



Sequence Analysis using VectorNTI

An introduction to VNTI Advance v10 on the PC

Version 1.2 (public)



Licence

This manual is © 2007-8, Simon Andrews.

This manual is distributed under the creative commons Attribution-Non-Commercial-Share Alike 2.0 licence. This means that you are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution. You must give the original author credit.
- Non-Commercial. You may not use this work for commercial purposes.
- Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

Please note that:

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Full details of this licence can be found at

<http://creativecommons.org/licenses/by-nc-sa/2.0/uk/legalcode>



Introduction

VectorNTI is a mature package which was originally developed by a company called InforMax, but was more recently purchased by Invitrogen.

This manual does not cover every function within VectorNTI but rather provides an overview of the most commonly used parts to get you started.

One module of VectorNTI is completely omitted from the manual. This is the ContigExpress assembly module. The recommended package for sequence assembly is Staden which is a free package which is also available on site (and for which a separate training course exists).



Managing Molecules with VectorNTI Explorer

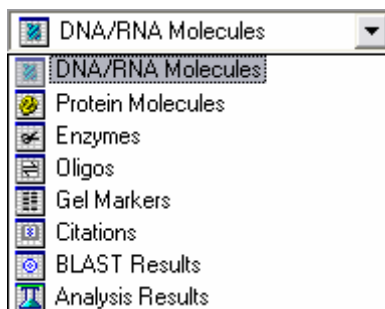
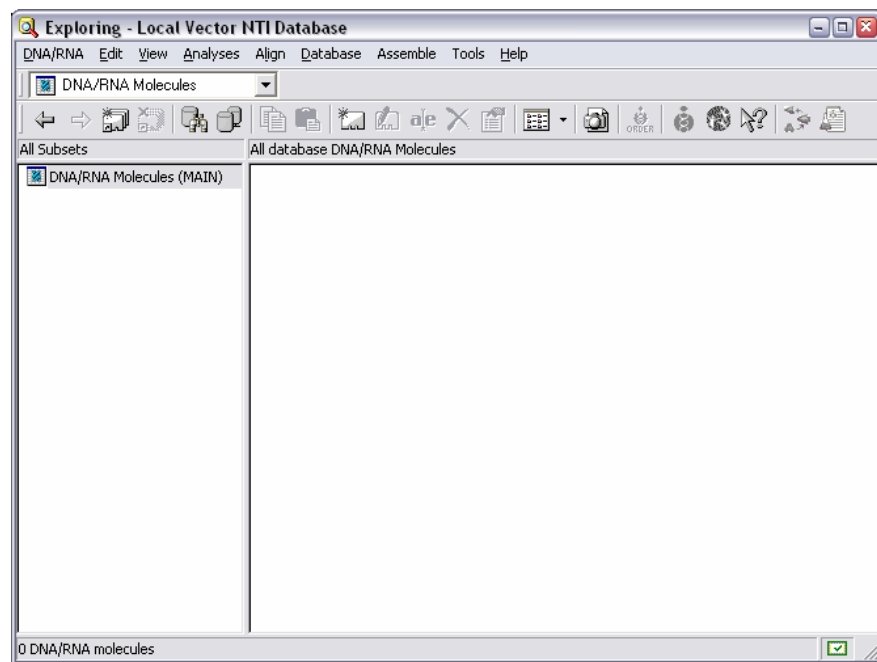
VectorNTI explorer is a database application which you can use to store, organise and query the set of sequences which are of use to you. The Explorer can then be used to launch the other visualisation and analysis tools within the VectorNTI suite.

You start VNTI Explorer by selecting:

Start → Programs → Invitrogen → Vector NTI Advance 9 → Vector NTI Explorer

From the main Start menu.

The explorer window will then open up, and should look something like this:



The database is divided into 8 sections which represent the different types of biological objects it is able to store. You can switch between the views of the different sections using the drop-down selector just below the main menu bar.



Bringing Molecules into the Database

One of the first things you'll want to do when you start VNTI is to bring in one or molecules to work on. The best way to do this is to initially import them into the explorer and then work on them from there. You can import molecules either from a local text file or from an online database. You can also create your own molecules from scratch.

Importing from a text file

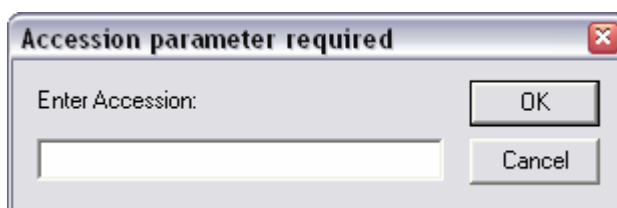
VectorNTI understands the GenBank, EMBL and FastA file formats for nucleotides and GenPept, SwissProt and FastA formats for peptides, and allows you to create a new molecule in the database by importing from any of these. The quickest way to import a sequence from a text file is simply to drag it onto the VNTI Explorer window. If the file type is recognised you will be asked to confirm the name of the sequence and it will be imported. Remember that DNA and protein sequences are stored separately in the database, so if you don't see your sequence straight away you could be looking at the wrong display.

Importing from a database

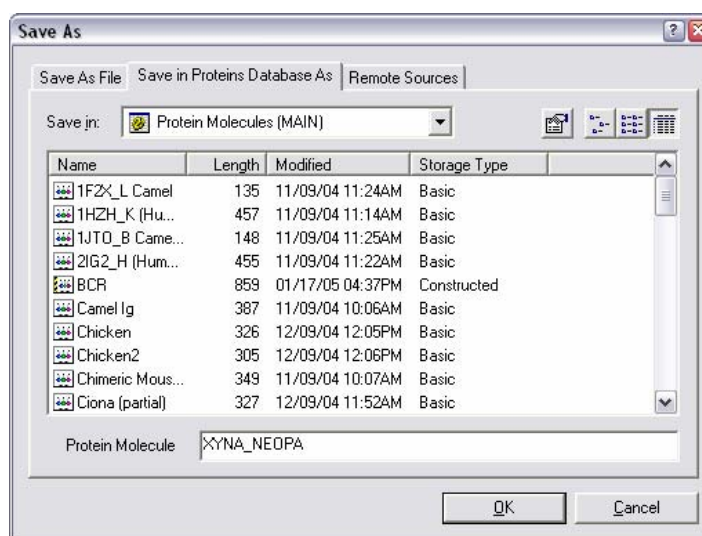
If you know the accession number for the sequence you want you can quickly import it from a local copy of the main sequence databases we hold on site. To do this select:

Tools → Open → Fetch Seq by Accession...

A dialog box should appear, into which you put your accession number:



The sequence should then be imported VectorNTI will open a new display window for it. You can save the molecule into your database by selecting File → Save (or clicking on the disk icon) and then selecting the tab which says "Save in Database As".

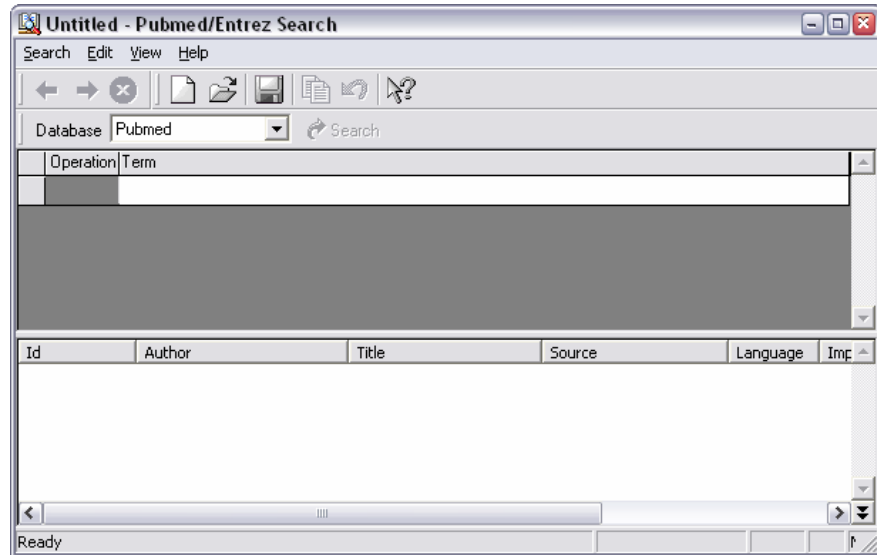




If you have multiple accession numbers you want to retrieve you can put them all into this box separated by commas. This allows you to fetch the sequences straight into the explorer without opening the viewer program first.

