the **Attach to 5' Terminus** settings, see page 83.

Advanced settings: Primers tab

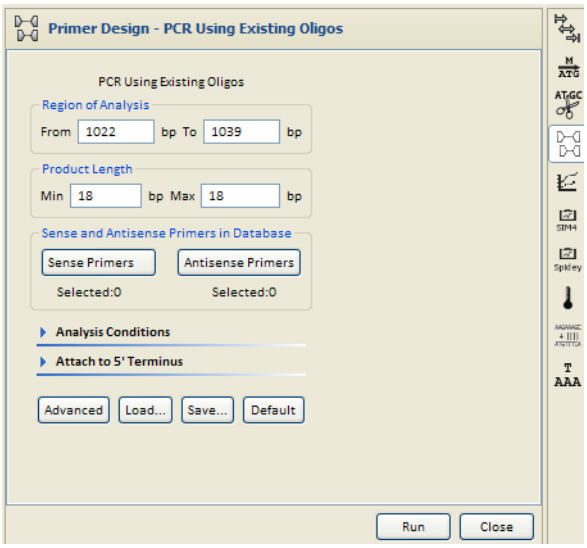
Click on the **Advanced** button below the main settings to open the Advanced settings dialog. For information about the **Advanced** settings on the Primers tab, settings, see page 83.

Additional Advanced settings

Additional advanced settings are described in "Shared Advanced settings" on page 88

PCR Using Existing Oligos settings

Select **PCR Using Existing Oligos** to search for suitable PCR primers from among those in the Vector NTI™ *Express* Oligo Database. With the molecule open in the Molecule Editor, select the region to be sequenced, then select **Primer Design > PCR Using Existing Oligos** from the Molecule Editor toolbar.



These settings are similar to the Find PCR Primers settings with a few exceptions. The following are the unique settings for **PCR Using Existing Oligos**. These are also available under the Primers tab if you click on the **Advanced** button:

PCR Using Existing Oligos setting	Description
Sense Primers	Click the **Sense Primers** button and select the desired primer(s) from the Oligo Database.
Antisense Primers	Click the **Antisense Primers** button and select the desired primer(s) from the Oligo Database.

Note: Since you can choose the number of 3′ and 5′ primers, these settings effectively enable you to analyze one 3′ primer against an array of 5′ primers or vice-versa.

Analysis Conditions settings

For information about the **Analysis Conditions** settings, see page 82.

Attach to 5′ Terminus settings

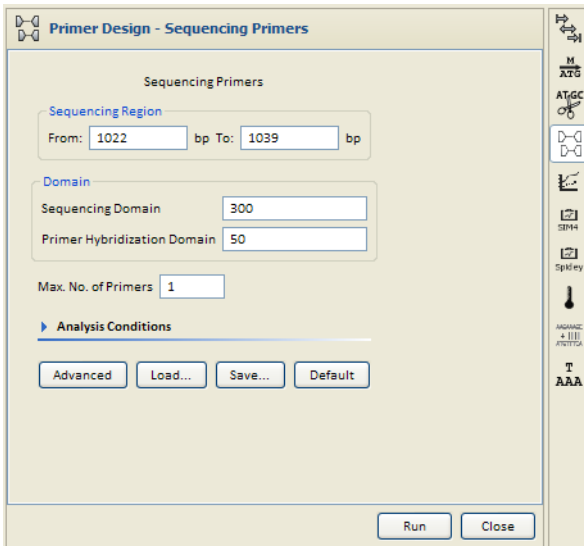
For information about the **Attach to 5′ Terminus** settings, see page 83.

Advanced settings

Click on the **Advanced** button below the main settings to open the Advanced settings dialog. Advanced are described in "Shared Advanced settings" on page 88

Sequencing Primers settings

Select **Sequencing Primers** to find primers for sequencing a DNA molecule fragment. With the molecule open in the Molecule Editor, select the region to be sequenced, then select **Primer Design > Sequencing Primers** from the Molecule Editor toolbar.



Primers will be generated anywhere within the designated "Primer Hybridization Domain" (upstream and downstream).

If the sequencing region is long enough, it is divided by Vector NTI™ *Express* into smaller sequencing domains, areas in which a single sequencing reaction will take place. The size of the primer hybridizing domain may then be set, as well as other primer parameters. Several primer options are evaluated and sorted from best to worst.

Note: Sequencing primers are designed for a sequence region, not an entire molecule.

The following are the unique settings for the **Sequencing Primers** tool. These are also available under the Primers tab if you click on the **Advanced** button:

Sequencing Primers setting	Description
Sequencing Region	Region that you want to sequence. Enter the start and end coordinates of the region to be sequenced.
Sequencing Domain	Enter the number of bases to be sequenced in a single sequencing reaction.
Primer Hybridization Domain	Enter the length of region where primers for each sequencing domain should be sought. Primers are generated within the set domain.

Sequencing Primers setting	Description
Maximum Number of Primers	Enter the number of primers to be found for each sequencing domain. (The actual result may contain fewer primers than this number if there are not enough possible primers.)
DNA/RNA	Select the type of nucleotide sequence.
Complementary Strand	Select if you are sequencing the complementary strand.
User-Defined First Primer	Enter a user-defined nucleotide sequence to be evaluated as a primer for the FIRST sequencing domain, instead of leaving the primer search to Vector NTI™ *Express* Software.

Analysis Conditions settings

For information about additional **Analysis Conditions** settings, see page 82.

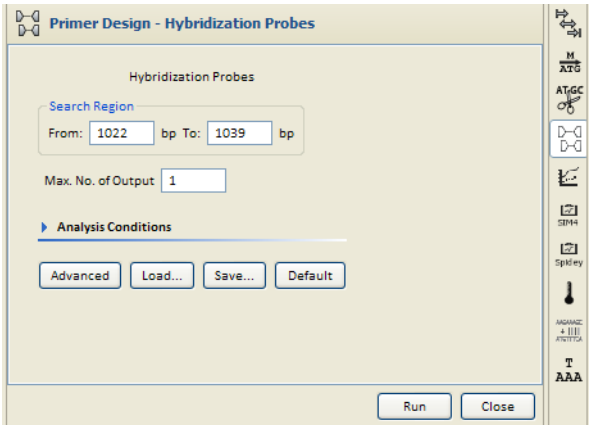
Advanced settings

Click on the **Advanced** button below the main settings to open the Advanced settings dialog. Advanced settings are described in "Shared Advanced settings" on page 88.

Hybridization Probes

Select the **Hybridization Probe** settings to find oligonucleotides that will hybridize to a selected molecule sequence. Vector NTI™ *Express* can generate a set of oligos or use user-defined or database-stored oligos to test for hybridization efficiency with a target molecule.

With the molecule open in the Molecule Editor, select the hybridization region, then select **Primer Design > Hybridization Probes** from the Molecule Editor toolbar.



The following are the unique **Hybridization Probes** settings. These are also available under the Primers tab if you click on the Advanced button:

Hybridization Probes setting	Description
Search Region	Region for which you want to design the probe. Enter the start and end coordinates.
Maximum Number of Output	Enter the number of probes to be found for the region. (The actual result may contain fewer probes than this number if there are not enough possible designs.)
DNA/RNA	Select the type of nucleotide sequence.
Complementary Strand	Select if you are sequencing the complementary strand.
User-Defined Oligo	Enter a user-defined nucleotide sequence to be evaluated as a probe instead of leaving probe search to the software.

Analysis Conditions settings

For information about additional **Analysis Conditions** settings, see page 82.

Advanced settings

Click on the **Advanced** button below the main settings to open the Advanced settings dialog. Additional advanced settings are described in "Shared Advanced settings" on page 88.

Shared Advanced settings

Click on the **Advanced** button below the main settings to open the Advanced settings dialog. The following Advanced settings are the same for all the primer/probe design tools.

Amplicon tab

Click the **Amplicon** tab to customize parameters relating to the resulting PCR product. %GC content for the product or a portion of the product and allowed bases adjacent to the primer annealing site can be specified.:

Amplicon setting	Description
Amplicon %GC	Enter the minimum and maximum for the desired %GC content in the PCR product.
Next to Primer Annealing Site	Choose accepted bases for the four successive bases adjacent to the primer annealing site. Set minimum and maximum %GC range for a specified length of the amplicon adjacent to the primer annealing site.

Structure tab

Click the **Structure** tab to set acceptable limits for nucleotide repeats, palindromes and hairpin loops for the primers. You can also check your primers/product for a selected group of restriction sites from this tab.

Structure setting	Description
Nucleotide Repeats	Enter the maximum permitted length of nucleotide repeats in primers.
Palindromes	Enter the maximum permitted length of palindromes in primers.
Hairpin Loops	Stem Length: Enter the minimum number of base pairs in a hairpin stem. (This value is also used as a minimum stacking length for primer-primer complementarity and primer-primer 3' end complementarity.) Permitted with dG: Check the Permitted box for hairpin loops; enter the minimum permitted value for free energy of hairpin loops. Primers with hairpin loops which have free energy values > /= to this number will be accepted.
Check Hairpin Loops, Palindromes, Nucleotide Repeats and Dimers Only Within 3' Region of ...	Check this box and enter the length of a 3' region if all of a primer's features (repeats, palindromes, hairpin loops, dimers) should be checked only within that 3' region.(If this box is empty, the whole primer will be evaluated.)
Check Primers For Restriction Sites From	Check to find possible cloning sites inside primers and attached nucleotide sequences (if any). In the drop-down menu, specify the REN subset. Enzymes will be checked for the presence of their sites in the primers and attached sequences, and within the PCR product.

Pairs tab

Click the **Pairs** tab to specify how closely parameters such as Tm and %GC, etc. must match between two primers in a generated primer set.

Pairs setting	Description
Tm Difference	Enter difference in degrees Celsius between T_m for sense and antisense primers.
%GC	Enter the difference between GC percentages for sense and antisense primers.
Primer-Primer Complementarity	Check the Permitted box for primer-primer complementarity; enter the minimum permitted value for duplex free energy.
Primer-Primer 3' End Complementarity	Check the Permitted box for primer-primer 3' end complementarity; enter the minimum permitted value for duplex free energy.

Similarity tab

Click the **Similarity** tab to determine the similarity relationship between the primers and the target sequence.

Similarity setting	Description
Best Fit	Check this button to specify the search for site(s) with maximum similarity with no set threshold.
With Similarities > = Similarity Threshold	Check this button to indicate similarity site search above the specified similarity threshold.
Similarity Threshold	Enter the percentage of minimally acceptable similarity.
Last ... Nucleotides Must Have 100% Similarity	Check and specify the number of nucleotides necessary to have 100% complementarity with the target sequence at the 3' end.
Similarity Between Ambiguous Nucleotides	Specifies acceptable similarity between ambiguous nucleotides (if any). The **Average, Minimum,** and **Maximum** buttons indicate that the average, minimum, and maximum possible similarity will be calculated respectively for any nucleotide pair. For instance, if you are calculating similarity between N and A, then the average similarity is 25%, the minimum similarity is 0%, and the maximum similarity is 100%. In case of R and A they are 50%, 0%, and 100%; in case of R and T—0%, 0%, and 0%.

The similarity table used by Vector NTI™ *Express* is as follows:

	N-N	N-R	N-A	R-W	R-A	R-T
Maximum	100	100	100	100	100	0
Average	25	25	25	25	50	0
Minimum	0	0	0	0	0	0

3' End tab

Click the **3' End** tab to set specifications for the 3' end of the primers generated by Vector NTI™ *Express*. Parameters such as dG and specific nucleotide content for the 3' end of both sense and antisense primers can be set here.

3' End setting	Description
dG <=	Specify the maximum permitted value of 3' end free energy.
Length for Analysis	Enter the length of the primer's 3' region that should be analyzed.
Sense Primer 3' Nucleotides	Check the nucleotide boxes to specify permitted last primer nucleotides for the sense primer.
Antisense Primer 3' Nucleotides	Check the nucleotide boxes to specify permitted last primer nucleotides for the antisense primer.

Uniqueness Tab

Click the **Uniqueness** tab to select settings to determine the uniqueness of the generated primers. These parameters can be used to help ensure that generated primers bind to the desired template area with greater specificity than to the rest of the PCR product.

Uniqueness tab setting	Description
Uniqueness Checks for	Choose the area of the molecule to check for primer uniqueness. Either the entire molecule or the Amplicon only can be selected for the uniqueness check.
Max Allowed Similarity	Check this box and enter the similarity threshold to check primer uniqueness on the molecule. Primers which have parasitic hybridization with similarity > /= this threshold will be rejected. Note: this similarity threshold must be </= the minimum similarity required for hybridization of user-defined primers (if any).
Max Consecutive Match for Entire Primer	Check this box and enter the maximum acceptable match of consecutive bases for the entire primer and the Amplicon.
Primer 3' End	Check the first box and enter the number of consecutive 3' bases that must match the amplicon with 100% similarity. Check the second box and specify the maximum acceptable % match between the Amplicon and the designated number of bases on the 3' end of the primer.

Qualities tab

Click the **Qualities** tab to set parameters that govern primer quality by determining how much weight should be assigned to the parameters on other tabs. These values affect scoring functions that evaluate the quality rating of the primer sets generated.

The importance factors are integers between 1 and 10 used in calculating the score evaluating primer/oligo quality. The lower the factor, the less weight given in the calculation. For example, for minimal importance, enter 1 in the appropriate box. For maximum importance, enter 10.

4 BioAnnotator

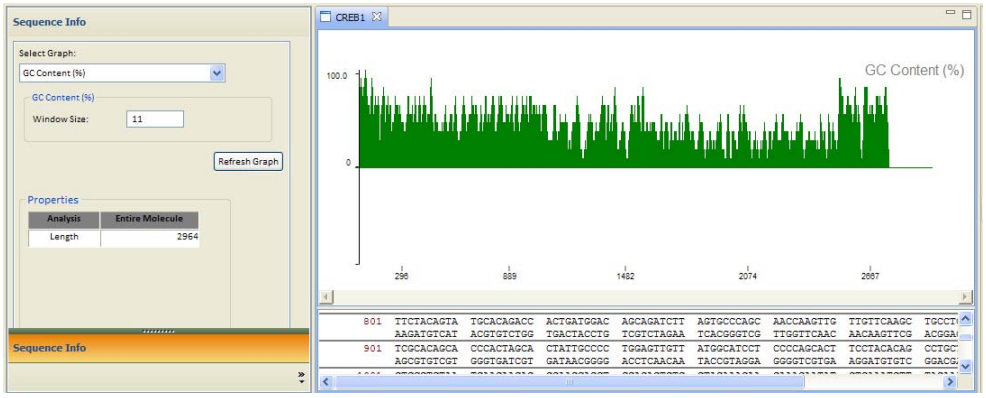
BioAnnotator is a sequence analyzer that performs certain types of DNA/RNA sequence analyses and displays the results as linear graphics.

Launching BioAnnotator

To open BioAnnotator:

- Right-click on a DNA/RNA molecule in the Database Explorer and select **Bioannotator,** or
- With a molecule open in the Molecule Editor, click on the **Bioannotator** button on the right-side toolbar.

The BioAnnotator window contains three panes: a Sequence Info pane, Graph pane, and Sequence Pane.



Selecting an analysis

In the Sequence Info pane:

1. Select the graph to display from the **Select Graph** dropdown list.
2. Select any additional settings for the specific graph (see **Analysis parameters** on page 94).
3. Click on **Refresh Graph** to display the graph.

Graph and Sequence panes

The graphs in BioAnnotator display physiochemical properties of the DNA molecule. The Graph pane consists of the graphical analyses region: a vertical (Y) axis, showing minimal and maximal values of analysis results, and an individual horizontal (X) axis displaying numerical positions in the sequence, scrollbars, and the legend that displays the name of each analysis.

At any point along the sequence (X-axis), the value (Y-axis) for the property is derived not just from the specific base at that point, but from adjacent bases as well. For each property, the algorithm determines the optimum window of adjacent bases to be considered when calculating the value for a point. For instructions in modifying the Window Size parameter, see **Window size** on page 95.

Highlight sequence region in the graph

To highlight a particular data region, drag your cursor in the Graph or Sequence pane and the corresponding region will be highlighted in both panes.

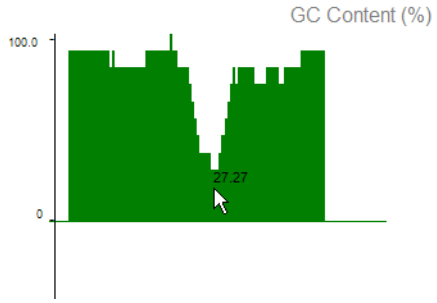
Magnify the graph

In Windows, click in the Graph pane and use the mouse wheel to change the horizontal scale of the graph to zoom in on particular features. A scroll bar will appear below the graph as you magnify.

Determine the value at a point in the graph

To determine the value at a particular point on the graph, hold your cursor over the bar in the graph at that point. A popup number will be displayed, reflecting the value at that point.

If you have magnified the graph and highlighted a region in the sequence, you can see the value calculated for that specific region.



Note: The Window Size parameter described below means that there is not necessarily a 1:1 correlation between a particular base or small group of bases in the sequence and a value in the graph. Due to physiochemical interactions within a sequence, each analysis algorithm necessarily calculates graph values from a "window" of surrounding data points.

Analysis parameters

The following parameters and setting are available in the Sequence Info pane.

Window size

All the graphs have a **Window Size** parameter. This is the number of adjacent data points used in averaging values for each displayed data point, and the optimal value is analysis dependent. The default window size has been optimized based on the selected analysis, but this value can be changed. You can experiment with different window sizes and their effect on the resulting graph by entering a different number in this field.

After any change, click on **Refresh Graph**.

Analyses descriptions and parameters

Melting temperature and free energy are calculated using the "nearest neighbors" method. For constants and algorithms used to calculate thermodynamic parameters, see Appendix B.

The available DNA/RNA analyses and their specific parameters are:

- Free Energy (dG) (kcal/mol)
  - This can be recalculated for a different temperature by entering a new value in the **Temperature** field (°C).
- Melting Temperature (GC Content) (°C)
  - This can be recalculated for a different **Salt Concentration (mMol)** and **% Formamide** of the solution by entering new values in the appropriate fields.
- Sequence Complexity
- GC Content (%)
- Nucleic Acid Distribution (%)
  - Select the bases for which you want to calculate the percent distribution using the checkboxes under **Sequence includes**.
- Melting Temperature (Thermodynamic) (°C)
  - This calculation is dependent on the **Salt Concentration (mMol)** and **% Formamide** of the solution, as well as the concentration of any probe in the solution (**Probe Concentration** in pMol).
- Entropy (dS) (cal/K/mol)
- Enthalpy (dS) (kcal/mol)

After any change, click on **Refresh Graph**.

5 Regenerator

The Regenerator application in Vector NTI™ *Express* Software enables you to back-translate a DNA sequence from a protein sequence, or modify an existing DNA sequence for a given expression system. You can introduce mutations such as insertions, deletions, and substitutions, or modifications like restriction or Gateway® sites, in the back-translated DNA sequence, which can then be optimized for a specific expression system.

Regenerator Workflow

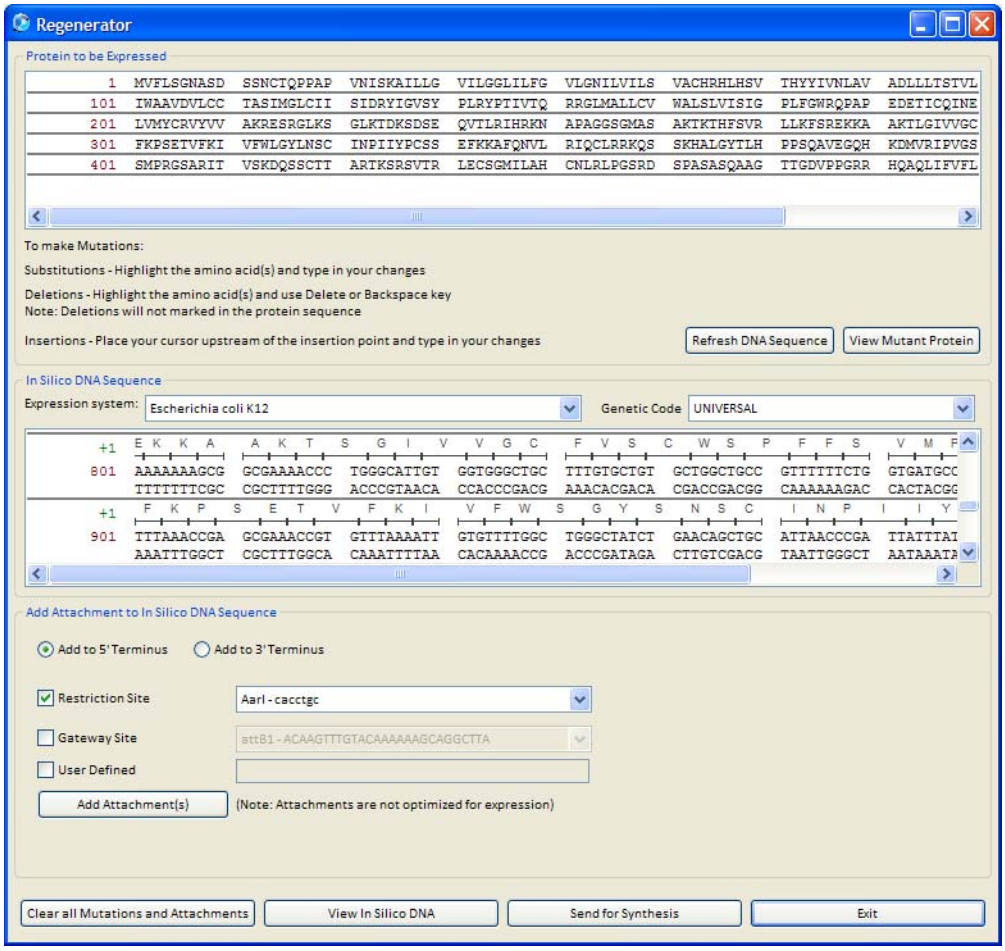
1. Load a protein or DNA sequence into Regenerator from the Molecule Editor.
2. If desired, introduce insertions, deletions, or substitutions into the sequence.
3. Select the desired expression system.
4. If desired, add attachments relevant to downstream cloning to the back-translated DNA sequence.
5. Save the newly created DNA sequence as a molecule in the database and/or send it for synthesis.

Open Regenerator

To launch Regenerator:

1. Load a protein or DNA molecule into the Molecule Editor.
2. Select part of the sequence or make no selection to select the entire sequence.
3. Right-click in the Graphics pane and select **Regenerator** or click on the **Regenerator** button in the Molecule Editor toolbar. 

The Regenerator tool will open.



Regenerator tool features

The Regenerator tool has two main sequence panes, as well as various controls for editing the sequence:

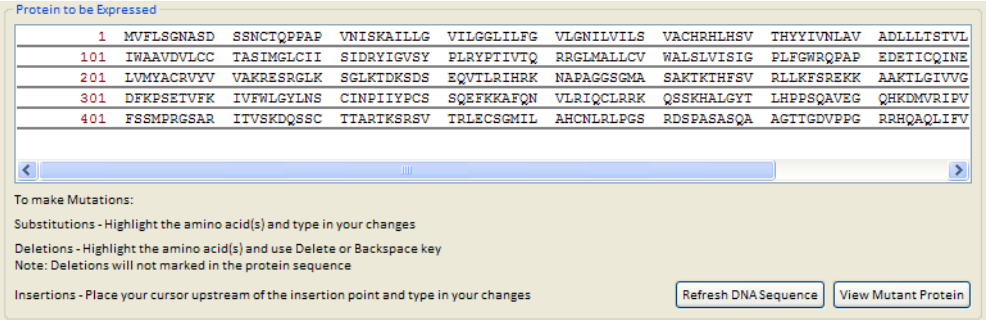
- The top pane contains the **input amino acid sequence or DNA sequence** and is editable, allowing you to add or delete amino acids or bases directly in the sequence.
- The middle pane contains the ***in silico* DNA sequence** generated from the mutations selected or entered in the tool.

Create mutations in the input sequence

You can insert, delete, or substitute amino acids (for proteins) or bases (for DNA) in the input sequence directly.

- To insert amino acids or bases, place your cursor upstream of the insertion point and type in your desired changes.

- To substitute one or more amino acids or bases, highlight the desired part of the input sequence and type in your desired changes.
- To delete part of the input sequence, highlight the sequence and use the **Delete** or **Backspace** key.



Clear mutations

Click the **Clear all Mutations and Attachments** to remove mutations and restore the original sequence.

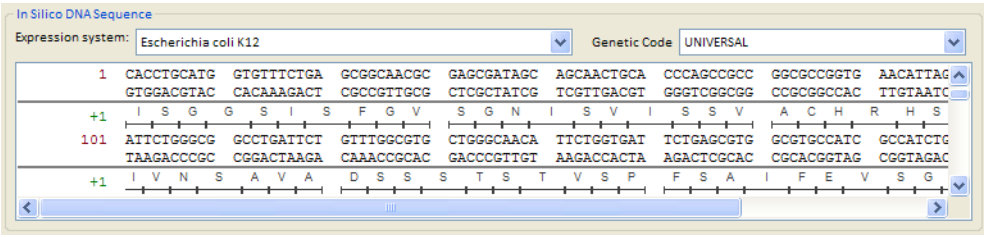
View the mutated sequence

The **View Mutant Protein/DNA** button is enabled once any mutation is introduced into the input sequence pane. Click this button to view the mutated protein or DNA sequence in the Molecule Editor.

The protein is given a name using the following convention **VNTI_<Protein Name>_mutated**.

Refresh the *in silico* DNA sequence

To update the **In Silico DNA Sequence** pane with any mutations, click on the **Refresh DNA Sequence** button.



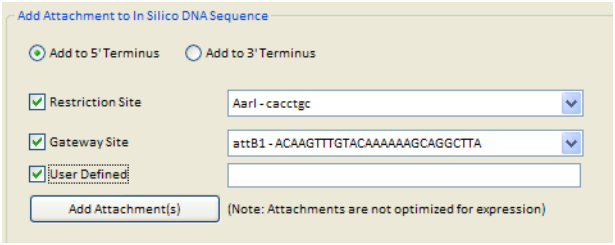
Optimize the expression system and genetic code

- Select the desired **Expression system** from the dropdown list. The *in silico* DNA sequence will automatically update based on an internal codon usage table for the expression system in Vector NTI™ *Express*.
- Select the desired **Genetic code** from the dropdown list. The *in silico* DNA sequence will automatically update based on the selection.

Add attachments

You can add the following attachments to the **5′** and **3′** ends of the *in silico* DNA sequence:

Note: The **3′ end** attachment will always be the **Reverse Complement** of the sequence displayed in the attachments dialog box.



- Select the **Restriction Sites** checkbox and then select from the dropdown list to add a restriction site to the selected end(s) of the *in silico* DNA sequence. The selected Restriction Site appears in the DNA sequence pane at the 5′ and 3′ end, based on your selection.
- Select the **Gateway® Sites** checkbox to add attB sites for Gateway® cloning to the back translated molecule. In the **Choose attB Extension** dialog box that appears based on the terminus you have selected (5′ or 3′), select a fragment from the list and click **OK.** The selected Gateway® site appears in the DNA sequence pane at the 5′ or 3′ end, based on your selection.
- Select the **User Defined** checkbox to add special sequences (promoter, tags, etc.) to the back translated molecule. Enter the sequence in the field; the The sequence will appear in the DNA sequence pane at the 5′ or 3′ end, based on your selection.

Note: The attachments made to the DNA sequence will not appear in the protein sequence pane.

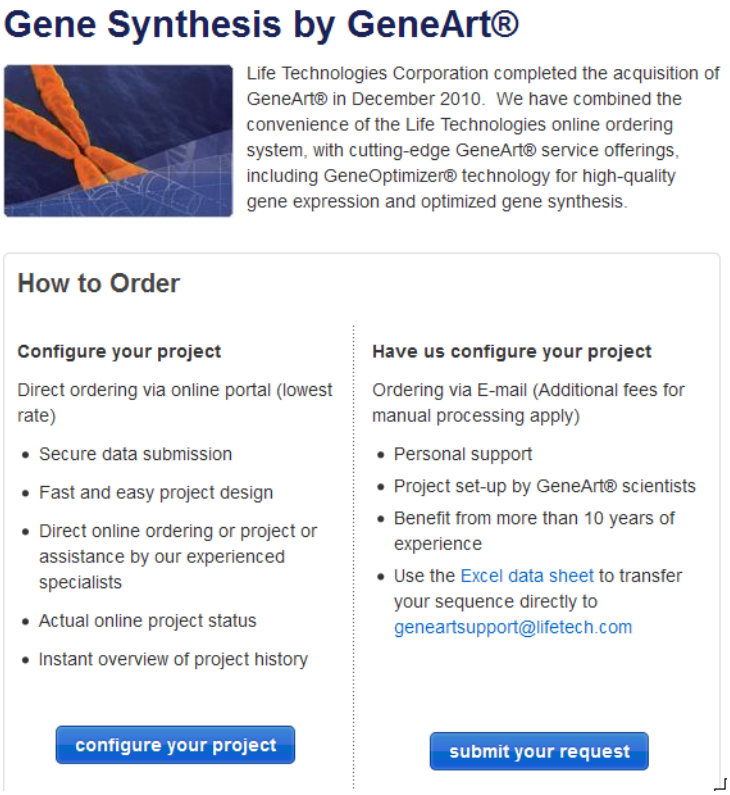
Clear attachments

Click on **Clear All Mutations and Attachments** to clear all attachments and restore the original sequence.

Generate a new sequence and send for synthesis

- Click on **View In Silico DNA** to open the newly created DNA molecule in the Molecule Editor You can then save the molecule, and copy and submit the sequence for synthesis as described below.

- Click on the **Go to Gene Synthesis** button to go to the Thermo Fisher Scientific GeneArt® gene synthesis service website. There you can create an account and submit the *in silico* DNA sequence for synthesis.



6 BLAST and Entrez Searches

Vector NTI™ *Express* Software includes the search engines BLAST and Entrez Query for querying sequences.

BLAST search

BLAST (Basic Local Alignment Search Tool) is a search engine for exploring available public sequence databases for DNA or protein sequence similarities to a query sequence. BLAST programs have been designed for speed, with a minimal sacrifice of sensitivity to distant sequence relationships. BLAST scores have a well-defined statistical interpretation, making real matches easy to distinguish from random background hits. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity.

For detailed information on BLAST search types, settings, parameters, search databases, etc., visit **http://blast.ncbi.nlm.nih.gov/Blast.cgi**.

Open the BLAST search tool

- To BLAST search an entire molecule in the database:
  - Select the molecule in Database Explorer, right-click, and select **BLAST Search**.
  - Open the molecule in the Molecule Editor and click on the **BLAST** button on the main toolbar.

- To BLAST search part of a sequence, open the sequence in the Molecule Editor, select the desired part of the sequence in the Graphics or Sequence pane, and right-click to select **BLAST sequence**.
- To open a blank BLAST search window and type and paste a sequence directly into the tool, click on the **BLAST** button on the main toolbar without selecting a molecule first.

BLAST search settings

The BLAST search tool settings are displayed in the right-hand pane of the tool; the results of a search are displayed in the left-hand pane.

Select the following settings to define your search:

Program

In the Program dropdown menu, specify the type of database search to be performed:

blastn	Compares a nucleotide query sequence against a nucleotide sequence database.

blastp	Compares an amino acid query sequence against a protein sequence database.
blastx	Compares a nucleotide query sequence translated into all reading frames against a protein sequence database.
tblastn	Compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).
tblastx	Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. *This program cannot be used with the nr database.*

Algorithm

In the Algorithm drop-down menu, specify the type of algorithm search to be performed. In addition to blastn and blastp, algorithm searches include:

megablast	Concatenates many queries to save time spent scanning the database. It is optimized for aligning sequences that highly similar and is up to 10 times faster than more common sequence similarity programs. It can be used to quickly compare two large sets of sequences against each other. MEGABLAST permits searching with batches of ESTs or with large cDNA or genomic sequences.
PHI blast	Pattern Hit Initiated BLAST. A program for searching a protein database using a protein query; it seeks only alignments that preserve a specified pattern contained within the query.
PSI blast	Position Specific Iterated BLAST. A program for searching protein databases using protein queries to find other members of the same protein family.

Database

In the dropdown list, select the GenBank database to query:

Menu item	Description
nr/nt	Peptide Sequence Database: All non-redundant GenBank sequences and CDS translations. Nucleotide Sequence Database: All GenBank+EMBL+PDB sequences (no EST, STS, GSS or phase 0, 1 or 2 HTGS sequences). No longer non-redundant.
Refseq RNA	Human Genome BLAST databases: human RefSeq mrna with NM_#### or XM_#### accessions
Refseq Genomic	Human Genome BLAST databases: human genomic contig sequences with NT_#### accessions
Chromosome	Nucleotide Sequence Database: complete chromosomes
EST	Nucleotide Sequence Database: EST (Expressed Sequence Tags)—mouse and human.
EST-others	Nucleotide Sequence Database: EST (Expressed Sequence Tags)—other than mouse and human.

Menu item	Description
GSS	Nucleotide Sequence Database: Genome Survey Sequence, includes single-pass genomic data, exon-trapped sequences, and Alu PCR sequences.
HTGS	Nucleotide Sequence Database: Unfinished High Throughput Genomic Sequences.
PAT	Protein sequences from the Patent division of GenBank.
PDB	Peptide Sequence Database: *Saccharomyces cerevisiae* protein sequences—genomic CDS translations. Nucleotide Sequence Database: *Saccharomyces cerevisiae* genomic nucleotide sequences.
Alu repeats	Peptide Sequence Database: Translations of select Alu repeats from REPBASE. Nucleotide Sequence Database: Select Alu repeats from REPBASE.
DBSTS	Nucleotide Sequence Database: Database of GenBank+EMBL+DDBJ sequences from STS Divisions.
ENV-NT	Sequences from environmental samples, such as uncultured bacterial samples isolated from soil or marine samples. The largest single source is the Sargasso Sea project. This does not overlap with the Nucleotide Sequence Database.

Limit by Entrez Query

This checkbox lets you limit the BLAST search to the results of an Entrez query against the database chosen. This can be used to limit searches to subsets of the BLAST databases.

Select the checkbox and in the field below enter terms that would normally be allowed in an Entrez search session. For example:

protease NOT hiv1[Organism] — this form of a search will limit the search to all proteases, except those in HIV 1.

biomol_mrna[PROP] AND brain — this form of a search can be used to limit searches to a particular molecule type

Mus musculus[Organism] — this form of a search will limit the search to a specific organism. Enter the name of the organism in the Entrez Query field with the [Organism] qualifier.

General Parameters

The parameters here are almost identical to parameters for the various BLAST searches at the NCBI website. For more information regarding these parameters, visit http://www.ncbi.nlm.nih.gov/blast/html/blastcgihelp.html.

Max target sequences—The maximum number of aligned sequences to display.

Expect threshold—The statistical significance threshold for reporting matches against database sequences. The default value of 10 means that in a database of the current size, 10 matches would be expected merely by chance (stochastic model of Karlin and Altschul, 1990.) Hits showing a statistical significance greater than the Expect threshold are not reported. Increasing the E value above 10 produces a larger list with more low-scoring hits (chance matches). Lower expectation value thresholds are more stringent, leading to fewer chance matches being reported.

If your query peptide or nucleotide sequence is short, you might want to increase the Expect value. Because a short query is more likely to occur by chance in the database, even a perfect match can have low statistical significance and may not be reported. Increasing the E value lets you look farther down the hit list and see matches that would normally be discarded because of low statistical significance.

Word size—Word size is roughly the minimal length of an identical match an alignment must contain if it is to be found by the algorithm. Mega BLAST is most efficient with word sizes 16 and larger, although word size as low as 8 can be used. If the value W of the word size is divisible by 4, it guarantees that all perfect matches of length W + 3 will be found and extended by Mega BLAST search, however perfect matches of length as low as W might also be found, although the latter is not guaranteed. Any value of W not divisible by 4 is equivalent to the nearest value divisible by 4 (with 4i+2 equivalent to 4i).

Scoring Parameters

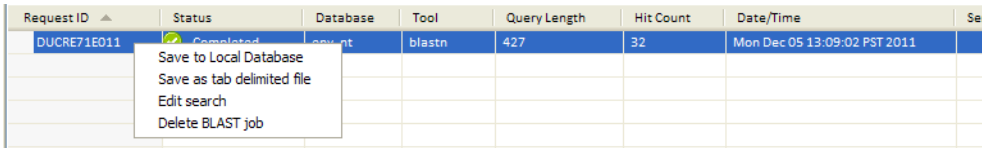
- **Match/Mismatch Scores**—Reward and penalty for matching and mismatching bases. Many nucleotide searches use a simple scoring system that consists of a "reward" for a match and a "penalty" for a mismatch. The (absolute) reward/penalty ratio should be increased as one looks at more divergent sequences. A ratio of 0.33 (1/-3) is appropriate for sequences that are about 99% conserved; a ratio of 0.5 (1/-2) is best for sequences that are 95% conserved; a ratio of about one (1/-1) is best for sequences that are 75% conserved.
- **Gap costs**—The penalty to open a Gap and penalty to extend a Gap. Increasing the Gap Costs decreases the number of Gaps introduced in the alignment.

Filters

- **Low-complexity**—This filter masks off segments of the query sequence that have low compositional complexity (as determined by the SEG program of Wootton & Federhen, Computational Chemistry, 1993). Regions with low-complexity sequence can create problems in sequence similarity searching by producing artificial hits, sequences that are not truly related. Such hits can produce high scores because of the presence of low-complexity regions.
- **Human Repeats**—This option masks Human repeats and is especially useful for human sequences that may contain these repeats.
- **Mask for Lookup**—This option masks only for purposes of constructing the lookup table used by BLAST. The BLAST extensions are performed without masking.
- **Mask lower case characters**—With this option selected you can cut and paste a FASTA sequence in upper case characters and denote areas you would like filtered with lower case. This allows you to customize what is filtered from the sequence during the comparison to the BLAST databases.

Perform the BLAST search

1. To perform the search, click on **Submit**.
2. The search progress will be displayed left-hand pane of the BLAST search viewer. When the search is complete, the search will be flagged as **Complete**.
3. Right-click on the search result and select from the following options:
   - **Save to Local Database**—saves the search in the local database.
   - **Save as tab delimited file**—save as a text file.
   - **Edit search**—change the search parameters and re-**Submit**.
   - **Delete BLAST job**—clears the results from the list.

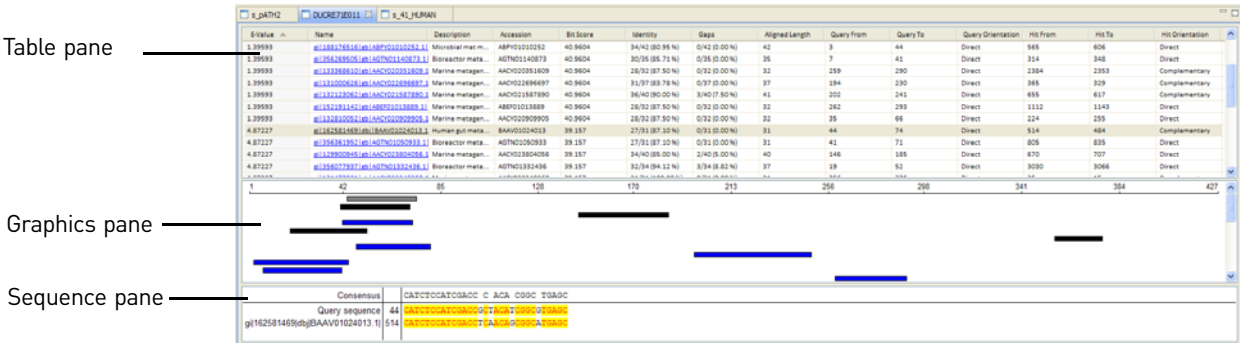


4. Double-click on the search result to open the result in the BLAST Result Viewer.

Note: BLAST results that have been saved to the database may be opened in Database Explorer; click on **Results** and open the **BLAST Results** folder.

BLAST Result Viewer

Double-click on a BLAST search result in Database Explorer or the BLAST tool to open the result in the BLAST Result Viewer.



The Viewer contains three main panes:

- The **Table pane** displays a textual list of query "hit" molecules for the sequence
- The **Graphics pane** displays the corresponding sequence of each hit in graphical form
- The **Sequence pane** displays the Query sequence, "hit" sequence (with Accession number), and Consensus sequence

Click on a row in the Table pane or a feature in the Graphics pane to display the corresponding sequence in the Sequence pane.

Table pane columns

- **E-Value:** This value reflects the likelihood that the similarity between the sequences would occur by chance when searching a database of a particular size. A zero or extremely low number suggests that the match is so perfect that it is extremely unlikely that the similarity would occur randomly.
- **Name:** The NCBI sequence identifier of the query hit.

- **Accession:** The GenBank Accession number of the hit.
- **Bit Score:** A measure of how close the identity of the match is to the query sequence.
- **Identity:** The ratio (and percentage) of matching residues in the hit elements. The numbers n/n refer to the number of identical residues out of the number of matches in the hit element. This is important to consider when determining the significance of this statistic. A high identity percentage may mean nothing if a low number of nucleotides is being compared.
- **Positives:** the ratio (and percentage) of similar residues in the hit elements
- **Query From/To/Orientation:** The start and end position numbers in the query sequence matching that of the hit element, and the strand that corresponds to the hit element.
- **Hit From/To/Orientation:** The start and end position numbers in the hit sequence matching that of the query element, and the strand that corresponds to the query element.

Download sequence features and save them as molecules

In the Table pane, click on the link in the **Name** column to download the sequence and database information for that query molecule from the NCBI database and save it as a separate molecule in Vector NTI™ *Express* Software. The database information for the sequence will be populated in the Properties of the created molecule.

Note: You will be prompted to submit your email address when connecting to the NCBI database.

Entrez Search

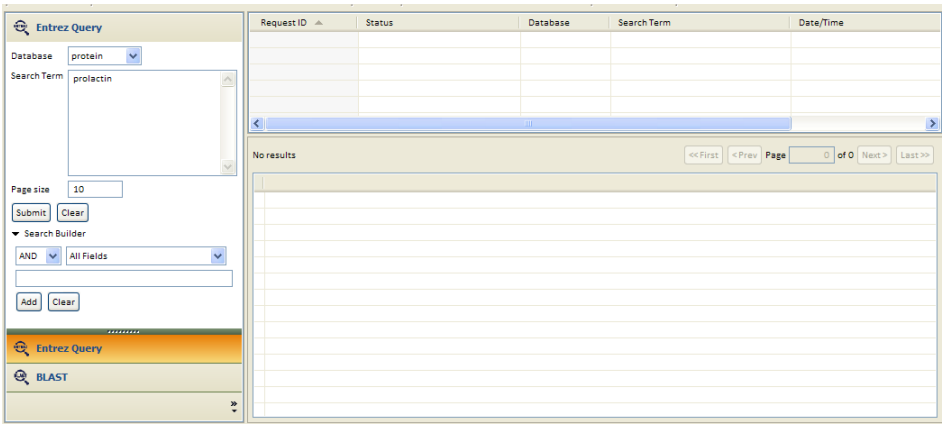
You can use the NCBI Entrez cross-database search engine directly from Vector NTI™ *Express* Software. For more information about this search engine, visit **www.ncbi.nlm.nih.gov/sites/gquery**.

Note: An Internet connection is required to perform this search.

Open Entrez search tool

To perform an Entrez search, click on the **Public Database Search** button on the main toolbar.

The Entrez search tool will open.



The Entrez search tool includes the following panes:

- **Query pane**—Contains settings for formulating a query.
- **Query List pane**—Lists the status of each query request.
- **Results pane**—Displays the results of a query.

Entrez search settings

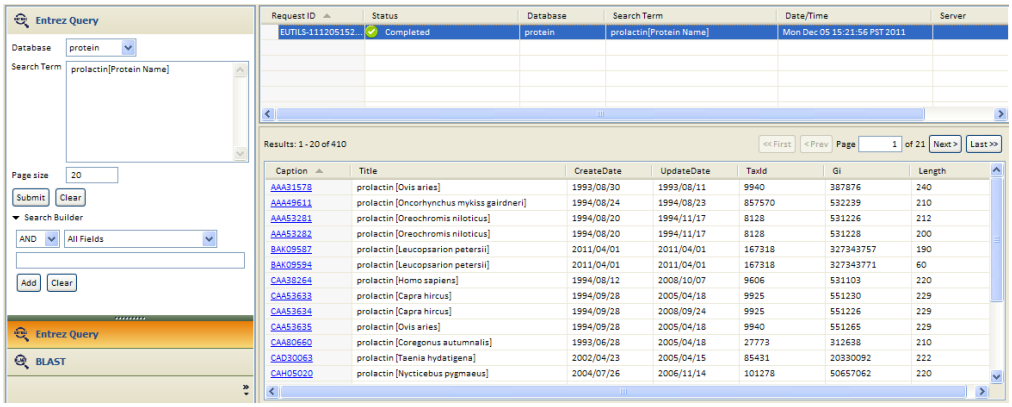
The Entrez search engine allows you to search across a wide variety of NCBI databases.

1. To begin, select the desired NCBI database from the **Database** dropdown list:
   - **PubMed:** A database of biomedical literature citations and abstracts from MEDLINE, life science journals, and online books
   - **Gene:** Wide-ranging database containing nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.
   - **OMIM:** Database of human genes and genetic phenotypes.
   - **Protein:** Database of translations from annotated coding regions in GenBank, RefSeq and TPA, as well as records from SwissProt, PIR, PRF, and PDB.
   - **Structure:** Database of three-dimensional genetic and protein structures.
   - **Popset:** Database of DNA sequences that have been collected to analyze the evolutionary relatedness of a population.
   - **Nucleotide:** Database of genome, gene and transcript sequences from several sources, including GenBank, RefSeq, TPA and PDB.
   - **SNP:** Database of single nucleotide polymorphisms (SNPs).
2. Search terms will vary by database. You can enter known search terms directly in the **Search Term** field, or you can use the **Search Builder** function to construct your search from known database elements and Boolean operators. For a description of search terms and operators, visit **www.ncbi.nlm.nih.gov/books/NBK3837/**.
3. Select the desired **Page size**.
4. When you have made your selections, click on **Submit**.

5. The search progress will be displayed Query List pane. When the search is complete, the search will be flagged as **Complete** and the search results will be displayed in the Results pane.

Entrez search results

Entrez search results are displayed in tabular form in the Results pane.



The table of results includes columns appropriate for the database you are querying.

Click on the link in the Results pane to open the online database record for each result.

Note: Entrez queries are not saved in the Vector NTI™ *Express* Software database.

Editing and deleting queries

- To delete an Entrez query, right-click on it in the Query List pane and select **Delete search job**.
- To edit a query, right-click on it in the Query List pane and select **Edit search**.

7 GenomeBench

GenomeBench can be used to download, view, analyze, annotate, and save local copies of reference genomic DNA sequences from several principal Distributed Annotation System (DAS) servers.

GenomeBench was designed to support the following workflows:

- Retrieval of genomic data from public DAS servers.
- Annotating and creating a local copy of a genomic region.

In support of these workflows, it accepts megabase-sized genomic sequences and all attendant track annotations. All annotated features are linked to their corresponding region of the genomic sequence backbone; additional information for any sequence-based annotation (for example, mRNAs, ESTs, and STSs), including other defined features from the GenBank record, is easily downloadable. Annotated sequences can be aligned with the genomic backbone simply by dragging and dropping into an alignment window.

Proprietary DNA sequences stored in the Vector NTI™ *Express* local database can also be positioned along genomic backbones using Spidey or Sim4 alignment algorithms to determine the intron-exon structure of genes of interest. Genomic sequences in GenomeBench can be exported to Vector NTI™ *Express* for further sequence analysis, such as PCR primer design or the creation of spliced transcripts.

GenomeBench also provides the capability to query public databases and perform sequence and feature searches.

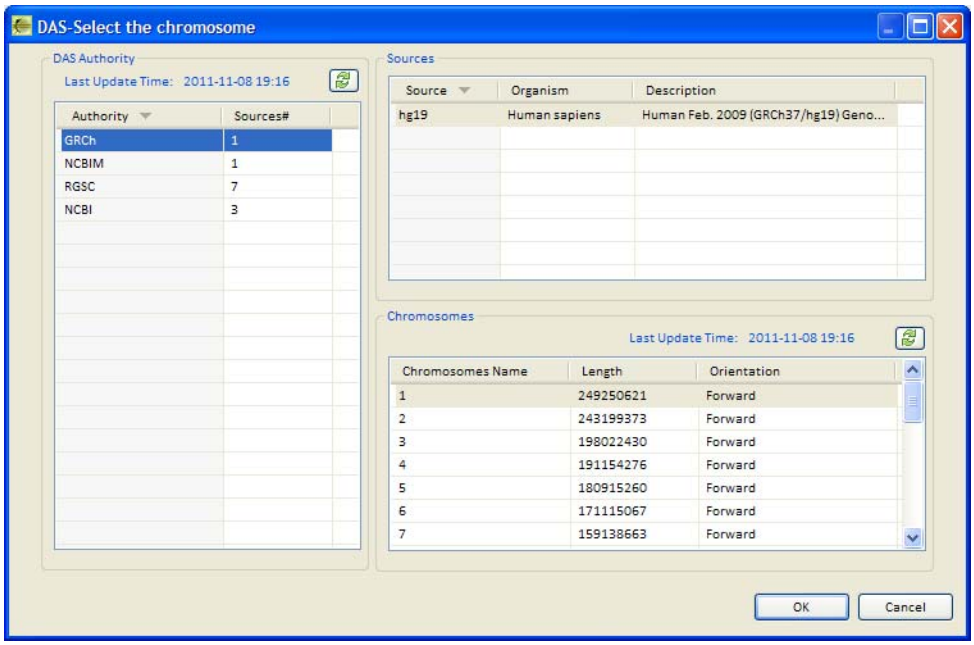
Download data from public DAS servers

Vector NTI™ *Express* has preconfigured DAS servers

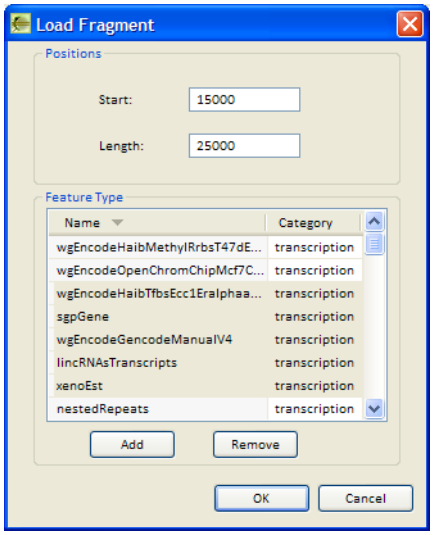
To launch GenomeBench and download data:

1. Click on the **GenomeBench** button on the main toolbar.
2. When you first open the application, an empty project will be displayed.
3. In the GenomeBench toolbar (on the right side of the window), click on the **Load from DAS Server** button to download genomic data from one of the pre-configured public DAS servers:
   - Genome Reference Consortium: human (GRCh)
   - National Center for Biotechnology Information mouse genome assembly (NCBIM)
   - Rat Genome Sequencing Consortium (RGSC)

– NCBI's 32K BAC array, BAC end-sequence pairs, and Fosmid clone end-sequence pairs



4. In the DAS Server dialog, select the server from the DAS Authority list.
5. Select the genomic source from the **Sources** list.
6. Select the desired chromosomes from the **Chromosomes** list, then click on **OK**.
7. In the Load Fragment dialog, specify the base range of the sequence to download using the **Start** and **Length** fields
8. Select the feature information to download from the **Feature Type** list, which displays a selected list of features available in the DAS databases.



– Use **Ctrl+click** and **Shift+click** to make multiple selections.
– Click on **Remove** to remove selected features from the download.

– Click on **Add** to open the **Feature Types Filter** and select from all available feature types to add them to the **Feature Type** list.

9. Click on the **OK** in the Load Fragment dialog to begin the download.

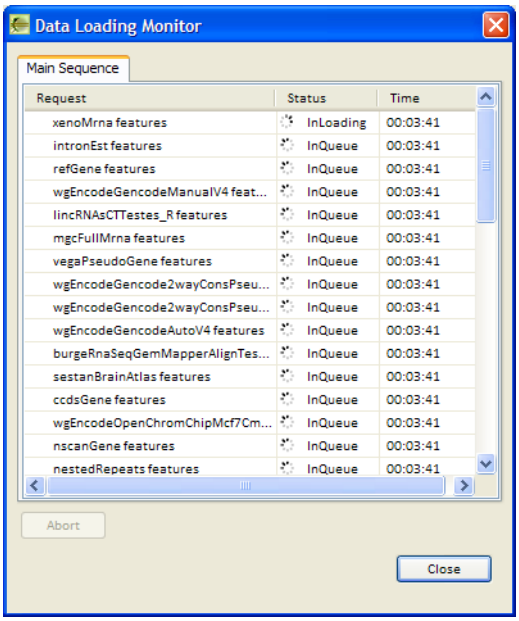
Note: It can take some time for all the features to load, but you can start working on features that have loaded before the remaining features are loaded.

10. The download time can vary considerably depending on the server connection and data volume.

11. When the download is finished, the GenomeBench Project viewer will display the project.

Data Loading Monitor

You can monitor the download progress by clicking on the **Download Progress** button on the GenomeBench toolbar.



In the **Data Loading Monitor**, features are listed as “Loading” or “In Queue.” To delete an “In Queue” feature in the list, select it and click on **Abort**.

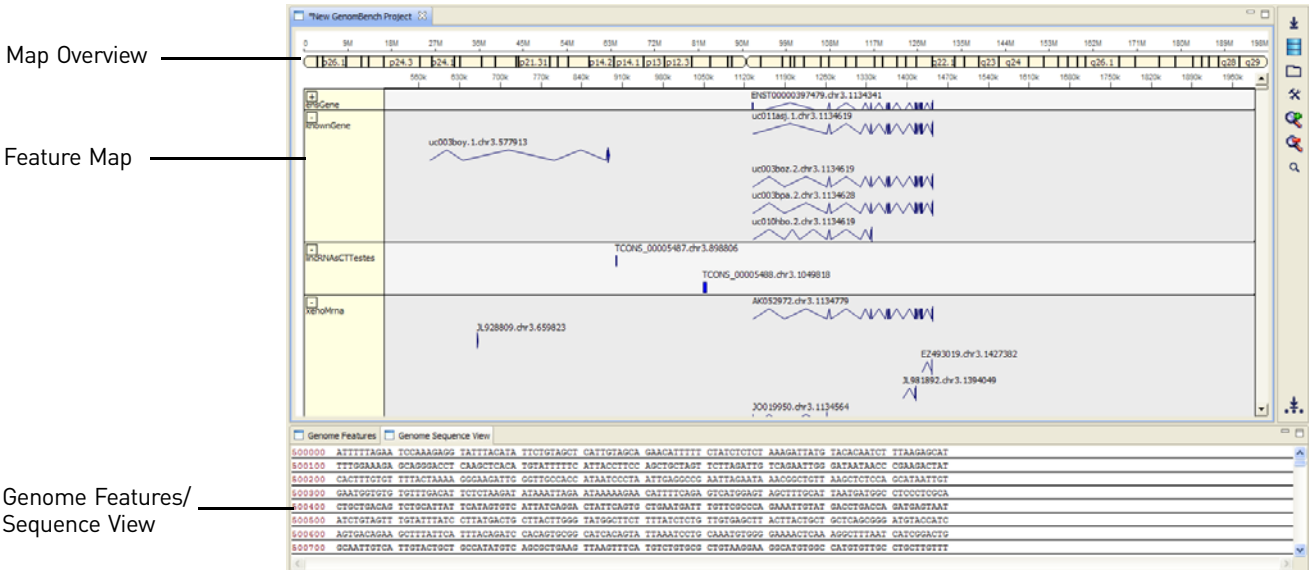
Local GenomeBench Projects

GenomeBench Projects that have been downloaded can be saved in the local Vector NTI™ *Express* database for future access:

- To save a project to the local database (i.e., after downloading the data from the DAS server), select **File > Save As** and specify a name for the project in the dialog box.
- To open a saved project, open GenomeBench and click on the **Load from Local Database** button. Then select the project name from the **Existing Genome Projects** dialog.

GenomeBench Project Viewer

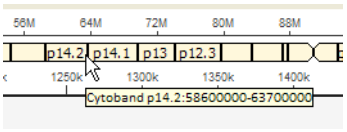
The GenomeBench Project Viewer contains the following panes: Map Overview, Feature Map, Genome Features, and Genome Sequence View.



Map Overview

The Map Overview pane presents a graphical depiction of the entire chromosome and has the following features:

- A ruler marks the length of the chromosome.
- The rod-shaped band depicts the sequence/chromosome.
- Narrowed areas on the rod indicate centromeric regions.
- Cytoband patterns are indicated by the vertical dark lines that run the length of the rod. Pause the cursor over a particular cytoband to display a tool tip with information about the cytoband.



Centromeric regions and cytoband patterns are shown only for those DAS servers that provide the information in the correct format.

Feature Map

The Feature Map Pane shows a graphical depiction of the features for the currently loaded fragment.

Like the Overview Pane, the Feature Map Pane has a ruler at the top to help you visualize the orientation of features along the length of the fragment.

Feature types or tracks display along the left side of the Feature Map Pane. Different features are depicted with various graphical representations and are labeled with the feature name. Shown features are listed along with relevant feature information in the Genome Features pane.

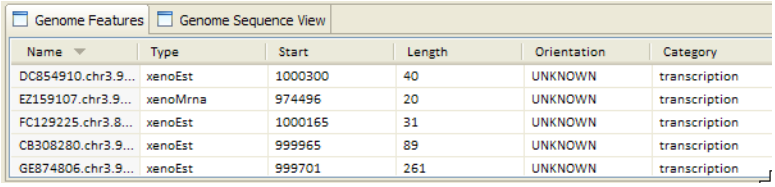
The associated GenBank accession number is derived from the Target Accession number of the feature.

Magnifying tools

Use the magnifying tools to the right of the Feature Map to magnify features in the pane horizontally.

Genome Features and Genome Sequence

The **Genome Features** pane lists the features in the map by name, and indicates their type, location, orientation, and feature category.



Name	Type	Start	Length	Orientation	Category
DC854910.chr3.9...	xenoEst	1000300	40	UNKNOWN	transcription
EZ159107.chr3.9...	xenoMrna	974496	20	UNKNOWN	transcription
FC129225.chr3.8...	xenoEst	1000165	31	UNKNOWN	transcription
CB308280.chr3.9...	xenoEst	999965	89	UNKNOWN	transcription
GE874806.chr3.9...	xenoEst	999701	261	UNKNOWN	transcription

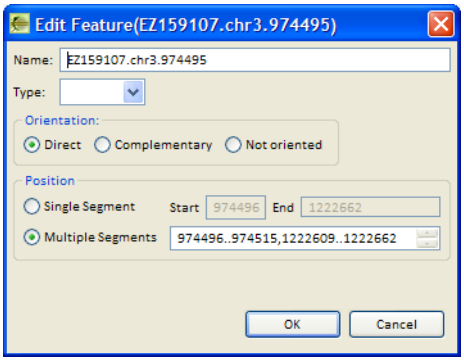
The **Genome Sequence** pane displays the entire downloaded sequence

Editing features

In the Feature Map and Genome Features list, the right-click menu allows you to create, edit, or delete features.

Right-click on a feature in the map or list and select:

- **Edit the selected feature:** Edit basic settings and information for the feature, including feature name, type, orientation, and position in the segment or multiple segments.



- **Create New Feature**: Create a new feature, using the same settings and fields as in the **Edit Feature** dialog.

- **Delete the selected feature**: Deletes the selected feature.

8 Align Multiple Sequences

The simultaneous alignment of multiple nucleotide or amino acid sequences is an essential tool in molecular biology. Alignment enables you to design PCR primers for amplifying a region of aligned DNA/RNA molecules. Multiple alignments are used to find diagnostic patterns, characterize protein families, as well as detect or demonstrate a similarity between new sequences and existing families of sequences. They are also useful in predicting secondary and tertiary structures of new sequences, suggesting oligonucleotide primers for PCR and serving as an essential prelude to molecular evolutionary analysis.

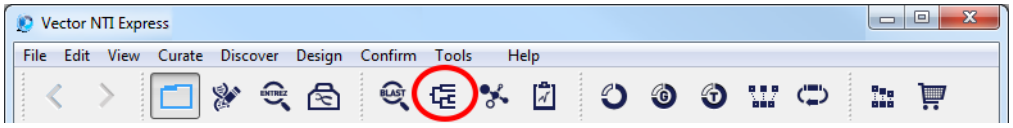
AlignX® is the multiple sequence alignment tool of Vector NTI™ *Express* Software. In addition to aligning sequences, it can be used to create and manage sequence alignment projects. It uses a modified Clustal W algorithm and incorporates the following features:

- Phylogenetic tree construction, displayed in graphical representation
- Use of residue substitution matrices
- Secondary structure consideration
- Multicolored alignment presentation
- Automatic consensus calculation
- Full alignment editing capabilities

Open AlignX®

There are several ways to open the AlignX® tool:

- Click on the **AlignX** button on the main toolbar.



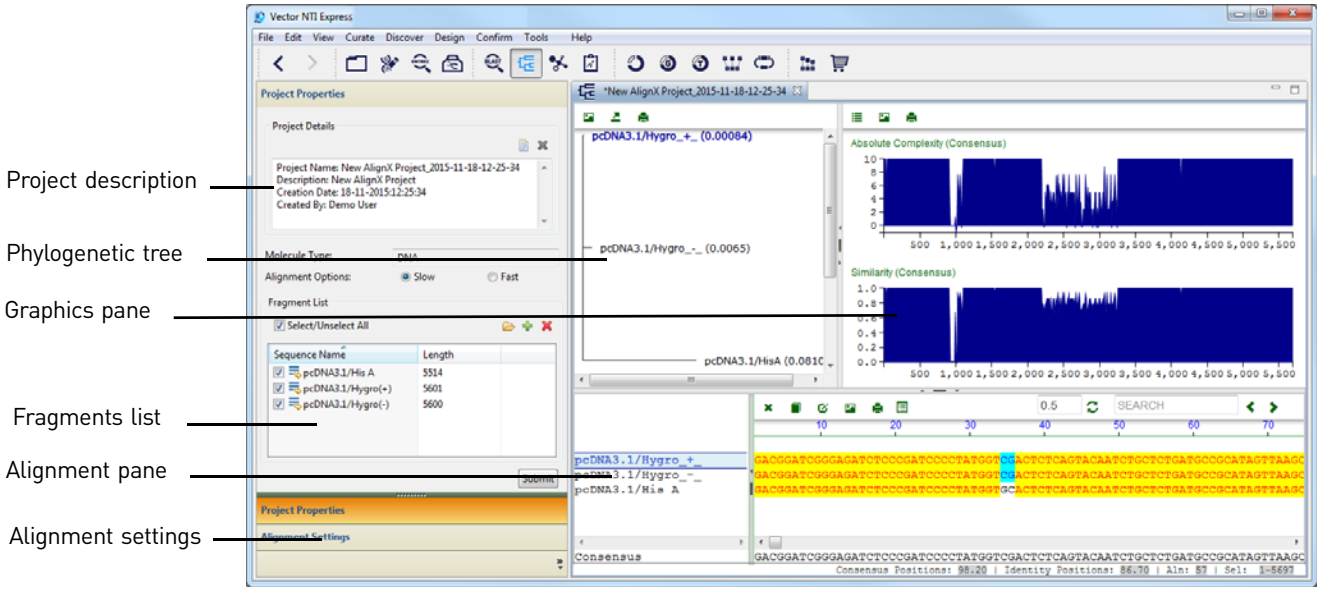
- In Database Explorer, right-click on a molecule or Ctrl+click on multiple molecules and select **AlignX - Align selected Molecule(s)** to load those molecules into the tool.
- Load an existing AlignX® project as described in "Manage AlignX® projects" on page 118.

AlignX® window

The AlignX® window consists of the following panes:

- **Project Properties pane**: Contains the Project Description and the Fragments list. As molecules are added to an AlignX® project, they are listed in the Fragments list.
- **Alignment Settings pane:** Contains the settings used to perform the alignment.
- **Graphs pane:** Displays a graphical representation of the aligned molecules, showing both similarity and absolute complexity for consensus.
- **Alignment pane:** Displays the aligned sequences, with aligned regions highlighted

- **Phylogenetic Tree pane**: A graphical representation of the degree of similarity among sequences (for three more sequences).



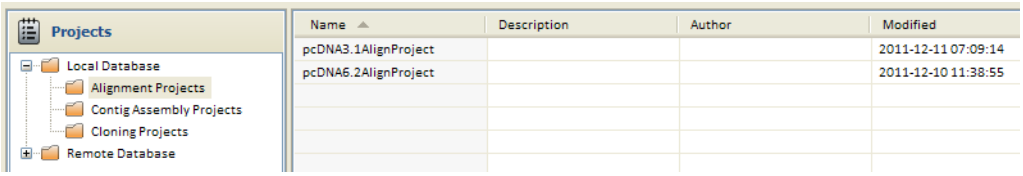
Manage AlignX® projects

Save and rename a project

- Following analysis, a new AlignX® project is automatically named `New AlignX Project_<timestamp>`. To edit the name or description, click **Edit Project Properties**.
- To save changes to a project, select **File > Save** or **File > Save As**. Save the project as either an Alignment Project (*.aprx) or Multiple Sequence Formats (*.msf).

Open a project

- To load an existing project in the database, in Database Explorer, go to the **Projects** list, double-click on the **Projects** folder, select the **Alignment Projects** folder from the Local Database, and double-click on the AlignX® project in the list to open it
  Use this method to load projects saved in earlier versions of VectorNTI. While you can open projects from the database, you cannot save to the database from Vector NTI Express.



- To load an existing project that has been saved as a **.apr, .aprx,** or **.msf** file, select **File > Open > Alignment Project**, thenchoose between the supported formats.

Close a project

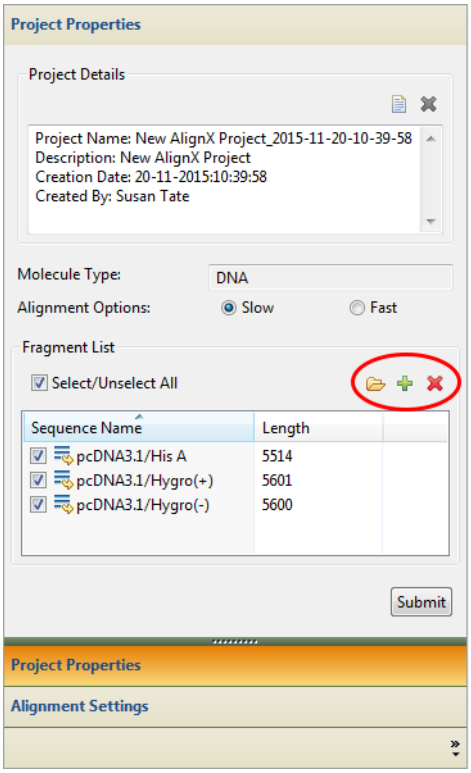
To close a project, click **Close Project**. If there are unsaved changes, you will be prompted to save the project before closing.

Save consensus sequence

After you perform an alignment, you can save the consensus sequence as a separate molecule in the database. Right-click over the consensus sequence, and select **Save As Molecule**.

Select fragments to align

The tools for selecting the fragments to align are located in the Fragment List region of the Project Properties pane.

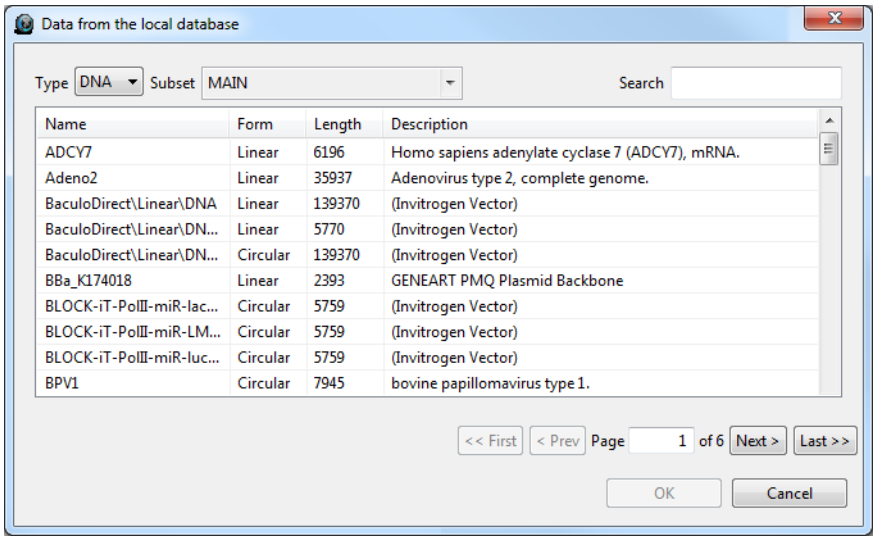


Add fragments

To add fragments:

- Click **Add molecule(s) from file(s) to this list for alignment** to select **.gb, .gp,** or **fasta** sequence files.
- Click **Add molecule(s) from the database to this list for alignment** to select DNA or Protein molecules from the local database.

Note: Use the drop-down menu to toggle between molecule types (DNA vs. Protein). Press the **Ctrl** or **Shift** keys to select multiple files/molecules.



Remove fragments

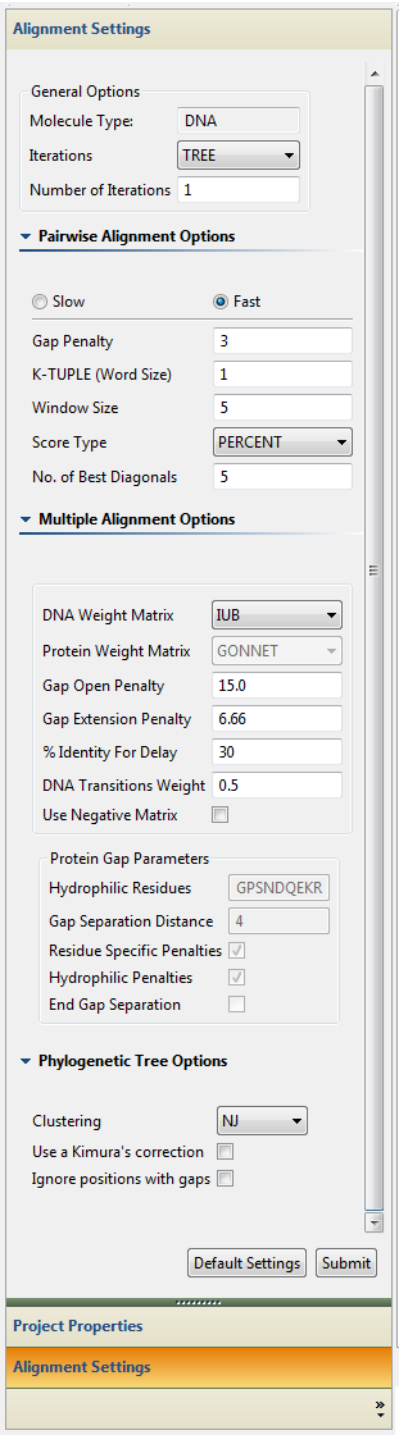
To remove a fragment from the AlignX® project, select the check box next to the sequence name in the Fragment List and click on the **Remove selected molecule(s) from alignment**. This will remove the fragment from the project, not delete it from the database.

Select fragments to align

To select fragments for alignment, select the check boxes next to the fragment names in the Fragments List.

Alignment settings

Click on the **Alignment Settings** button below the Project Properties pane to view the settings for performing an alignment.



DNA weight matrix programs

Clustal W v2.1

The local Clustal W algorithm uses matrices to determine alignments of protein and DNA sequences with each other. The program allows you to select different matrices to identify related sequences. For sequences with high similarity, it is best to use a more strict matrix so the algorithm will favor conserved substitutions. For more divergent sequences, more loosely related matrices score frequent substitutions highly. The lower numbered matrices are more strict, the higher numbers favor more divergent matches.

The following matrices are available in the AlignX program.

Protein weight matrix programs

IUB

This is the best default matrix for the BESTFIT comparison of two nucleic acid sequences. X's or Ns in a nucleic acid sequence are treated as matches to any IUB ambiguity symbol and score a mismatch of 0. Matches score 1.9.

PAM (Dayhoff)

These matrices are the original matrices, developed in the 1970s, that are used in many applications. The matrices used are PAM 20, 60, 120 and 150.

GONNET

These matrices are calculated in a similar way to the PAM matrices but they are calculated using more recent snapshots of the data available to us now. They appear to be more sensitive than the PAM matrices. We provide the GONNET 80, 120, 160, 250 and 350 matrices.

BLOSUM (Henikoff)

These matrices appear to be the best available for performing database homology searches. The matrices used as BLOSUM 80, 62, 45 and 30.

Identity Matrix (ID)

This matrix gives a score of 1.0 for two amino acid matches or a 0 otherwise. It is good for quick screens of conserved homology between sequences.

CLUSTALW

This is the second nucleic acid matrix. It is similar to the identity matrix, in that matches score 1.0 and mismatches or IUB symbols score 0. It is good for a fast comparisons between a set of sequences.

Molecule type

DNA or **Protein** is automatically detected and displayed.

Iterations

A remove first iteration scheme has been added. This can be used to improve the final alignment or improve the alignment at each stage of the progressive alignment. During the iteration step each sequence is removed in turn and realigned. If the resulting alignment is better than the previous alignment it is kept. This process is repeated until the score converges (the score is not improved) or until the maximum number of iterations is reached. The user

can iterate at each step of the progressive alignment by setting the iteration parameter to TREE or just on the final alignment by setting the iteration parameter to ALIGNMENT. The default is no iteration. The maximum number of iterations can be set using the numiter parameter. The default number of iterations is 3.

Choose an option from the pull-down menu, then enter the Number of Iterations in the box below.

Slow options	Description
NONE	Do not perform any additional iterations when performing alignment.
TREE	Do an iteration at each step of the progressive alignment.
ALIGNMENT	Iterate only on the final alignment.

Pairwise alignment

These settings control pairwise distances based on the Clustal W algorithm selected.

The **Alignment Type** options control the speed/sensitivity of the initial alignments:

- **Fast** (approximate) method
- **Slow** (more accurate) method uses two gap penalties (for opening or extending gaps) and a full amino acid weight matrix. This is the default setting.

There are different sets of parameters, depending on the alignment type:

Slow options	Description
DNA Weight Matrix	IUB (default) or Clustal W
Gap Open penalty	The penalty for the first residue in a gap
Gap Extension penalty	The penalty for additional residues in a gap
Protein weight matrix	BLOSUM, PAM, GONNET, and ID

Fast (approximate) alignment parameters

These similarity scores are calculated from fast, approximate, global alignments, which are determined by four parameters. Two techniques are used to make these alignments very fast: Only exactly matching fragments (k-tuples) are considered, and only the best diagonals (the ones with most k-tuple matches) are used.

Fast options	Description
K-TUPLE (Word size)	Change the K-tuple value to limit the word-length the search should use. A word-length of 2 is sensitive enough for most protein database searches. The general rule is that the larger the word length, the less sensitive, but faster the search will be. INCREASE this value for speed, or DECREASE for sensitivity. For longer sequences (e.g. >1000 residues) you may need to increase the default. The maximum value is 2 for proteins and 4 for DNA.
Window Size	The number of diagonals around each of the best diagonals used. Decrease this value for speed or increase it for sensitivity.

Fast options	Description
Score Type	Percent or absolute
No. of Best Diagonals	The number of k-tuple matches on each diagonal (in an imaginary dot-matrix plot) is calculated. Only those with the most matches are used in the alignment. Decrease the value for speed or increase it for sensitivity.
GAP PENALTY	This is a penalty for each gap in the fast alignments. It has little affect on the speed or sensitivity except for extreme values.

Multiple Alignment options

These parameters control the final multiple alignment.

Option	Description
DNA/Protein Weight Matrix	All algorithms designed to evaluate pairwise sequence alignment are based on systems which rank aligned residues. Nucleotides or amino acids that are identical or similar in alignment score higher than those less similar. Matrices generated with these assigned scores are used to detect similarities between differing sequences. The most common of many different scoring systems are based on substitutions of amino acids in related proteins.
Gap Open Penalty	The penalty for the first residue in a gap
Gap Extension Penalty	The penalty for extending a gap by one residue.
% Identity For Delay	The DELAY DIVERGENT SEQUENCES switch delays the alignment of the most distantly related sequences until after the most closely related sequences have been aligned. The setting shows the percent identity level required to delay the addition of a sequence; sequences that are less identical than this level to any other sequences will be aligned later.
DNA Transition Weight	The TRANSITION WEIGHT gives transitions (A to G or C to T, i.e. purine-purine or pyrimidine-pyrimidine substitutions) a weight between 0 and 1. A weight of zero indicates that the transitions are scored as mismatches, while a weight of 1 gives the transitions the match score. For distantly related DNA sequences, the weight should be near to zero. For closely related sequences it is useful to assign a higher score.
Use Negative Matrix	In the weight matrices, you can use negative as well as positive values. However, the matrix will be automatically adjusted to all positive scores, unless the NEGATIVE MATRIX option is selected.
Num of Iteration	1(default) to 10

Protein Gap Parameters	Description
Hydrophilic Residues	NEED INFO
Gap Separation Distance	GAP SEPARATION DISTANCE tries to decrease the chances of gaps being too close to each other. Gaps that are less than this distance apart are penalized more than other gaps. This does not prevent close gaps; it makes them less frequent, promoting a block-like appearance of the alignment.
Residue Specific Penalties	RESIDUE SPECIFIC PENALTIES are amino acid specific gap penalties that reduce or increase the gap opening penalties at each position in the alignment or sequence. As an example, positions that are rich in glycine are more likely to have an adjacent gap than positions that are rich in valine.
Hydrophilic (gap) Penalties	HYDROPHILIC GAP PENALTIES are used to increase the chances of a gap within a run (5 or more residues) of hydrophilic amino acids. These are likely to be loop or random coil regions where gaps are more common.
End Gap Separation	END GAP SEPARATION treats end gaps just like internal gaps for the purposes of avoiding gaps that are too close (set by GAP SEPARATION DISTANCE above). If you turn this off, end gaps will be ignored for this purpose. This is useful when you wish to align fragments where the end gaps are not biologically meaningful.