Importing a sequence from an Entrez text search

To launch the Entrez search tool select Tools → Open → Retrieve from NCBI Entrez server

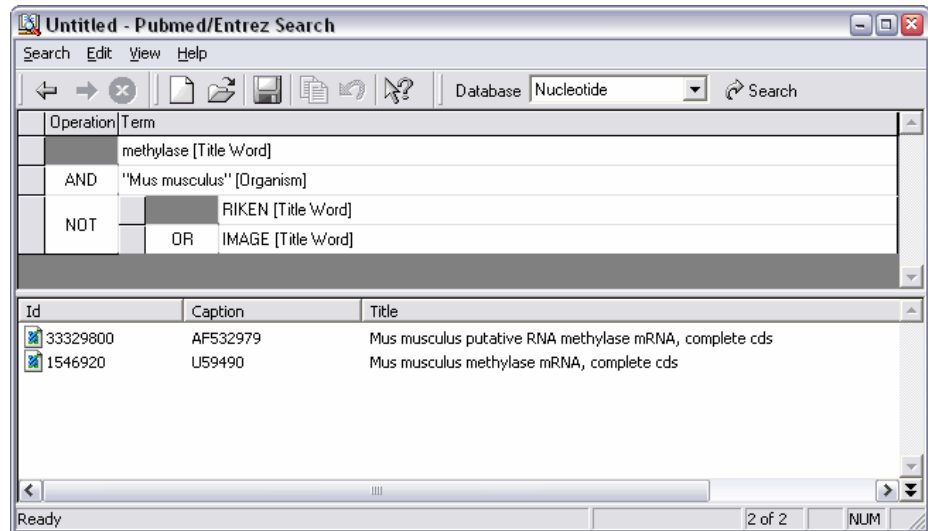


The Entrez search tool in VNTI works in the same way as the Entrez search site itself. The first thing you need to do is to select the database you want to query using the pull-down menu underneath the toolbar. The default is to search PubMed, but for importing molecules you will want either the nucleotide or protein databases.

Next you need to specify one or more search terms to use. You put these into the table underneath the database selector. For each search term you need to provide a piece of search text and also to select from a drop down list the field in the entry you want this search to apply to. A new search term will be started whenever you put a space in a search condition (unless you enclose it in double quotes). For subsequent search terms you also need to say how this term is to be combined with the existing results (using OR, AND or NOT).

If you want to create more complex queries, you can also insert subconditions by selecting Edit → Subcondition.

The following example shows a search for Methylases in Mouse which do not come from either the IMAGE or RIKEN clone collections.



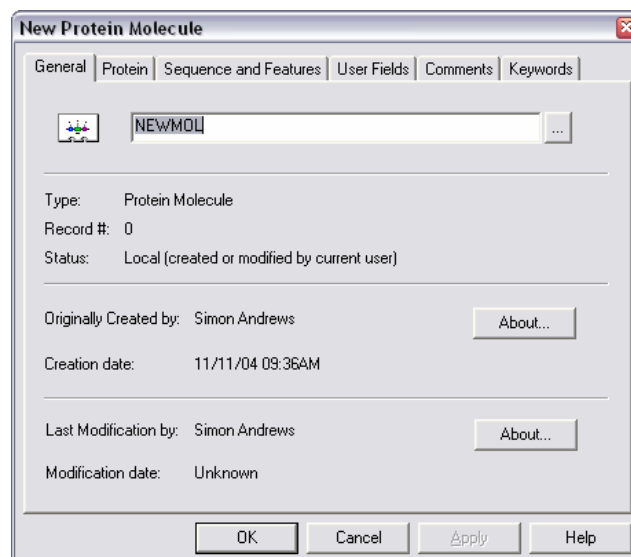
When you've finished you simply press Return (or the Search button) to run the search. Results appear in the bottom half of the window.

To put one or more of the hits into your explorer database you can either open the molecules in the VectorNTI viewer (shown later) and then save them from there, or you can simply drag the molecules from the Entrez results into the Explorer where they will be saved.

Manually creating a new molecule

If you want to create a new molecule manually (say you want to copy/paste the sequence from a web page) then you can also do this from the explorer. Firstly you need to choose the correct type of molecule (protein / DNA) from the drop down selector, then select:

Table → New → Molecule (using sequence editor)

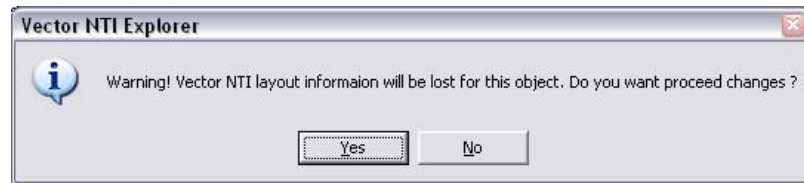


You can then use the text box on the initial screen to give your molecule a name.

You should now move through the tabs at the top of this box from left to right, entering the information required at each stage as you go. The only things you have to fill in are the name of the molecule and its sequence. Everything else is optional.



When you've finished annotating your sequence select OK to create your new molecule. You will see the following warning message appear:

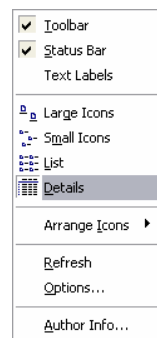


..and if you say "Yes" then your molecule will be created and will appear in the Explorer.

Organising the Explorer

You'll quickly find that it won't take long for the number of molecules in your database to build up to such an extent that you have difficulty finding the one you want. There are therefore a couple of additional things you can do to organise the molecules in your database to keep things a bit tidier.

Column Formatting



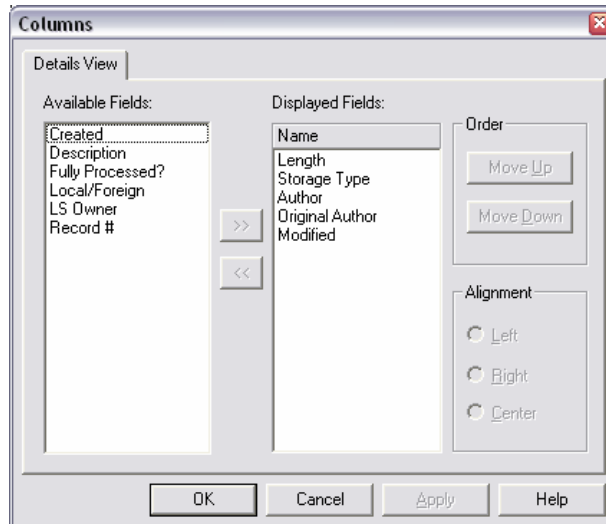
The list of molecules in your database can be displayed in a number of ways. The options actually mirror those you get when looking at files in Windows Explorer. To change the style of list you see use the options you have in the View menu. The most useful option is to view the details of the sequences (this is not the default so you have to change it).

When you look at a series of molecules in the main window there are a series of data fields which are shown alongside each entry. The default selection of fields chosen isn't particularly useful (it doesn't include the molecule description!) so you may find it useful to alter the information shown to something more to your liking.

To change the columns of information shown you need to right mouse click on the toolbar above the main sequence list. You should see a menu appear which says "Columns", and you should select this.

Name	Length	Storage Type	Author	Original Aut...	Modified
BAB76147	382	Basic	UNKNOWN	UNKNOWN	11/10/04 03:08PM
BAB80533	305	Basic	UNKNOWN	UNKNOWN	11/10/04 03:08PM
BAB81183	238	Basic	UNKNOWN	UNKNOWN	11/10/04 03:08PM
BAB81213	320	Basic	UNKNOWN	UNKNOWN	11/10/04 03:08PM

You will then get a new tool appear which you can use to change the list of columns shown in this display.



You can therefore set up a more useful set of information to be displayed.

Name	Length	Description	Modified
BAB76147	382	endo-1,4-beta-xylanase [Nostoc sp. PCC 7120].	11/10/04 03:08PM
BAB80533	305	probable endo-1,4-beta-xylanase [Clostridium perfringens str. 13].	11/10/04 03:08PM
BAB81183	238	probable endo-1,4-beta-xylanase [Clostridium perfringens str. 13].	11/10/04 03:08PM
BAB81213	320	probable endo-1,4-beta-xylanase [Clostridium perfringens str. 13].	11/10/04 03:08PM
BAB81257	329	probable endo-1,4-beta-xylanase [Clostridium perfringens str. 13].	11/10/04 03:08PM
BAB81316	327	probable endo-1,4-beta-xylanase [Clostridium perfringens str. 13].	11/10/04 03:08PM
BAC09454	385	tlr1902 [Thermosynechococcus elongatus BP-1].	11/10/04 03:08PM

Subsets

The main thing you can do to organise your database is to arrange it into subsets. These are groups of sequences which you can define and which you can later select to restrict your database view to show only these sequences. Subsets work like filters – showing you only a selected set of sequences. A single sequence can exist in more than one subset.

As an example if you were working on the ABC1 gene you might have a subset which contained the cDNA sequences for this gene from a range of organisms.

To create a new subset simply select Table → New → Subset from the main Explorer menu.

Once you have your new subset visible you can move files into it simply by dragging them into the subset from the main sequence window view.

If you want to delete a subset then you have two choices and you should be careful that you pick the right one. You get to these options by right-clicking on the subset icon.

Dismiss subset: This is the safe option. The subset is removed but all of the sequences stay in the main list and any other subsets which contain them. You might end up with a few sequences not in any subset, but you definitely won't lose anything.

Delete contents: **This can be dangerous.** All of the sequences in that subset are permanently removed from the database. Please note that if those sequences are also in other subsets they will be deleted from them as well (without a specific warning). **Use this option with care!**



Molecule Display with VectorNTI

VectorNTI (as its name suggests) was originally developed as a package to draw plasmid maps, so it has a very capable visualisation tool included with it. This allows you to see a graphical representation of your sequence annotated with the features the program knows about. You can also link this to the raw sequence.

To open up the main VectorNTI window for a molecule from the Explorer you simply need to double click on the molecule (or right click and select "Open").