Phylogenetic Tree options

The Phylogenetic Tree is calculated and displayed after three or more sequences are aligned. It can be built using two different methods; NJ or UPGMA. The calculated distance values are displayed in parenthesis following the molecule name in the Phylogenetic Tree pane..

Phylogenetic Tree Option	Description
Clustering: NJ	The Neighbor Joining method (NJ) of Saitou and Nei works on a matrix of distances between all pairs of sequence to be analyzed. These distances are related to the degree of divergence between the sequences.
Clustering: UPGMA	The Unweighted Pair Group Method with Arithmetic Mean (UPGMA) is a simple agglomerative or hierarchical clustering method in which, at each step, the nearest two clusters are combined into a higher-level cluster.
Use a Kimura's correction	This option attempts to correct divergence for the fact that observed distances underestimate actual evolutionary distances. See papers on Motoo Kimura for more information.
Ignore positions with gaps	With this option enabled, any alignment positions where any of the sequences containing a gap will be ignored and ambiguous parts of the alignment which are usually concentrated around gaps will be discarded.

Perform the alignment

When you have selected the fragments you want to align and the proper alignment settings, click **Submit**.

Graphs pane

The Graphs pane provides a graphical representation of sequence similarity and complexity among the aligned sequences.

The graphs in this pane interact with the sequences as listed in the Alignment pane.

Similarity (Consensus) graph

To generate the Similarity graph (upper graph), specific values (in a 0-1 range) are assigned to each residue at a given alignment position in each aligned sequence, depending on whether the residue is identical, similar, or weakly similar to the corresponding residue of the consensus sequence. The values (1 (identical), 0.5 (similar), and 0.2 (weakly similar) for each residue at a given position are totaled; the sum is divided by the number of the sequences in the alignment, "normalizing" the resulting value.

Absolute Complexity (Consensus) graph

The Complexity graph (lower graph) is calculated as a sum of all pairwise residue substitution scores at a given alignment position divided by the number of pairs in the alignment. The scores are taken from the residue substitution matrix used for alignment calculation.

Select a region

Drag your cursor over a region of a graph to highlight the corresponding sequence in the Alignment pane.

Identify a data point

Mouse over a region of a graph or alignment pane to display data point information such as the position, weight, symbols, identity, or positive status.

Alignment pane

The Alignment pane displays aligned sequences and the resulting consensus sequence. The bottom row in the pane consists of the alignment consensus. Consensus residues are those that appear most commonly at a particular site.

Note: You can save the consensus sequence as a separate sequence molecule; click on **Save Consensus** in the Project Properties pane.

Select a region

Drag your cursor over a sequence region in the Alignment pane to highlight that same region in the Graphs pane.

Alignment toolbar

Button	Description
	Delete the selected fragment

Button	Description
	Copy the selected sequence
	Edit alignment
	Camera
	Print
	Alignment setup

9 3D Molecule Viewer

The 3D Molecule Viewer is a tool for visualizing protein structures in three dimensions. It can open Protein Data Bank (PDB) files with the file extension *.pdb.

The 3D Molecule Viewer uses **JMol**, an open-source Java viewer for chemical structures in 3D that has been integrated into Vector NTI™ *Express*. For more information about JMol, visit **www.jmol.org**.

Download 3D Structure Files

You can search for PDB files (with the *.pdb file extension) in public databases using the Entrez query tool in Vector NTI™ *Express*. Click on the **Public Database Search button** on the main toolbar, and for the database type, enter "structure."

Save downloaded *.pdb files in an appropriate file location and access them using the tool as described below.

Open a molecule in 3D Molecule Viewer

1. To open a PDB molecule in the 3D Molecule Viewer, click on the **3-D Molecule Viewer** tool on the main toolbar, or select it from the **Discover** menu.

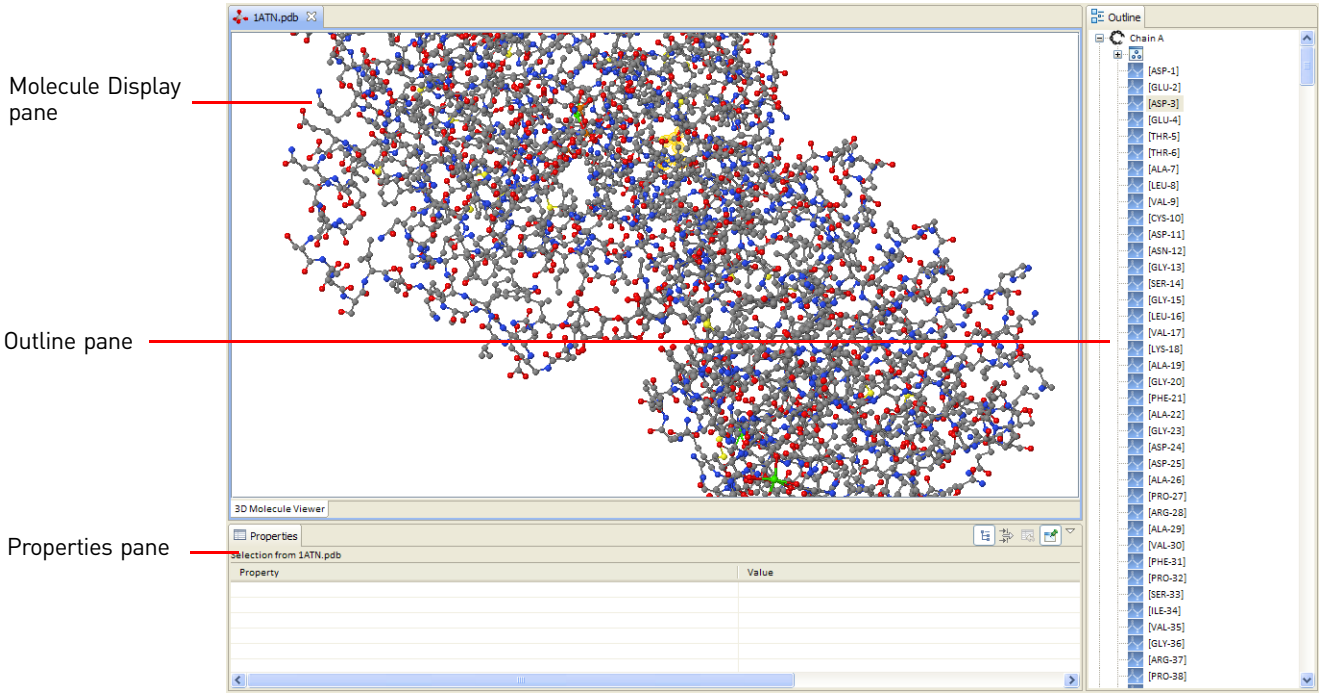
2. In the Open dialog, select a *.pdb file to open it.

To open a different molecule, select **File > Open > 3D Molecule**.

Note: The PDB file you open will remain open in Vector NTI™ *Express* until you close the window, and you can navigate back to it by clicking the **3-D Molecule Viewer** button.

Elements of the 3D Molecule Viewer window

The 3D Molecule Viewer window consists of three panes:



- **Graphics pane**—displays the three-dimensional structure of the molecule.
- **Properties pane**—lists any properties associated with the molecule
- **Outline pane**—lists the amino acids in the sequence

The right-click menu in the Molecule Display pane contains information about the molecule and tools for manipulating the graphic representation of the structure, exporting molecule data in various formats, and optimizing the display.

Magnify and rotate the molecule

In the Molecule Display pane:

- To zoom in or out, rotate the wheel on your mouse or select **Zoom** from the right-click menu.
- To rotate the molecule in three-dimensional space, drag your cursor in the pane.
- To reposition the molecule within the pane without rotation, hold the **Ctrl + Alt** keys and drag your cursor in the pane.
- To begin spinning the molecule in three-dimensional space, select **Spin > On** from the right-click menu, and adjust the settings on the **Spin** menu. Select **Spin > Off** to stop spinning.

Highlight an amino acid or a chain

- Click within the structure in the Molecule Display window to highlight a particular amino acid. Click on it again to un-highlight it.
- Click on an amino acid in the Outline pane to highlight it in the Molecule Display pane.
- Click on a chain in the Outline pane to highlight that entire amino acid chain in the structure.

Additional menu operations

The right-click menu in the Molecule Display window contains all the JMol commands for viewing and manipulating the structure. For additional information about these commands, visit **www.jmol.org**.

10 Sim4 and Spidey Analysis

Sim4 and Spidey are tools for aligning expressed nucleic acid sequences (e.g., mRNA) with genomic sequences in online databases. Vector NTI™ *Express* includes these tools for analyzing molecules stored in the database.

Sim4

Sim4 is a similarity-based tool for aligning an expressed nucleic acid sequences (EST, cDNA, mRNA) with a genomic sequence. For more information about Sim4, see http://www.hgmp.mrc.ac.uk/.

Sim4 employs the following multi-stage BLAST-based technique:

- Sim4 detects all possible exact matches of W-mers between the two sequences and extends them to maximal scoring gap-free segments (exon-cores).
- Exon-cores are extended into the adjacent, as-yet-unmatched fragments using greedy alignment algorithms. Heuristics are used to favor configurations that conform to the splice-site recognition signals. If necessary, the process is repeated with less stringent parameters on the unmatched fragments.

Sim4 functions similarly to BLAST, but performs a more thorough mRNA alignment search.

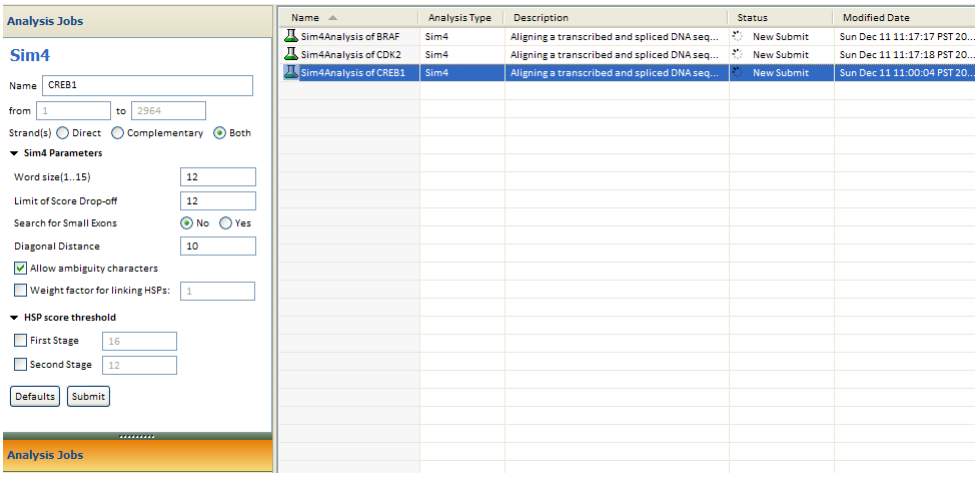
Launch Sim4 analysis tool

To open the Sim4 Analysis tool:

- In Database Explorer, right-click on a molecule or use **Ctrl+click** or **Shift+click** to select multiple molecules and right click. In the right-click menu, select **Analysis Monitor > Sim4 Analysis.**
- With a molecule open in Molecule Editor, click on the **Sim4 Analysis** button in the Molecule Editor toolbar.

The Analysis Monitor will open, with the Sim4 Analysis settings selected and the selected molecule(s) listed.

The analysis window has an **Analysis Jobs pane** containing the analysis settings, and an **Analysis List pane** listing the molecules loaded in the tool.



Analysis Jobs settings

Molecule and strand to analyze

In the **Analysis Jobs** pane:

- Select the molecule from the **Name** dropdown list.
- Specify the **Strand** to analyze. The **Both** option searches both strands and reports the best result.

Sim4 parameters

- **Word Size** – The number of DNA bases in a size unit. The larger the word size, the faster and less sensitive Sim4 becomes.
- **Limit of Score Drop-off –** The trigger for stopping ungapped extension.
- **Search for Small Exons –** If **Yes** is selected, an additional search for small exons is performed.
- **Diagonal Distance –** The upper boundary of diagonal distance within consecutive HSPs in an exon.
- **Allow Ambiguity Characters –** If checked, allows the following ambiguity characters: ABCDGHKMNRSTVWXY. If unchecked, only the following characters are allowed: ACGTNX.
- **Weight Factor for Linking HSPs –** The multiplication factor used when calculating the score of a chain of HSPs.
- **Defaults –** Resets Sim4 parameters to the original default values.

HSP Score Threshold

- **First Stage –** The threshold for HSPs for the first stage of comparison. If no value is specified, the default value is used.
- **Second Stage** – The threshold for HSPs when aligning the extended-and-unmatched portions. If no value is specified, the default value is used.

Submit the job

When you have made your selections, click on **Submit.**

When analysis is complete, a **Completed** checkbox will appear next to the job name in the Analysis Monitor.

View analysis results

The Analysis Monitor contains a list of all the analyses performed by Vector NTI™ *Express*, including Sim4 analyses. Click on **Analysis Monitor** on the main toolbar to open it.

To view the analysis for a particular molecule, you can:

- Open the molecule in the Molecule Editor, and click on **Analysis Results**.
- Open the Analysis Monitor and double-click on the particular analysis.

Spidey

Spidey is an mRNA-to-genomic alignment program. It takes as input a single genomic sequence and a set of mRNA accessions or FASTA sequences. All processing is done one mRNA sequence at a time. The first step for each mRNA sequence is a high-stringency BLAST against the genomic sequence. The BLAST alignments are sorted by score and then assigned into windows by a recursive function. For more information about Spidey, see www.ncbi.nlm.nih.gov/spidey/.

After the genomic windows are constructed, the initial BLAST alignments are freed and another BLAST search is performed, this time with the entire mRNA against the genomic region defined by the window, and at a lower stringency than the initial search. Spidey then uses a greedy algorithm to generate a high-scoring, non-overlapping subset of the alignments from the second BLAST search. This consistent set is analyzed carefully to make sure that the entire mRNA sequence is covered by the alignments.

Once the mRNA is completely covered by the set of alignments, the boundaries of the alignments are adjusted so that the alignments abut each other precisely and so that they are adjacent to good splice donor and acceptor sites. To position the exon boundaries, the adjacent exon alignment overlap region plus a few base pairs on each side is examined for splice donor sites, using functions that have different splice matrices depending on the organism chosen. The top few splice donor sites (by score) are then evaluated as to how much they affect the original alignment boundaries. The site that affects the boundaries the least is chosen, and is evaluated as to the presence of an acceptor site. The alignments are truncated or extended as necessary so that they terminate at the splice donor site and so that they do not overlap.

For details on Spidey analysis, see http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html.

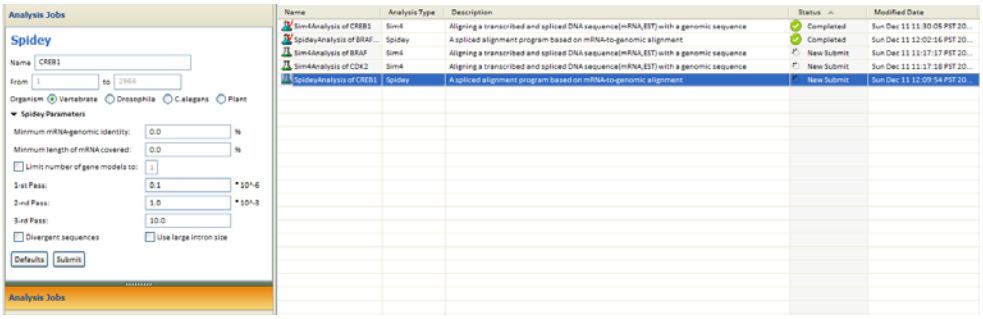
Launch Spidey analysis tool

To open the Spidey Analysis tool:

- In Database Explorer, right-click on a molecule or use Ctrl+click or Shift+click to select multiple molecules and right click. In the right-click menu, select **Analysis Monitor > Spidey Analysis.**
- With a molecule open in Molecule Editor, click on the **Spidey Analysis** button in the Molecule Editor toolbar.

The Spidey Analysis tool will open with the selected molecule(s) listed.

The analysis window has an **Analysis Jobs pane** containing the analysis settings, and an **Analysis List pane** listing the molecules loaded in the tool.



Analysis Jobs settings

Molecule and organism to analyze

In the **Analysis Jobs** pane:

- Select the molecule from the **Name** dropdown list.
- Specify the **Organism** to analyze for your genomic sequence.

Spidey Parameters

- **Minimum mRNA Genomic Identity** – Percent identity cutoff for gene models.
- **Minimum Length of mRNA Covered –** mRNA length coverage cutoff in percent for gene models.
- **Limit Number of Gene Models to –** Sets the maximum number of gene models Spidey analysis returns.
- **E-value**
  - **1st Pass** – The E-value cutoff for the initial high-stringency alignment. The higher the value, the less stringent and faster the run.
  - **2nd Pass –** The E-value cutoff for the low-stringency BLAST search within a genomic window, based on the high-stringency result.
  - **3rd Pass –** The E-value cutoff for a very low-stringency BLAST search to find hits for mRNA gaps.
- **Divergent Sequences –** If checked, search parameters are adjusted to tolerate mismatches and gaps for inter-species alignment.
- **Use Large Intron Sizes** – If checked, much larger maximal intron sizes are allowed. Checking this option increases computation time significantly.
- **Defaults –** Sets Spidey parameters to original default values.

Submit the job

When you have made your selections, click on **Submit.**

When analysis is complete, a **Completed** checkbox will appear next to the job name in the Analysis List.

View analysis results

The Analysis Monitor contains a list of all the analyses performed by Vector NTI™ *Express*, including Spidey analyses. Click on **Analysis Monitor** on the main toolbar to open it.

To view the analysis for a particular molecule, you can:

- Open the molecule in the Molecule Editor, and click on **Analysis Results**.
- Open the Analysis Monitor and double-click on the particular analysis.

11 Clone2Seq

Clone2Seq™ simplifies the **two fragment** restriction and ligation cloning process. Using Clone2Seq™, you will be able to clone a single DNA or RNA insert into a vector without having to go through multiple steps of the molecule construction wizard.

Launch Clone2Seq™

To launch Clone2Seq™:

- Click **Clone2Seq** on the main toolbar.
- Select **File** ▸ **New** ▸ **Clone2Seq Project** from the main menu.

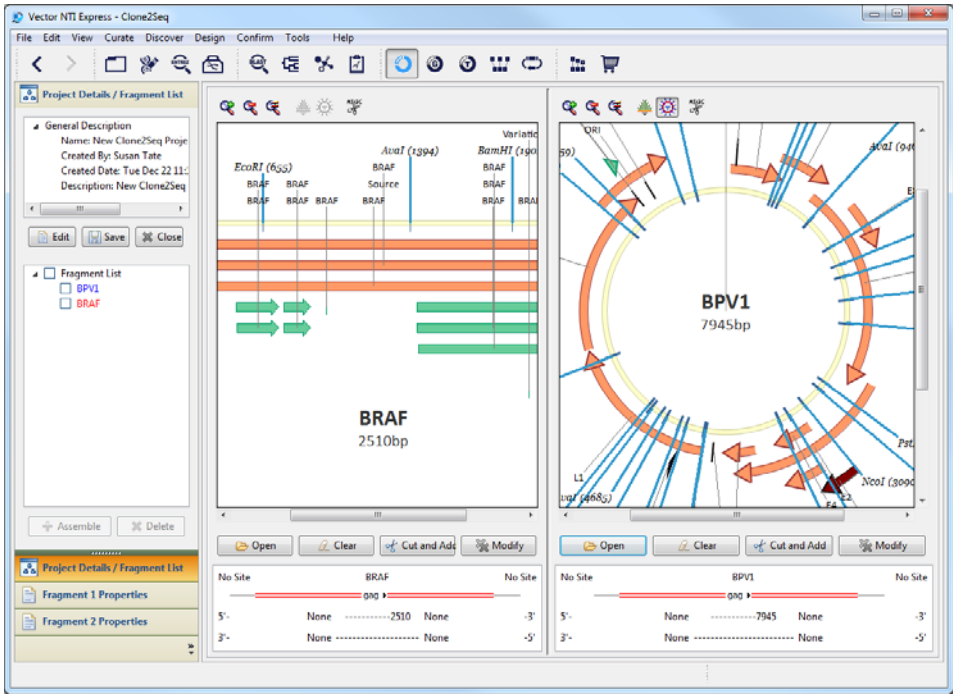
Clone2Seq™ window

The Clone2Seq™ window consists of two molecule panes, and a left pane containing the Fragment List and Molecule Properties lists. In the Fragment list, a red font indicates the selected molecule on the left, adn a blue font indicates the molecule on the right.

The left-most molecule pane contains the DNA/RNA molecule that is first fragment in the clone, and the right-hand pane contains the second fragment.

For each molecule, the restriction sites (selected by Restriction Analysis in the Molecule Editor) will be displayed in the molecule pane, and sequence ends of linear molecules will be shown in the View Sequence field below the pane.

Figure 3 Clone2Seq™ Window



Select molecules

1. Click on **Open** below the left-hand pane to select the first molecule.

2. Repeat this operation in the right-hand pane to select the second molecule.

Note: Click on **Clear** to remove a molecule from the pane.

Note: The molecule in the left-hand pane will always be the **first fragment** of the clone and the molecule in the right-hand pane will always be the **second fragment** of the clone. This maintains the directionality of the resulting construct.

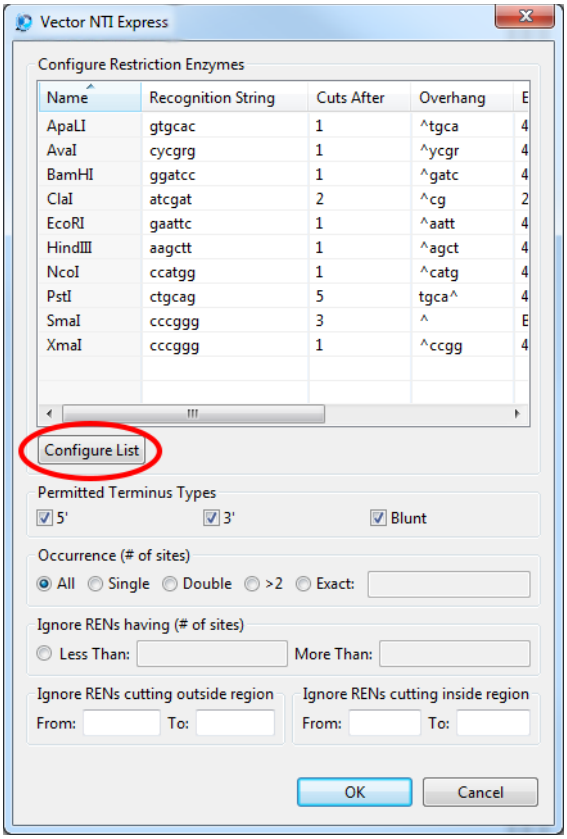
Molecule requirements

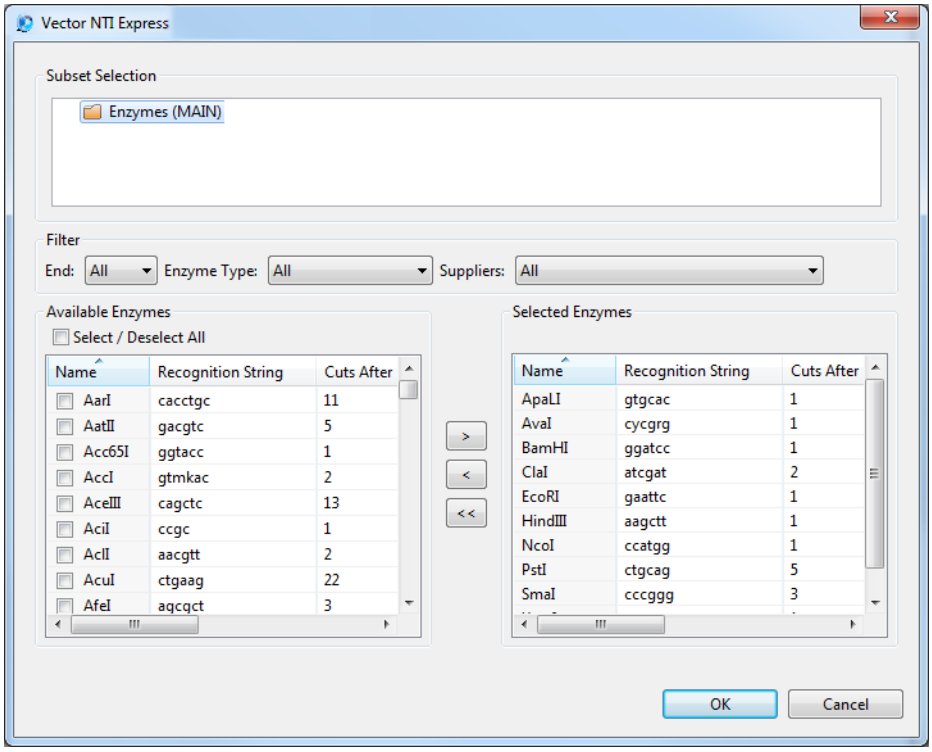
Notes about Clone2Seq™ assembly:

- Two linear molecules with blunt ends may be cloned. These blunt ends are created via restriction digest of sites that generate blunt ends.
- Two linear molecules with overhangs must have matching overhangs. This can be accomplished via restriction digestion, as described in **Generating molecule fragments** below, or by modifying the fragment ends, as described in **Modifying fragment ends**.
- Circular molecules must be linearized on a cut site, as described in **Generating Molecule Fragments** below.

Show or hide restriction sites in molecules

Click **Restriction Analysis** in the panel toolbar to view and select restriction sites displayed in the molecule pane. Click **Configure List** to adjust the list of displayed sites. Changes to the selected restriction enzymes will trigger a restriction analysis and update the reflected sites on the pane.

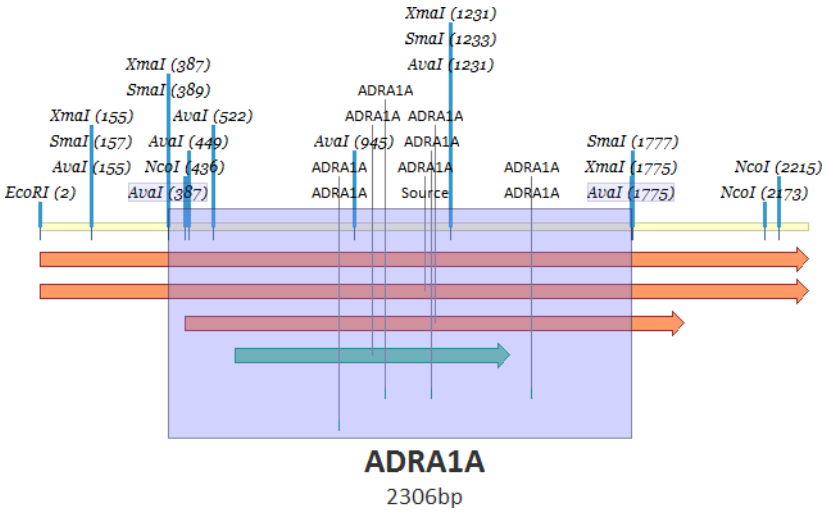




Generate molecule fragments

To generate a molecule fragment from a linear or circular molecule in Clone2Seq™:

1. Click on a restriction site in the desired molecule to generate a single cut site (in the case of circular molecules), or shift-click on two restriction sites to select the region between them (the region will appear selected in the window, as shown below).

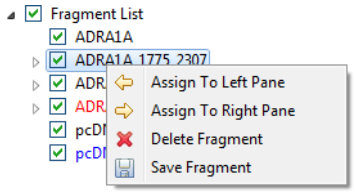


Note: When selecting restriction enzymes that generate overhangs, be sure to select enzymes in both molecules that will generate complementary overhangs. Alternatively, you can modify the fragment ends as described below.

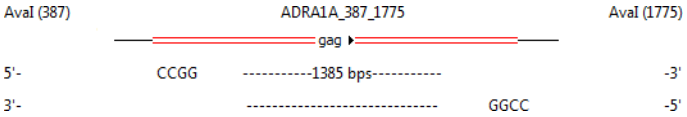
2. Click on **Cut & Add** below the molecule pane.
   The fragment will be added to the Fragment List.

Note: Fragments in the list are named for the molecule, as well as the region that was cut. For example, ADRA1A_1_608 is a fragment of the ADRA1A molecule taken from the 1–608 bp region.

3. To select the fragment for cloning, right-click on the fragment in the list and select **Assign To Left Pane** or **Assign To Right Pane**.



4. The fragment will be added to the appropriate pane, and the cut ends will be displayed below that pane.



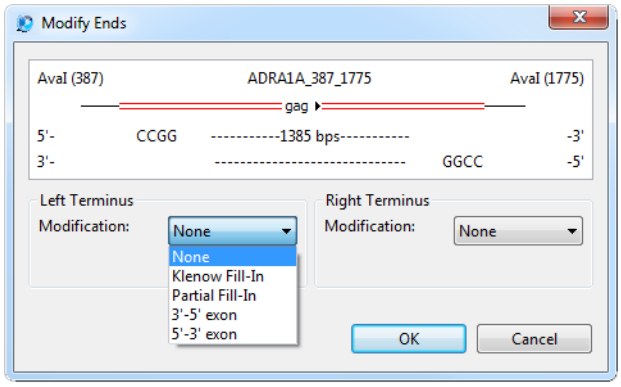
5. Repeat this procedure for the second molecule.

Edit the Fragment List

Right-click on a fragment in the Fragment List and select **Delete Fragment** to delete it, or **Save Fragment** to save it as a molecule in the database.

Modify fragment ends

1. To modify the ends of a molecule or fragment, click **Modify**, then select the desired modification for the Left and Right Terminus using the drop-down list.

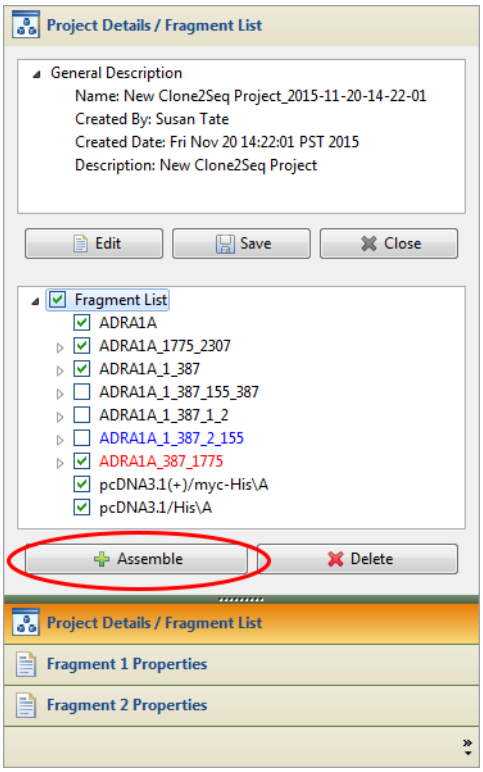


   The modification will be previewed in the dialog.

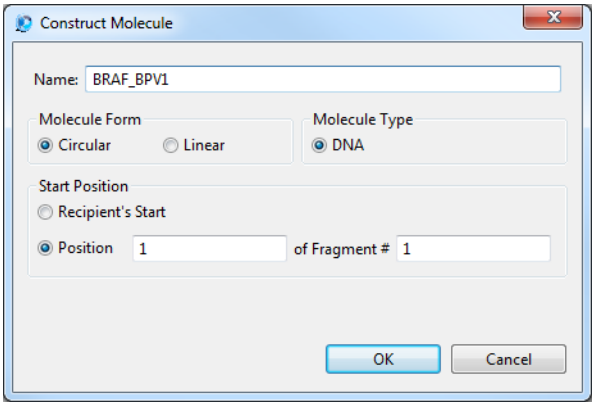
2. Click **OK** to make the modification.

Assemble the molecule

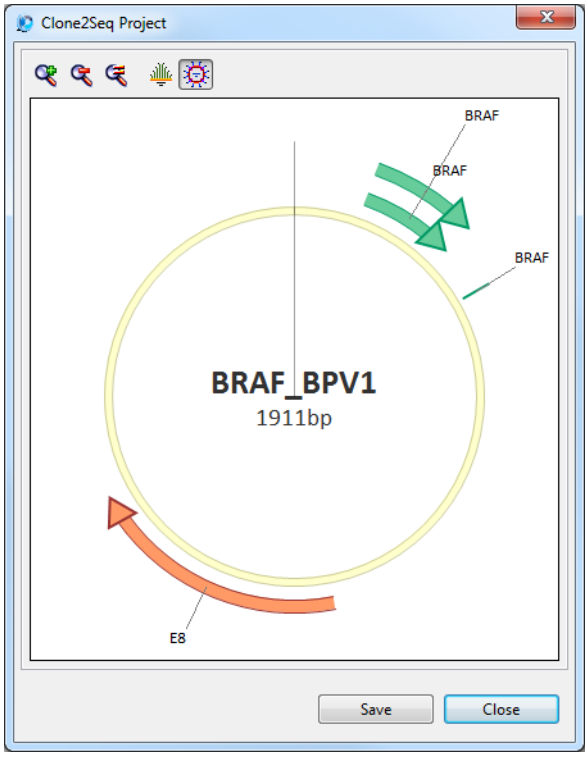
1. When you have made your selections, click **Assemble**. This button will not be available unless both molecules are linear.



2. In the **Construct Molecule** dialog box, enter a name for the assembly and select the desired Molecule Form, Type, and Start Position.



3. Click **OK**. The assembled molecule displays in a preview window.



Note: Use the tools in the window to magnify the molecule or display it as linear or circular.

4. Click **Save** to save the molecule to the database.

Multiple fragment cloning

To clone multiple fragments using Clone2Seq™, first create a clone from two fragments as described above, then generate a clone from the cloned molecule and the next fragment. Repeat as necessary.

12 Gateway® Cloning

Gateway® Cloning Technology is a rapid and highly efficient method for the cloning and subcloning of DNA segments. This system is based on the well-characterized bacteriophage lambda-based site-specific recombination system (*att*L x *att*R ↔ *att*B x *att*P).

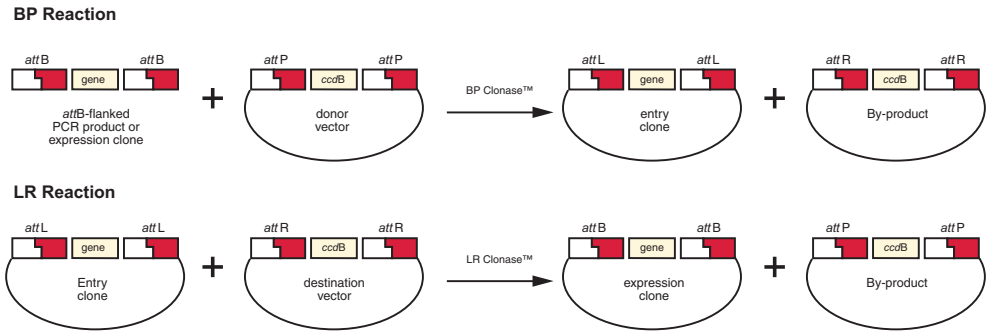
Gateway® Cloning is a 2-step process. In the first step a sequence of interest containing *att*B sites is recombined with a donor vector containing *att*P sites into an entry clone, creating *att*L sites in the process. The second step recombines the *att*L-containing entry clone with a destination vector containing *att*R sites, generating an expression clone that can be propagated and expressed in a range of host cells for a given experiment.

For more information

For detailed information on Gateway® Cloning, see the *Gateway® Technology User Guide* or the *Gateway® Technology with Clonase™ II User Guide*, available for download from **www.lifetechnologies.com/manuals**.

Gateway® Cloning Workflow

Workflow diagram



Step 1. Create an entry clone

The standard method for creating an entry clone in the Vector NTI™ *Express* Gateway® Cloning Tool involves amplifying a sequence or molecule of interest with *att*B-containing primers designed by the software, then performing BP recombination with a donor (pDONR) vector to generate an entry clone.

> *PCR product (flanked by attB sites) + pDONR vector (with attP) ↔ Entry Clone (with attL)*

You can also create an entry clone by the following methods:

- **Select an existing *att*B-containing DNA molecule**– From within the Gateway® Cloning Tool, you can select an already existing *att*B-containing DNA molecule in the database, such as a Gateway® Expression Clone or a pCMVSPORT6 library, for recombination with a donor vector, to create the entry clone

- **Construct an entry clone by alternative molecule construction methods**– You can construct your own entry clone using other Vector NTI™ *Express* molecule construction methods, ensuring that the clone contains the required *att*L1 and *att*L2 sites (labeled as features as described in "Add Entry Clones to Use in LR Reaction"), then save the molecule in the database and select it directly in the Gateway® Cloning tool.

Step 2. Create an expression clone

Using the Gateway® Cloning Tool, entry clones created from any of the above methods are recombined with destination (pDEST) vectors in an LR recombination reaction to generate expression clones, which can drive expression of the sequence of interest when transformed into host cells.

Gateway® Cloning Tool

Open the Gateway® Cloning Tool

The Gateway® Cloning Tool contains settings and functions for assembling a Gateway® construct using the workflow described above, and for creating and managing Gateway® Cloning projects.

There are several ways to open the tool:

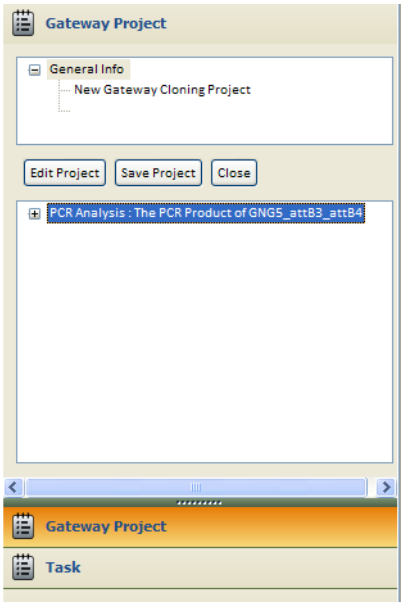
- Click on the **Gateway Cloning** button on the main toolbar.
- Select **File > New > Gateway Cloning Project** from the main menu.
- With a molecule open in Molecule Editor, right-click in the Graphics or Sequence pane and select **Launch Gateway**.
- Load an existing Gateway® Cloning project as described in "Create, save, and load projects" on page 148.