Vector NTI - [AF532979]

File Edit View Analyses Cloning Gel List Align Assemble Tools Window Help

Active Pane: [Icons]

AF532979

- General Description
- Standard Fields
- References
- Comments
- Annotations
- Feature Map
- Imported Features Not Shown
- Restriction/Methylation Map

1501 GGAGCGTTTT ATAAAAAGGC TTGGATGGG AATTGTGCA TATGATAAA
CCTCGCAAAA TATTTTCCG AACCTACCC TTAAACACGT ATACTAATTT

1551 AAACCTGTGC TGATTACAA TTTTATATA AATAACTGCT ATTAATCAAA
TTGGACACG ACTAAATGTT AAAAATATAT TTATGACGA TAATTAGTTT

1601 ATTGGATAAT TTCAAAATGG TCTTTGTTTT AGAGACTATT TTGATGTATA
TAACCTATTA AAGTTTTACC AGAAACAAA TCTCTGATAA AACTACATAT

1651 TGTCTTTTGA GCCTTGTTAA GTGTCAAGGA CTGGTATATA TATATATAGA
ACAGAAAACG CGGAACAATT CACAGTCCCT GACCATATAT ATATATATCT

1701 GAGAGAAAAGT GTTTATATTT TTACAAAAGAC AATATTGTAT CTGCCACATA

Ready 1673 bp - 1677 bp (5 bp) 1673 bp (5)

The default display mode is a 3 panel layout showing:

- A series of folders containing text based information about your sequence
- A graphical representation of the sequence annotated with its various features
- A sequence editor showing the raw sequence

Adjusting the display



You can drag all of the dividers between the panels so that they are the size you prefer. You can also change the order in which the windows are laid out. You do this by pressing the "Toggle layout" button on the toolbar. If you're working with linear sequences you may find the display clearer if you have the main graphics panel running along the whole of the top of the application with the other panels arranged underneath.



The screenshot shows the VectorNTI software window for file AF532979. The main window displays a graphical representation of a DNA molecule with various features like 'ORI', 'oriT', 'oriV', and 'oriP'. Below the graphics is a sequence viewer showing the following DNA sequence:

```

801 CATATGTGTG TGGACAGAC ATAGACTACA ATACAGTTCA TGGCTTGGGA
GTATACACAC ACCCTGTCTG TATCTGATGT TATGTCAAGT ACCGAACCCT
851 AAGGCTAGTA GAAAAAACA GAAATGGCGT GGACCAGATG AAAACATCAG
TTCCGATCAT CTTTTTGGT CTTTACCGCA CCTGGTCTAC TTTGTAGTC
901 GGCAAATCTT CGTCAGTATG GCTTAGAGAA GTTTTACCTT GATGCTTTGG
CCGTTTAGAA GCAGTCATAC CGAATCTCTT CAAAATGGAA CTACAGAACC
951 TTTCAGATGC GTCTAAGCCT TCCTGGAGGA AGGGAACATA TTTTGTATGCA
AAAGTCTACG CAGATTCGGA AGGACCTCCT TCCCTTGAT AAAACTACGT
1001 ATCATCACAG ATCCTCCATA TGGTATCAGA GAATCTACAA GAAGATCAGG
  
```

The bottom status bar shows 'Ready' and '830 bp'.

Adjusting the Zoom level

There are two methods for restricting the view to only one part of a sequence. You can either zoom in and out of the whole sequence, or you can change the view to show just a specified subsequence. Both of these can be useful in different situations.

The zoom controls are 3 buttons on the toolbar, which can also be accessed by right-clicking on an empty part of the graphics panel.

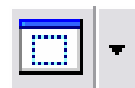
	This button will zoom in on the current display.
	This button will zoom out on the current display.
	This button will adjust the zoom level so that the full molecule exactly fits within the display area currently available.

You should note that the zoom tool operates on both the x and y axes at the same time, so that by the time you've zoomed in a lot you may find that your features are very tall. In these cases you may want to use the subsequence view instead.

Viewing a subsequence

If you want to focus on just one part of a molecule then you can use the subsequence display to restrict the view to just a selected region.

First you need to select the region of sequence you're interested in. You can do this by clicking and dragging in either the sequence or graphical view, or by clicking on any feature to select the sequence beneath it. If you want to select a particular sequence range you can use Edit → Set Selection to manually make the selection.



Once you've done this you can then press the "View Molecule Fragment" button to restrict the view to just the selected region. When you want to move back to the whole molecule simply press the same button again with nothing selected and the view will reset to the default.



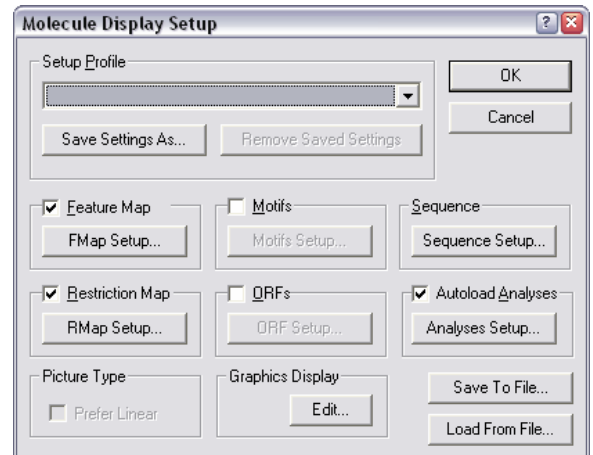
Changing the Graphics Display

VectorNTI gives you pretty much complete control over how the graphics panel is configured and laid out. With a bit of time and effort you can have things looking any way you want them.

VectorNTI has a default way of displaying sequences. This display is controlled by a series of preferences which apply across all of the sequences in your database. If you want to change how sequences are displayed you can do this from within the program.

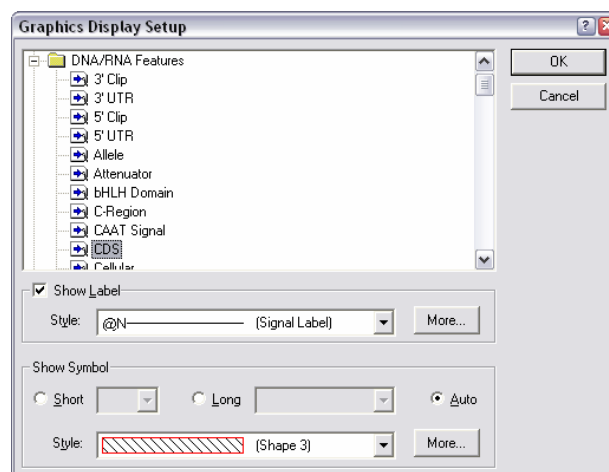
To see the options you have for changing the display of sequences you should right mouse click on the graphics panel and then select Display Setup. This will open up a new window which shows all of the ways in which you can customise the graphics display.

At the bottom of this window are a series of tick boxes which indicate the elements to be shown in the graphics display. Once a box is checked you can click on the button next to it to set your preferences for how that element should be displayed.



Feature Display

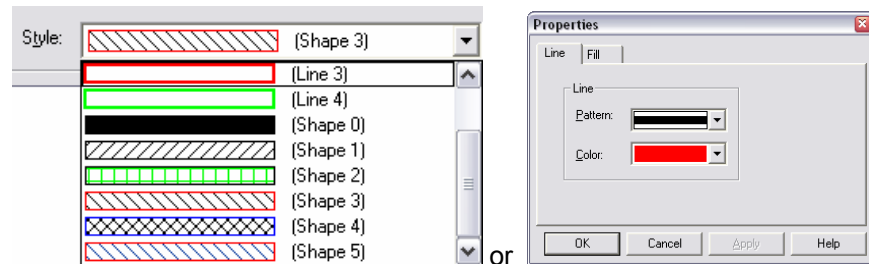
At the bottom of the display settings window is a box which says "Graphics Display" and which contains an Edit button. This tool allows you to change how different types of features are displayed by VNTI. If you click the edit button you will see a list of features appear. Clicking feature type brings up the details of how that feature is displayed.



In this example we can see that the CDS feature is outlined in red with black hatching inside it. It has a label on it. Most features can have two shapes depending on how much of the screen they occupy. If they're very small they get the "short" symbol, and if they have more space they get the long symbol.



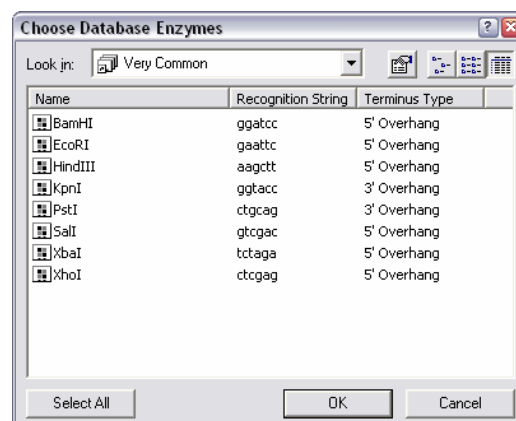
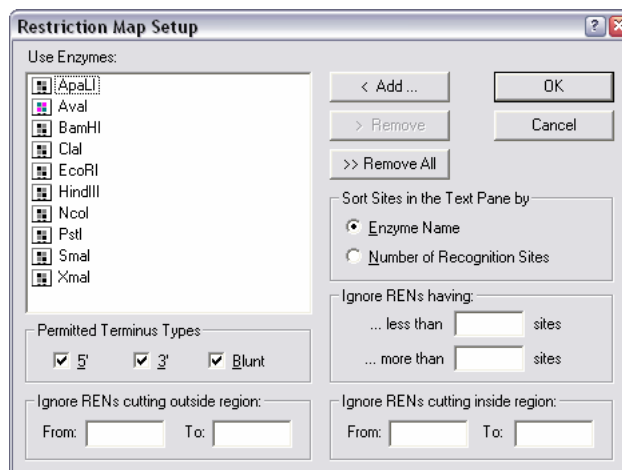
To change the colouring of a feature you can either select from a predefined list of styles by using the pull down style list, or you can create your own by clicking on the "More" button and making up your own style.