The tool window is composed of the following panes:

Gateway® Project pane

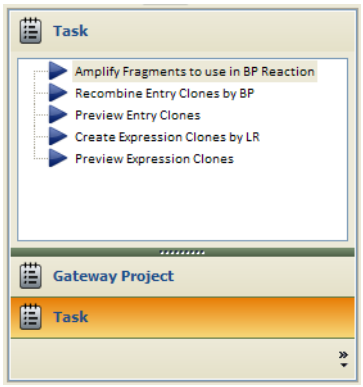
This pane displays the name of the current project, and includes controls for editing, saving, and closing projects. It also lists any generated molecules for the current project.

Click on the **Gateway Project** button to display this pane.



Task List pane

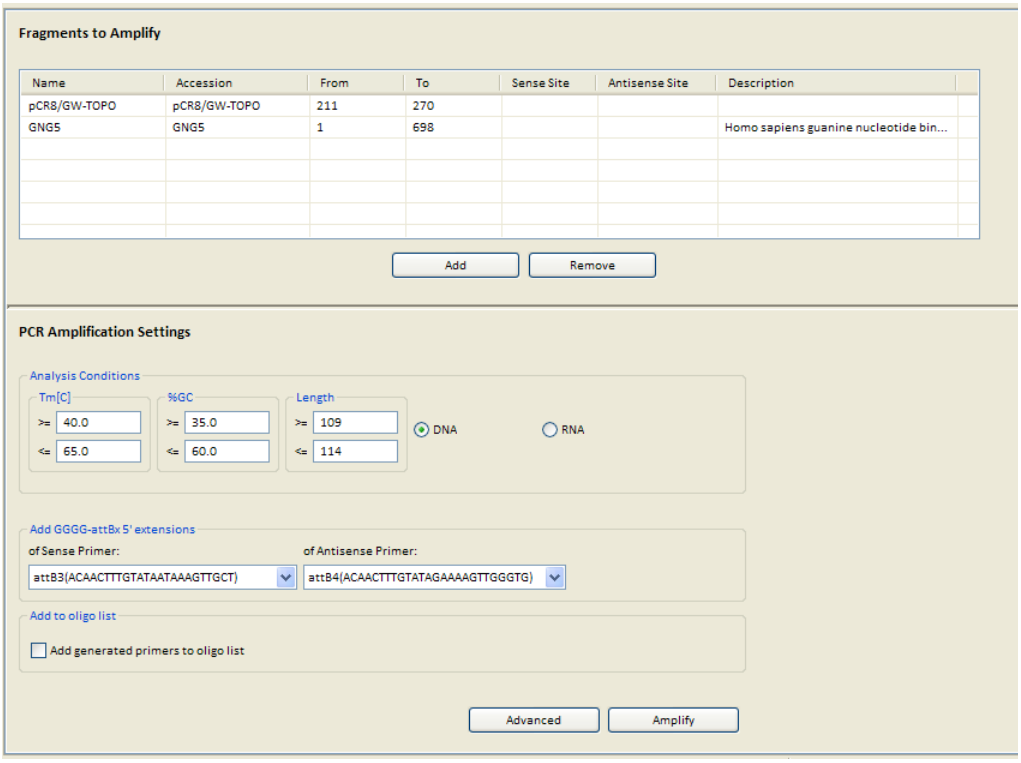
This pane displays the list of tasks in the selected project. Click on the **Task** button to display this list.



As you navigate through the Gateway® Cloning workflow, the Task List displays the current task, and allows you to move between tasks by clicking on a different task.

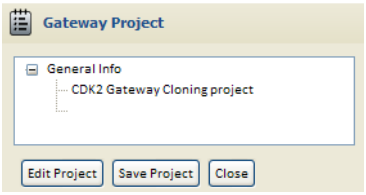
Current Task pane

This pane displays the commands and settings for the currently selected task in the project. As you navigate through the Task List workflow, the functions in this pane will change.



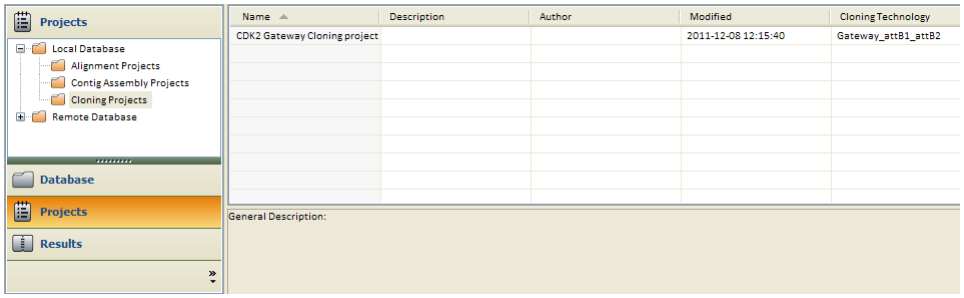
Create, save, and load projects

The tools for saving, editing, and closing projects are located in the Gateway® Project pane:



- To save a new Gateway® Cloning project, click on **Edit Project** in the Gateway® Project pane. In the dialog, enter a name and any description for the project.
- To save changes to a project, click on the **Save Project** button.
- To close a project, click on the **Close Project** button. If there are unsaved changes, you will be prompted to save the project before closing.

- To load an existing project, in Database Explorer, go to the **Projects** list, double-click on the **Projects** folder, select the **Cloning Projects** subfolder from the Local Database, and double-click on the Gateway® Cloning project in the list to open it.



Gateway® Cloning workflow

The Task List displays the default Gateway® Cloning workflow:

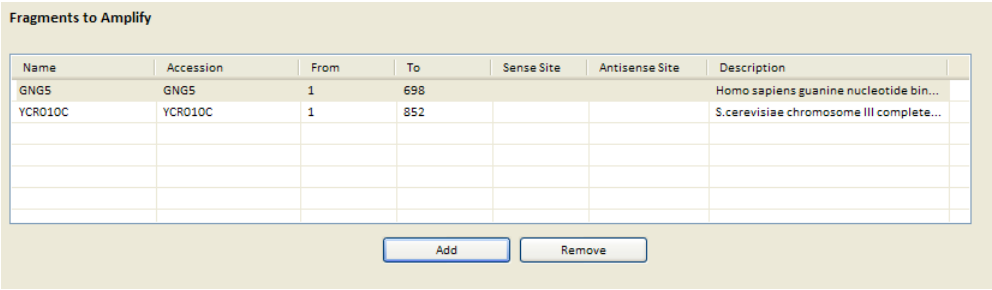
- Amplify Fragments to Use in BP Reaction
- Recombine Entry Clones by BP
- Preview Entry Clones
- Create Expression Clones by LR
- Preview Expression Clones

This section describes each task in the workflow.

Amplify fragments to Use in BP reaction

The **Amplify Fragments to Use in BP Reaction** task is the default task displayed when you first open the Gateway® Cloning Tool.

The first step in this task to select the fragment(s) you want to amplify by PCR for use in a BP cloning reaction. These fragments will be listed in the **Fragments to Amplify** list.



Load molecules or fragments in the Fragments to Amplify list

- With a molecule open in Molecule Editor, right-click in the Graphics or Sequence pane and select **Launch Gateway** to load the entire sequence in the list.
- With a molecule open in Molecule Editor, select a portion of the sequence in the Graphics or Sequence pane, right-click, and select **Launch Gateway** to load only that part of the sequence in the list.

- With the Gateway® Cloning window open, make sure the **Amplify Fragments to Use in BP Reaction** task is selected and click on the **Add** button under **Fragments to Amplify** to select a complete molecule from the database.

To change the regions to amplify in the selected molecules, type a new range in the **From** and **To** fields in the Fragments to Amplify subpane.

PCR Amplification settings

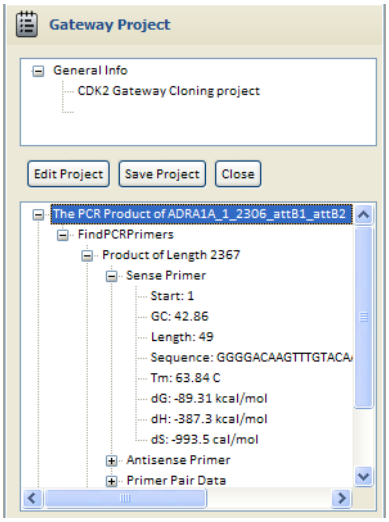
With the fragment(s) loaded, select the desired amplification settings under **PCR Amplification Settings**. The standard options are described below.

PCR Amplification Setting	Description
Tm(C)	Enter limits in degrees Celsius for primer melting temperature (T_m) (temperature at which 50% of primer is a duplex) and the difference between T_m for sense and antisense primers.
%GC	Enter limits in degrees Celsius for primer melting temperature (T_m) (temperature at which 50% of primer is a duplex) and the difference between T_m for sense and antisense primers.
Primer Length	Defaults to 20-25, recommended for Gateway® Primers
DNA/RNA button	Select the type of nucleotide sequence.
Add GGGG-*att*Bx 5' Extensions	The default *att*B extensions are for single fragment cloning: *att*B1 for the sense primer and *att*B2 for the antisense primer. Select from the dropdown list to replace the defaults with other *att*B sequences for creating Entry Clones for MultiSite Gateway® Cloning projects.
Add generated primers to oligo list	Select this checkbox to add the primers you generate to the oligo list

Note: For additional amplification settings, click on **Advanced**. The advanced amplification settings are identical for all PCR primers, and are described in "Amplify Selection settings" on page 83 and "Shared Advanced settings" on page 88.

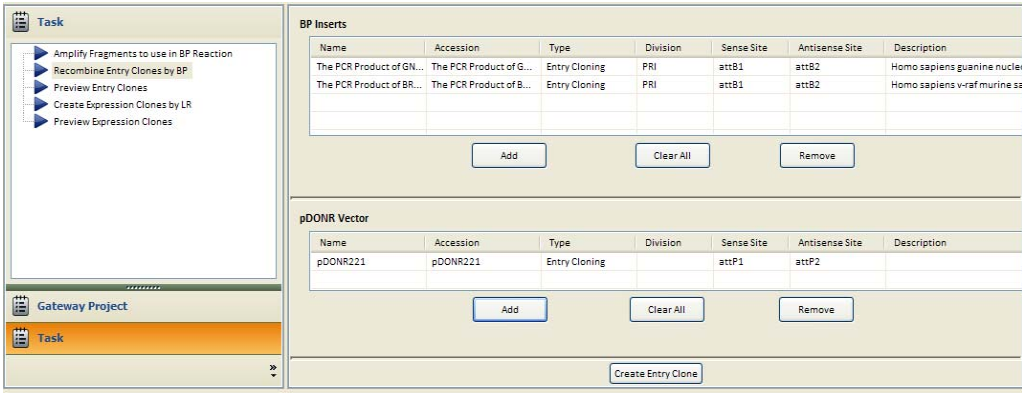
Amplify

When you have made your selections, click on the **Amplify** button. The next task pane will be displayed, and the generated PCR product(s) will appear listed in the **BP Inserts** subpane, and also listed in the Gateway® Project pane.



Recombine Entry Clones by BP

In the **Recombine Entry Clones by BP** task pane, you can modify the list of fragments with *att*B sites and select a donor (pDONR) vector or vectors with which to create entry clones.



BP Inserts

The fragments you amplified with *att*B sites are listed in the **BP Inserts** list at the top of the pane.

- To add a previously amplified fragment with *att*B sites, or a fragment designed with *att*B sites by another means (e.g., restriction-ligation), click on the **Add** button and select the molecule from the database.
- To remove a molecule from list, select it and click on **Remove**.
- To clear the entire list, click on **Clear All**.

pDONR Vector

The pDONR vector is a type of Gateway® Cloning vector that contains *att*P sites, which are recombined with the fragments containing *att*B sites to create entry clones.

A variety of pDONR vectors are sold by Thermo Fisher Scientific, and *in silico* sequences for these are installed as part of the default Vector NTI™ *Express* installation.

IMPORTANT! To be recognized as a pDONR vector in Vector NTI™ *Express*, a molecule must contain the correct *att*P1 and *att*P2 sequences, and these sites must be labeled as features in the molecule with the feature names **attP1** and **attP2**. See Chapter 2, "Molecule Editor" on page 39 for information on identifying and naming features.

To select a pDONR vector:

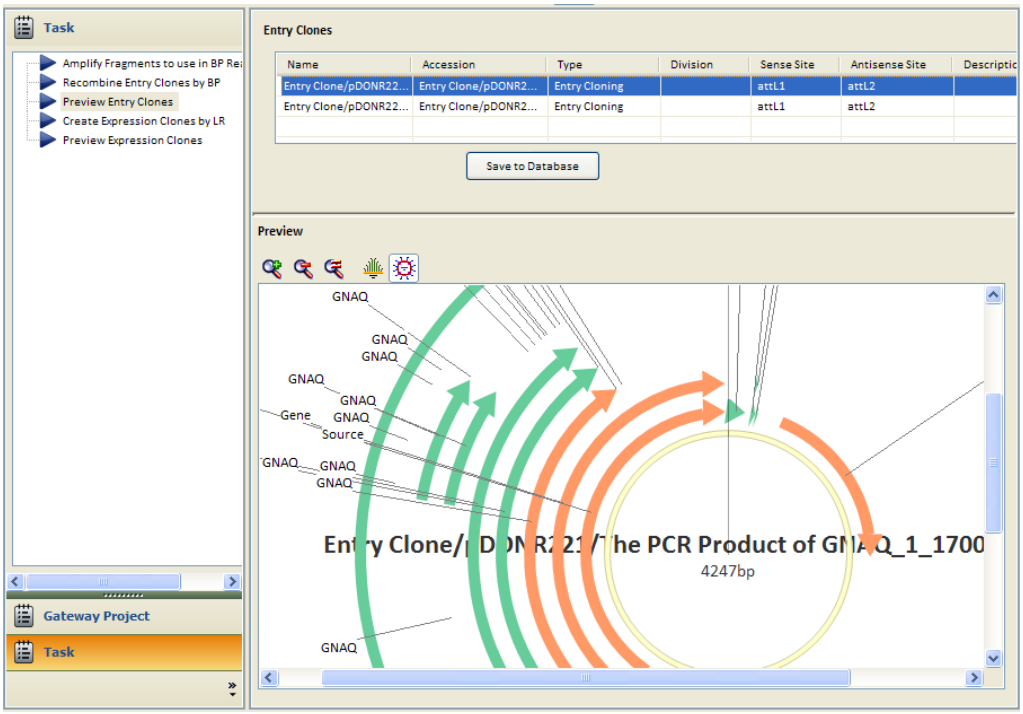
- Click on the **Add** button and select the pDONR vector from the database.
- To remove a vector from list, select it and click on **Remove**.
- To clear the entire list, click on **Clear All**.

Create the Entry Clone

When you have made your selections, click on **Create Entry Clone**. The **Preview Entry Clones** task pane will open.

Preview Entry Clones

The **Preview Entry Clones** task pane lists all the entry clones created from the *att*B-containing fragment(s) and the donor vector(s) you selected, and includes a preview window for viewing an entry clone.



Entry clones contain *att*L1 and *att*L2 sites, and are used to generate expression clones via an LR reaction.

In the **Entry Clones** list:

- Select a clone and click on **Save to Database** to save it as a DNA molecule in the database.
- Click on a clone to display it in the **Preview** window.

In the Preview window:

- Using the magnifying tools to zoom in and out of the molecule.

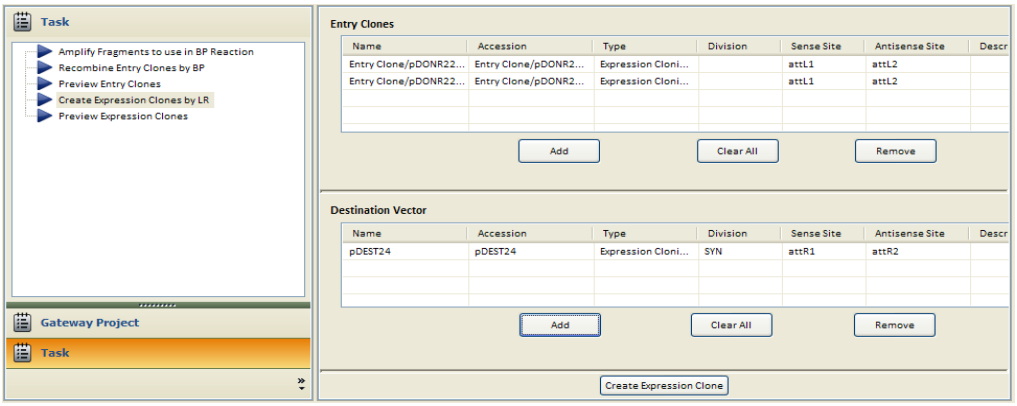
- Display the molecule as linear or circular by clicking on the appropriate button.

Create Expression Clones by LR

After creating entry clones, click on **Create Expression Clones by LR** in the Task List to proceed to the next task in the workflow.

Any entry clones that you generated from the previous tasks in the workflow will be listed in this window.

In the **Create Expression Clones by LR t**ask pane, you can modify the list of entry clones and select a destination (pDEST) vector or vectors with which to create expression clones.



Entry Clones

Any entry clones that you generated from the previous tasks in the workflow will be listed in the **Entry Clones** list.

- To select new or additional entry clones in the database, click on the **Add** button and select from the dialog box.

IMPORTANT! To be recognized as an entry clone in Vector NTI™ *Express*, a molecule must contain the correct *att*L1 and *att*L2 sequences, and these sites must be labeled as features in the molecule with the feature names **attL1** and **attL2**. See Chapter 2, "Molecule Editor" on page 39 for information on identifying and naming features.

- To remove an entry clone from the list, select it and click on the **Remove** button.
- To clear the list, click on **Clear All**.

pDEST Vector

The pDEST vector is a type of Gateway® Cloning vector that contains *att*R sites, which are recombined with the fragments containing *att*L sites to create expression clones.

A variety of pDEST vectors are sold by Thermo Fisher Scientific, and *in silico* sequences for these are installed as part of the default Vector NTI™ *Express* installation.

IMPORTANT! To be recognized as a pDEST vector in Vector NTI™ *Express*, a molecule must contain the correct *att*R1 and *att*R2 sequences, and these sites must be labeled as features in the molecule with the feature names **attR1** and **attR2**. See Chapter 2, "Molecule Editor" on page 39 for information on identifying and naming features.

To select a pDEST vector:

- Click on the **Add** button and select the pDEST vector from the database.

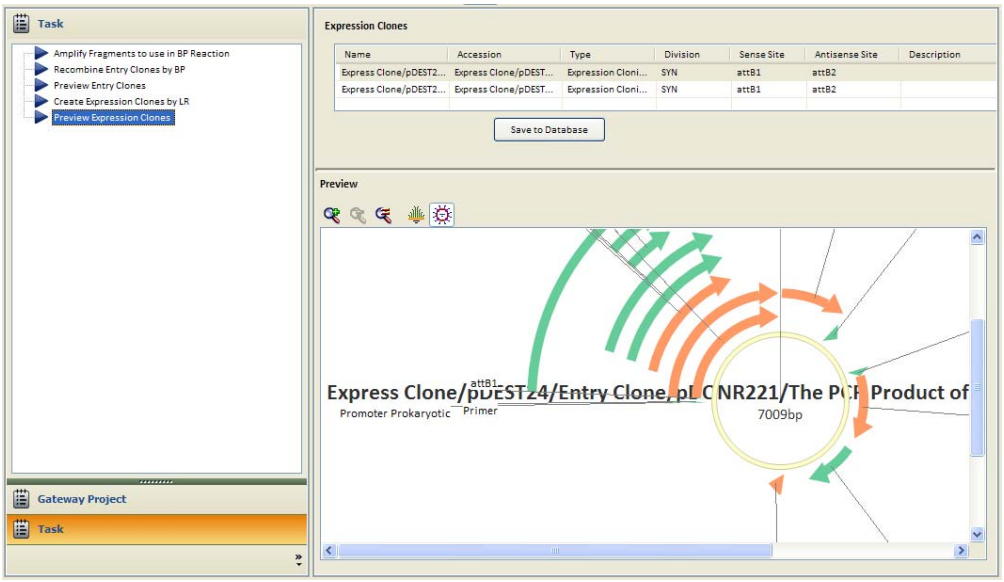
- To remove a vector from list, select it and click on **Remove**.
- To clear the entire list, click on **Clear All**.

Create the Expression Clone

When you have made your selections, click on **Create Expression Clone**. The **Preview Expression Clones** task pane will open.

Preview Expression Clones

The **Preview Expression Clones** task pane lists all the expression clones created from the entry clones and the destination vector(s) you selected, and includes a preview window for viewing an expression clone.



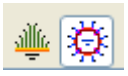
In the **Expression Clones** list:

- Select a clone and click on **Save to Database** to save it as a DNA molecule in the database.
- Click on a clone to display it in the **Preview** window.

In the Preview window:

- Using the magnifying tools to zoom in and out of the molecule.

- Display the molecule as linear or circular by clicking on the appropriate button.



13 TOPO® Cloning

TOPO® Technology is a fast, efficient way to clone. The key to TOPO® Cloning is the enzyme DNA topoisomerase I, whose biological role is to cleave and rejoin DNA during replication. To harness this activity, vectors are linearized and each end is conjugated with topoisomerase on the 3' phosphate. This enables fast ligation of DNA sequences with compatible ends. After 5 minutes at room temperature, the enzyme is released, the ligation is complete and the recombinant molecule is ready for transformation into *E. coli*.

Many Thermo Fisher Scientific expression vectors are adapted for one-step TOPO® Cloning of PCR products in both directional and non-directional formats. Other vectors contain *att* recombination sequences exterior to the TOPO® cloning sites so that cloned inserts are ready for entry into the TOPO® system.

TOPO® vectors can be grouped into three categories, based on the nature of their ends:

- **Zero-Blunt** vectors have two blunt ends and can accept blunt-ended DNA fragments, including amplicons produced by a proofreading polymerase. Inserts are cloned in both orientations.
- **T-A** vectors have two ends with 3′-T overhangs. They can accept products of PCR amplification with a *Taq* polymerase, whose terminal transferase activity adds 3′-A overhangs to the amplicon. Inserts are cloned in both orientations.
- In **directional** vectors one terminal is blunt ended and the other has a 5′-GGTG overhang on the bottom strand. PCR products are generated with a 5′-CACC extension on one end and this strand when unwound is preferentially annealed to the vector overhang. More than 90% of the clones are in the correct orientation and the time spent in screening colonies is thereby reduced.

The presence of topoisomerase enzyme also helps protect vector ends from degradation, particularly from contaminating nucleases that may be present in ligase preparations. Moreover, the avoidance of restriction site cutback for cloning PCR products means that internal cleavage sites are not a problem.

TOPO® Cloning in Vector NTI™ *Express*

Any linear, double-stranded DNA sequence of interest may be cloned into a TOPO® vector using Vector NTI™ *Express*. In addition, linear sequences with 3′-A overhangs, the products of PCR amplification with a *Taq* DNA polymerase, may be cloned into TOPO®-TA vectors. Such *Taq*-generated molecules can be generated *in silico* using the TOPO® Cloning tool.

TOPO® Cloning Tool

Open the TOPO® Cloning Tool

The TOPO® Cloning Tool contains settings and functions for assembling a TOPO® construct using the workflow described above, and for creating and managing TOPO® Cloning projects.

There are several ways to open the tool:

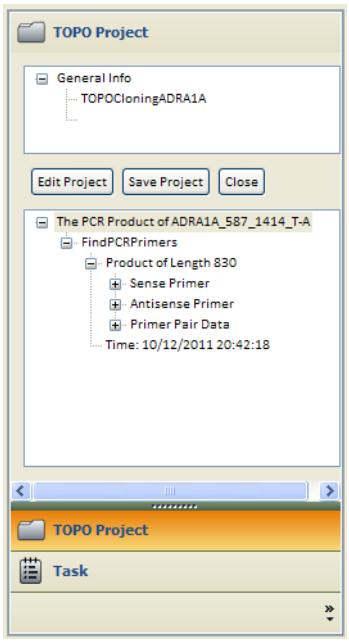
- Click on the **TOPO Cloning** button on the main toolbar.
- Select **File > New > TOPO Cloning Project** from the main menu.
- With a molecule open in Molecule Editor, right-click in the Graphics or Sequence pane and select **Launch TOPO Cloning**.
- Load an existing TOPO® Cloning project as described in "Create, save, and load projects" on page 160.

The tool window is composed of the following panes:

TOPO® Project pane

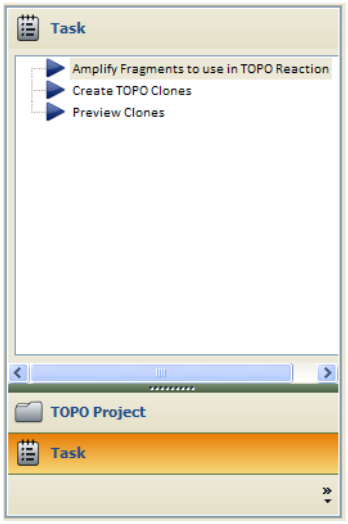
The **TOPO® Project** pane displays the name of the current project, and includes controls for editing, saving, and closing projects. It also lists any generated molecules for the current project.

Click on the **TOPO Project** button to display this pane.



Task List pane

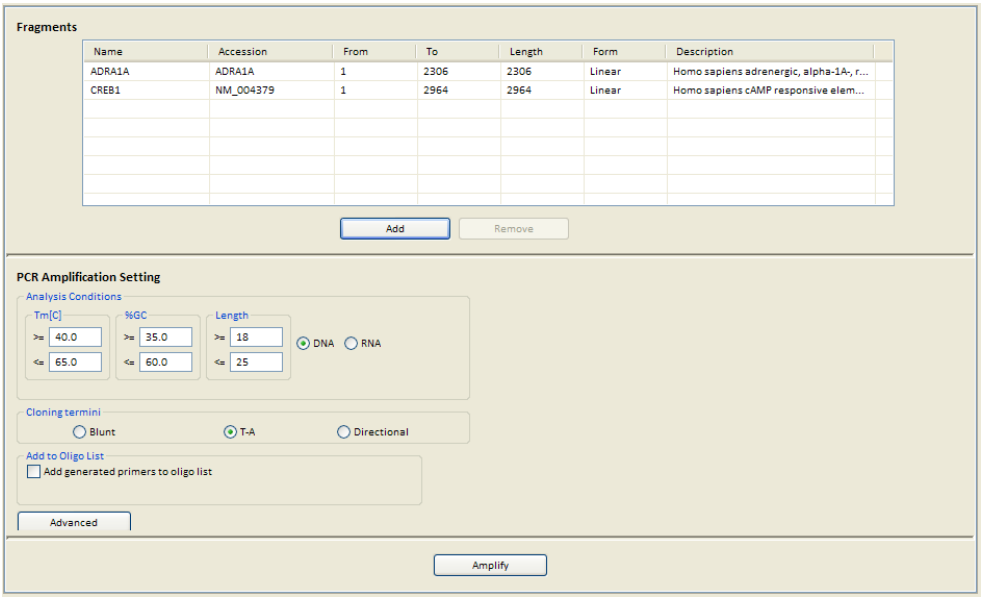
The **Task List pane** displays the list of tasks in the selected project. Click on the **Task** button to display this list.



As you navigate through the TOPO® Cloning workflow, the Task List displays the current task, and allows you to move between tasks by clicking on a different task.

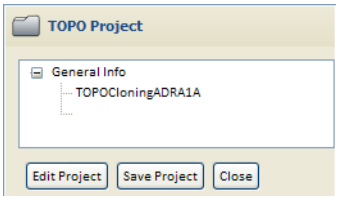
Current Task pane

This pane displays the commands and settings for the currently selected task in the project. As you navigate through the Task List workflow, the functions in this pane will change.

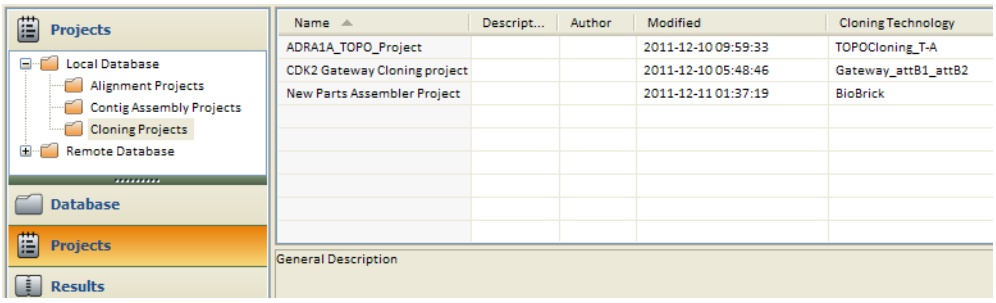


Create, save, and load projects

The tools for saving, editing, and closing projects are located in the TOPO® Project pane:



- To save a new TOPO® Cloning project, click on **Edit Project** in the TOPO® Project pane. In the dialog, enter a name and any description for the project.
- To save changes to a project, click on the **Save Project** button.
- To close a project, click on the **Close Project** button. If there are unsaved changes, you will be prompted to save the project before closing.
- To load an existing project, in Database Explorer, go to the **Projects** list, open the **Cloning Projects** folder in the Local Database, and double-click on the TOPO® Cloning project in the list to open it.



TOPO® Cloning workflow

The Task List displays the default TOPO® Cloning workflow:

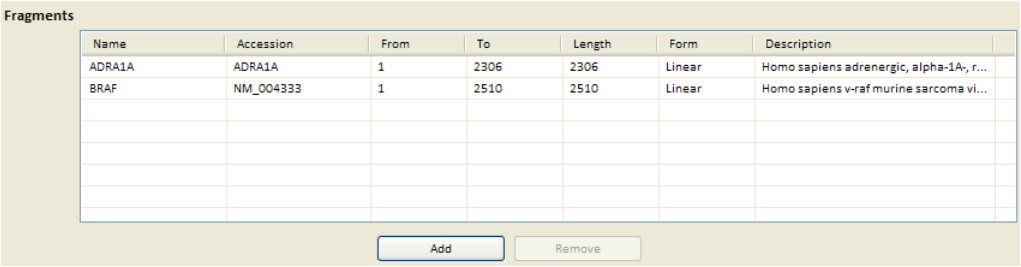
- Amplify Fragments to Use in TOPO® Reaction
- Create TOPO® Clones
- Preview Clones

This section describes each task in the workflow.

Amplify fragments to Use in TOPO® reaction

The **Amplify Fragments to Use in TOPO Reaction** task is the default task displayed when you first open the TOPO® Cloning Tool.

The first step in this task to select the fragment(s) you want to amplify by PCR for use in a TOPO® cloning reaction. These fragments will be listed in the **Fragments** list.



Load molecules or fragments in the Fragments list

- With a molecule open in Molecule Editor, right-click in the Graphics or Sequence pane and select **Launch TOPO Cloning** to load the entire sequence in the list.
- With a molecule open in Molecule Editor, select a portion of the sequence in the Graphics or Sequence pane, right-click, and select **Launch TOPO Cloning** to load only that part of the sequence in the list.
- With the TOPO® Cloning window open, make sure the **Amplify Fragments to Use in TOPO Reaction** task is selected and click on the **Add** button under **Fragments** to select a complete molecule from the database.

To change the regions to amplify in the selected molecules, type a new range in the **From** and **To** fields in the Fragments to Amplify subpane.

PCR Amplification Settings

With the fragment(s) loaded, select the desired amplification settings under **PCR Amplification Settings**. The standard options are described below.

Standard settings	
Tm(C)	Enter limits in degrees Celsius for primer melting temperature (T_m) (temperature at which 50% of primer is a duplex) and the difference between T_m for sense and antisense primers.
%GC	Enter limits in degrees Celsius for primer melting temperature (T_m) (temperature at which 50% of primer is a duplex) and the difference between T_m for sense and antisense primers.
Primer Length	Defaults to 20-25, recommended for TOPO® Primers
DNA/RNA button	Select the type of nucleotide sequence.
Add generated primers to oligo list	Select this checkbox to add the primers you generate to the oligo list

Cloning termini

In the options under **Cloning termini**:

- Choose **Blunt** to generate an amplicon with 2 blunt ends. These will be the exact boundaries of the selection. These amplicons are best used with Zero-Blunt TOPO® vectors.

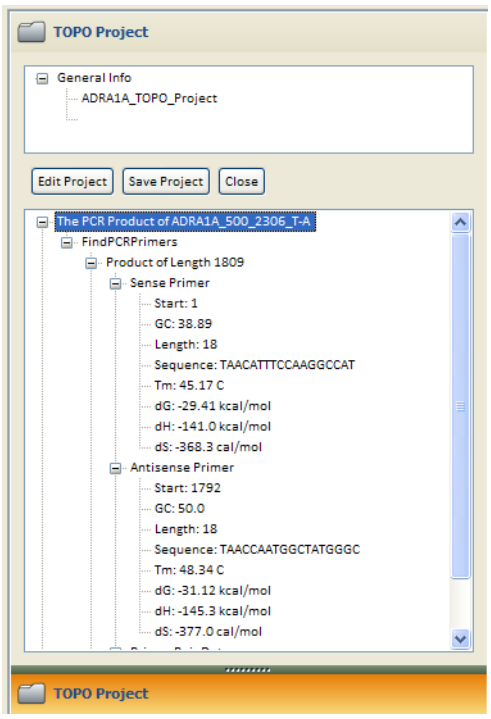
- Choose **T-A** to generate an amplicon as would be produced by amplification with a *Taq* polymerase. The primers in such a case will anneal to the exact boundaries of the selection. However, terminal transferase activity in the enzyme will add 3'-A overhangs to each end of the amplicon.
- Choose **directional** for cloning in TOPO® directional vectors, e.g. pENTR-D/TOPO®. The amplicon will be generated using primers, one of which includes a 5'-CACC extension.

Advanced

For additional amplification settings, click on **Advanced**. The advanced amplification settings are identical for all PCR primers, and are described in Chapter 3, Primer Design.

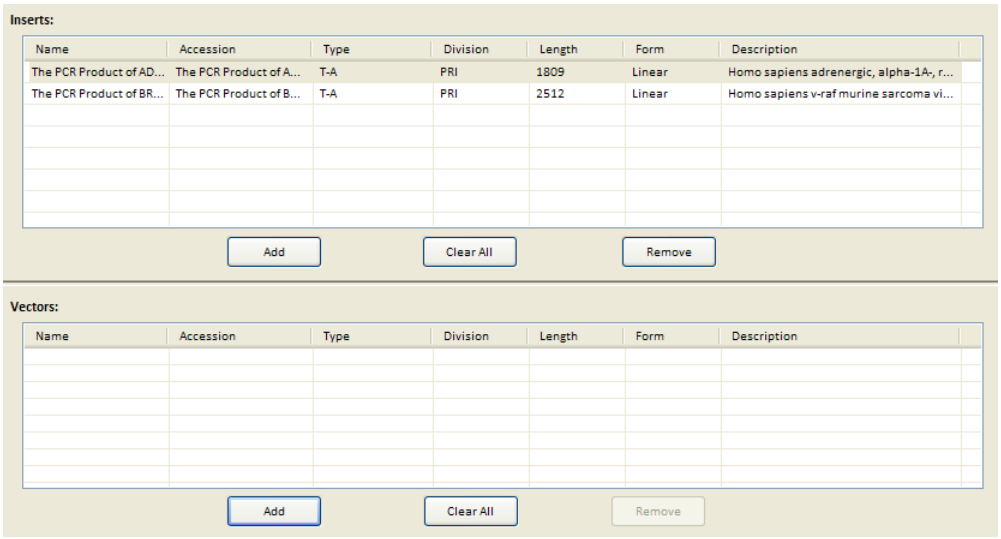
Amplify

When you have made your selections, click on the **Amplify** button. The next task pane will be displayed, and the generated PCR product(s) will appear listed in the **Inserts** subpane, and also listed in the TOPO® Project pane.



Create TOPO® Clones

In the **Create TOPO® Clones** task pane, you can modify the list of fragments and select a vector or vectors with which to create TOPO® clones.



Inserts

The fragments you amplified are listed in the **Inserts** list at the top of the pane.

- To add a previously amplified fragment or a fragment designed with the necessary overhangs by another means, click on the **Add** button and select the molecule from the database.
- To remove a molecule from list, select it and click on **Remove**.
- To clear the entire list, click on **Clear All**.

Vectors

To select a vector:

- Click on the **Add** button and select the vector from the database.
- To remove a vector from list, select it and click on **Remove**.
- To clear the entire list, click on **Clear All**.

Create the TOPO® Clone

When you have made your selections, click on **Create TOPO® Clone**. The **Preview Clones** task pane will open.

14 GeneArt® Cloning

Introduction

Using Vector NTI™ *Express*, you can create GeneArt® assemblies from DNA molecules in the database. Simply select the fragments to be assembled and the software will:

- Analyze the sequences for homologies between the fragments
- Design PCR primers to create the necessary end homologies
- Design stitching oligos for use in GeneArt® High Order assemblies
- Display the assembled molecule with the specified primers and/or stitching oligos in a single molecule file

For more information about GeneArt® technology, visit our web site at **www.lifetechnologies.com** and search for "GeneArt." Detailed technical information for each type of GeneArt® assembly method is available in the following user guides: *GeneArt® Seamless Cloning and Assembly Kit User Guide* and *GeneArt® High Order Genetic Assembly System User Guide*. These are available for download from **www.lifetechnologies.com/manuals** and are supplied with each kit.

GeneArt® Seamless Cloning Overview

GeneArt® Seamless Cloning Technology is a highly efficient, vector-independent system for the simultaneous and seamless assembly of up to four DNA fragments plus a vector totaling up to 13 kb in length (including the vector). The system allows the cloning of the DNA fragments into virtually any linearized *E. coli* vector, does not require pre-existing recombination sites or any extra DNA sequences, and eliminates the need for extensive enzymatic treatments of the DNA such as restriction and ligation. A single proprietary enzyme mixture recognizes and precisely assembles the DNA fragments sharing a 15-base pair (bp) end homology that you can create by PCR amplification.

GeneArt® High Order Assembly Overview

The GeneArt® High-Order Genetic Assembly System is a highly efficient, vector-independent system for the simultaneous assembly of up to 10 DNA fragments plus a vector totaling up to 110 kb in length (including the vector). The system relies on yeast's ability to take up and recombine DNA fragments with high efficiency. This process, termed transformation-associated recombination, greatly reduces *in vitro* handling of DNA and eliminates the need for enzymatic treatments of DNA such as restriction and ligation while allowing precise fusions of DNA sequences.

How GeneArt® Assembly Works

In GeneArt® Assembly, multiple DNA fragments plus a vector can be joined using overlapping sequence homology between fragment ends to splice the fragments together. If homology does not already exist between fragment ends, Vector NTI™ *Express* will automatically design PCR primers that you can use to add homology to the fragments via PCR amplification. Alternatively, for High Order Assembly, up to three sets of "stitching" oligos may be designed to create splices across fragments without homology.

When the fragments and vector are assembled in Vector NTI™ *Express*, the software performs a homology check based on the rules for the GeneArt® Assembly method you are using. The homology of the sequence ends will be analyzed, and in High Order Assembly the internal sequences of the fragments will also be analyzed to ensure that any internal homologies will not interfere with fragment splicing. Then any required PCR primers—and/or, in the case of High Order Assembly, stitching oligos—will be designed.

The rules for homology are described in detail in the *GeneArt® Seamless Cloning and Assembly Kit User Guide* and *GeneArt® High Order Genetic Assembly System User Guide*.

Seamless Cloning: Workflow diagram and an example of PCR primers designed to generate a 15-bp fragment end homology:

High Order Assembly: Workflow diagram and an example of PCR primers designed to generate a 30-bp fragment end homology:

Open the GeneArt® Assembly Tool

You can open the GeneArt® Assembly Wizard from the Molecule Editor with a molecule loaded or from the main toolbar with no molecule selected.

Note: All molecules used in GeneArt® assembly must be linear.

From the molecule editor

1. With a linear DNA molecule loaded in the Molecule Editor, select the part of the sequence you want to clone or make no selection to clone the whole molecule.

Note: The selected molecule or sequence must be >100 bp.

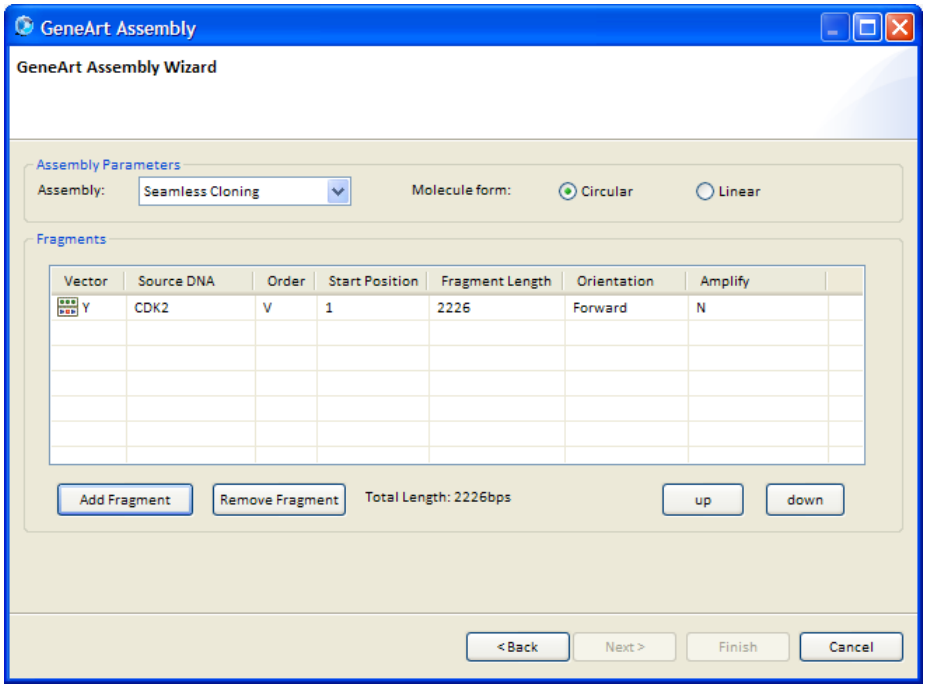
- Right-click and select **Load in GeneArt** or click on the **GeneArt Cloning** button on the main toolbar.

2. Select the strand you want to analyze (**Direct** or **Complementary**).
3. The **GeneArt Assembly Wizard** will open with the molecule or sequence loaded.

From the main toolbar with no molecule selected

1. With no molecule selected, click on the **GeneArt Cloning** button on the main toolbar, or select **Design > GeneArt Cloning**.
2. The **GeneArt Assembly Wizard** will open with no sequence loaded.

GeneArt® Assembly Wizard with molecule selected



GeneArt® Assembly Wizard

The GeneArt® Assembly Wizard window lists the molecules and fragments to be assembled and includes settings for assembly.

Assembly settings

The tool includes different options for **Seamless Cloning** or **High Order Assembly.**

1. Select the appropriate **Assembly** option from the dropdown list.
2. Select the **Molecule Form** you want for the final assembly: **Circular** or **Linear**.

Add and organize fragments

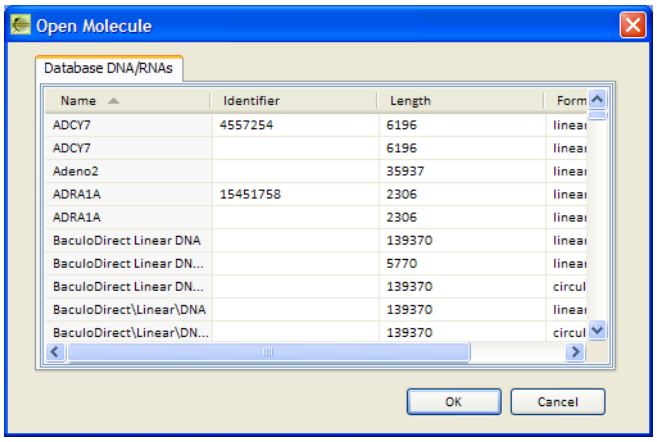
The Assembly Wizard has tools for adding and removing fragments and designating their order of assembly.

Fragment requirements

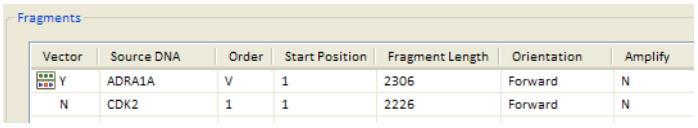
A fragment for GeneArt® Assembly can be any molecule in the database that is ≥100 base pairs in length.

Add fragments

To add a fragment, click on the **Add Fragment** button and select a molecule from the database.



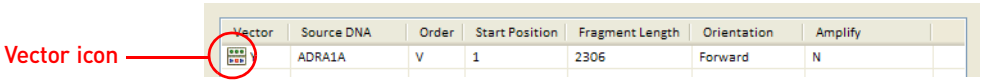
The molecule will appear added to the list in the tool.



Select the vector

One fragment in each GeneArt® assembly must designated as the vector. The vector forms the base fragment onto which other fragments are added. The vector is always the first fragment listed in the Wizard, and is flagged with a “V” in the Order column.

To change the fragment designated as the vector, click inside the Vector column and that fragment will be moved to the top of the list.



Re-order fragments

The fragments will be assembled in the order in which they're listed in the GeneArt® tool, starting with the vector. The order of the fragments is reflected in the Order column, which designates the vector with a "V" and numbers the rest of the fragments in order.

To re-order the numbered fragments, select a fragment in the list and then click on the **Up** or **Down** button to change its position. Note that you cannot move a fragment into the Vector position using the buttons—select the vector as described above.

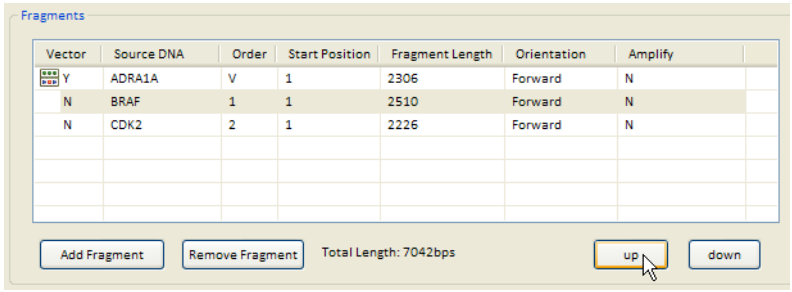


Figure 4 Clicking on the Up button to move a selected fragment up in the list.

Fragment orientation

The orientation of each fragment in the final assembly is indicated in the **Orientation** column (**Forward** or **Reverse**). You can change the orientation by clicking in this column.

Remove a fragment

To remove a fragment, click on it in the list to select it and then click on **Remove Fragment**.

Design PCR primers to create end homology

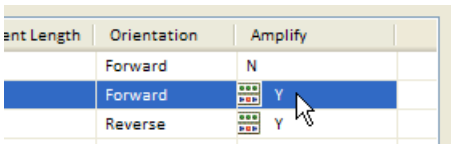
Vector NTI™ *Express* Software can automatically design PCR primers for adding homology to the ends of a fragment for GeneArt® assembly. In the PCR reaction, each primer will add bases to the amplified fragment to create the necessary homology with the adjacent fragment.

Note: If the fragments already have the required end homology, primer design is unnecessary.

This primer design option is available for both Seamless Cloning and High Order Assembly, though the length of the PCR primers varies between assembly methods, based on their different end homology requirements. The primer design rules are described in detail in the GeneArt® user guides. The PCR primer designs will be saved with the final assembled molecule.

To design PCR primers to create the necessary homology for a fragment:

1. Click in the **Amplify** column for that fragment.
2. The **N** will change to a **Y** for that fragment.



Design stitching oligos (High Order Assembly only)

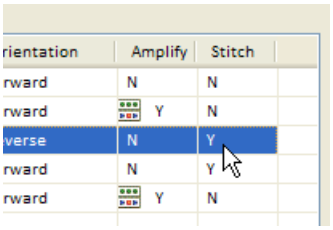
For High Order Assembly, in addition to primer design to create end homology, up to three sets of stitching oligos may be designed to create splices across fragments **without** homology. These oligo linkers create a "bridge" across adjacent fragments to promote recombinational joining in yeast.

Note: A limit of three sets of stitching oligos may be used in a single High Order Assembly, and if stitching oligos are used, a total of five fragments plus vector may be included in the assembly. (This differs from a High Order Assembly without stitching oligos, which can include up to 10 fragments plus vector.)

The stitching oligo rules are described in detail in the *GeneArt® High Order Genetic Assembly System User Guide*. The oligo designs will be saved with the final assembled molecule.

To select stitching oligos for a particular fragment:

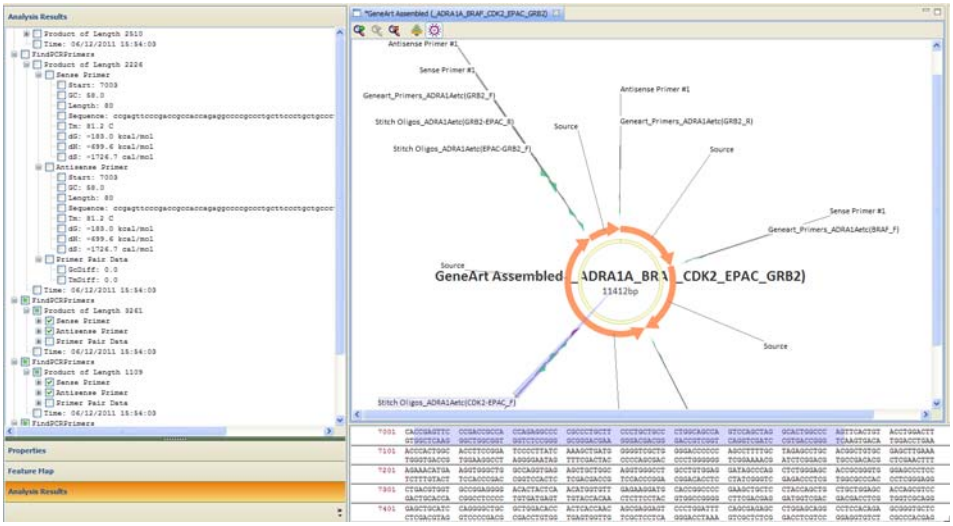
1. With **High Order Assembly** selected, click in the **Stitch** column for that fragment.
2. The **N** will change to a **Y** for that fragment.



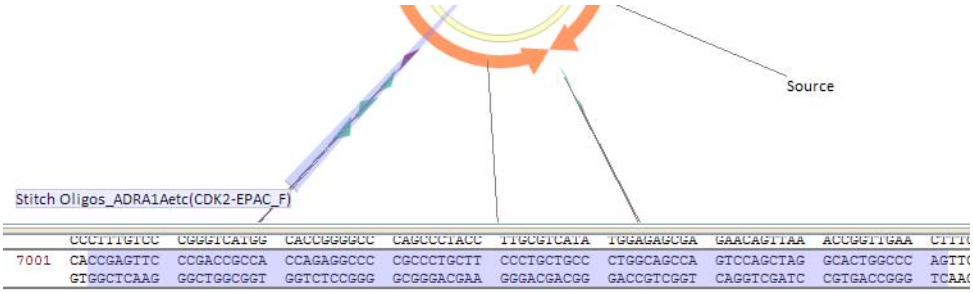
Create the assembled molecule

1. When you have selected your desired settings in the GeneArt® Assembly Wizard, click on the **Next** button.
2. Specify a name for the molecule in the database, as well as names for any PCR primers and/or stitching oligos.

3. The final assembly will be displayed in the Molecule Editor.



- The source fragments will be listed in the Feature Map under **Misc Feature**.
- Click on a PCR Primer or Stitch Oligo in the Graphics pane to highlight that feature in the sequence.



- The PCR primers and stitching oligos are listed in the Analysis Results pane, and will be added to the Ordering list.

15 Parts Assembler

Using Vector NTI™ *Express* Software, you can assemble standard DNA parts that encode basic biological functions from molecules using defined assembly standards. The standard DNA sequences defined in Vector NTI™ *Express* have been developed via an open technical standards-setting process led by the BioBricks Foundation.

At its most basic level, a part is any DNA sequence with a defined biological function (e.g., a promoter region, a sequence encoding a protein, etc.). To join two parts together, upstream and downstream sequences containing sites for specific restriction enzymes are added to each part. This allows for the creation of larger parts by chaining together smaller ones in any desired order. In the process of chaining parts together, the restriction sites are removed, allowing the further use of those restriction enzymes without breaking the new, larger assembly apart. To facilitate this assembly process, each part itself may not contain any of these restriction sites.

Additional Information about Parts and Standards

Vector NTI™ *Express* Parts Assembler is compatible with BioBrick™ part standards. For general information about the BioBricks Foundation, visit **www.biobricks.org**. For information about parts and assembly standards, including instructions and tutorials, visit **http://openwetware.org/wiki/The_BioBricks_Foundation:Standards/Technical/Resources**.

The DNA sequences of thousands of public domain standard biological parts are available through the Registry of Standard Biological Parts at **http://partsregistry.org**.

For a detailed description of the assembly standards, visit **http://openwetware.org/wiki/The_BioBricks_Foundation:Standards/Technical/Formats**.

Note: BioBrick is a trademark of the BioBricks Foundation, Inc. The BioBrick trademark is used herein merely for the purpose of fair use. The BioBricks Foundation, Inc. is not affiliated with Thermo Fisher Scientific. The BioBricks Foundation, Inc has not authorized, sponsored or otherwise endorsed this product or our use of the BioBrick trademark herein. Information about BioBrick™ and the BioBricks Foundation, Inc. can be found at **www.biobricks.org**.

Using the Parts Assembler

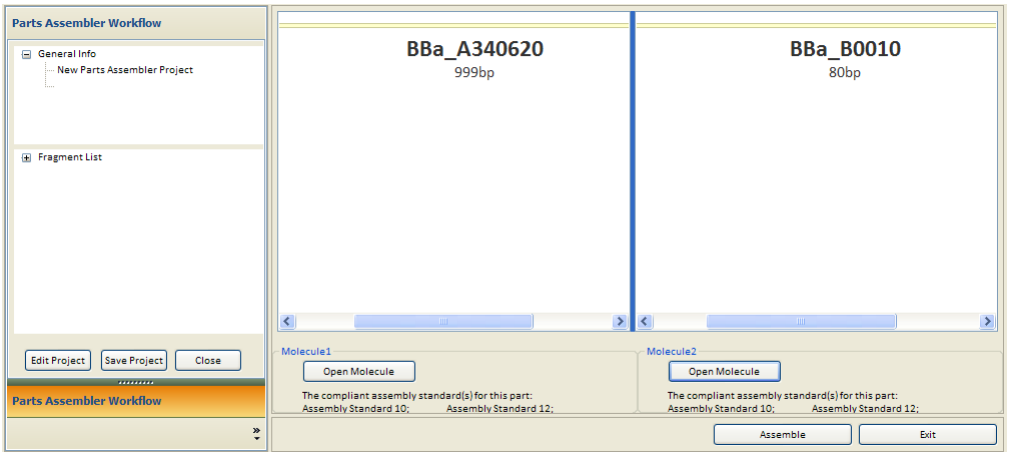
To open the Parts Assembler:

- Click on the **Parts Assembler** button on the main toolbar.
- Select **File > New > Parts Assembler Project** from the main menu.

Selecting Parts

The Parts Assembler window is divided into two panes—one for each part you want to combine.

1. To select the first part that you want to assemble, click on the **Open Molecule** button beneath the left pane, and select the molecule from the database. The molecule must be compatible with a defined assembly standard, as described in the following section.
2. Select the second molecule in the right pane.



When you select a molecule, its sequence will be displayed in the pane and the assembly standards that are compatible with that sequence will be listed at the bottom of the pane. If the molecule does not conform to any of the standards, you will receive a warning message.

Restrictions on Parts

Any molecule in the database can be selected as a part, as long it conforms one of the assembly standards shown in the following table. To conform to an assembly standard, the molecule must not contain the restriction enzyme digestion sites listed in the table for that standard. Some standards also include additional rules. For more information about assembly standards, visit http://openwetware.org/wiki/The_BioBricks_Foundation:Standards/Technical/Formats.

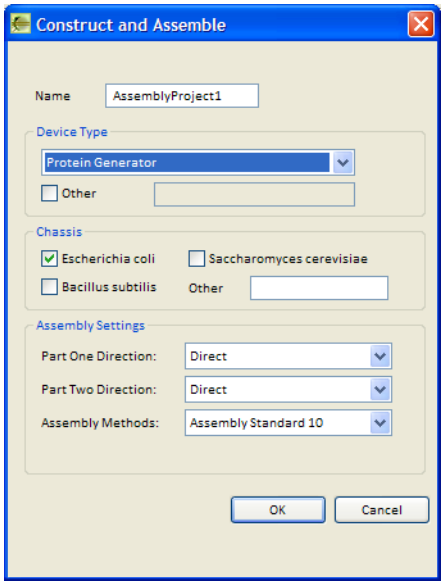
Assembly Standard	Parts must not contain the following restriction sites
10	*Eco*R I, *Not* I, *Xba* I, *Spe* I, *Pst* I, *Nhe* I, *Pvu* II, *Xho* I, *Avr* II, *Sap* I
12	*Eco*R I, *Spe* I, *Nhe* I, *Not* I, *Pst* I
20	*Eco*R I, *Xba* I, *Spe* I, S*bf* I
21	*Eco*R I, *Bgl* II, *Bam*H I, *Xho* I
23	*Eco*R I, *Not* I, *Xba* I, *Spe* I, *Pst* I (in addition, sequences must be in frame without start or stop codons, and may not begin with "TC")
25	*Eco*R I, *Not* I, *Xba* I, *Ngo*MI (aka *Ngo*MIV), *Age*I, *Spe* I, *Pst* I

Note: If you are planning to create and share parts with other individuals and groups using these standards, we recommend designing them so that they contain none of the restriction sites listed for any of the assembly standards, to ensure maximum portability.

Assembly Settings

When you have selected two molecules, click on the **Assemble** button at the bottom of the window.

The **Construct and Assemble dialog** will open with options for assembly.



Enter a name for your new construct in the **Name** field.

Device Type

A "device" in is any construction that performs a particular biological function. The **Device Type** section of the dialog allows you to characterize the type of device you are constructing. This information will be included in the General Description information of the assembled molecule.

The dropdown list includes standard types of devices. Alternatively, you can select **Unassigned** from the list, or enter your own device description by selecting the **Other** checkbox and typing in the field.

Chassis

Parts are typically integrated into the genome of a particular organism—a.k.a. a "chassis"—for propagation and functionality. You can specify the genome of the organism that you will be using by selecting the appropriate checkbox or entering text in the **Other** field. This selected chassis will be listed in the General Description information of the assembled molecule.

Assembly Direction and Standard

Select the direction of each molecule to use for the final assembly, and select the assembly standard to use from the dropdown list.

Note: If there is no shared assembly standard that will work with both molecules, this dropdown list will be blank and no assembly will be possible.

Completing and previewing the assembly

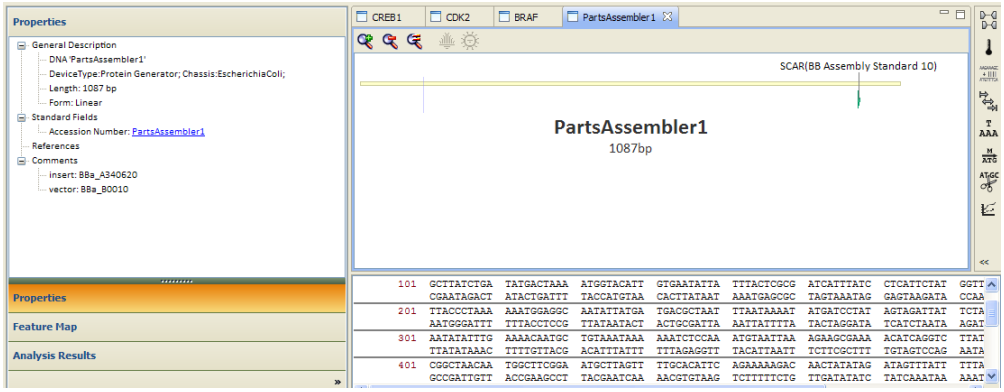
When you click on **OK** in the Construct and Assemble dialog, the assembled molecule will open in a **Preview** window.

You can use the tools in the window to magnify the assembly, or view it as a linear or circular molecule.

Click on **Save** in the Preview window to save the assembly as a molecule in the database.

Viewing the assembly in Molecule Editor

To view the assembled molecule in the Molecule Editor, open the Database Explorer and open the molecule.



The assembled molecule consists of the two original sequences joined together with a standard bridging sequence or “scar” created by the overlap of the restriction digested ends. The scar sequence is determined by the restriction enzymes used in the assembly standard, as described on the web pages mentioned earlier.

- The original parts are listed in the Components list in the Properties pane.
- The scar is listed in the **Feature Map**, and displayed as feature in the Graphics pane.
- The General Description includes the information you entered about the assembled part, including device type and chassis.

16 Contig Assembly using ContigExpress® program

ContigExpress is a program for assembling and editing sequencing fragments, either in the form of text sequences or chromatograms from automated sequencers, into longer contiguous sequences or "contigs."

The ContigExpress® program uses CAP3 to drive the assembly process. This widely-used sequence assembly program can use quality value scores (QVs) in ends trimming, contig construction and consensus calculation. It also produces excellent results when QVs are unavailable. It is capable of using forward-reverse constraints to evaluate contigs, a feature that helps in accurate placement of repetitive sequence fragments. It can also identify and discard chimeric sequencing reads that frequently result from lane tracking errors. One of the major strengths of CAP3 is its consensus generation algorithm based on weighted sum of QVs.

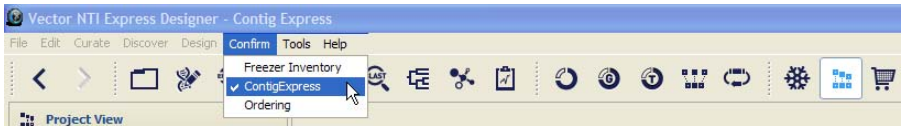
The ContigExpress® program analysis can be saved as a ContigExpress Project, which contains the fragments, their assemblies, and assembly options. In the ContigExpress® program, fragments can be edited directly, with the chromatograms in full view. Changes are tracked and a history is maintained. The contigs generated can be saved to theVector NTI™ *Express* database.

Launch the ContigExpress® program

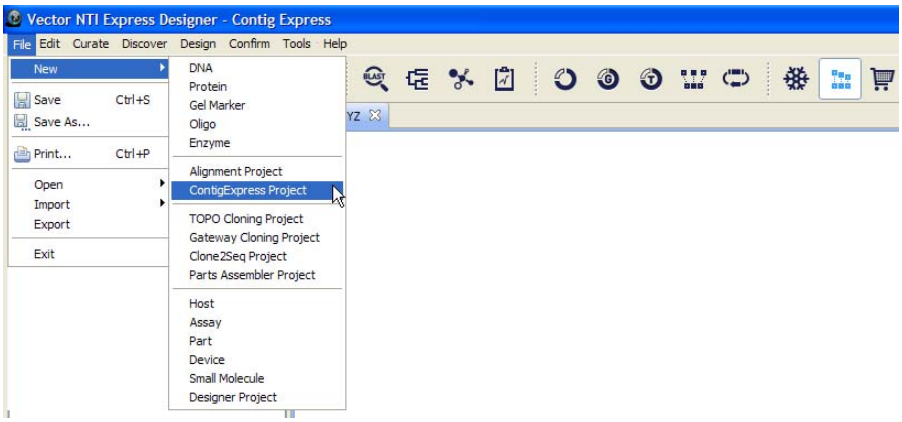
Launch the ContigExpress® program tool in one of the following ways.

- Click on the **ContigExpress** button on the main toolbar.

- In the menu bar, go to **Confirm** ▸ **ContigExpress**.

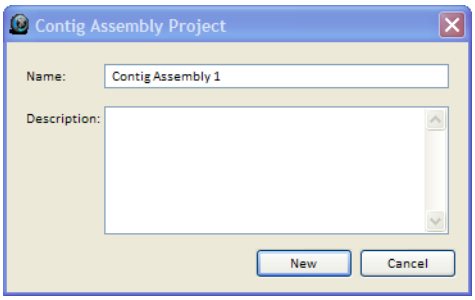


- Go to **File** ▸ **New** ▸ **ContigExpress Project**.



Irrespective of the option you choose to open the ContigExpress® program tool, you will need to enter a Name and Description in the Contig Assembly Project dialog box. Click **New** to create a new project or **Cancel** to exit the dialog box.

Note: The Contig Assembly project takes a default name Contig Assembly ‘count’, where count corresponds to the Contig Assembly project number.



Open a Contig Assembly Project and add fragments to the project

Open the demo Contig Assembly Project

1. Open ContigExpress® program and launch the default demo project.

   The Contig Assembly DemoProject contains 12 fragments, from the installation folder, listed in the Project View pane as well as the Project Editor pane. (Figure 5).

   Example of the demo project folder: C:\Program Files\Life Technologies\Vector NTI Express Designer\Demo Projects\ ):

   The tools for assembling and managing the fragments are located in the Project Editor pane.

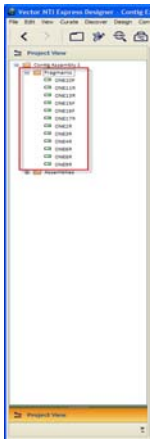


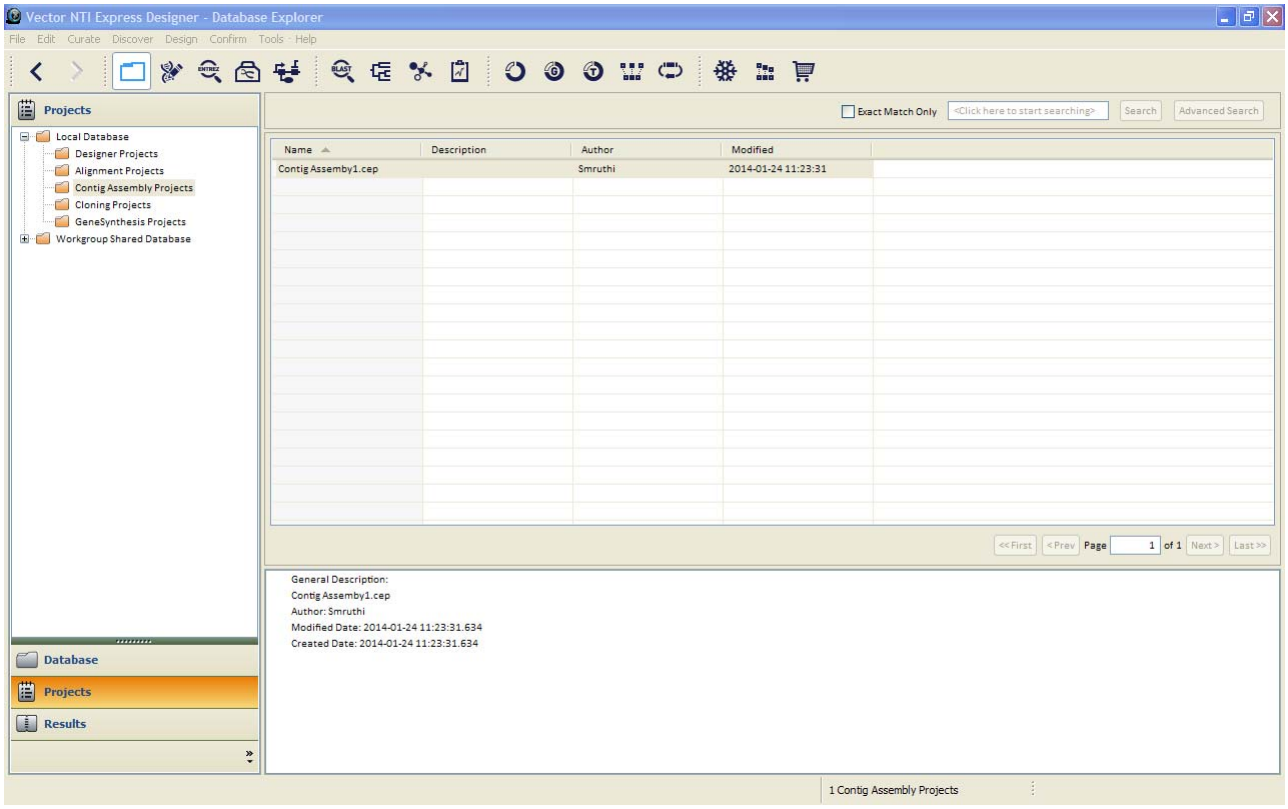
Figure 5 Fragments added to a Contig Assembly project

As you add fragments, they are listed in the Project Viewer pane as well as in the Fragment Viewer. You can then select the fragments for assembling. Refer to "Managing fragments in the Fragments list in the Contig Editor" on page 190 for information on managing and assembling the fragments.

Open an existing Contig Assembly project

To open an existing Assembly Project in the database:

1. In Database Explorer, click on the Projects folder at the bottom-left.
2. Expand the Local Database Projects folder and select **Contig Assembly Projects**.
3. Double-click on the ContigExpress® program project in the list in the right pane to open it.



To load an existing project that has been saved as a **.cep** file:

1. Go to **File ▸ Open ▸ Assembly Project**.
2. In the Open File dialog box, select the project file you want, and click **Open**.

You can also import an existing project saved as a .cep file.

1. Go to **File ▸ Import ▸ Assembly Project**.
2. In the Import into Database dialog box, select the project file you want, and click **Open**.

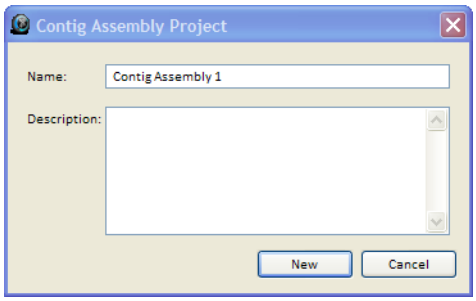
   The imported project will reside in the Database within the Contig Assembly project folder.

Create a new Contig Assembly Project

1. To create a brand-new ContigExpress® program Project, go to **File ▸ New ▸ ContigExpress Project**.

   Note: You can open more than one Contig Assembly project at a time in the ContigExpress® program. Each Contig Assembly project is represented by a tab at the top of the Project Editor pane.

2. In the Contig Assembly Project dialog box, enter a name and description for the project, and click **New**.



Add the fragments

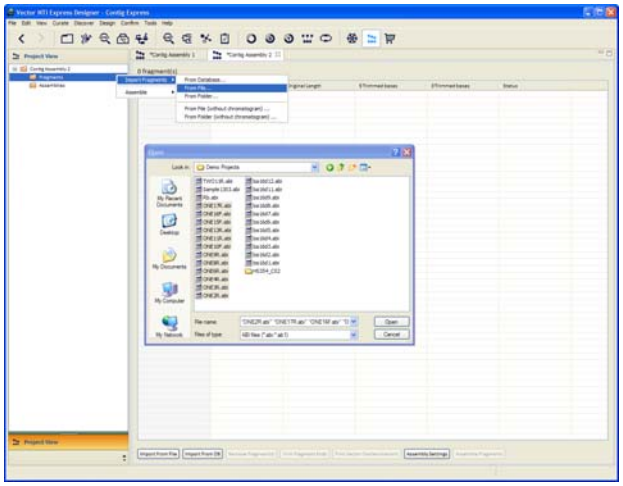
You can add individual fragments to a Contig Assembly project in the Project Viewer and Contig Editor, in one of the following ways:

- **Import from file**
  - In the Project Viewer pane, right-click on the Fragments folder, and go to **Import Fragments ▸ From File...**.
  - In the Contig Editor pane, click **Import Fragment(s) ▸ From File...** to select a supported file type.

  In the Open dialog box, browse to the location of the fragment files, and click **Open**.

  Note: In both the panes, you can choose to add fragments without the chromatogram from a file. To add fragments without the chromatogram, go to **Import Fragments ▸ From File (without chromatogram) ...**

  The supported file types include
    - GenBank® (*.gb)
    - FASTA (*.fasta; *.txt)
    - ABI (*.abi)
    - AB1 (*.ab1)
    - Staden SCF (*.scf)
    - EMBL files (*.txt)
    - Text files (*.txt)
    - Phred files (*.phd.1; *.seq)
    - Phrap ACE files (*.ace)

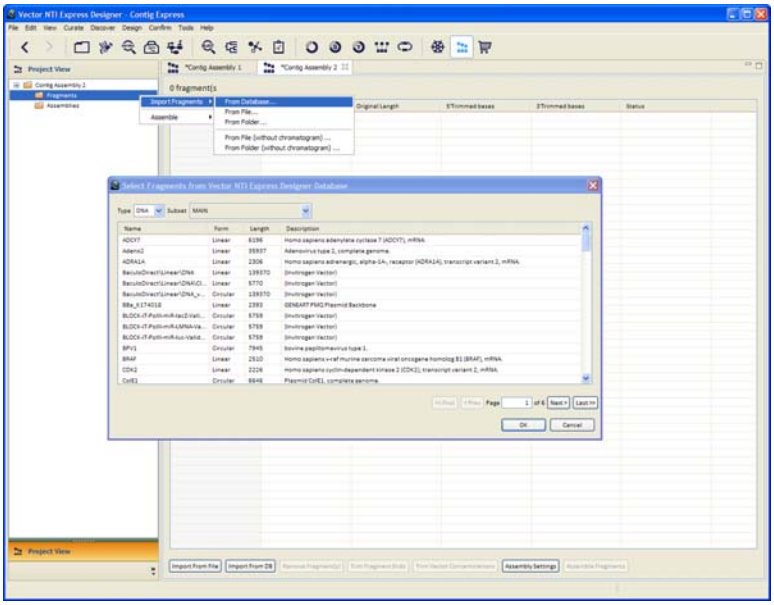


- **Import from database**
  - In the Project Viewer pane, right-click on the Fragments folder, and got to **Import Fragments ▸ From Database...**.
  - In the Contig Editor pane, click **Import From DB** to select DNA molecules from the local database.

    In the Select Fragments from Vector NTI Express Designer Database dialog box, select the DNA molecule you want to add, and click **OK**.

    Use the <<First , <Prev , Next> , and Last>> to scroll through the fragment list.

    Note: Press the **Ctrl** or **Shift** keys to select multiple files/molecules.

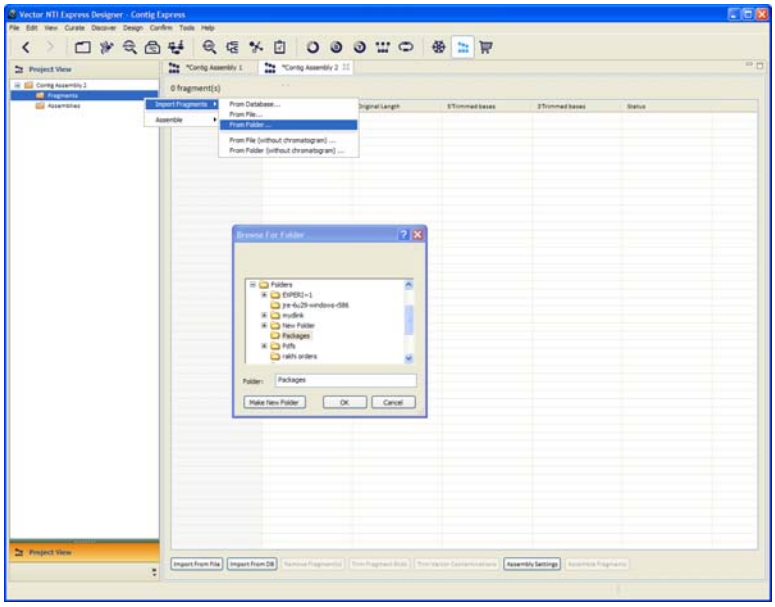


- **Import from folder**
  - In the Project Viewer pane, right-click on the Fragments folder, and got to **Import Fragments ▸ From Folder...**.
  - In the Contig Editor pane, click **Import Fragment(s) ▸ From Folder...** to select DNA molecules from the local database.

    Note: Press the **Ctrl** or **Shift** keys to select multiple files/molecules.

  In the Browse for Folder dialog box, select the folder which contains the fragments to be added, and click **OK**. Click Make New Folder to create a new folder to hold the fragments to be added.

Note: In both the panes, you can choose to add fragments without the chromatogram from a file. To add fragments without the chromatogram, go to **Import Fragments ▸ From Folder(without chromatogram) ...**



After you import the fragments, a dialog box appears, confirming the number of fragments imported during import.

Export Fragments

Use the Export Fragments feature in the Project Viewer pane or the Contig Editor pane to export the fragment(s) and save them as .gb, .fasta, or .txt files. To export fragment(s):

1. Select one or more fragments in the Project Viewer pane or Contig Editor pane.
2. Right-click on the fragment(s) and select **Export Fragments** from the drop-down menu.

   The fragments can be exported as a molecule or saved as .gb, .fasta, or .txt files.

Remove fragments

You can remove fragments from a Contig Assembly project in one of the following ways:

- Right-click menu

  Right-click on a fragment name in the Project Viewer pane or the Contig Editor pane, and select **Delete**.