A useful shortcut to know about is that you can change the display preferences of any feature type directly from the graphics display by right clicking on any graphical feature.

Restriction Maps

By default all new sequences are annotated with a restriction map using a very common set of enzymes. If you want to see a different list of restriction sites you can specify this in the display setup by clicking on the "RMap Setup" button.



Here you can select which enzymes you want to include in the restriction map, you can also choose to filter this list according to how many times they cut your sequence. If you choose to add enzymes to the list you will see the enzyme list stored in your explorer database. Restriction enzymes can be organised into subsets in exactly the same way as sequences. In this example I've made a subset containing very common enzymes. Other subsets which are useful to keep around might be all non-ambiguous enzymes or all the enzymes you have in your freezer.



Open Reading Frames (ORFs)

If you're working with a new or unannotated DNA sequence it is often useful to be able to see the positions of the open reading frames within that sequence. These are not shown in the default VNTI display setup, but can be easily added by ticking the ORFs box in the display setup.

When viewing ORFs you will almost certainly want to make some changes to the default setup, so you should press the "ORF Setup" button.



The default ORF search in VNTI allows start codons from both eukaryotes and prokaryotes so if your sequences doesn't come from a species which allows a GTG start codon you should remove this. Also, by default only complete ORFs which include a start and a stop are shown. If you are likely to have partial ORFs in your sequence then you should tick the boxes to allow incomplete ORFs to be shown.

One quirk of the ORF display tool is that although you can specify to show incomplete ORFs without a start or a stop VNTI won't show you ORFs which have neither a start nor a stop (ie a complete ORF running right through your sequence).

Display Profiles

Once you've made the changes you'd like to see to the graphics display these will be maintained within the sequence you have edited. However if you would like to use this display setup again on other sequences then you can save these preferences as a display profile.

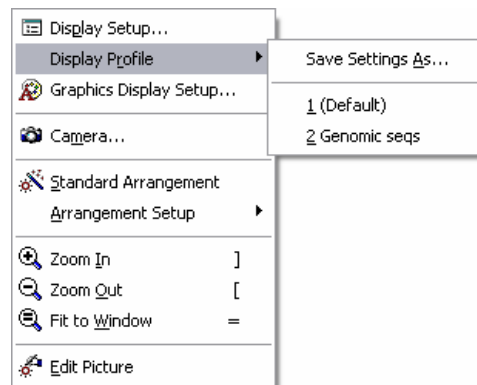
At the top of the display settings window is a section titled Setup Profile which contains a drop down list box (initially empty). Next to it is a button saying "Save settings as". Once you've finished making your changes press this save button and you will be prompted for which profile name you want to use.

Unless you specify a new profile name the system will alter the default profile and all of your sequences will be changed!



A better idea is to create a new profile with a name of your choosing. Here I've created a profile for genomic sequences. You can have as many different display profiles as you like.

Once you've created a new display profile it's easy to switch between them in the main display. Simply right-click on the graphics window and select "Display Profile" and then pick whichever of your profiles you want to use.



Manually Editing the Graphics Display

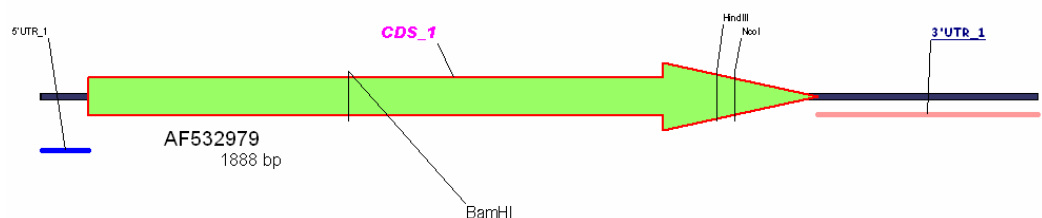
The tools described up until now are useful for changing some aspects of the way your molecules are shown, but often you will have a particular requirement which isn't served by the tools available in the display setup.

To get more complete and individual control over the display of a sequence you can use the Graphics Editor. The changes you make with this tool won't apply outside of the current sequence, but you can make almost any change you want.

To begin editing your display select Edit → Edit Picture

You can now let the mayhem commence!

- You can move anything by clicking and dragging it (where the cursor is a white cross).
- You can resize anything by clicking on it and then dragging from the corners (where the cursor is a black cross)
- You can change the shape, colour and shading of anything by double-clicking on it. These changes are per feature, so if you have two different CDS's in your sequence you can use this method to make them look different from each other.



Please note that we are in no way suggesting that you should make your molecules look anything like the one above, merely that you could if you wanted to.

One important thing to note about changes made whilst editing the picture is that these changes are **not** stored in the database and will be lost when you next open the molecule. If you want to preserve these changes you can save the molecule as a separate file, and this will cause them to be remembered. To do this select File → Save As and then click the tab at the top which says "Save as File".



Sequence Manipulation / Editing

As well as changing the way a molecule is displayed you can also edit sequences within VNTI and it has some advanced tools allowing you to do virtual cloning experiments.


Making Selections

Since a lot of editing relies on being able to make and work with selections we may as well start by showing a few techniques for getting to the bit of sequence you want.

Manually

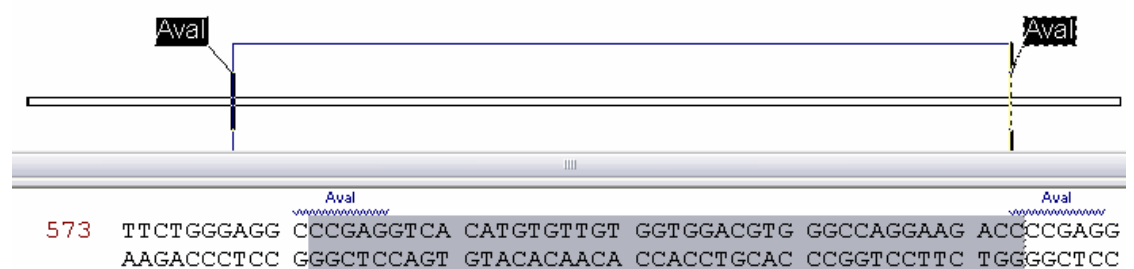
The quickest and easiest way to make a selection is to do so manually. Within either the sequence or the graphical display panels you can click and drag with your mouse to select a region of sequence. As you make a selection in one panel you should see it become selected in the other. Also if you look at the bottom of the VNTI window you should see a small display giving the exact details of the region you have selected.

314 bp - 359 bp (46 bp)

Once you have made a selection you can adjust the ends of it by moving your mouse over them so that the cursor looks like this. 

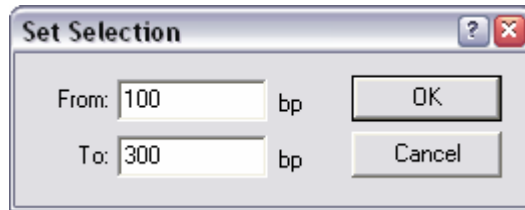
You can then click and drag to move the end of the selection. Clicking away from the ends causes the selection to be reset and a new selection to start.

In addition to dragging a selection you can also click on any feature in the graphical display to select the region of sequence covered by that feature. If you want to select the region between two features (say restriction sites) then you can click on the first, then shift+click on the second one, and the region between them will be selected.



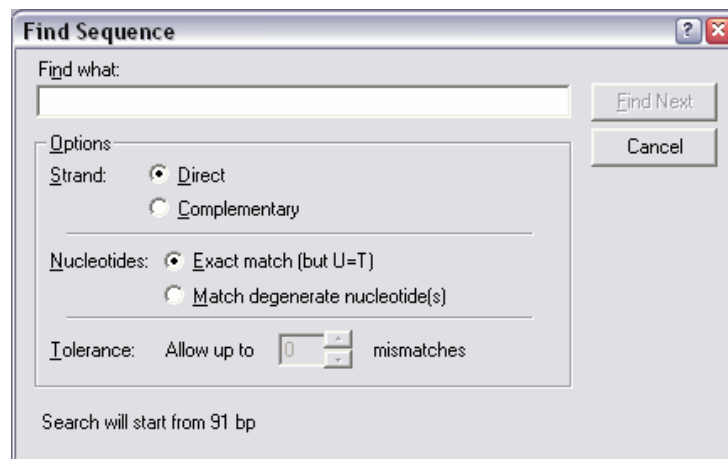
By BasePair Position

If you know the basepair range you want to select then you can do this directly, either by selecting Edit → Set Selection, or more quickly by pressing Control+G. This opens up a small dialog box into which you can enter a basepair range and have that set as the current selection.



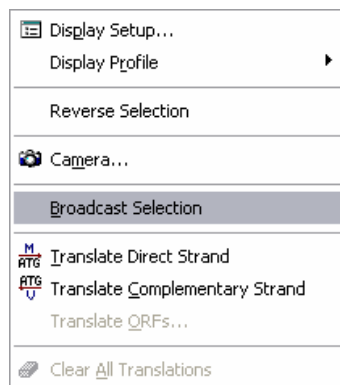
By Sequence

If you have a piece of sequence you would like to find in a larger VNTI entry then you can use the Find Sequence tool to locate it. You can launch this by selecting Edit → Find Sequence.