  Note: In the Project Viewer pane or Contig Editor pane, you can delete multiple fragments in a single right-click action.

- From the Fragment/s management button in the Contig Editor pane

  In the Contig Editor pane, select a single or multiple fragments and click **Remove Fragment(s)** at the bottom of the Contig Editor pane.

The above methods remove fragment(s) from the project only and do not delete them from the database.

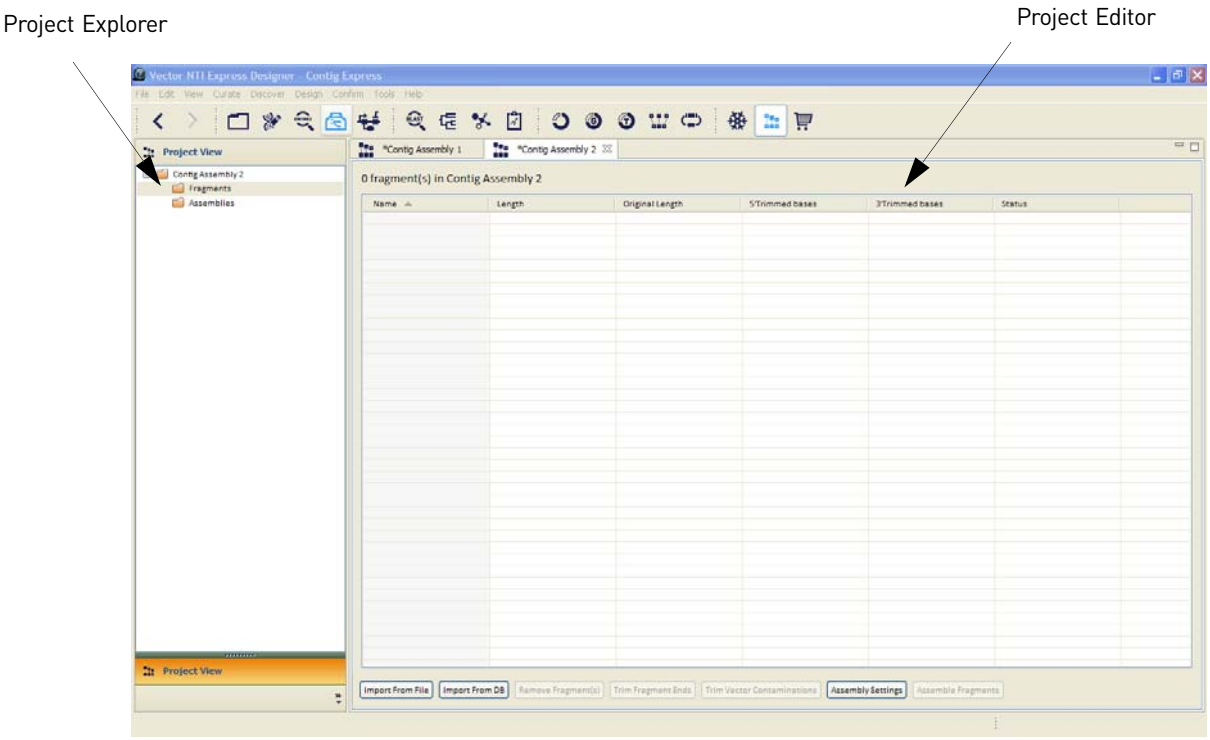
Rename fragments

1. To rename a fragment, right-click on the fragment name in the Project Viewer pane or Contig Editor pane, and select **Rename**.

2. In the Rename dialog box, enter a new name for that fragment and click **OK** to implement the change or **Cancel** to exit the dialog box.

Examine the ContigExpress® program Project Explorer Window

The ContigExpress® program opens to an empty Project Explorer and Project editor window where you can view and edit fragments, assemble and dissolve contigs, and dismiss assemblies. The window is similar to other Vector NTI™ *Express* Software windows with a Project View pane and the Contig Editor pane.



Note: Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

The ContigExpress® program window consists of the following panes:

- **Project View pane:** Located at the left-hand side of the ContigExpress® program window. The Project View pane displays the Contig Assembly project and the fragments and assemblies it is made of.
  - **Fragment Viewer**: Displays the sequence of the selected fragment as well as a chromatogram of the fragment when the fragment is selected in the Project View pane. The Fragment Viewer includes the following sub-panes:
    - Sequence Pane
    - Properties Pane

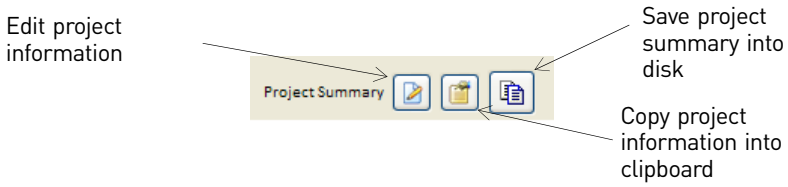
– **Contig Viewer**: Displays the Contig Assembly in graphical form, with the overlapping assembled fragment sequences and chromatograms when the contig is selected in the Project View pane. The Contig Viewer includes the following subpanes:

  • Sequence Pane
  • Properties Pane
  • Alignment Pane

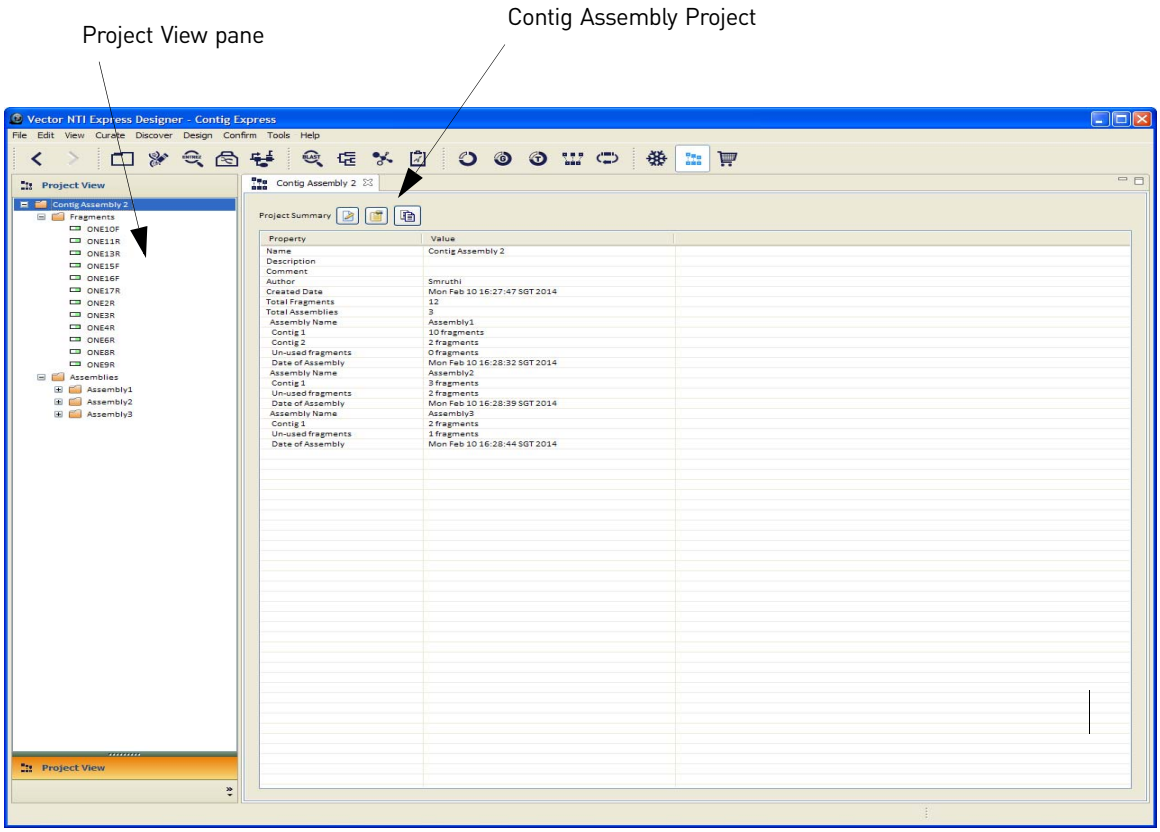
  You can view the Fragments and Assemblies summary when you select the Fragments and Assemblies folders, respectively, in the Project View pane.

• **Contig Editor**: Located at the right-hand side of the ContigExpress® program window. When you select a Contig Assembly project in the Project View pane, the Contig Editor displays the Project Summary that includes the following:

  – Edit project information
  – Save project summary into disk
  – Copy project information into clipboard



Selecting a fragment within the Fragments folder in the Project View pane displays the **Fragment Viewer** in the Contig Editor. Selecting a contig within the Assembly folder in the Project View pane opens a **Contig Viewer** in the Contig Editor.



Fragment viewer

When you select a fragment within the Fragments folder in the Project View pane the **Fragment Viewer** is displayed in the Contig Editor.

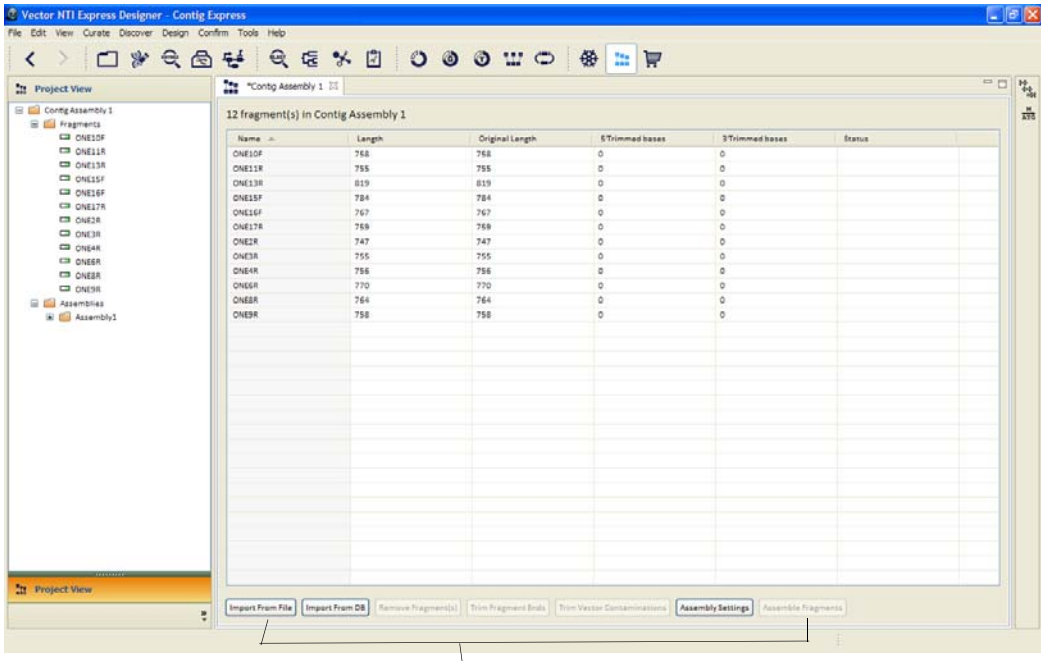
Viewing Fragments summary

To view a summary of the Fragments added to a Contig Assembly Project, select the Fragments folder in the Project View pane.

Note: The Fragment list always contains the original fragments imported from files or the database. The information on trimmed fragments is displayed under the assembled contigs.

By default, the Project Explorer is configured to display six columns listing the following:

- Fragment identification code of each fragment
- Current length
- Original length
- 5’ trimmed bases
- 3’ trimmed bases
- Fragment Status.



Manage Fragments/Assembly Settings/Trim Settings

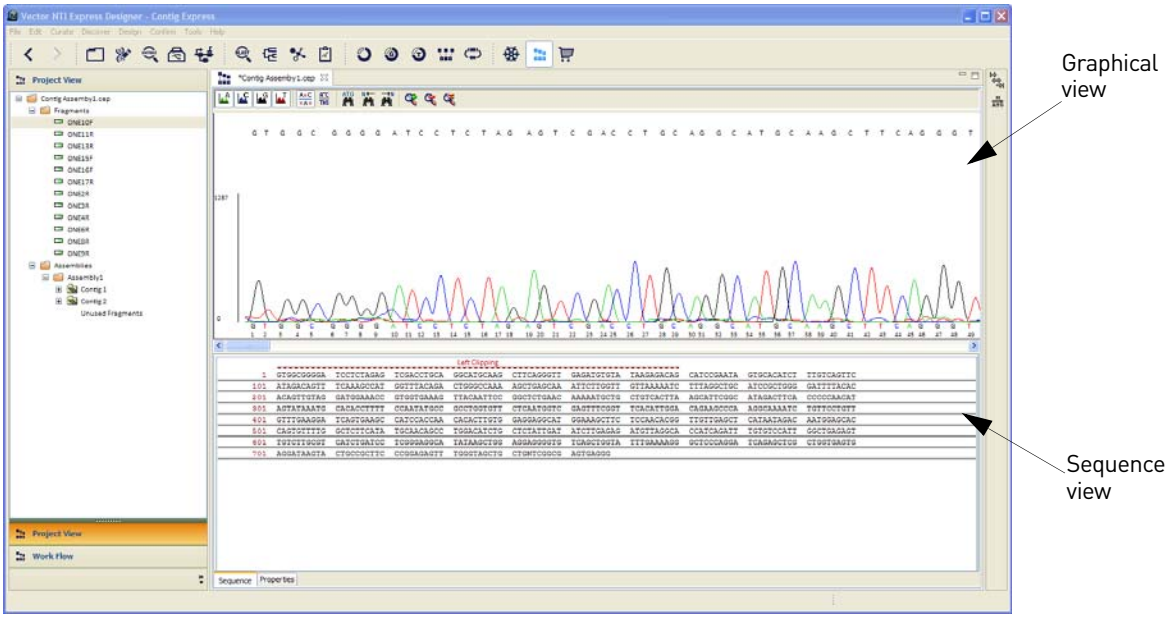
The following table includes the features within the right-click drop-down menu for individual fragments in the Project Viewer pane as well as the Contig Editor:

Feature	Sub-menu	
Load Fred Quality Values		
Discard Chromatogram		

Feature	Sub-menu	
Make Reverse Compliment		
Trim Fragment Ends		
Trim Vector Contamination		
Call Secondary Peaks for the selected fragments		
Set As Reference		
Analyses ▸	Analysis Monitor ▸	Sim4 Analysis
		Spidey Analysis
Alignment ▸	AlignX - Align Selected Molecule(s)	
	AlignX - Open New Alignment Project	
Assemble ▸	ContigExpress -Assemble Selected Molecule(s)	
	ContigExpress - Open New ContigExpress Project	
	Send to GeneArt Assembly	
Tools ▸	BLAST Search	

Viewing individual Fragment details

You can view the details of each Fragment added to a Contig Assembly Project in the Contig Editor. To view the details, select a Fragment in the Project View pane.



The Contig Editor displays two views of the fragment selected in the Project View pane.

- **Graphical view** in the upper half of the Contig Editor
- **Sequence view** in the lower half of the Contig Editor

The lower half of the Contig Editor has two tabs:

- **Sequence**
- **Properties**

Graphical view

The graphical view pane toolbar includes the following features:

- Six toggle buttons to allow clutter-free viewing of the fragment chromatogram.
  - **Show/Hide A - trace**: Click  to show/ hide the Adenine nucleic acid base
  - **Show/Hide C - trace**: Click  to show/ hide the Cytosine nucleic acid base
  - **Show/Hide G- trace**: Click  to show/ hide the Guanine nucleic acid base
  - **Show/Hide T - trace**: Click  to show/ hide the Thymine nucleic acid base
  - **Show/Hide Two Strands**: Click  to display only two out of the four nucleic acid base strands at one time.
  - **Show/Hide Deleted Peaks**: Click  to show/hide the trimmed/clipped traces of fragment. The Show mode appears grey in color.
- Three Find buttons to find a particular strand.
  - **Find**: Click  to find a particular fragment. In the Find dialog box, you can insert sequence and click **Find Next** to look for a match. You can find ambiguous character as well as Complementary sequence.
  - **Find Previous Ambiguous**: Click  to find a previous ambiguous fragment.
  - **Find Next Ambiguous**: Click  to find a next ambiguous fragment.
- Three Zoom buttons.
  - **Zoom In**: Click  to zoom in to the graphical view pane.
  - **Zoom Out**: Click  to zoom out from the graphical view pane.
  - **Fit to Window**: Click  to display the default view of the graphical view pane.

Note: To take a screenshot of the current view of the fragment in the graphical view pane, right-click anywhere in the graphical view pane and select **Camera**. You can then 'paste' the information in MS Word, MS Paint, or any other graphic-editing tool for future reference.

Print Report

You can print a report of the graphical view of the fragment chromatogram.

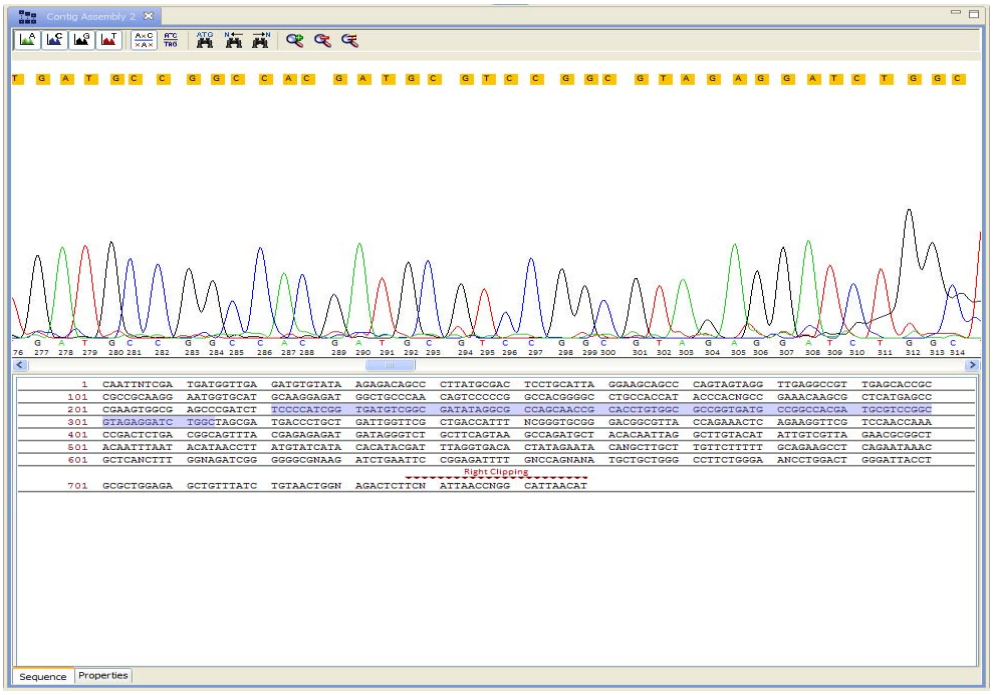
Go to **File** ▸ **Print**; in the Print dialog box, select **PDF Creator** from the Select Printer Menu, and click **Print**.

You can save the Print Report at a location of your choice on your computer.

Sequence tab

The sequence view of a fragment is displayed in the lower half of the Contig Editor pane.

You can select a particular sequence of the fragment in the Sequence view pane and view the corresponding chromatogram in the graphical view pane.



Use the right-click drop-down menu to perform tasks in the Sequence view pane.

1. Select a sequence in the Sequence view pane.
2. Right-click and select from:
   a. **Copy**: To copy the selected sequence.

      Note: Keyboard shortcut key to copy the selected sequence is Ctrl +C.

   b. **Paste**: To paste the copied sequence
   c. **Delete**: To delete the selected sequence
   d. **Select All**:To select the entire fragment (if you have not selected a part of it)

      Note: Keyboard shortcut key to select the entire fragment is Ctrl +A.

   e. **Clear Analysis Results**: Activated when you perform analysis from the right hand side toolbar pane

Print Report

You can print a report of the sequence view of the fragment chromatogram.

Go to **File ▸ Print**; in the Print dialog box, select **PDF Creator** from the Select Printer Menu, and click **Print**.

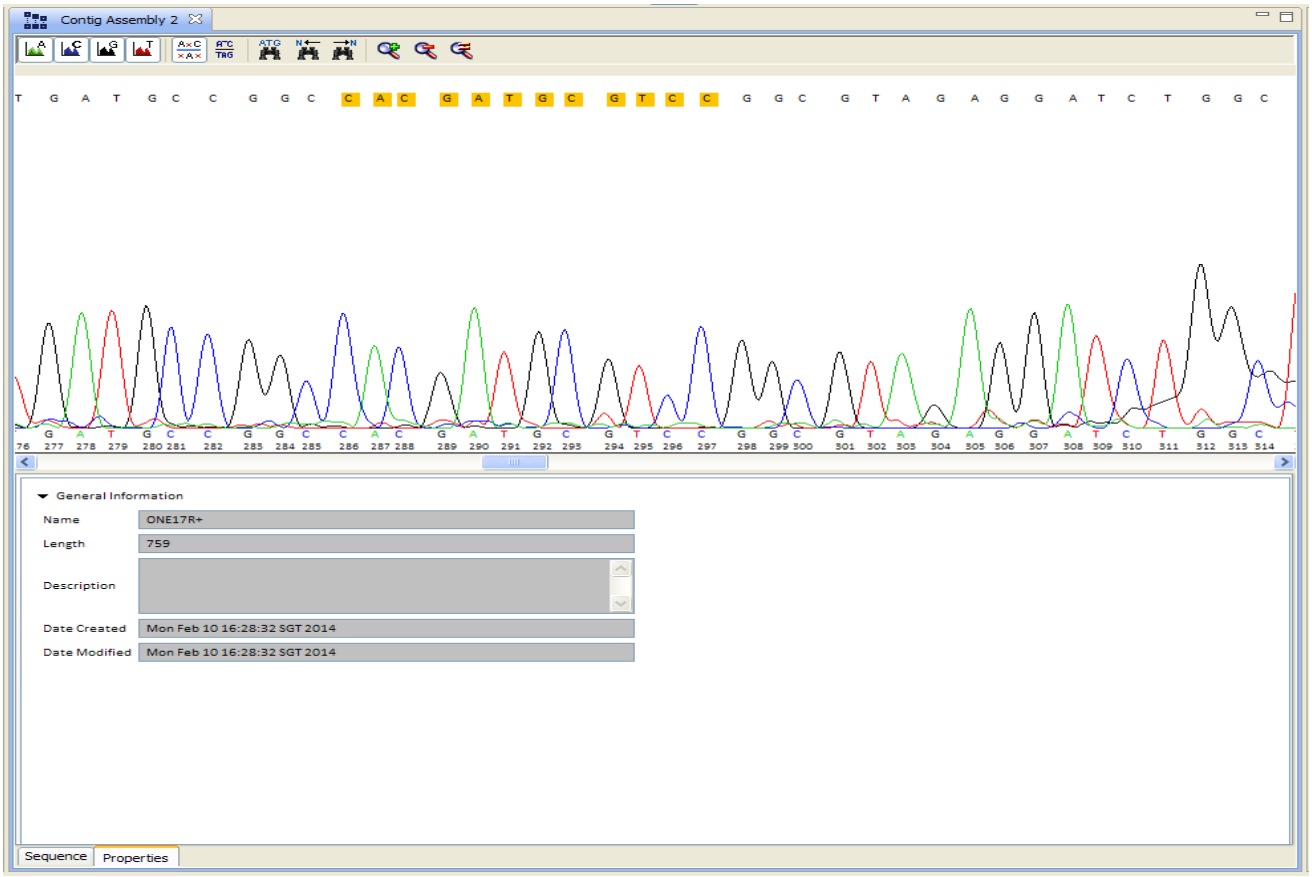
You can save the Print Report at a location of your choice on your computer.

Properties tab

The Properties tab in the lower half of the Contig Editor, displays the general information of the fragment selected in the Project View pane.

The **General Information** includes:

- Name of the selected fragment
- Length of the fragment
- Description
- Date on which the fragment was created
- Date on which the fragment was modified

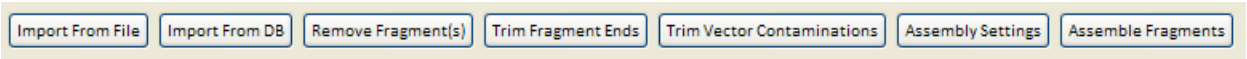


Managing fragments in the Fragments list in the Contig Editor

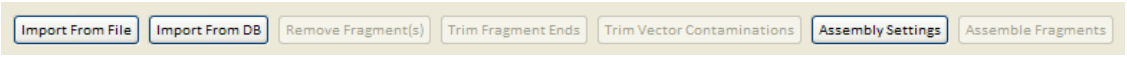
To manage the fragments in the Contig Editor, use the features provided at the bottom of the Contig Editor pane. You can choose from:

- Import Fragment(s): to import fragments saved as files, in a folder, or in the local database on your computer
- Import from DB: to import fragments saved in your database
- Remove Fragments: to remove one or more fragments added to a project
- Trim Fragment Ends: to trim the 5′ or 3′ ends of a fragment
- Trim Vector Contaminations: to trim the contaminated fragments
- Assembly Settings: to take you to the Contig Assembly Setup dialog box to edit the assembly settings. For more information on Assembly Settings, see "Assembly Settings" on page 196.

- Assemble Fragments: to create an assembly from the fragments added to a Contig project.



Note: When there are no fragments selected in the Contig Editor, the **Remove Fragment(s)**, **Trim Fragment Ends**, **Trim Vector Contaminations**, and **Assemble Fragments** are disabled. The Assemble Fragments button is also disabled when only one fragment is selected in the Contig Editor.



Import Fragment(s)

Use this feature to add fragments, as files saved on your computer, to a Contig Assembly project.

1. Click **Import from File**, and in the Open dialog box, browse to the location where you have saved the fragment files.

   Note: Compatible formats include GenBank® (*.gb), FASTA (*.fasta*.txt), ABI (*.abi), AB1 (*.ab1), Staden SCF (*.scf), EMBL files (*.txt), Text files (*.txt), Phred files (*.phd.1; *.seq), and Phrap ACE files (*.ace).

2. Click **Open** to import the fragment files or **Cancel** to exit the dialog box.
3. Click **OK** in the pop-up message confirming the number of fragments added to the Contig Assembly Project.

Export fragments

Use the Export Fragments feature in the Contig Editor pane to export the fragment(s) and save them as .gb, .fasta, or .txt files. To export fragment(s):

1. Select one or more fragments in the Contig Editor pane.
2. Right-click on the fragment(s) and select **Export Fragments** from the drop-down menu.

   The fragments can be exported as a molecule or saved as .gb, .fasta, or .txt files.

Import from DB

Use this feature to add fragments, saved in the database on your computer, to a Contig Assembly project.

1. Click **Import from Database**, and in the Select Fragments from Vector NTI™ *Express* Software Database dialog box.
2. Use the 

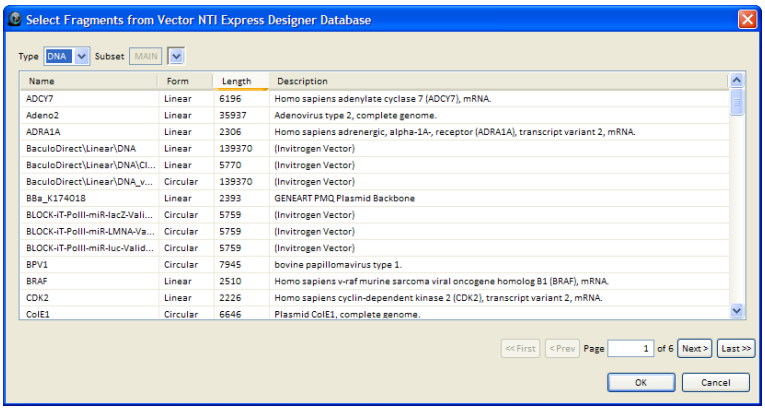
, 

, 

, and 

to scroll through the fragment list.
3. Select the fragment from the list, and click **OK** to import that fragment into the Contig Assembly Project or Cancel to exit the dialog box.

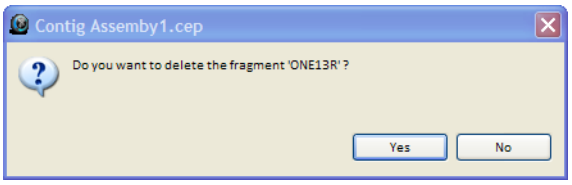
4. Click **OK** in the pop-up message confirming the number of fragments added to the Contig Assembly Project.



Remove Fragments

Use this feature to delete fragments from a Contig Assembly project.

1. In the Contig Editor, select one or more fragments.
2. Click **Remove Fragment(s)**, and in the 'Contig Assembly project name' dialog box, click **Yes** to confirm if you want to delete the selected the fragment(s).
3. Click **No** to exit the dialog box.



Trim Fragment Ends

Use this feature to trim the 5′ or 3′ ends of a fragment added to a Contig Assembly project.

Note: You can also access this feature from the right-click drop-down menu of individual fragments.

1. In the Contig Editor, select one or more fragments, and click **Trim Fragment Ends**.
2. In the ContigExpress® program - Fragment Ends Trimmer dialog box,
   a. Use the 5′ end tab to trim the 5′ end of the fragment, and select or deselect the following checkboxes:

      Note: You can also change the default values if necessary.

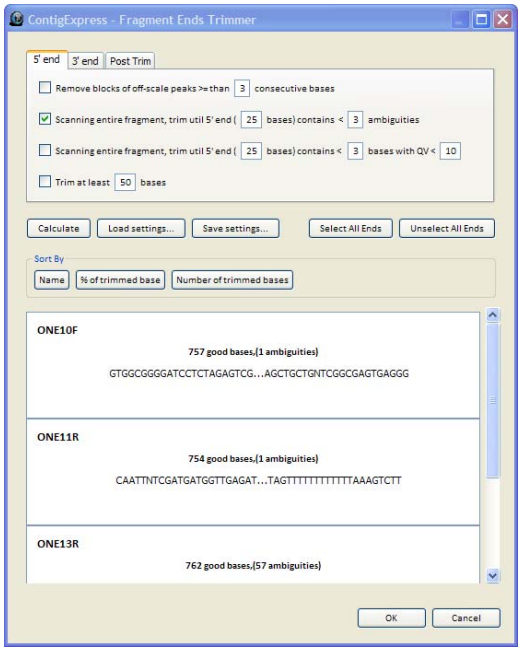
      - **Trim __ consecutive off-scale bases** removes the defined # of consecutive bases that are below acceptable criteria
      - **Trimming until __ 5′ bases** can be based upon chromatogram quality
      - **Trim at least __ 5′ bases** is an arbitrary setting that may be based upon the fact that your primers have "tails"

   b. Use the 3′ end tab to trim the 3′ end of the fragment, and select or deselect the following checkboxes:

Note: You can also change the default values if necessary.

- **Trim until __ 3′ bases** % of maximum value removes bases whose peaks do not meet the % value you define
- **From position… trim first** removes these bases if you can see that quality well into the sequence is not good
- **Trim until __ 3′ bases… ambiguities** trims poorly resolved bases
- **Trim at least…** removes bases that begin to be poorly resolved at the 3′ end Post trimming
- **Define maximum…** set the length of the fragment that must be left after trimming
- **Leading and trailing…** removes poorly resolves nucleotides that may be left after trimming
- **Remove poly A/T…** removes these nucleotides that may be present if the sequence was flipped, producing a poly-T 5′ end

3. Click,
   - **Calculate** - to preview the trimmed results. Click **OK** to propagate trimmed results to fragments
   - **Load settings...** - to use pre-loaded Fragment Ends Trimmer Setup Files
   - **Save settings...** - to save the new settings
   - **Select All Ends** - to select all ends of the fragment
   - **Unselect All Ends** - to unselect all ends of the fragment
4. If you have selected multiple fragments, use the following Sort By buttons to sort the fragments:
   - **Name** - to sort the fragments by their name
   - **% of trimmed base** - to sort the fragments by the percentage of the trimmed base
   - **Number of trimmed bases** - to sort the fragments by the number of trimmed bases
5. Click **OK** to implement the edits or **Cancel** to exit the dialog box.



Trim Vector Contaminations

Use this feature to trim from sequenced fragments any residues from your frequently used cloning vectors that may have been amplified in the sequencing process.

Note: You can also access this feature from the right-click drop-down menu of individual fragments.

1. In the Contig Editor, select one or more fragments, and click **Trim Vector Contaminations**.
2. In the ContigExpress® program - Fragment Vector Contamination Trimmer dialog box, click,
   - **Calculate** - to preview the trimmed results. Click OK to propagate trimmed results to fragments
   - **Settings...** - to access the Fragment Vector Contamination Trimming Settings dialog box. The vectors or polylinker sites you selected in the Local Database should be listed in the List Pane at the left. The Edit Pane where you can edit the vector sequence is on the right.

     In the Fragment Vector Contamination Trimming Settings dialog box:

   a. Edit the following Common Settings:
      - Minimum Vector overlap is the minimum number of bases in the fragment that overlap with those on the clone. This setting must be 5 or greater.
      - Remove additional bases from contaminated 3′ end refers to the additional bases to be removed.
      - Minimum Vector overlap with ambiguities includes poorly resolved residues
      - Remove additional bases from contaminated 3′ end
      - Vector match threshold is the percentage required to match

   b. Select one or more polylinkers from the Polylinker List and click
      - Remove Polylinker(s) to remove the selected polylinker(s)
      - Add Insertion Point to add an insertion point to indicate location where the fragment was inserted in the vector
      - Remove Insertion Point to remove the insertion point from the location where the fragment was inserted as a vector.
      - Select All to select all the insertion points
      - Remove Selected to remove the selected insertion points

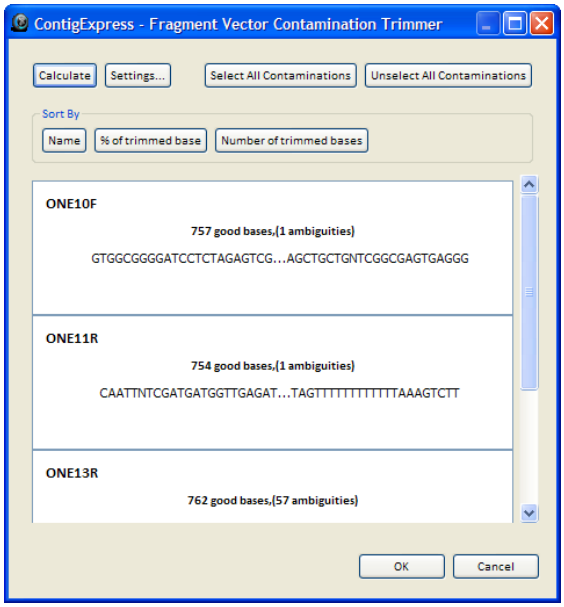
      Note: To add a polylinker to the Polylinker List, go to **Database Explorer ▸ DNA/ RNA Molecules**. Select a DNA molecule from the list in the right-hand side pane. Right-click on the DNA Molecule name, and select **ContigExpress ▸ Polylinker to ContigExpress**. The DNA molecule will be added as Polylinker to the Polylinker List in ContigExpress® program.

      IMPORTANT! You can send only one vector at a time from the local database as a polylinker.

      ---

   c. Click
      - **Load settings...** - to use pre-loaded Fragment Vector Trimmer Setup Files
      - **Save settings...** - to save the new settings
      - **OK** to implement the changes or **Cancel** to exit the dialog box

3. If you have selected multiple fragments, use the following Sort By buttons to sort the fragments:
    - **Name** - to sort the fragments by their name
    - **% of trimmed base** - to sort the fragments by the percentage of the trimmed base
    - **Number of trimmed bases** - to sort the fragments by the number of trimmed bases
4. Click **OK** to implement the edits or **Cancel** to exit the dialog box.



Calling Secondary Peaks for Fragments

If conflicting residues appear on a chromatogram, you can alter the nucleotides as you wish.

1. With one or more fragments selected in the Project Explorer List Pane, select **Edit ▸ Call Secondary Peaks for Selected Fragments**. The Call Secondary Peaks dialog box opens, listing all selected fragments.

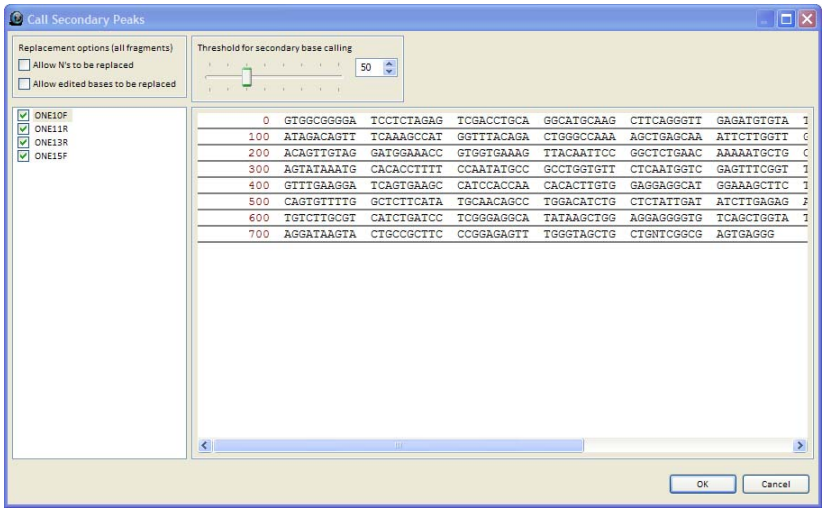
    Note: You can also access this feature from the right-click drop-down menu of individual fragments.
2. Select the check box for replacement options. The options include:
    - Allow N's to be replaced
    - Allow edited bases to be replaced
3. Use the Threshold for secondary base calling scroll bar to adjust the threshold.
4. Click **OK** to continue.

Note:

- Because this operation applies only to fragments, if contigs are selected, they will be skipped.

- If a fragment selected for secondary peak search is currently open in its Fragment Window, initiating this operation in Project Explorer automatically forces the fragment into the read-only mode in the Fragment Window to avoid possible inconsistencies in the project after a successful search.
- If any of these fragments were already changed in their respective windows, select **Yes** to save the changes or **No** to lose the changes or **Cancel** to skip the fragment in question.

The Call Secondary Peaks for Selected Fragments dialog box looks like this:



Assembly Settings

Use this feature to take you to the Contig Assembly Setup dialog box to edit the assembly settings. The Contig Assembly Setup dialog box has the following tabs:

- Assembly
- Clipping
- Overlap
- Contig
- Description
- Lite Settings Tab

Assembly

There are three main check boxes to control some important aspects of the sequence assembly process, as shown in the table below:

	Assembly Options
Use Quality Values	When available, base quality values (QVs) are used to trim poor-quality ends, compute overlaps between reads, construct multiple sequence alignments, and generate a consensus sequence. Use of such scores is optional; when they are unavailable, assembly will still proceed. Default ON.

	Assembly Options
Detect Chimeric Reads	Chimeric reads consist of pieces from different parts of the sequence region, usually generated as artifacts. They are identified based on overlap conflicts and are excluded from the construction of contigs. The mechanism of detecting chimeric reads can be enabled or disabled with "Detect Chimeric Reads" check box. Default: ON
Use Forward-Reverse Constraints	When reads from both ends of subclones are available, constraints are satisfied if they lie on opposite strands of a double-stranded DNA molecule and within a specified minimum and maximum range. This corrects assembly errors due to misplacement of reads containing repeat sequences and minimizes occurrence of singletons. A few unmet constraints are allowed. However, if a sufficient number of constraints are not satisfied by a join AND they all suggest an alternative one supported by the overlap information, the alternative join will be made. For most small- and moderate-sized projects, it is not necessary to use this feature, unless in a situation involving large region of repeats. Default: OFF.

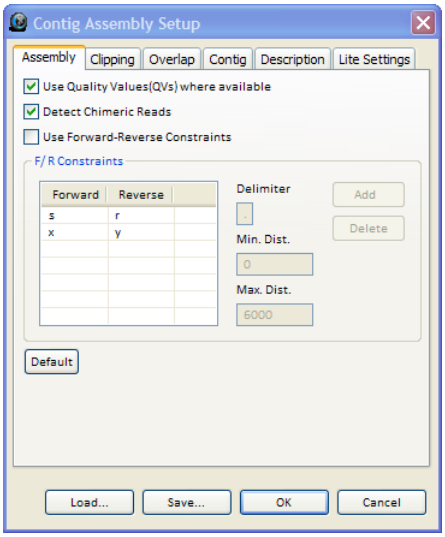
The Assembly Tab contains some detailed settings related to the use of forward-reverse constraints. All these settings are disabled when **Use Forward-Reverse Constraints** is unchecked, and are enabled and editable when checked.

Forward and reverse reads must be named identically up to the delimiter (default = dot), and must contain paired suffixes thereafter (e.g., .s & .r or .x & .y). Suffixes may be added and deleted. Only the letter immediately following the delimiter is recognized as the suffix. Other letters following it, if any, do not contribute to the identity as forward versus reverse end sequence; however, they do distinguish one forward/reverse pair from another, thus making two sequence pairs with the matching reverse/forward end sequence. Minimum and maximum distances may also be edited.

Note: Take care when defining file names at the sequencer, especially when a project is loaded in batches over several days. Entries with misplaced suffixes (e.g. <filename>F.<projectName>, instead of <filename>.F<projectName>) will be overlooked by the constraints feature, although they may be included in an assembly.

- Min. Dist. – Minimum Distance between the forward and reverse reads. Default: 0.
- Max. Dist. – Maximum Distance between the forward and reverse reads. Default: 6000.
- Delimiter – the letter separating main part and suffix part in read names. In cases that there are multiple occurrences of the designated delimiter, the rightmost delimiter is selected.

The values of minimum and maximum distance are uniformly applied to all sequence pairs.



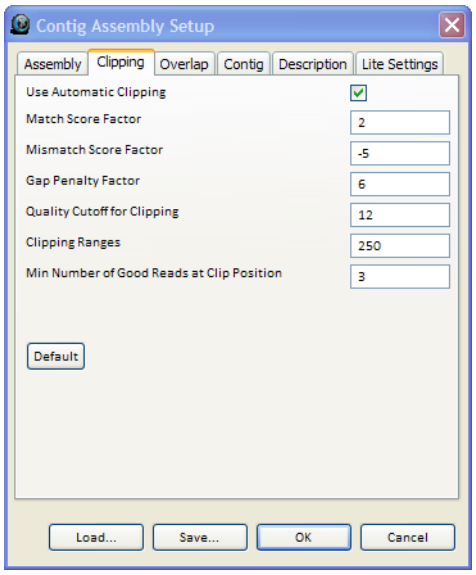
Clipping

Sequence clipping based on quality and similarity is an essential step in the overlap computation and assembly process and ensures validity and correctness. You can control and fine-tune the clipping process through adjusting the parameters on the Clipping tab.

Note: The clipping is irreversible and iterative within each project. When multiple assemblies are made within a project, each round of clipping applies to the sequences that had been clipped in the last round of assembly. This implies that a later attempt to assemble may affect the existing assemblies.

- **Use Automatic Clipping** –Ends of sequencing reads are usually unreliable and low in QV values. CAP3 comes with a mechanism to detect and clip these poor end regions based on sequence similarities with or without QVs. Clipping is done before the computation of overlaps. Note: trimming is available in the ContigExpress® program, in addition to CAP3 clipping, and can be carried out before loading input sequences into assembly projects. Default ON.
- **Match Score Factor** – Also used in calculating Similarity Score for computing overlaps, but only editable in the Clipping Tab. Match Score Factor is a positive integer to award each match between two bases from the pair of sequences being compared during the banded Smith-Waterman alignment. Default: 2.
- **Mismatch Score Factor** – Also used in calculating Similarity Score for computing overlaps, but only editable in the Clipping Tab. Mismatch Score Factor is a negative integer to penalize each mismatch between two bases from the pair of sequences being compared during the banded Smith-Waterman alignment. Default: -5.
- **Gap Penalty Factor** –Also used in the computation of overlaps, but only editable in the Clipping Tab. Gap Penalty Factor is a positive integer to penalize each gap extended during the banded Smith-Waterman alignment. Default: 6.
- **Quality Cutoff for Clipping** – Quality Cutoff for Clipping applies to the clipping of a poor end region for each read when QVs are provided. It is not used when QVs are not available. The specified value is used to find the low-quality ends of reads, where the quality value of a base is considered low if it is less than this value. Default: 12.

- **Clipping Range** – Clipping Range applies to the clipping of a poor end region for each read regardless of whether QVs are available. The value is used to extend the ranges for clipping further away from the ends based on the low-quality positions at each end as determined with the Quality Cutoff for Clipping value. The larger the value of Clipping Range, the more extensive the clipping for poor end regions. Default: 250.
- **Min Num Good Reads at Clip Position** – This is the depth of good coverage at the clip position. It applies to the clipping of a poor end region for each read regardless of whether QVs are available. Depending on the actual depth of coverage, the Minimum Number of Good Reads at Clip Position parameter determines the exact clipping position within the clipping region. The larger this value is, the more extensive the clipping becomes. Default: 3.

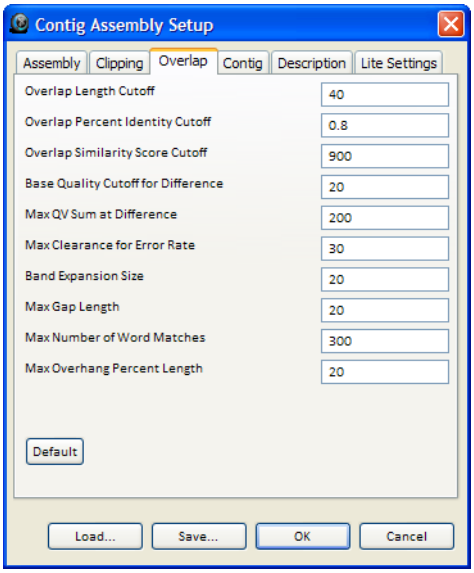


Overlap

Overlaps between reads are computed immediately following sequence clipping. This step serves as the basis for contig construction. In order to ensure the quality, each overlap is evaluated with a few measures. You can change the rigor of overlap computation. The parameters are explained below:

- **Overlap Length Cutoff –** Minimum length required for the length in base pair of all the overlaps. Default: 20.
- **Overlap Percent Identity Cutoff –** Minimum percent identity of the overlap. Default: 0.80.
- **Overlap Similarity Score Cutoff –** Minimum similarity score for an overlap. The score is computed as the sum of match, mismatch or gap scores for each pair of bases weighted by QVs. Default: 900.
- **Base Quality Cutoff for Difference –** Determines the minimum overlap quality by examining the differences of the overlap at bases of high quality values. This is useful only when QVs are available. The Base Quality Cutoff for Difference value defines “high quality” bases. Default: 20.
- **Max QV Sum at Difference –** Applies only when QVs are available. Each overlap is given a “difference score” based on the QVs and the Base Quality Cutoff for Difference. An overlap with a difference score over the Max QV Sum at Difference value is excluded from contig construction. Default: 200.

- **Max Clearance for Error Rate –** Applies even and especially when QVs are not available. If the error rate of the overlap is greater than the sum of those of the overlapping fragments plus this value, the overlap is not used for assembly. The smaller the value, the better the quality control of the overlaps. Default: 30.
- **Band Expansion Size –** This parameter specifies band expansion size. The program automatically determines a minimum band of diagonals for an overlapping alignment between two sequence reads. The band is then expanded, in each direction, by a number of bases specified here. This affects the computation of both potential overlaps and true overlaps. Default: 20.
- **Max Gap Length –** This is the maximum gap length allowed in an overlap. Default: 20.
- **Max Num Word Matches –** This parameter controls the fast method for finding potential overlaps between a pair of sequences. For each word in one read, at most Max Num Word Matches occurrences of the word in the other read are considered for non-gapped extension. A larger value forces the program to consider more word matches at the expense of time and computer memory. Default: 300.
- **Max Overhang Percent Length –** This parameter controls the different overhang regions before or after the aligned region. It is defined as 100 times the total length of the different overhang regions divided by the length of the overlap. Overlaps with a value greater than the maximum cutoff are not used for assembly. Default: 20.

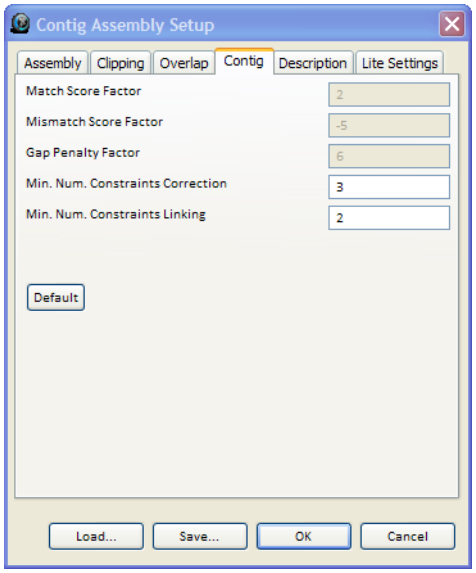


Contig

Contig Tab includes parameters affecting construction of contig and multiple sequence alignments, thus the consensus sequences. QVs are used extensively during this process, if they are available. If unavailable, the program assigns a QV of 10 for each base. Three of the parameters, Match Score Factor, Mismatch Score Factor and Gap Penalty Factor, are editable only on Contig Tab.

- **Match Score Factor** –Match Score Factor is a positive integer to award each match between the existing alignment and the sequence being added when calculating the score of global alignments. Default: 2.
- **Mismatch Score Factor** –Mismatch Score Factor is a negative integer to penalize each mismatch between the existing alignment and the sequence being added when calculating the score of global alignments. Default: -5.

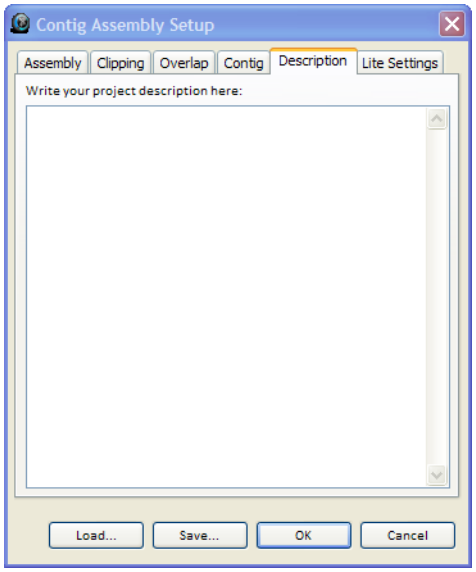
- **Gap Penalty Factor** –Also Gap Penalty Factor is a positive integer to penalize each gap extended when calculating the score of global alignments between the existing alignment and the sequence being added. Default: 6.
- **Min Num Constraints Correction –** This is the minimum difference between the numbers of constraints satisfied in the current assembly and in the alternative assembly. A difference greater than this value, if the contig is also supported by an alternative set of overlaps, results in the alternative join. Default: 3.
- **Min. Num Constraints Linking** – This is the minimum number of constraints for reporting a link between two contigs. Default: 2.



Description

Use this tab to enter the description of the Contig Assembly project you are working on. Click:

- **Load...** to use an existing description from the Assembly Setup Files
- **Save...** to save the new description
- **OK** to implement the edits or **Cancel** to exit the dialog box

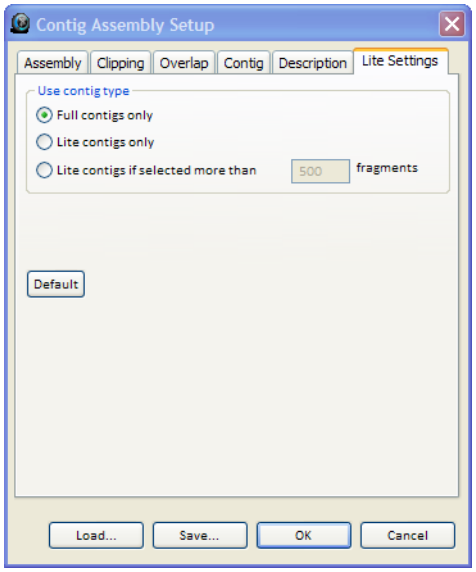


Lite Settings

In ContigExpress® program, you can create two different types of contigs, Full Contigs or Lite contigs. On the **Lite Settings** tab, you can specify the type of contigs you want to assemble in the project.

- In Full contigs mode, chromatogram data are retrievable and-sequence editing performed in the Contig Viewer are reflected in the individual fragment files.
- Editing done on Lite Contigs is NOT reflected in the original fragment sequences (original sequences remain unedited). Assembly in Lite Contig mode reduces memory consumption and is, therefore, the preferred contig type for assembling large projects.

	Lite Settings Parameters
Full Contigs only	Check this box to perform assembly in Full Contig mode. You can retrieve chromatograms and maintain dynamic links between the contigs and their component reads.
Lite Contigs only	Check this box to perform assembly in Lite Contig mode. All linkage between the contigs and sequence reads will be lost and no chromatogram is retrievable in the Contig Viewer.
Lite contigs if selected more than <#> fragments	Creates Lite Contigs during the assembly process only if the selected number of fragments is greater than the number specified; otherwise creates full/ regular contigs.



Assemble Fragments

Use the Assemble Fragments button to assemble the selected fragments into contigs using Cap3. See "Perform an assembly" on page 203 for information on assembling fragments.

Perform an assembly

Before performing an assembly, you can go through the Assembly Settings to set the correct parameters for assembling fragments into contigs.

Assembly setup

To see the assembly parameters available for control:

1. In the Contig Editor, right-click on a Fragment and select **Assemble ▸ Assembly Setup...**.

   Alternatively, select the Fragments main folder in the Project View pane, and from the right-click drop-down menu, select **Assemble ▸ Assembly Setup...**.

   The Assembly Setup dialog box (Figure 6) appears. For more details about Assembly Setup, see "Assembly Settings" on page 196.

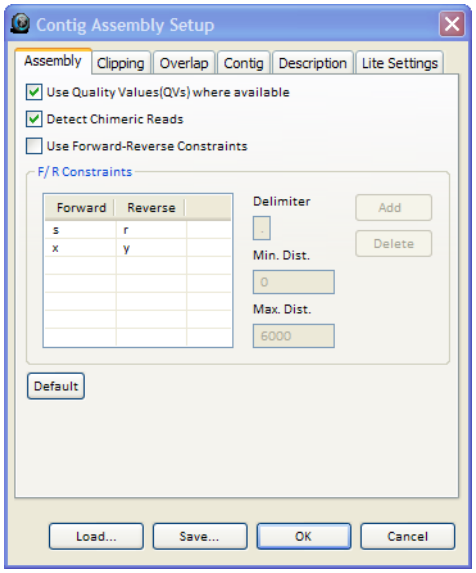


Figure 6 Contig Assembly Setup dialog box

2. To proceed with the assembly using default values, click **Default** in the Contig Assembly Setup dialog box. Click **OK** to implement the changes you made to the settings or **Cancel** to exit the Contig Assembly Setup dialog box.

Perform the assembly

After you have selected the trimming and assembly settings, select the fragments you want to assemble in the Contig Editor, and click **Assemble Fragments**.

Assembly may take some time, depending on how many fragments you selected.

When assembly is complete, the assembled contig(s) and any unassembled fragments will be listed in the Project View pane and can be viewed in the Contig Viewer.

Note: If a contig cannot be created from the fragments, an alert box will be displayed.

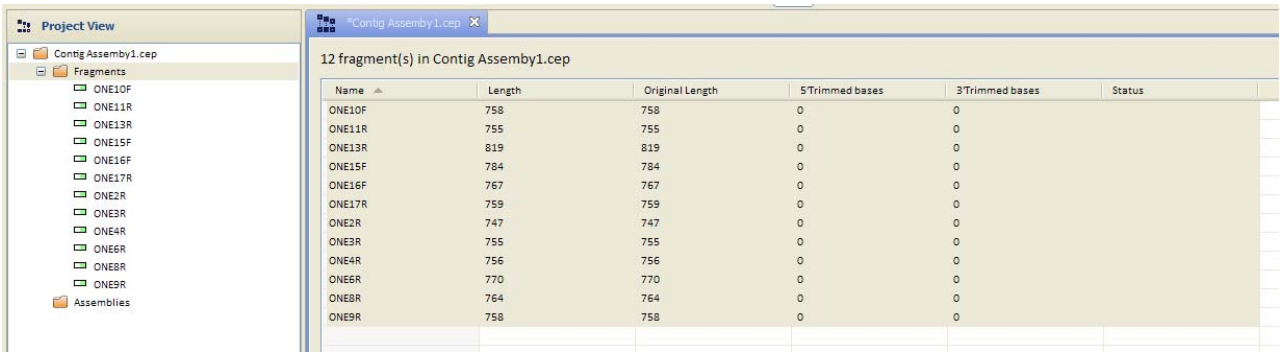
IMPORTANT! Although Vector NTI™ *Express* Software does not have an exact upper size limit for a ContigExpress® program project, your project size may be limited by available computer resources. If you do encounter an "out of memory" situation, you should consider assembly in Lite Contig Mode (see "Lite Settings Tab" on page 204). Often the "out of memory" problem occurs due to the presence of too many assemblies in the ContigExpress® program Project. In

this case, you are advised to delete some of these assemblies as described below, save the project, and restart the ContigExpress® program. Limiting the number of assemblies in a project is always a good idea with large projects. If you decide to use Lite Contig Mode but don't want to lose the link between your contigs and fragments, you can discard chromatogram data (see "Lite Settings Tab" on page 204) and trade the ability to invoke chromatograms for a larger capacity.

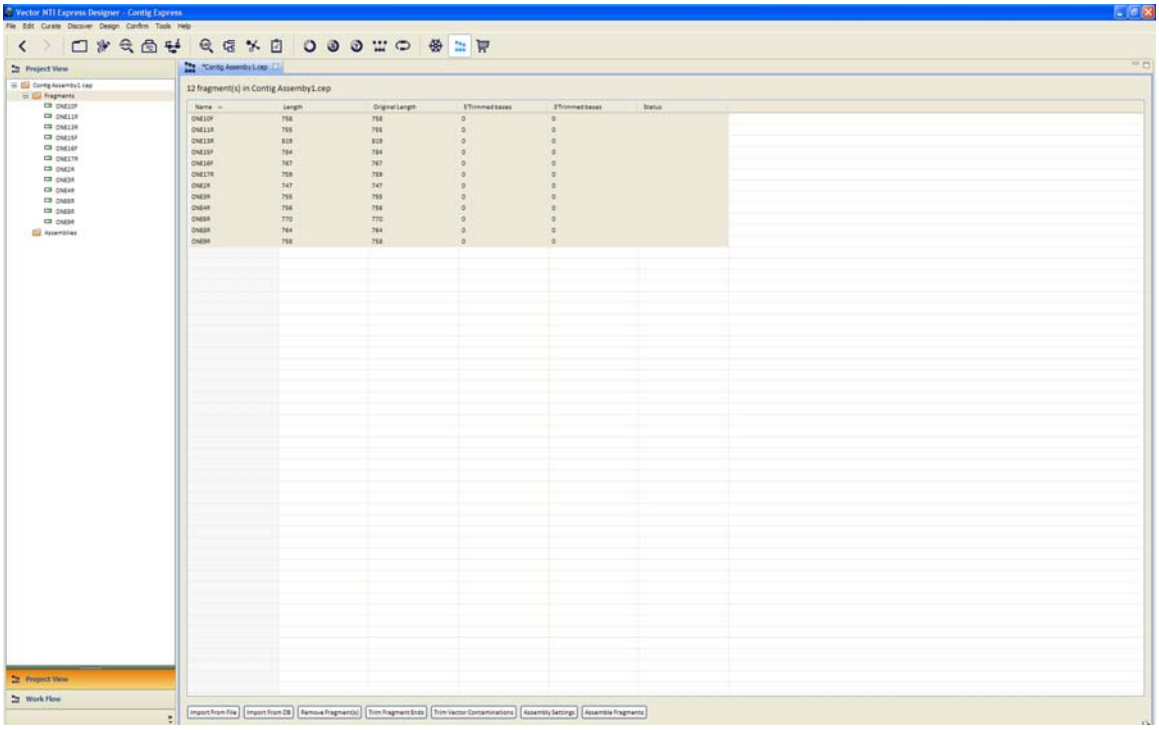
To start an assembly:

1. In the Contig editor pane, select all the fragments from the demo project, by **SHIFT+ CLICK** on all the names on the list.

   Note: You can also use the keyboard shortcut command Ctrl +A to select all the fragments in the Contig Editor pane.



2. Click **Assemble Fragments** at the bottom of the Contig Editor pane.



Note: You can also assemble fragments using the right-click menu of the fragments. Select the fragments by **SHIFT+ CLICK** on all the names on the list, right click on any of the fragment names and select **Assemble Selected Fragments** from the right-click drop-down menu.

After the assembly process is finished, a summary dialog box displays the number of contigs assembled (Figure 7).

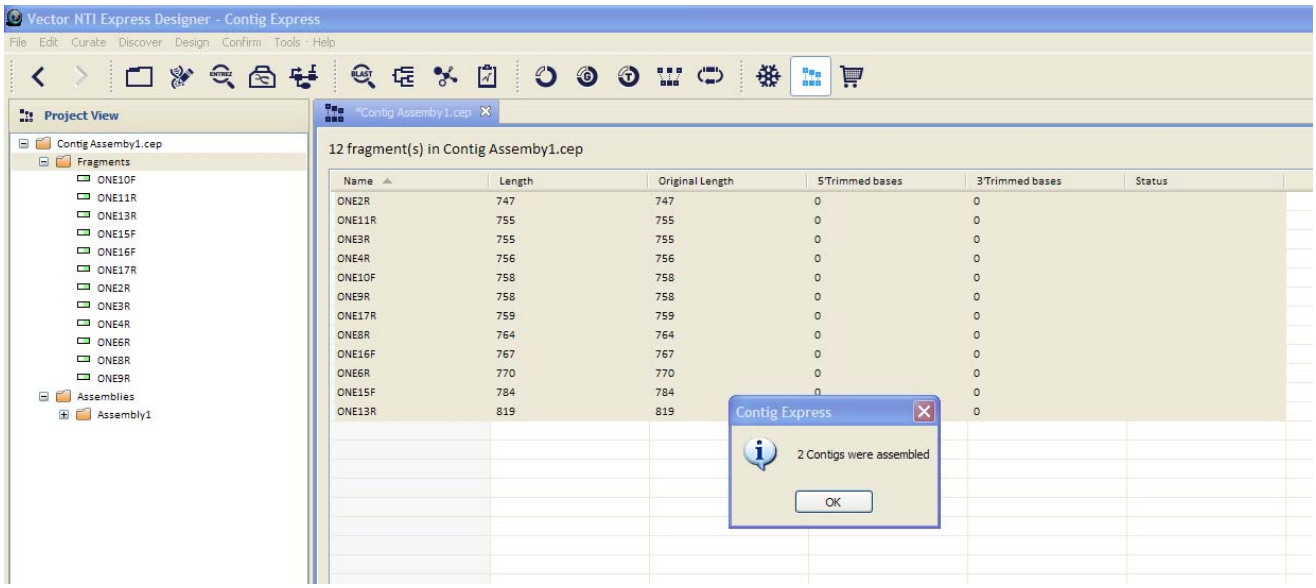


Figure 7 Contig Assembly summary

Examine Assembly Results

Once you have performed the Contig Assembly, the Project View pane in the ContigExpress® program Project Explorer shows the new assembly, Assembly 1 (Figure 8).

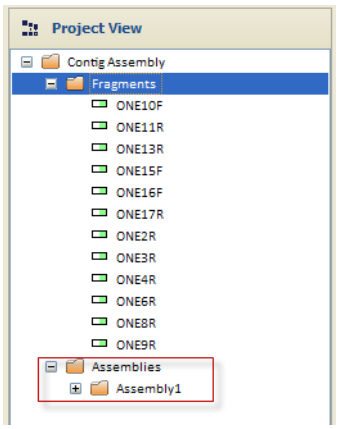


Figure 8 Contig Assembly

An assembly can be made up of one or more contigs. Click **Assembly 1** to select and expand it. All of the project fragments and contigs are listed in the Project View pane. Assembly 1 includes two assembled contigs, with Contig 1 made from eleven fragments and Contig 2 composed of two. Unassembled fragments are shown below the contigs. Note that many of the fragments that are included in contigs underwent trimming at their 5' or 3' ends or both. This results from the operation of the assembly algorithm, although it can be turned off if so desired.

Contig viewer

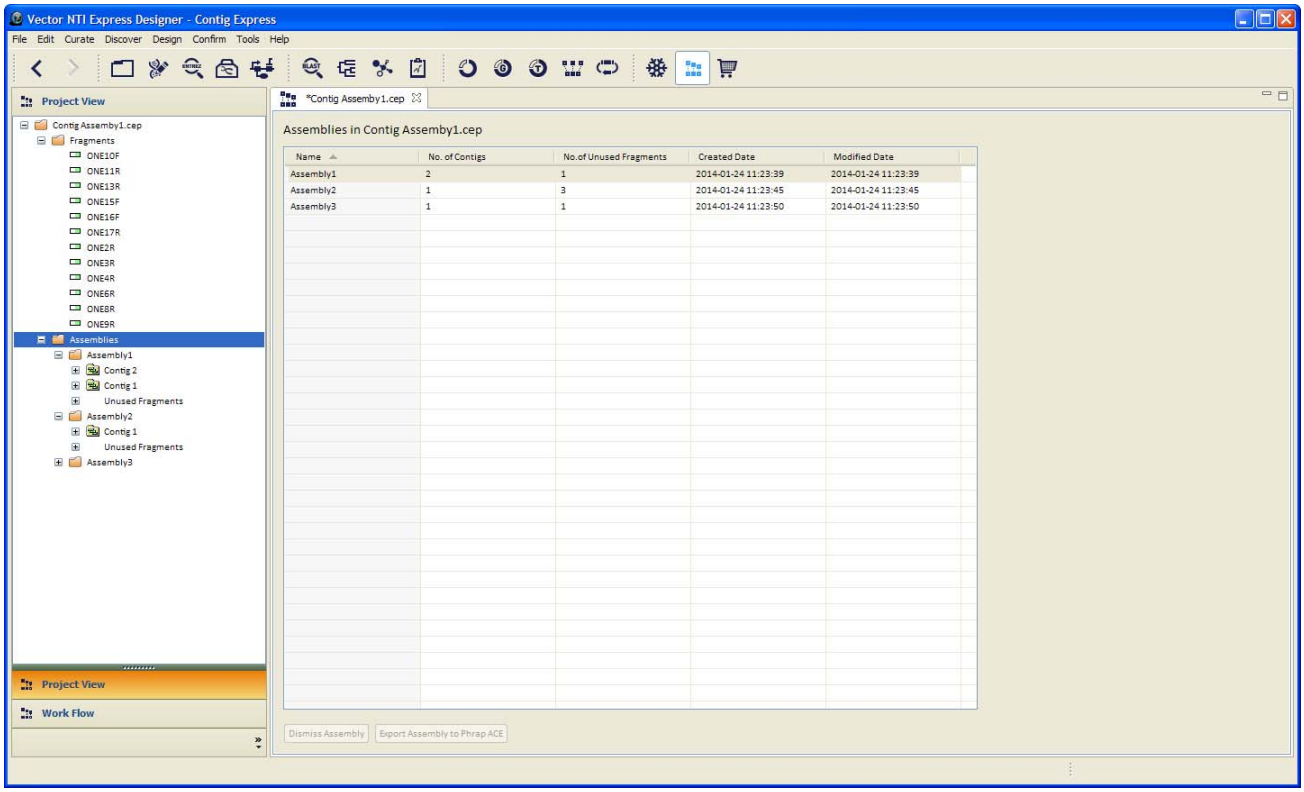
When you select a contig within the Assemblies folder in the Project View pane the **Contig Viewer** is displayed in the Contig Editor.

Viewing Assemblies summary

To view a summary of the total number of assemblies created by the assembling process in a Contig Assembly project, select the Assemblies folder in the Project View pane.

By default, the Project Explorer is configured to display five columns listing the following:

- Assembly Name
- No. of Contigs
- No. of Unused Fragments
- Created Date
- Modified Date



At the bottom of the Assemblies Summary pane, you will find the Dismiss Assemblies button.

Use this feature to remove or delete assemblies from a Contig Assembly project.

Viewing Assembly summary

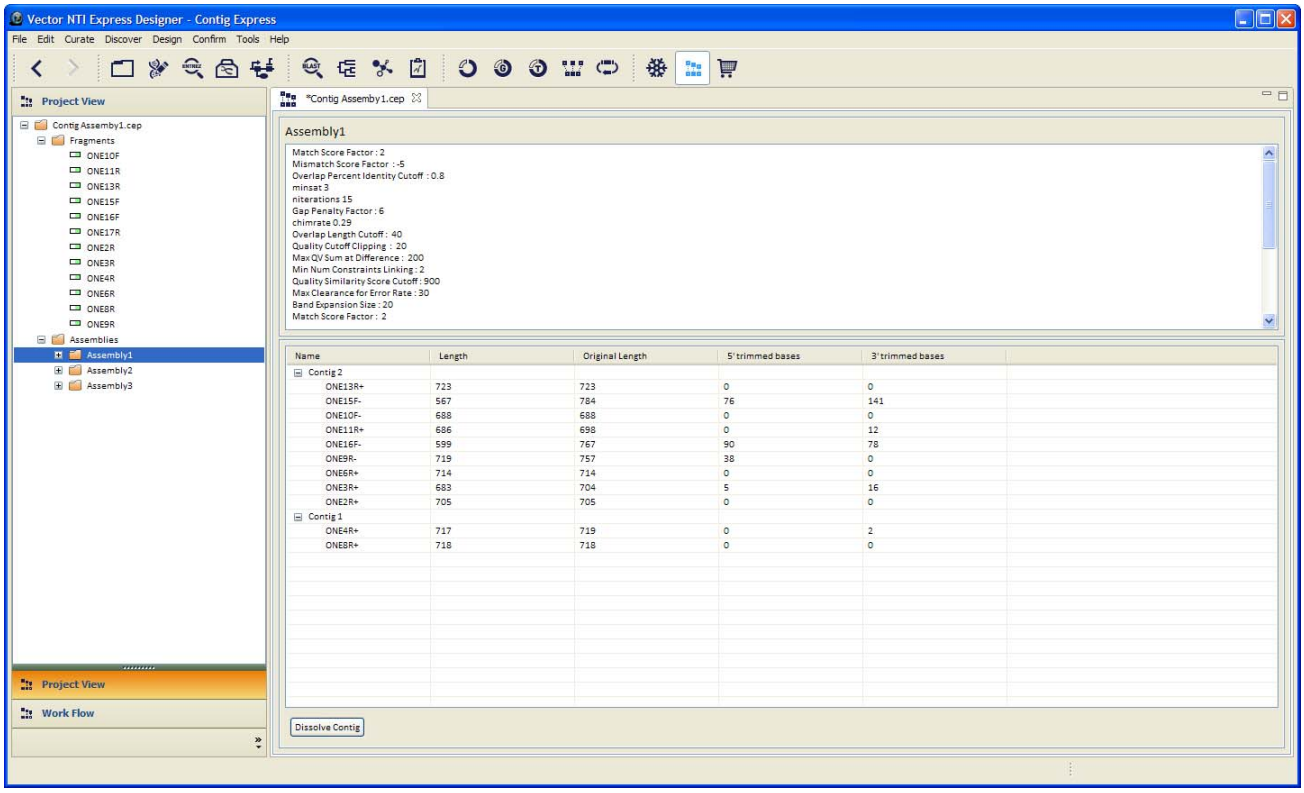
To view the summary an individual assembly created by the assembling process in a Contig Assembly project, select the Assembly1 folder in the Project View pane.

Note: You can view the summary of each assembly created by the assembling process by selecting the name of that assembly in the Project View pane.

On selecting an Assembly in the Project View pane, the Contig Editor displays two panes. The upper pane displays the Assembly Settings values used in performing an assembly while the lower pane displays the summary of the Contigs (and the Fragments) in that Assembly in the Contig Assembly Project.

The lower pane includes the following columns

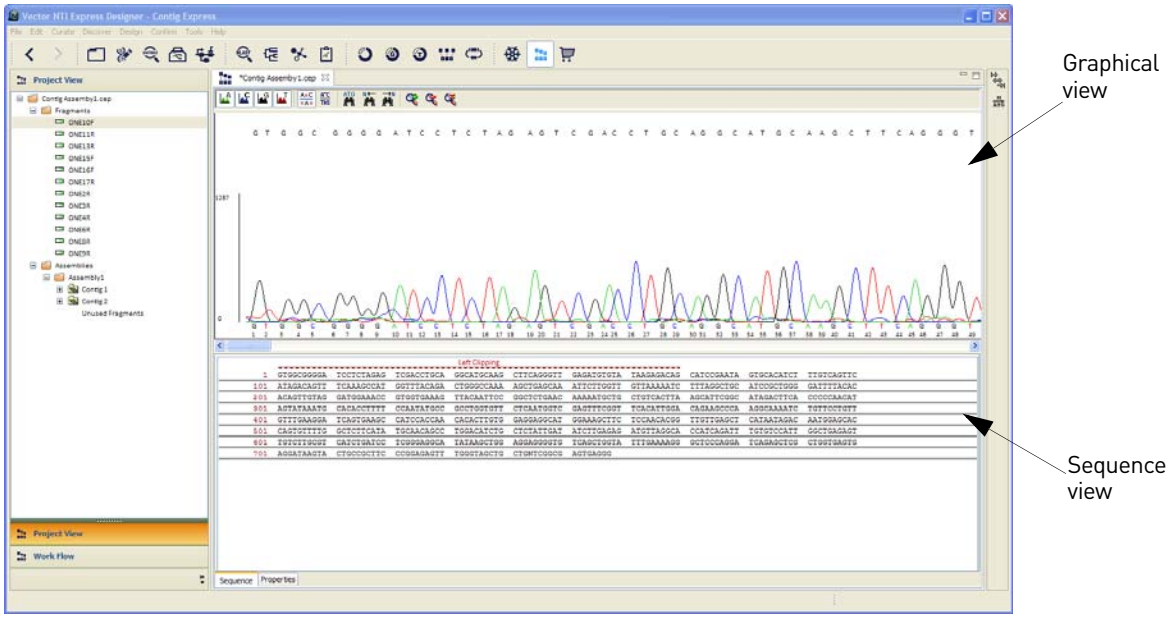
- Name
- Length of the fragment
- Original Length of the fragment
- 5′ trimmed bases
- 3′ trimmed bases



Use the Dissolve Contig feature at the bottom of the Assembly Summary pane to dissolve the contig.

Viewing Contig details

You can view the details of each Contig in an Assembly added to the Contig Assembly Project in the Contig Editor. To view the details, select a Contig in the Project View pane.



The Contig Editor displays two views of the contig selected in the Project View pane.

- **Graphical view** in the upper half of the Contig Editor displays a graphical representation of the Contig Assembly
- **Sequence view** in the lower half of the Contig Editor shows the assembled fragment sequences

The lower half of the Contig Editor has three tabs:

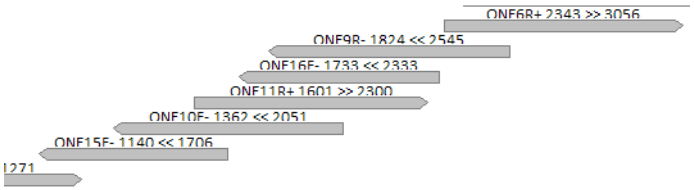
- **Alignment**
- **Sequence**
- **Properties**

Graphical view

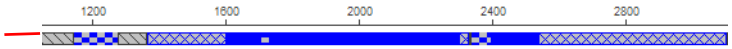
The Contig Viewer Graphical Pane contains horizontal arrows representing the relative positions of the fragments forming the contig. The arrowheads indicate whether the respective fragment is in the direct or complementary strand, with the names of the fragments displayed above the fragment lines.

Contig coverage bar

The Contig Coverage bar spans the length of the contig and contains segments of varying patterns/colors that represent the amount and type of fragment coverage in that segment



Contig Coverage Bar



The patterns and colors are:

- **Single fragment –** gray bar with slants
- **Two fragments in the same direction –** red cross-hatching
- **Two fragments in different directions –** blue checkerboard
- **Multiple fragments in both directions –** solid blue bar

The graphical view pane toolbar includes the following features:

- Three Zoom buttons.
  - **Zoom In**: Click to zoom in to the graphical view pane.
  - **Zoom Out**: Click to zoom out from the graphical view pane.
  - **Fit to Window**: Click to display the default view of the graphical view pane.
- Six toggle buttons to allow clutter-free viewing of the fragment chromatogram.
  - **View Chromatogram**: Click to show/ hide the corresponding Chromatogram in the Alignment pane in the lower half of the Contig Editor.
  - **Consensus Translation 1st frame**: Click to show/ hide the first frame translation consensus in the Alignment pane in the lower half of the Contig Editor.
  - **Consensus Translation 2nd frame**: Click to show/ hide the second frame translation consensus in the Alignment pane in the lower half of the Contig Editor.
  - **Consensus Translation 3rd frame**: Click to show/ hide the third frame translation consensus in the Alignment pane in the lower half of the Contig Editor.
  - **Show/Hide Two Strands**: Click to display only two out of the four nucleic acid base strands at one time.
  - **Show/Hide ORF**: Click to show/hide the Open Reading Frames of the sequence. To execute this feature, you first need to click **ORF Finder** on the right-hand side of the Contig Editor. The ORF sequence is only displayed in the sequence pane.
- Three Find buttons to find a particular strand.
  - **Find**: Click to find a particular fragment. In the Find dialog box, you can insert sequence and click **Find Next** to look for a match. You can find ambiguous character as well as Complementary sequence.
  - **Find Previous Ambiguous**: Click to find a previous ambiguous fragment.
  - **Find Next Ambiguous**: Click to find a next ambiguous fragment.
- Two Move buttons to find a move a particular Fragment.