You can choose which strand of your sequence you want to search in, and can also allow for some degree of degeneracy in your search.

Broadcasting Selections



Within VNTI it is possible to have several different views of the same molecule open at the same time (say a single sequence view and an alignment). Within a single VNTI window a selection will automatically extend to all of the views of that sequence, but it will not pass to other windows which also have that molecule open.

If you want to spread your selection to all views of that sequence then you must use the "Broadcast selection" tool. You access this by right-clicking over a sequence view (but not a graphical view) and selecting "broadcast selection" from the popup menu which appears. All other views of that sequence should now show the same selection.

Editing a sequence

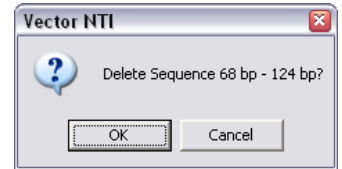
Unlike some other sequence analysis applications the editor in VNTI doesn't work like a normal text editor. Instead it splits all edits into defined insertions and deletions. In practice this way of editing more closely resembles how you're likely to modify a sequence in a lab, and therefore doesn't prove too confusing.



You should note that any edits you make to a sequence will not be made permanent until you save the sequence back to your database. If you do not explicitly save your sequence, then when you close it you will be prompted that it has changed since you opened it.

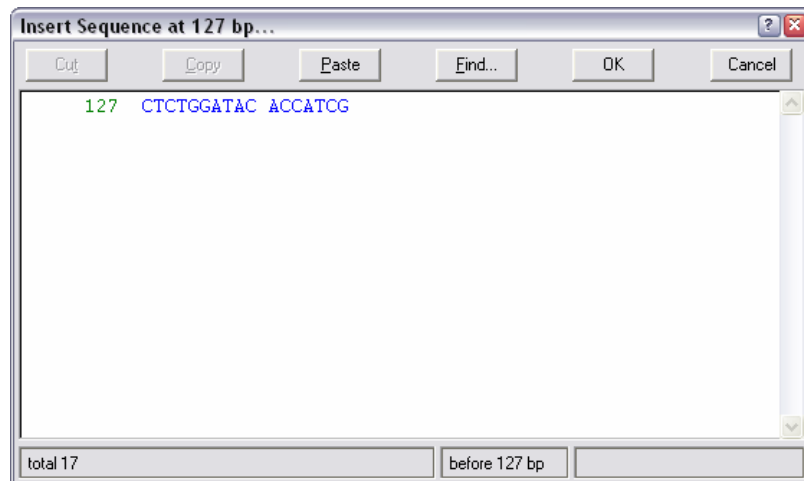
Deleting sequence

First select the region you want to delete (even if it's only 1bp), then press the delete key. You will get a warning message asking if you really want to delete the region and if you say yes it will be removed.



Inserting Sequence

First select the point at which you want to insert the sequence and then start to add some sequence. You can do this either by starting to type sequence or by pasting it from the clipboard. A new sequence editor window will now open in which you can edit the sequence to be inserted.



This edit window does operate like a normal sequence editor and once you're happy with the sequence you wish to insert just press OK to update the original sequence.

Adding / Removing Features

As well as editing the sequence itself you may also wish to change the features which are annotated onto it. When you import a sequence into VNTI it may well have some features already defined, but if not you can soon add your own.

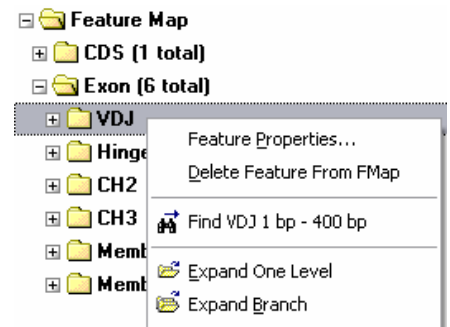


Editing / Deleting an existing feature

To change an existing feature first locate it either in the graphical display, or in the Feature Map section of the sequence property browser window. If you now right mouse click over the feature you will see a list of options

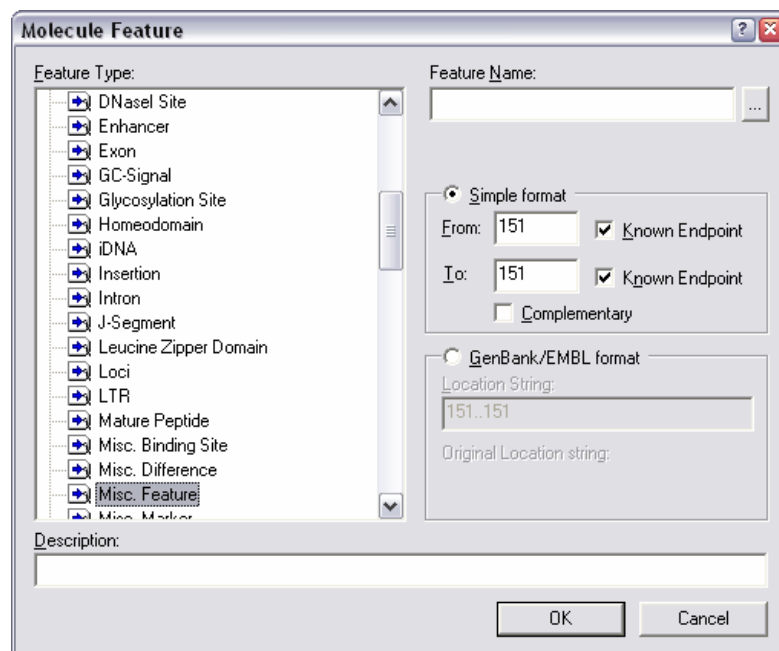
Delete Feature will remove that feature from the sequence.

Feature properties allows you to change the type of the feature, its position within the sequence and the text label which is associated with it.



Creating a new feature

To create a new feature select Edit → New → Add Feature to FMap. You will then see the following dialog appear:



You can select the type of feature you want to add from the list on the left, and select the position of the feature using the boxes on the right. If you have a region of sequence selected when you launch this tool, these coordinates will automatically be filled in for you.

If you want to insert a feature with a complex, multi-position location (say a multi-exon mRNA feature on a genomic sequence), then you can use the GenBank/EMBL format position option to enter a location string for your feature.

Finally you can give your feature a name. This will be the text associated with your feature in the graphical view of your sequence.



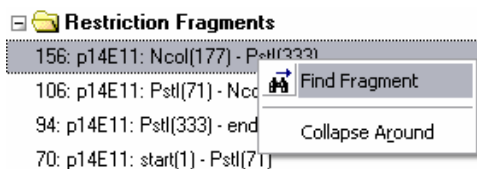
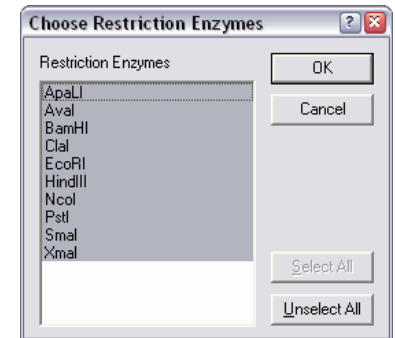
Analysing a Molecule

Finding Restriction Fragments

The basic restriction map will annotate the positions of enzymes onto your sequence, but it will not provide you with a list of the sizes of fragments which would be generated if you were to restrict your sequence. To get this information you need to select:

Analyses → Restriction Analysis → Restriction Fragments

You should then see a box showing the currently selected restriction enzymes. You can select one or more enzymes from this list. Press OK to do the fragment calculation. If you have performed a previous fragment calculation on this sequence you will be warned that the new data will overwrite the old fragment information.



Your list of fragments will now appear in the text pane window. They will be ordered by size with the largest first. If you want to see where any of these fragments are in your sequence you can right click on them and select "Find Fragment" from the menu which appears.

Translating a Sequence

In order to translate a sequence you first need to select the region you wish to translate. If you have created a feature for your translated region (either a CDS from a database entry, or and ORF from an ORF search), you can click on that in the graphical view to select it. Alternatively you can manually select the region you're interested in using the selection methods mentioned previously.

There are 3 different types of translation you can apply in VNTI:

1) A single frame translation which appears in the sequence editor

If you are translating from an existing CDS feature you can simply right-click on the feature and select

Translate → Feature into Sequence Pane

If you have manually selected a region you need to select

Analyses → Translation → Into Sequence Pane → Direct / Complementary Strand

Your sequence pane should now show the appropriate translation.

```

+1  Asp Val Gln Leu Gln Ala Ser Gly Gly Gly Ser Val Gln Ala Gly Gly Ser
1   GATGTGCAGC TGCAGGCGTC TGGAGGAGGC TCGGTGCAGG CTGGAGGGTC
   CTACACGTCG ACGTCCGCAG ACCTCCTCCG AGCCACGTCC GACCTCCCAG
+1  Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Tyr Thr Ile Gly Pro Tyr Cys Met
51  TCTGAGACTC TCCTGTGCAG CCTCTGGATA CACCATCGGT CCCTACTGTA
   AGACTCTGAG AGGACACGTC GGAGACCTAT GTGGTAGCCA GGGATGACAT

```

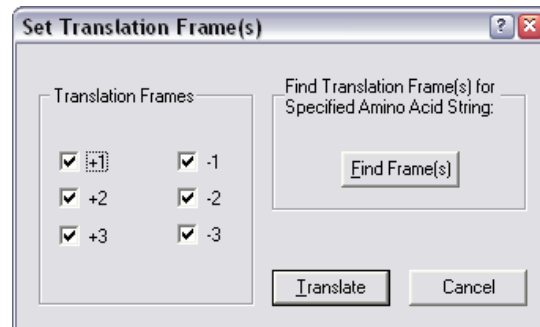


If you want to change the translation to a 1 letter rather than a 3 letter code you can do this by right-clicking on a sequence display and selecting Display Setup → Sequence Setup.

2) A multi-frame translation which appears in the editor

To perform a multi-frame translation select the region you want to translate and then select:

Analyses → Translation → Into Sequence Pane → 6 Frame Translations



You should then see this dialog box appear. From here you can select which of the 6 frames you actually want to see.

A useful addition is that if you know part of the translation you're expecting you can press the "Find Frame(s)" button and enter the sequence and the program will tell you which frame(s) contain that translation.

```

+3   Cys Ala Ala Ala Gly Val Trp Arg Arg Leu Gly Ala Gly Trp Arg Val
+2   Met Cys Ser Cys Arg Arg Leu Glu Glu Ala Arg Cys Arg Leu Glu Gly Leu
+1   Asp Val Gln Leu Gln Ala Ser Gly Gly Gly Ser Val Gln Ala Gly Gly Ser
1    GATGTGCAGC TGCAGGCGTC TGGAGGAGGC TCGGTGCAGG CTGGAGGGTC
    CTACACGTCG ACGTCCGCAG ACCTCCTCCG AGCCACGTCC GACCTCCCAG

```

3) A translation to create a new protein sequence

If you want to translate a region of DNA sequence to make a new protein you really should create a CDS feature on the sequence first. This will make the translation easier and it will also provide you with a record of which region you translated to make the protein.

Once you have a CDS feature you can create a new protein by right-clicking on it in the graphical display and selecting Translate → Feature into New Protein.

When you translate a sequence to make a protein VectorNTI will keep track of the relationship between these sequences so that should you later edit the DNA sequence you will get a warning when you next view the protein that the sequence from which it derives has changed. In order to establish this connection the program requires that every sequence used for translation first be saved into your database. Therefore if the sequence you're using has had any changes made to it since the last time it was saved you will be prompted to save it before continuing.

You will then be asked which Protein subset you want your newly created sequence to be put into, and finally asked to provide a name for the new sequence. Once this is done the new protein sequence will be opened in its own VNTI window (although your original DNA sequence will also remain open).



Primer / Probe searches

VNTI has a series of tools for primer design. They come with a default set of criteria which describe their idea of what makes a good primer, but you may want to adjust some of the parameters to suit your own preferences.

PCR Primers

There are two tools for PCR primer design, both of which are very similar. One is used when you want to make primers to identify the presence of a particular sequence, the other is used when you want to amplify up a defined region.

The first thing you need to do is to select the region of sequence you want to work with. Then you need to do one of the following:

- To design primers within this region which will amplify a subsection of it for identification purposes, select Analyses → Primer Design → PCR Primers.
- To design primers which will amplify up all of the selected region select Analyses → Primer Design → Amplify Selection.

Either of these options produce a window which looks very similar to the one shown above. This window is split into several sections in which you can specify the criteria you want the program to use when designing your primers.

The standard criteria used by VNTI are fairly relaxed and will produce suggested primers from most templates. If you would like to be more selective about the primers designed then a few things you could consider changing would be:

- In the primer tab you could narrow the %GC allowed, and maybe the Tm. You could increase the allowed length of the primers, but if you do this you should also increase the maximum Tm.
- In the structure tab you could reduce the number of allowed repeats to 3.
- In the pairs tab you could reduce the allowed difference in Tm or %GC between the primers.



If you need to add extensions to your primers (for example restriction sites) then you can specify this by pressing the "More" button on the primer tab. This reveals an extra set of options in which you can include any 5' extensions.

Attach to 5' terminus
of Sense Primer: ... of Antisense Primer: ...

One thing to realise about the VNTI primer tools is that although you set preferences for the qualities you'd like to see in a good primer, if the program can't find any primers which match those qualities it will relax your restrictions until it does find something. Under the qualities tab you can specify how important each of the limits you have set are (ie which ones to ignore first)

Find Primers in Selected Region of NM_000077

Primer Amplicon Structure Pairs Similarity 3' end Uniqueness **Qualities** Filters

"Importance" below must be between 1 and 10:
1 - not important (almost don't care) 10 - crucially important (disregard if not so)

Importance to be inside specified limits

Tm:	<input type="text" value="9"/>	Tm Difference:	<input type="text" value="9"/>
%GC:	<input type="text" value="9"/>	%GC Difference:	<input type="text" value="9"/>

Importance to satisfy specified conditions

Palindromes:	<input type="text" value="9"/>	Nucleotide Repeats:	<input type="text" value="9"/>
Hairpin Loops:	<input type="text" value="9"/>	Dimers:	<input type="text" value="9"/>
		3' End:	<input type="text" value="9"/>

Load... Save... OK Cancel Apply

If you want to ensure that all of your restrictions are obeyed you need to set all of your quality values to 10.

PCR Analysis

- #1: Product of length 1404 (rating: 171)
 - Contains region of the molecule from 729 to 2132
 - Tm: 77.8 C TaOpt: 53.8 C GC: 44.9
 - Sense Primer:
 - ATGGTGAAACCTGAGGAATGC
 - Similarity: 100.0%
 - Length: 20 Tm: 51.2 C GC: 50.0
 - dH: -148.1 kcal/mol dS: -381.4 cal/mol dG: -32.6 kcal/mol
 - Antisense Primer:
 - CGTAGGAAAAGACTGTGCTG
 - Similarity: 100.0%
 - Length: 20 Tm: 47.5 C GC: 50.0
 - dH: -146.6 kcal/mol dS: -381.6 cal/mol dG: -31.0 kcal/mol
 - Tm Difference: 3.7
 - GC Difference: 0.0

Once you run your primer search your results will appear as a list in the VNTI text window. For each primer you will see details of its length, Tm, %GC etc. You will also get details of the properties of the region amplified by the primer set.

If you want to locate the region amplified by a particular primer set you can right-click on the folder icon for the primer set and select "Find PCR product". The product will then become selected in the main view.



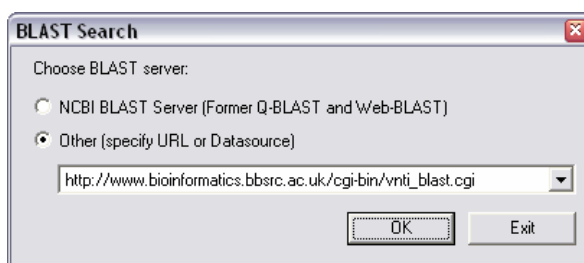
Database Searching with BLAST

VNTI has a built in tool for submitting jobs to a remote BLAST server, and viewing the results. By default it connects to the NCBI blast server, but we also now have a local BLAST server to which you can connect, which should be faster and offer a more useful set of databases.

Launching BLAST

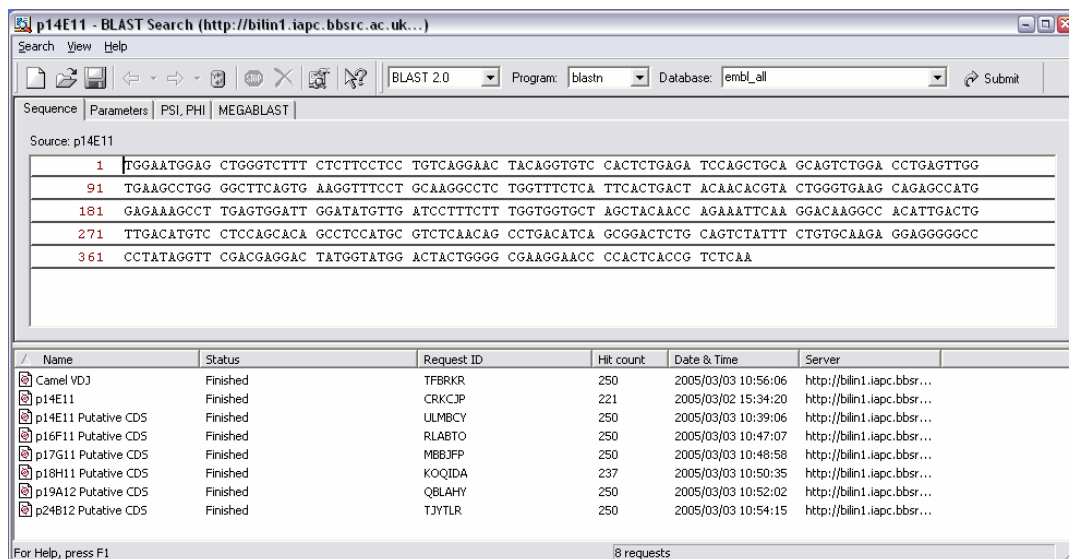
You can launch a BLAST search either from the Database Explorer, or from a molecule view window by selecting Tools → BLAST search. If you are in a molecule view and have part of the sequence selected you will be asked whether you want to search with the whole sequence, or just the selected region.