- **Move Fragment Left**: Click to move a particular fragment to the left of the sequence.
- **Move Fragment Right**: Click to move a particular fragment to the right of the sequence.

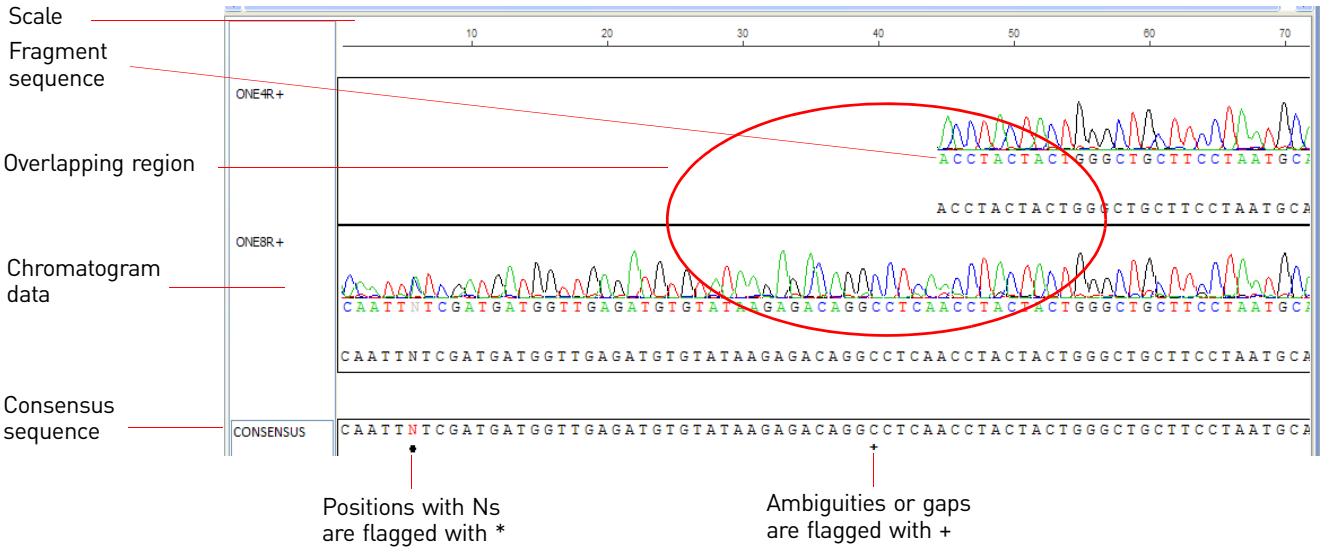
Note: To take a screenshot of the current view of the fragment in the graphical view pane, right-click anywhere in the graphical view pane and select **Camera**. You can then 'paste' the information in MS Word, MS Paint, or any other graphic-editing tool for future reference.

Alignment tab

The Contig Alignment Pane displays the nucleotide sequences of the fragments that form the contig, with overlapping regions aligned appropriately and displayed relative to their positions in the contig.

The consensus sequence is displayed below the fragments, and chromatograms for the fragments are displayed. Translations can be identified and displayed. You can edit the sequences here and see how your actions are reflected in the contig alignment and consensus.

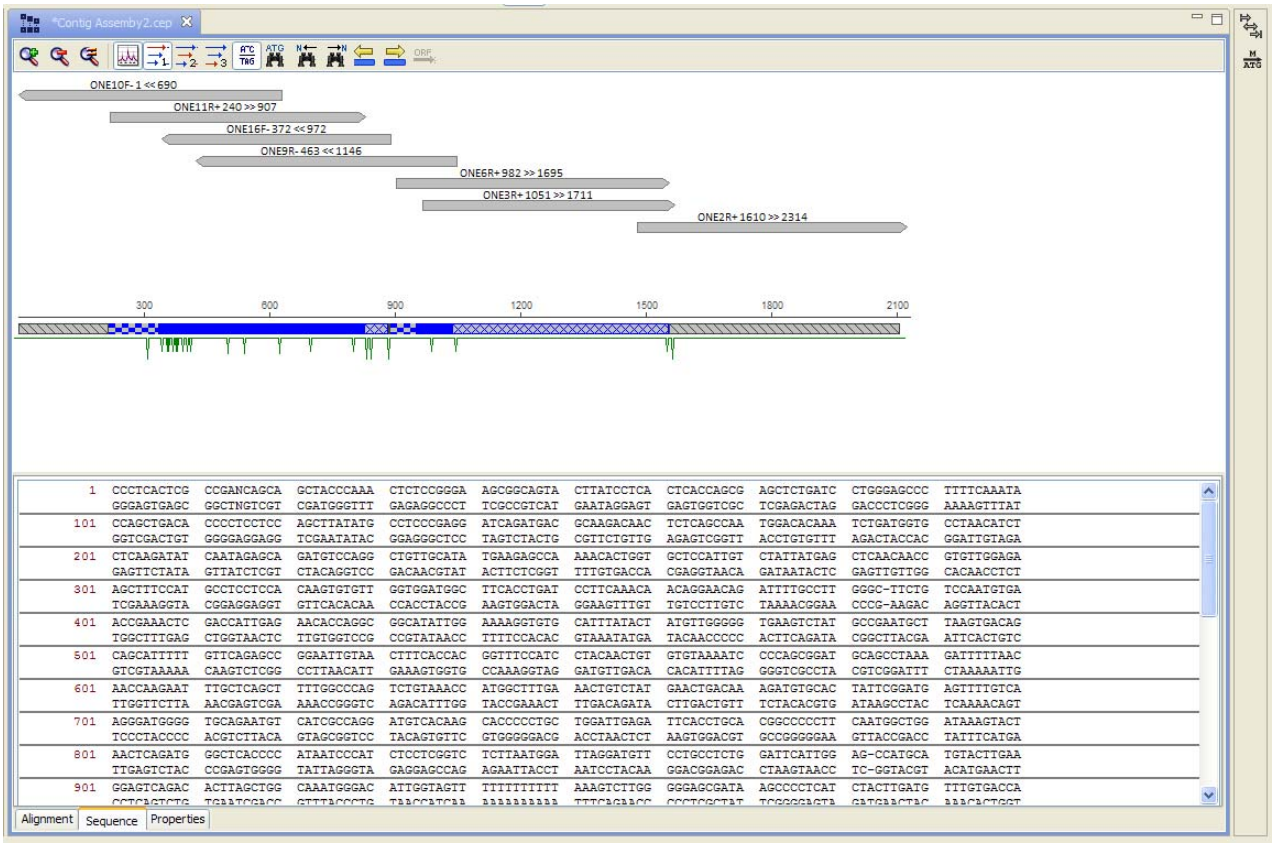
Use the horizontal and vertical scroll bars in the Alignment Pane to view the fragment sequences, overlapping regions, and consensus sequence.



Sequence tab

The sequence view of a contig is displayed in the lower half of the Contig Editor pane.

You can select a particular sequence of the contig in the Sequence view pane and view the corresponding chromatogram in the graphical view pane.



Use the right-click drop-down menu to perform tasks in the Sequence view pane.

1. Select a sequence in the Sequence view pane.
2. Right-click and select from:
   a. **Copy**: To copy the selected sequence

      Note: Keyboard shortcut key to copy the selected sequence is Ctrl +C.

   b. **Paste**: To paste the copied sequence
   c. **Delete**: To delete the selected sequence
   d. **Select All**:To select the entire contig (if you have not selected a part of it)

      Note: Keyboard shortcut key to select the entire sequence is Ctrl +A.

   e. **Clear Analysis Results**: Activated when you perform analysis from the right hand side toolbar pane

Properties tab

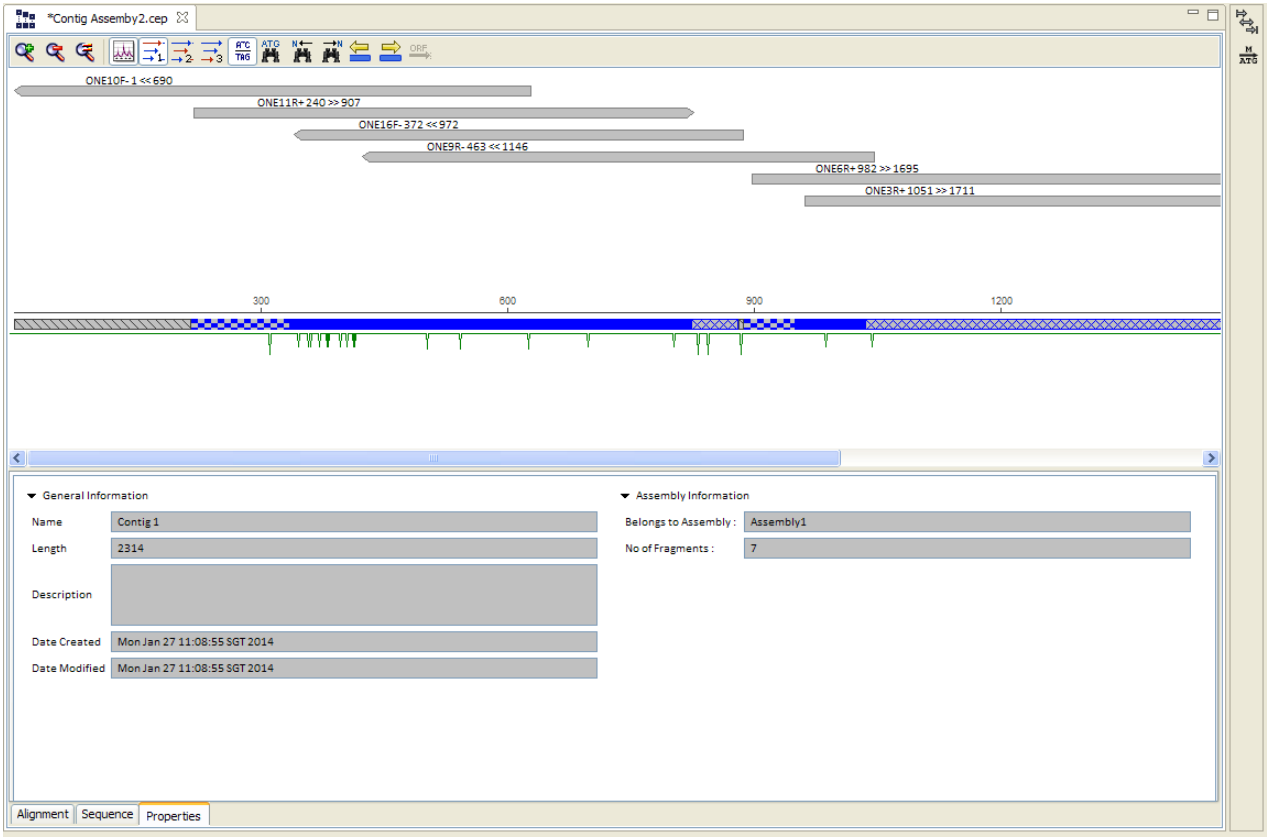
The Properties tab in the lower half of the Contig Editor, displays the general information and assembly information of the contig selected in the Project View pane.

The **General Information** includes:

- Name of the selected fragment
- Length of the fragment
- Description
- Date on which the fragment was created
- Date on which the fragment was modified

The **Assembly Information** includes:

- Belongs to Assembly
- No. of Fragments



Assemble by Reference

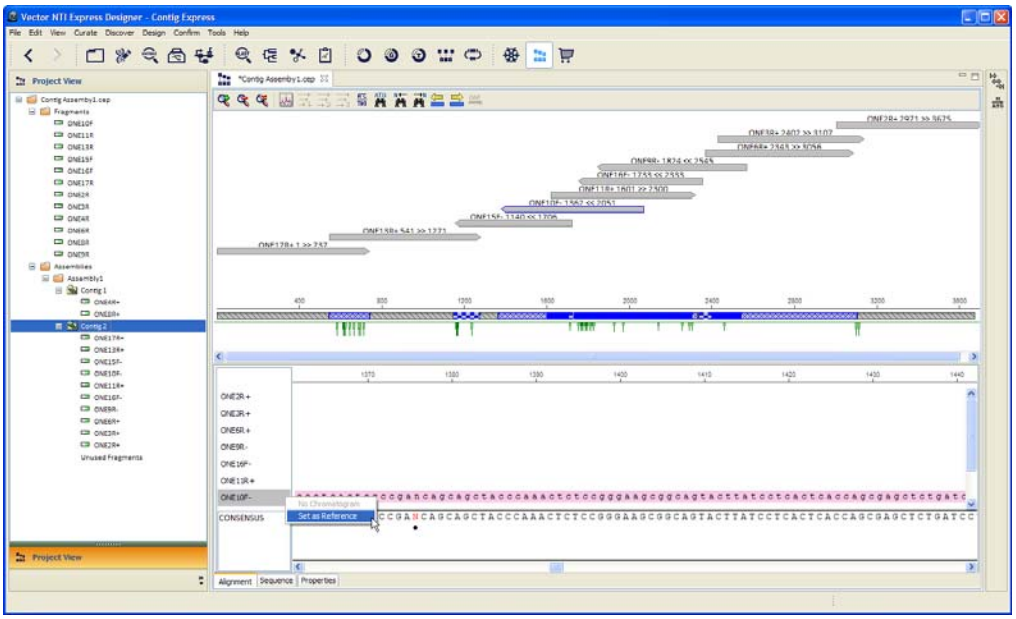
For workflows in applications such as comparative genomics or forensics, it is useful to include a reference sequence in the assembly. Such a sequence may be introduced into the ContigExpress® program in any supported file format and may or may not include a chromatogram.

1. Add the reference sequence to the project in the Project View Pane and include it in the assembly.
2. In the Contig Viewer, select the reference sequence in the Alignment Pane fragment names list, and click **Set as Reference** from the right-click menu (this action is not permitted unless the Viewer is in Edit mode.)

Only one sequence may be designated as reference at any one time in a contig. The selected sequence reappears in lower case letters and pink highlight. Generally, it no longer contributes to the consensus calculation or any analysis derived from it (ambiguity, protein translation, etc.). However if the reference constitutes single coverage anywhere in the contig, it will define the consensus for that domain.

Note: If any fragment in an assembly lacks QV scores (e.g. a .gb file intended as a reference sequence), QVs associated with all other fragments will be ignored during assembly. To avoid such omissions first exclude the reference from assembly, then re-assemble the resulting contig with the reference sequence.

Note: Fragment can be first selected as a reference before assembly. Right click on tree-pane and select assemble by reference



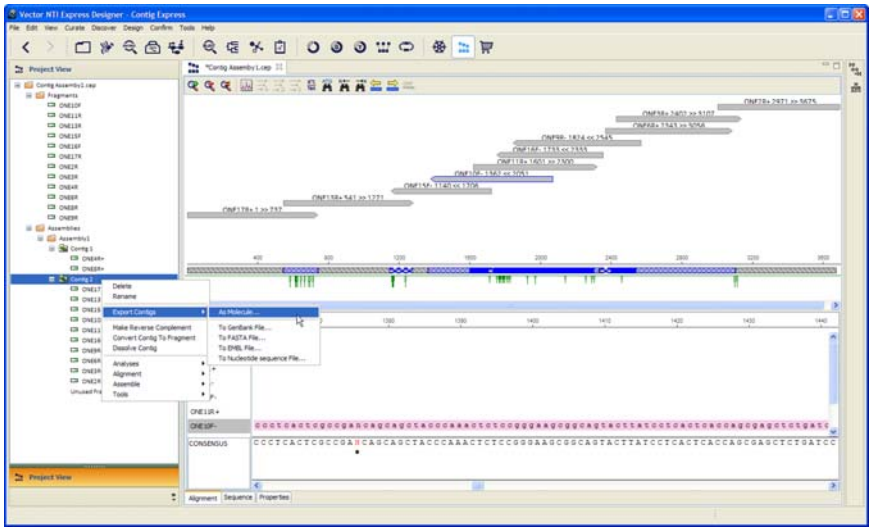
Export Contigs

Use the right-click drop-down menu to export contigs as molecules or to different file types.

To export a contig:

1. Select a contig in the Project View pane.
2. Right-click on the selected contig and select Export Contigs from the drop-down menu. You can choose to export the contig:
   - As a molecule
   - To a GenBank® file...
   - To a FASTA file...
   - To a EMBL file...

- To a Nucleotide sequence file...



Save and close a Contig Assembly Project

1. To save the project, select **File** ▸ **Save As**.
   Name the project and select the Demo Projects folder as the destination.
2. To close your project, click ⊠ on the top right of your project tab.
   The ContigExpress® program reverts to its initial state, with an empty workspace. You can open an existing project from the disk or start a new project:
   - From **File** ▸ **New** ▸ **ContigExpress Project**
   - By clicking the ContigExpress® program icon in the toolbar.

Delete and rename contigs

Most of these commands are located below the Contig/Fragment Tree.

Delete a Contig Assembly

To delete a particular Contig Assembly, select the Assembly name in the Project Explorer, and click **Dismiss Assembly**.

Re-assemble contigs

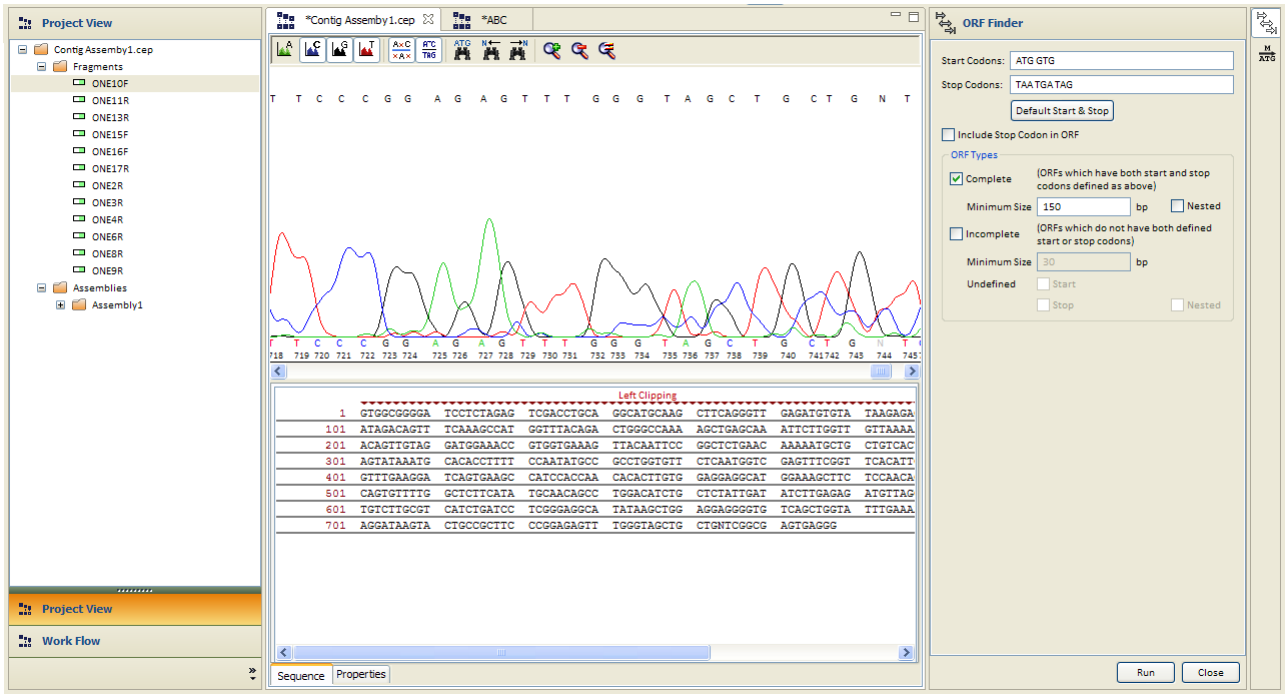
You can select fragments in contigs and re-assemble them. Select the desired fragments from the Fragment List in the Project Explorer, and click on **Assemble Fragments.**

Display ORFs and translate the nucleotide sequence

Display ORFs for the Sequence

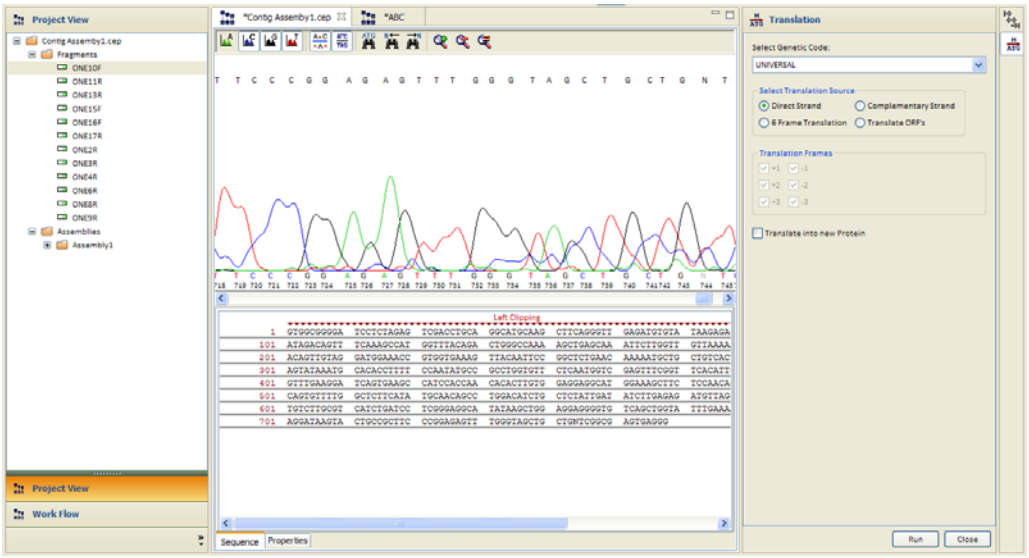
1. In the Contig Editor, click **ORF Finder** to run the ORF finder tool to display ORF Finder pane.

   ORFs are displayed for the direct and complementary strands. ORF finding criteria (for example, minimum length, inclusion/exclusion of stop codon, and so on) can be adjusted via the ORF finder dialog box from the ORF finder tool toolbar button.



Translate the Nucleotide Sequence

1. Click **Translation** on the Fragment viewer toolbar.
   The ContigExpress® program shows the amino acid abbreviations above the selected sequence.



2. To translate the complementary strand of the selection, choose **Complementary Strand** from the dialog box and click **Run** (add graphic) on the menu or toolbar.

Edit the nucleotide sequence in the Sequence Pane

You may freely edit a fragment in the Fragment Viewer only if it is not part of a contig. You may also edit a fragment that is part of a contig but they are handled separately from the fragments in the Fragment Viewer.

There are three ways to edit the fragment sequence. All are illustrated in Figure XXX.

ADD SCREENSHOT.

Figure 16.1 Editing a sequence in the Fragment Viewer

1. **Nucleotide Deletion**: In the Sequence Pane,
   a. Select nucleotides 316 through 319.
   b. Press the Delete key on your keyboard.
      The selected symbols move under the row, with downward pointing arrows (↓) taking their place.
   c. Move the cursor to the Chromatogram Pane.
      There is a sequence above and one below the graph. Before editing, they were identical. The sequence above the graph reflects all editing commands, the sequence below always remains in its original form, thus giving a useful reference to track and compare changes.
      Note: The deletion results in each deleted nucleotide being replaced by a dash (-) in this sequence.
2. **Nucleotide Insertion**: In the Sequence Pane,

a. Move the caret in the Sequence Pane to position 307, and type ***TTT***.

Three Ts with upward pointing arrows (↑) below them are inserted at the caret position.

The arrows indicate that symbols were inserted in a sequence ("added to" the sequence) into a position, not substituted for other nucleotides. The sequence above the graph in the Chromatogram Pane also shows the insertion (in another color).

Note: A gap appears in the graph because there is no data to build chromatograms for newly inserted symbols.

3. **Nucleotide Substitution:** Select nucleotide 303 and type *G*.

The selected base is replaced by G, and A is shown below G to help track the change. The *chromatogram* does not change because the G was not inserted; it just replaces the original base call. Replace the bases one by one.

Edit a sequence in the Chromatogram Pane

1. Activate the Chromatogram Pane by clicking **on the contig** in the Project Viewer.
2. In the top, editable sequence in the Chromatogram Pane, select **"T"** at bp 11 and type ***C***. Like the Sequence Pane, T is replaced by C.

Note: All editing operations are reflected in both the Sequence and Chromatogram Panes.

The editing operations can be summarized as follows (NTs = nucleotides):

Action	How to Perform	Sequence Pane Result	Chromatogram Pane Result
Delete	Select residues; press Delete	Replaces NT with ('---' in red color); original NT moves below strand	(- - - -) Appear in upper, editable sequence
Insert/ Replace	Place caret; type new NTs	For Insert, (NT with green color) appears below new NTs; new NTs are colored For Replace, (NT with blue color) appears below new NTs; new NTs are colored	A break appears in the chromatogram
Replace	Select existing NTs and type new NTs one by one	New NT appear in sequence; replaced NTs moved below strand	New NT appears in upper sequence; no break or change in chromatogram

Table 1 Editing operations for Sequence and Chromatogram Panes

A Symbols and Formats: IUB (IUPAC) Ambiguity Codes and ASCII Format

Format for ASCII Sequence Files

An ASCII sequence file must obey the following rules:

- It must be a plain (ASCII) text file.
- The file must contain the nucleotide (amino acid) sequence arranged in lines. Each line may contain the following:
- Nucleotide (amino acid) symbols and white space, or
- A number followed by white space and nucleotide (amino acid) symbols (therefore, similar to GenBank format), in which case the number will be ignored, or
- A number only, in which case the number will be interpreted as a block of unknown nucleotides (amino acids) of the corresponding length.

IUB Formats recognized by Vector NTI™ *Express*

The following characters, defined by the International Union of Biochemistry (IUB), are used to represent nucleotides throughout Vector NTI™ *Express*:

Symbol	Meaning
A	adenine
T	thymine
C	cytosine
G	guanine
R	purine(A or G)
Y	pyrimidine(C or T)
W	A or T
S	C or G
M	A or C
K	T or G
B	C, G, or T
D	T, G, or A
H	C, A, or T
V	C, G, or A
N	unknown nucleotide

Amino acids are represented by standard 1- (or 3-) letter codes:

1-Letter Symbol	3-Letter Symbol	Meaning
E	Glu	Glutamic acid
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Asparagine or Aspartic acid

B Primer Tm Calculations

General Information

Vector NTI™ *Express* calculates and reports two different melting temperatures for DNA/RNA oligonucleotides, Thermodynamic Tm (Therm. Tm) and %GC Tm .

Usefulness of Thermodynamic Tm Versus %GC Tm

Vector NTI™ *Express* reports both the Thermodynamic and %GC Tm values, regardless of the length of the oligo. However, generally only one of the reported ™ values should be considered useful, depending on the length of the oligo as follows:

- Therm. Tm – useful for oligos that are greater than about 7-10 residues and less than about 35 residues long
- %GC Tm – useful for oligos greater than about 35 residues long

Note: For oligos that are 7-10 residues or shorter (cutoff length depends on the base content of the particular oligo being analyzed), Vector NTI™ *Express* reports a Therm. Tm value of zero.

Effects of Primer (Probe) and Salt Concentration on Tm Calculations

Tm calculations are highly dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Therefore, it is important that you adjust the primer and salt concentrations appropriately so that accurate Tm values are generated.

Note: In Vector NTI™ *Express*, the default parameters for primer and salt concentration are 250 pM and 50 mM, respectively, for calculating Tm values. Other Tm calculators commonly use a default probe concentration of 50 nM. Because of this, Vector NTI™ *Express* default parameter Tm values may not correspond to the default Tm values calculated using other programs. Before comparing Vector NTI™ *Express* Tm values with those generated by other Tm calculators, make sure that the parameters are adjusted appropriately.

%GC Tm Calculation

The %GC Tm calculation1 does not rely on the thermodynamic properties of the oligo (i.e. dHo, dSo and dG° values). The formula for %GC Tm is as follows:

Tm = 81.5 + 16.6(log[Na+]) + 0.41(%GC) – 675/probe length

Note: [Na+] is in molar units.

Example: For oligo GTGCGAGGCAGCTGCGGTAA at 50mM salt:

Tm = 81.5 + 16.6(log(0.05)) + 0.41(65) – 675/20

= 81.5 + 16.6(–1.30) + 26.65 – 33.75

= 81.5 – 21.58 + 26.65 – 33.75

= 52.82 °C

Thermodynamic Tm Calculation

The Thermodynamic Tm calculation is based on the Nearest Neighbor theory of DNA/RNA duplex stability. Briefly, this theory states that the overall duplex stability (and, hence, the melting temperature) of an oligonucleotide can be predicted from the primary sequence based on the relative stability and temperature-dependent behavior of every dinucleotide pair in the oligo2. In practice, enthalpy (dH°) and free energy (dG°) values for each of the 10 possible Watson-Crick DNA pairwise interactions are used to calculate pairwise entropy (dS°) values via the following standard equation:

$dG° = dH° - TdS°$

Note: T is temperature in °K.

The pairwise dH° and dS° values are then summed to calculate overall values for the oligo under consideration. The overall values are used in the following formula3 to calculate the Thermodynamic Tm:

$\text{Therm. Tm} = dH°- 273.15 + 16.6(\log[Na+]) dS° + dS_o° + R(\ln(c/4))$

Notes:

- $dS_o°$ is the entropy associated with helix initiation (-10.8 cal/mol per °K).
- R is the Universal Gas Constant (1.987 cal/mol per °K).
- c is the concentration of the probe, in molar units.
- The factor –273.15 corrects for absolute temperature so that the final Tm is in °C.

The pairwise dH° and dS° values for DNA used in Vector NTI™ *Express* are taken from reference 2. Those values, along with the corresponding dG° values at 25°C, appear in the following table:

Interaction	dH° kcal/mol	dS° cal/mol per °K	dG° kcal/mol
AA/TT	-9.1	-24.0	-1.9
AT/TA	-8.6	-23.9	-1.5
TA/AT	-6.0	-16.9	-1.0
CA/GT	-5.8	-12.9	-2.0
GT/CA	-6.5	-17.3	-1.3
CT/GA	-7.8	-20.8	-1.6
GA/CT	-5.6	-13.5	-1.6

Table 2 DNA Nearest Neighbor thermodynamics

CG/GC	-11.9	-27.8	-3.6
GC/CG	-11.1	-26.7	-3.1
GG/CC	-11.0	-26.6	-3.1
XX/XX	-6.0	-16.9	-1.0

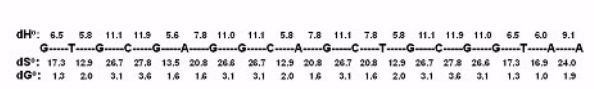
Table 2 DNA Nearest Neighbor thermodynamics *(continued)*

Notes:

- All values refer to the disruption of a duplex at 1 M NaCl, 25°C and pH 7.
- The units for dH° and dG° are kcal/mol of interaction, whereas those for dS° are cal/°K per mol of interaction.

Example

The oligo 5′-GTGCGAGGCAGCTGCGGTAA-3′ is parsed as follows:



- Total dH° = -164.7 kcal/mol
- The total dS° reported by Vector NTI™ *Express* is the sum of the pairwise values above and the entropy associated with helix initiation (dSo°). Thus, for the example oligo above:
- Total dS° = -405.7 + (-10.8) = -416.5 cal/mol per °K

The total dG° is the sum of the pairwise dG° values for the oligo plus a helix initiation free energy term (dGo°) that is added to better reflect experimentally determined free energy values for tested oligos. The value of the helix initiation free energy term (dGo°) depends on the base composition of the oligo2 as follows:

- +5.0 kcal/mol for oligos containing any G-C base pairs
- +6.0 kcal/mol for oligos composed exclusively of A-T base pairs

Therefore, for the example oligo:

Total dGo = -43.7 kcal/mol (sum of the pairwise dGo values) + 5 kcal/mol (free energy term) = -38.7 kcal/mol

The 3′ End dG° is calculated using the number of 3′ pairwise dG° values specified in the 3′ End Length (bp) box, and is not further adjusted.

Using Vector NTI™ *Express*'s default probe and salt concentrations (250 pM and 50 mM, respectively) and the values for dH° and dS° calculated above, Therm. Tm can be calculated as follows:

$$ThermTm = \frac{dH^\circ}{dS^\circ + dS_o^\circ + R\left(\ln\left(\frac{c}{4}\right)\right)} - 273.15 + 16.6(Log(Na))$$

$$\ldots = \frac{-164.7}{-0.4165 + (0.001987)\left(\ln\left(\left(\frac{250 \times 10^{-12}}{4}\right)\right)\right)} - 273.15 + 16.6(Log(0.05))$$

$$\ldots = \frac{-164.7}{-0.4165 + (0.001987)(-23.49)} - 273.15 + 16.6(-1.301)$$

$$\ldots = \frac{-164.7}{(-0.4165) - 0.04667} + (-273.15) - 21.60$$

$$\ldots = \frac{-164.7}{-4632} - 294.75$$

$$\ldots = 355.57 - 294.75$$

$$\ldots = 60.82°C$$

Vector NTI™ *Express* adjusts the %GC and Therm. Tm values accordingly, based on the input formamide concentration.

Oligos Containing IUB Ambiguity Characters

Vector NTI™ *Express* can analyze oligos that contain IUB nucleotide ambiguity characters (i.e. R, Y, W, S, M, K, B, D, H, V and N). In the case of ambiguity characters, Vector NTI™ *Express* uses average pairwise dH° and dS° values for calculating the Tm.

For example, for the dinucleotide pair CB, Vector NTI™ *Express* averages the CC, CG and CT thermodynamic parameters (Table <Blue><Emphasis> 2) to obtain average pairwise dH° and dS° values for CB. It then sums the average pairwise thermodynamic parameters and calculates the Therm. Tm values according to the equation described above (refer *Thermodynamic Tm Calculation* on page 222).

In the case of %GC Tm, Vector NTI™ *Express* applies the appropriate %GC contribution represented by each ambiguity symbol to the standard %GC Tm formula (see "%GC Tm Calculation" on page 221). For example, a B ambiguity symbol contributes only two-thirds the amount of a G or C residue to overall GC content.

RNA Oligos

RNA oligos use a different set of pairwise thermodynamic values than DNA oligos. Pairwise thermodynamic values for RNA are summarized in the following table:

Interaction	dH° kcal/mol	dS° cal/mol/K	dG° kcal/mol
AA/UU	-6.6	-18.4	-1.1

Table 3 RNA Nearest Neighbor thermodynamics

AU/UA	-5.7	-15.5	-1.1
UA/AU	-8.1	-22.6	-1.4
CA/GU	-10.5	-27.8	-2.2
GU/CA	-10.2	-26.2	-2.4
CU/GA	-7.6	-19.2	-1.9
GA/CU	-13.3	-35.5	-2.7
CG/GC	-8.0	-19.4	-2.2
GC/CG	-14.2	-34.9	-3.8
GG/CC	-12.2	-29.7	-3.3
XX/XX	-6.0	-16.9	-1.0

Table 3 RNA Nearest Neighbor thermodynamics *(continued)*

Notes:

- All values refer to the disruption of a duplex at 1 M NaCl, 25°C, and pH 7.
- The units for dH° and dG° are kcal/mol of interaction, whereas those for dS° are cal/°K per mol of interaction.
- The dS° value for RNA oligos is adjusted by –10.8 cal/°K per mol to reflect the entropy associated with helix initiation, as it is for DNA oligos.
- The dG° value is adjusted by +3.4 kcal/mol to account for helix initiation. Note that this adjustment is NOT dependent on the base composition of the RNA oligo as it is for DNA oligos (refer *Thermodynamic Tm Calculation* on page 222).

Primer/Probe Tm, Ta_{Opt} and Similarity Calculations

For oligos designed using Vector NTI™ *Express*'s PCR Primers, Sequencing Primers and Hybridization Probes features, the oligo Tm, product Ta_{Opt} and oligo percent binding similarity are reported in the Text Pane of the Molecule Viewing window.

Primer/Probe Tm Values

Tms for designed primers/probes are reported is as follows in Vector NTI™ *Express*:

- Therm. Tm is reported if the oligo is less than or equal to 35 residues
- %GC Tm is reported if the oligo is 36 residues or greater

For PCR products, Vector NTI™ *Express* reports the %GC; the assumption being that the majority of PCR products are larger than 35 residues.

Ta_{Opt} Values

The PCR product Ta_{Opt} (optimal annealing temperature for amplification of the fragment) in °C is calculated using the following formula3:

$$Ta_{Opt} = ((0.3)Tm_{primer} + (0.7)Tm_{product} - 14.9)\ °C$$

Notes:

- Tm_{primer} is the Tm of the less stable primer of the pair

- $Tm_{product}$ is the Tm of the PCR product

Primer/Probe Similarity Values

When designing PCR, sequencing or hybridization primers, Vector NTI™ *Express* reports the overall similarity in percent of an oligo to its binding site based on the oligo's nucleotide composition. For oligos containing IUB ambiguity symbols, three similarity values are reported:

- Minimum Similarity
- Maximum Similarity
- Average Similarity

For Minimum Similarity, all ambiguities are classed as complete mismatches (i.e., they are assigned values of 0 at each position). For example, a 20mer containing 2 Rs and 2 Ns has a Minimum Similarity of 80%.

For Maximum similarity, all ambiguities are considered identical to their cognate nucleotides (i.e. they are assigned values of 1 at each position), so the Maximum Similarity is always 100%.

For Average similarity, Vector NTI™ *Express* weights each ambiguous nucleotide depending on whether it represents 2, 3, or 4 possible nucleotides. For example, Ns have a score of 0.25, Rs of 0.5, Bs of 0.33, etc. Therefore, for the 20mer described above, the Average Similarity is 85%.

References

Baldino Jr., F., Chesselet, M.F., and Lewis M.E. (1989) High-resolution *in situ* hybridization histochemistry, *Methods Enzymol. 168:*761-777.

Breslauer, K.,J., Frank, R., Blocker, H., and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence, *Proc. Natl. Acad. Sci. USA 83:*3746-3750.

Rychlik, W., Spencer, W.J., and Rhoads, R.E. (1990) Optimization of the annealing temperature for DNA amplification in vitro, *Nucleic Acids Res. 18:*6409-6412.

Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996) Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes, *Nucleic Acids Res. 24:*4501-4505.

Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Nielson, T., and Turner, D.H. (1986) Improved free-energy parameters for predictions of RNA duplex stability, *Proc. Natl. Acad. Sci. USA 83:*9373-9377.

Index

D

E

F

G

H

I

J

L

M

N

O

P

R

S

T

U