You will then be asked which BLAST server you want to use:



The default will be to use the NCBI's public BLAST server, and this should be the one you use unless you know of a local BLAST server which is available to you.

Once you have selected a server, the BLAST submission window will then open.



The two main windows in this view show the sequence you are searching with at the top, and a list of recent BLAST searches you have run at the bottom.



Selecting a Database

To run a BLAST search you need to select an appropriate BLAST program and database, set the stringency of your search, and then run it.

To select the database you want to search you first need to select the correct blast program for the type of search you want to run. The program you want depends on the type (DNA or Protein) of query sequence and database you want to use.

Program	Query Sequence	Database
blastn	DNA	DNA
blastp	Protein	Protein
blastx	DNA	Protein
tblastn	Protein	DNA
tblastx *	DNA	DNA

* In tblastx searches both the query sequence and the database are translated into protein before searching. This is VERY slow (~36 times slower than a normal search), so don't use this unless you really, really need to!

The default program selected will be the one which lets you search the same type of database as your query sequence – which is usually what you want.

Once you have selected the appropriate blast program you can now select the database you want to search. Only databases appropriate to the search program you have selected will be shown (eg for blastp only protein databases will appear).

Setting BLAST options

Before you start your search there are a couple of options you may want to set. These can be found under the "Parameters" tab.

The screenshot shows the 'Parameters' tab of the BLAST interface. Key settings include: 'Expect' set to 10, 'Word Size' set to 3, 'Genetic codes' set to 'Standard', and 'Other advanced options' set to 'None'. Under 'Limit by entrez query', 'Limit' is checked and set to 'None', with 'Operation' set to 'OR'. The 'Filter' section has 'Low complexity' checked, 'Human repeats' unchecked, 'Mask for lookup table only' unchecked, and 'Mask lower case' unchecked. 'Composition based statistics' is checked. 'Matrix' is set to 'BLOSUM62' and 'Gap Costs' is set to '11, 1'.

Although there are a lot of options here, there are only two which will have any effect on your search.

The Expect value sets a cutoff for how significant a hit needs to be before it gets reported. The higher the value the more hits you get. If you want to increase the stringency of your search you should reduce this value to, say, 0.001. If you want to see anything and everything you can leave it at its default value of 10.



By default BLAST will filter any query sequence to remove areas of low sequence complexity (things like poly-A tails) as these can tend to bias the hits you get. In some circumstances (especially when searching with short sequences such as primers) you want to force BLAST to use all of your sequence. In these cases you should untick the box which says "Low complexity" in the filters section.



Running BLAST

Once you've set up all of your search options you can start your blast job running by pressing the "Submit" button at the top right of the blast tool window. Your job should then be transferred to the search list in the bottom half of the screen from where you can monitor its progress.

As soon as the job is submitted you should see that its status should change to "Not ready yet. Waiting...", and it should be assigned a job ID. If you're using the Babraham BLAST server then, should you wish, you can enter this ID code into the webblast tool on the intranet and view your BLAST results in a webbrowser.

The system will check every 20 seconds or so to see if your job has finished. If you want to you can force the system to check immediately by right-clicking on your job and selecting "Check Ready Status". Eventually the job should complete, VNTI will then download your results which changes your job status to "Finished" and causes the number of hits you got to be reported.

Viewing BLAST Results

Once a BLAST job has finished you can view the results by double clicking on the job in the job list. A separate BLAST result viewer will now open up.

The screenshot shows the BLAST Results Viewer window for a query named 'hGFP2 1'. The window is split into three main sections:

- Left Pane:** A tree view showing the query 'hGFP2 (molecule)' with a length of 717. It lists 4 hits from the database. The first hit is 'gi|AF324407|ref|AF324407 (1 element)' with a score of 342 (Bits: 678.0) and an expectation of 0. The second hit is 'gi|AF324408|ref|AF324408 (1 element)' with a score of 338 (Bits: 670.0) and an expectation of 0. The third hit is 'gi|AF324406|ref|AF324406 (1 element)'. Each hit includes the molecule name, length, score, bits, and expectation.
- Top Right Pane:** A sequence profile plot showing the distribution of hits across the query sequence. The x-axis represents the sequence position (1 to 717), and the y-axis represents the hit distribution. A yellow bar highlights the alignment of the first hit, and a red bar highlights the alignment of the second hit.
- Bottom Pane:** A sequence alignment view showing the consensus sequence and the alignment of the query sequence with the hit sequences. The consensus sequence is GA A C T T T C A C T G G G T G T C C C A A T T C T G T G A A T G A T G G G A T G T A A T G G G C A C A A A T T T C T G T C A G G G A G A G G G T G A A G G T G A T. The query sequence is GA A C T G T T C A C T G G C G T G T C C C A A T T C T G T G A A C T G G A T G G C G A T G T A A T G G G C A C A A A T T T C T G T C A C G G A G A G G G T G A A G G T G A T. The hit sequence is GA A C T T T C A C T G G A G T G T C C C A A T T C T G T T G A A T T A G A T G G T G A T G T A A T G G G C A C A A A T T T C T G T C A G T G G A G A G G G T G A A G G T G A T.

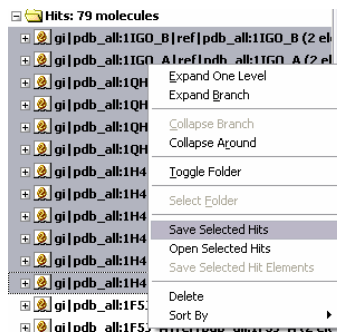
The window is split into three parts.

- The top left pane is a list of the hits. This provided a series of statistics about each hit. Various sections of this list can be expanded (so for instance you can see details of the individual hit segments), or collapsed (so that you can just see the sequence name). Within this list are links to the sequence which was hit. If you click on any of



these the sequence will be retrieved from the database and opened in a new VNTI window.

- The right hand side window shows a graphical view of your hits. The bottom part shows the extent of all of the hits, with the currently selected hit highlighted in red. The middle part shows which parts of your query and the hit matched with each other, and the top part is a graph showing how the similarity between the sequences varied over the length of the hit.
- Finally the bottom panel in this window shows a sequence level alignment of your hit.



If you want to do a bulk import of a lot of blast hits then you should select all the sequences you want from the hit list on the left (you can get a more compact view of this by selecting one hit and then right clicking and selecting "Collapse Around"). You can then right click and select "Save Selected Hits". This will import all of the sequences in one go.

Saving BLAST results

BLAST results will stay on the previous search list until you delete them by right-clicking and selecting delete – however results only stay on the server for 14 days. If you want a more permanent store for these results you can save them back into your explorer database. To do this first open up the results viewer for the search you wish to save. Then simply select BlastRestuls → SaveAs and then select the tab which says "Database BLAST Results". Your results should now be stored in your Explorer Database.



Multiple Sequence Alignments

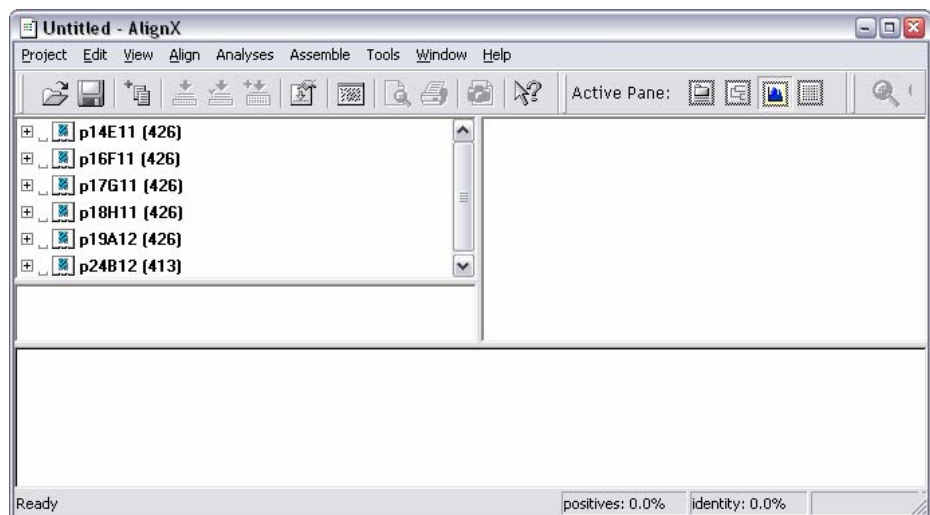
VectorNTI uses a separate program called AlignX to perform multiple sequence alignments. AlignX itself is just a graphical interface around ClustalW, so if you've ever used Clustal before you'll quickly see that the options in AlignX are the same.

AlignX

To begin creating a new alignment first select the sequences you wish to align from your Explorer window and then select:

Align → AlignX → Align selected molecules

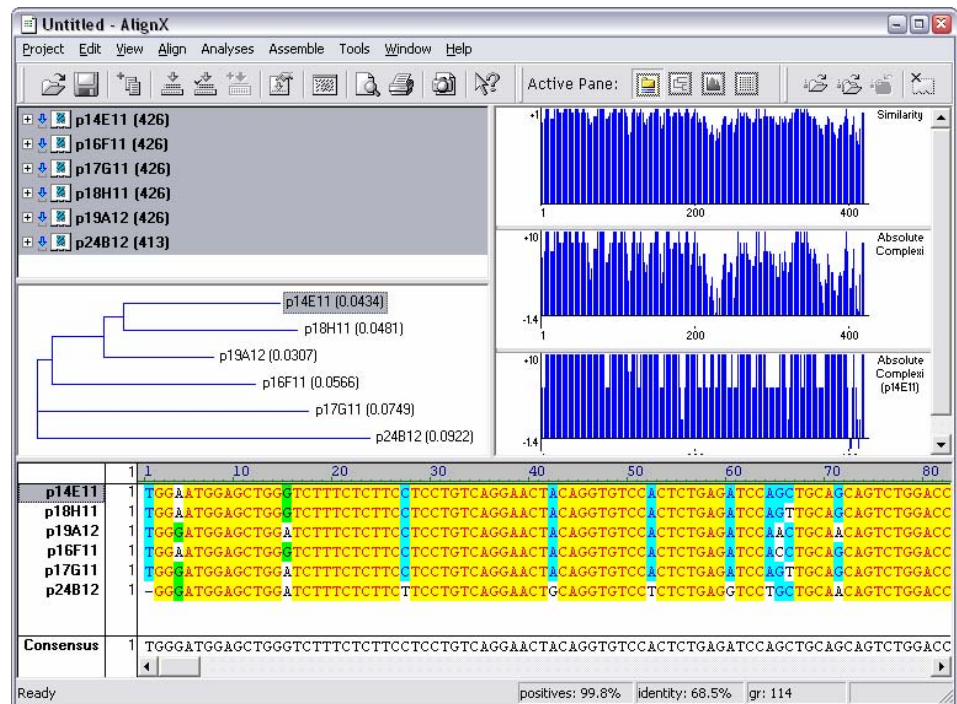
This should cause a new window to open up with your sequences listed in the top left hand corner.



To create an alignment all you need to do is to select the sequences you want to align and then choose:

Align → Align Selected Sequences

A progress box will appear and shortly after that the alignment should be produced.



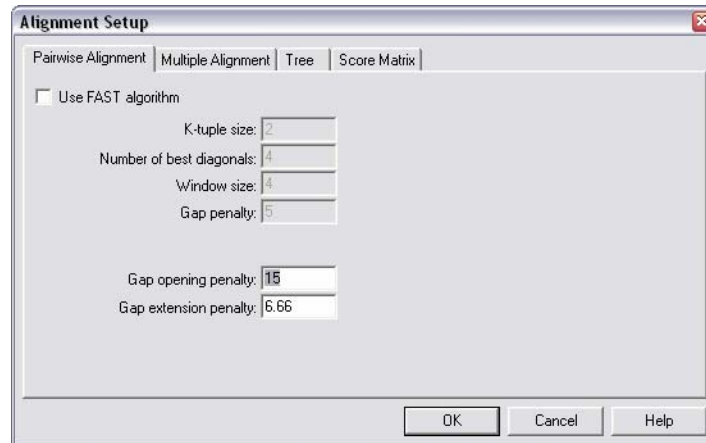
Along with the actual alignment, a few other views are also displayed.

- A dendrogram (**NOT** a phylogenetic tree!) showing the order in which the sequences were added to the alignment.
- A group of 3 graphs which show the patterns of conservation in the alignment.
 - (top) the level of similarity at each position in the alignment
 - (middle) the overall significance of the similarity at each position in the alignment
 - (bottom) the significance of the similarity of the currently selected sequence at each position in the alignment.

Changing Alignment Options

For most alignments the default setup in AlignX is fine. The only time you'll ever need to change anything is if you're aligning a large (say > 20) number of sequences, or if you have a requirement to use a particular similarity matrix (if you don't know what one of these is, you don't need to change it).

Options for changing the alignment parameters can be found in the Align → Alignment Setup menu item.

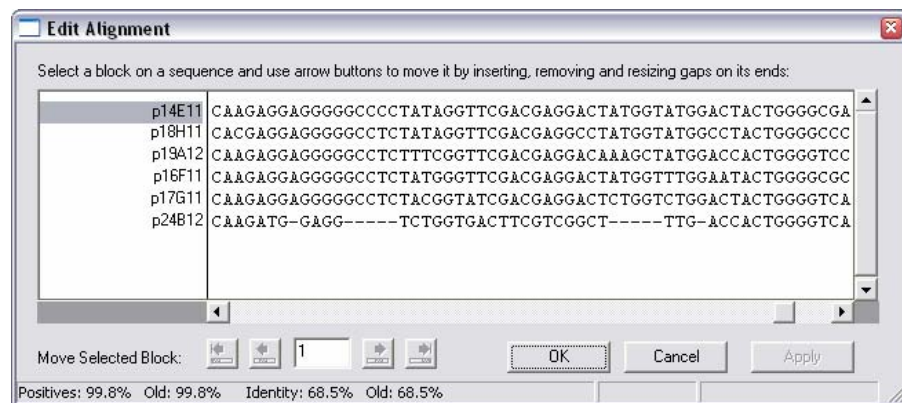


If you're aligning more than 20 sequences then you should tick the box which says "Use FAST algorithm" under the Pairwise Alignment tab and this will help things finish in a reasonable time.

If you need to change the default similarity matrix you can do this under the Score Matrix tab.

Editing your Alignment

Once you have constructed your alignment you may wish to make changes to it where you can see that the program has not produced an optimal alignment. To edit part of an alignment simply scroll through the alignment until you find the bit you want to change, then right-click on it and select Edit Alignment. A separate editor window should now appear.



Alignment editing in AlignX is performed through moving blocks of sequence. You cannot change any of the sequences in the alignment (if you need to do this you probably want to use a sequence assembly package such as Staden rather than an alignment tool). To make an edit you should therefore select a block of sequence and then use the buttons at the bottom of the edit window to move it left or right.

The arrow buttons will move your block either one position at a time, or as far as it is possible to move the block before it hits the adjacent block. When you have a block of sequence selected, only the buttons which it is valid to press will be highlighted.

If you want to insert a new gap into the alignment rather than moving an existing gap you do this by right-clicking where you want to do the insert and selecting "Insert 1 gap(s) at position".



When you've finished changing your sequence you can press the "Apply" button to update the main alignment, and OK to quit the editor window.

Saving your Alignment

Unlike individual molecule sequences, alignments cannot be saved back to your explorer database. Instead they must be saved to a file. You should therefore set aside some space on your fileshare for storing these files.

To save your alignment simply select Project → Save. Once saved you can reload a project using Project → Open.

If you want to export your alignment in a format which allows it to be transferred to other analysis programs (for example GeneDoc) you should select Project → Export MSF format. You should also see that there is an import function which allows you to bring in alignments constructed in other packages.

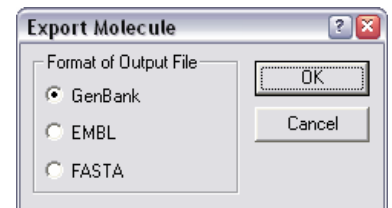


Exporting Information from VectorNTI

Exporting Sequences

VectorNTI allows you to export sequences from your Explorer Database in a number of common formats. To do this right-click on the molecule you want to export and then select Export → Molecule into Text file...

You should then see a dialog box appear which shows you the format options you have for that type of sequence. Once you have selected one of these you will see a file selector which will allow you to pick a filename for the newly created sequence file.



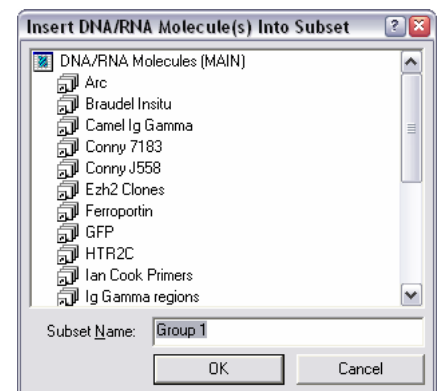
You can do a quick export of a single sequence in genbank format by dragging it out of your explorer window into an open folder, or onto your desktop.

VNTI Archive Files

If you want to pass several sequences to another VNTI user, you also have another option. You can create a VNTI archive file which contains multiple molecules. To do this first select the sequences you want to include in the archive, and then right click and select Export → Into Archive. You can then provide a filename to create your archive file.

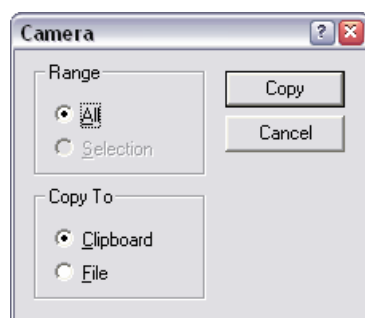
If you have been given an archive file by someone else all you need to do to bring those sequences into your own database is to double click on the file. VNTI will start up and will prompt you for which subset you want to include these sequences in. They will then all be transferred to your local database.

Once you've transferred the sequences to your database you do not need to keep the archive file to access these sequences in the future.



A quick way to create an archive of files is to select them in the explorer and then drag them into an open folder or onto your desktop.

Exporting Graphics



There are many displays within VNTI that you are likely to want to include in other documents. Fortunately there is a simple way to export graphical information from any VNTI window. Simply right-click on a blank part of the window and select the "Camera" option from the menu which appears. You will then be prompted for which part of the display you want to copy, and whether you want it to go to the clipboard or to a file.



If you've chosen to put your graphic on the clipboard (which is probably the easiest way) you can then paste this information into your Word Processor or Graphics package. The graphics file can usually be ungrouped to allow flexible editing if you wish.

The camera export function works in nearly all display windows, both graphical and textual.

