Bioinformatics Toolbox Release Notes

The Chapter 1, "Bioinformatics Toolbox 2.2 Release Notes" describe the changes introduced in the latest version of the Bioinformatics Toolbox. The following topics are discussed in these Release Notes:

- "New Features" on page 1-2
- "Major Bug Fixes" on page 1-3
- "Known Software and Documentation Problems" on page 1-4

If you are upgrading from a earlier version, you should see

- Chapter 2, "Bioinformatics Toolbox 2.1.1 Release Notes"
- Chapter 3, "Bioinformatics Toolbox 2.1 Release Notes"
- Chapter 4, "Bioinformatics Toolbox 2.0.1 Release Notes"
- Chapter 5, "Bioinformatics Toolbox 2.0 Release Notes"
- Chapter 6, "Bioinformatics Toolbox 1.1.1 Release Notes"
- Chapter 7, "Bioinformatics Toolbox 1.1 Release Notes"
- Chapter 8, "Bioinformatics Toolbox 1.0 Release Notes"

Printing the Release Notes

If you would like to print the Release Notes, you can link to a PDF version.

Contents

Bioinformatics Toolbox 2.2 Release Notes

1

Bioinformatics Toolbox 2.1.1 Release Notes

2

Bioinformatics Toolbox 2.1 Release Notes

3

Bioinformatics Toolbox 2.0.1 Release Notes

4

Bioinformatics Toolbox 2.0 Release Notes

5

Bioinformatics Toolbox 1.1.1 Release Notes

6

Bioinformatics Toolbox 1.1 Release Notes

7

Bioinformatics Toolbox 1.0 Release Notes

8

Bioinformatics Toolbox 2.2 Release Notes

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 2.2 (Web Release R14SP3+).

- New multiple sequence alignment viewer (`multialignviewer`) to interactively view, explore alignments, and make manual modifications.

- New function to read an Agilent Feature Extraction Software file (`agferead`) and utility function to extract data from a microarray structure (`magetfield`).

- New functions for the Gene Ontology database — Import the Gene Ontology database from the Web (`geneont`), get a subset of the ontology (`getancestors`, `getdescendants`, `getrelatives`), parse Gene Ontology Annotated files (`goannotread`), and convert numbers to Gene Ontology IDs (`num2goid`).

- New demonstrations for the new Gene Ontology functions (`geneontologydemo`) and working with whole genomes (`biomemorymapdemo`).

Major Bug Fixes

To view major bug fixes made in R14SP3+ for the Bioinformatics Toolbox Version 2.2, use the Bug Reports interface on the MathWorks Web site.

Note If you are not already logged in to Access Login, when you link to the Bug Reports interface (see below), you will be prompted to log in or create an Access Login account.

After you are logged in, use this `Bug Reports` link. You will see the bug report for the Bioinformatics Toolbox. The report is sorted with fixed bugs listed first, and then open bugs.

If you are viewing these release notes in PDF form on the MathWorks Web site, you can refer to the HTML form of the release notes on the MathWorks Web site and use the link provided.

For bug fixes added prior to Release 14SP2, see "Major Bug Fixes" on page 6-2 in the Bioinformatics Toolbox Version 1.1.1 release notes.

Known Software and Documentation Problems

To view important open bugs in R14SP3+ for the Bioinformatics Toolbox Version 2.2, use the Bug Reports interface on the MathWorks Web site.

Note If you are not already logged in to Access Login, when you link to the Bug Reports interface (see below), you will be prompted to log in or create an Access Login account.

After you are logged in, use this Bug Reports link. You will see the bug report for the Bioinformatics Toolbox. The report is sorted with fixed bugs listed first, and then open bugs. You can select the Status column to list the open bugs first.

If you are viewing these release notes in PDF form on the MathWorks Web site, you can refer to the HTML form of the release notes on the MathWorks Web site and use the link provided.

Bioinformatics Toolbox 2.1.1 Release Notes

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 2.1.1 (CD Release R14SP3).

There are no new features with this release.

Major Bug Fixes

To view major bug fixes made in R14SP3 for the Bioinformatics Toolbox Version 2.1.1, use the Bug Reports interface on the MathWorks Web site.

Note If you are not already logged in to Access Login, when you link to the Bug Reports interface (see below), you will be prompted to log in or create an Access Login account.

After you are logged in, use this `Bug Reports` link. You will see the bug report for the Bioinformatics Toolbox. The report is sorted with fixed bugs listed first, and then open bugs.

If you are viewing these release notes in PDF form on the MathWorks Web site, you can refer to the HTML form of the release notes on the MathWorks Web site and use the link provided.

For bug fixes added prior to Release 14SP2, see "Major Bug Fixes" on page 6-2 in the Bioinformatics Toolbox Version 1.1.1 release notes.

Known Software and Documentation Problems

To view important open bugs in R14SP3 for the Bioinformatics Toolbox Version 2.1.1, use the Bug Reports interface on the MathWorks Web site.

Note If you are not already logged in to Access Login, when you link to the Bug Reports interface (see below), you will be prompted to log in or create an Access Login account.

After you are logged in, use this `Bug Reports` link. You will see the bug report for the Bioinformatics Toolbox. The report is sorted with fixed bugs listed first, and then open bugs. You can select the Status column to list the open bugs first.

If you are viewing these release notes in PDF form on the MathWorks Web site, you can refer to the HTML form of the release notes on the MathWorks Web site and use the link provided.

Bioinformatics Toolbox 2.1 Release Notes

- "New Features" on page 3-2
- "Major Bug Fixes" on page 3-3
- "Known Software Problems" on page 3-3

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 2.1 (Web Release R14SP2+).

Sequence Alignment

- `multialign` — New function. Align multiple sequences using progressive method with Distributed Computing Toolbox support.

- `multialignread` — New function. Read multiple sequence alignment file.

- `profalign` — New function. Align two profiles using Needleman-Wunsch global alignment

- `showalignment` — Updated to now show multiply aligned sequences.

- `seqpdist` — calculate pairwise distance between observations with Distributed Computing Toolbox support.

Sequence Statistics

- `codonbias` — New function. Calculate condon frequency for each amino acid in a DNA sequence.

- `cpgisland` — New function. Locate CpG islands in a DNA sequence.

Sequence Utilities

- `rebasecuts` — New function. Find restriction enzymes that cut a protein sequence.

- `seqtool` — New GUI for sequence analysis.

Phylogenetic Tree Functions

- `dnds`, `dndsml` — New functions. Estimate synonymous and nonsynonymous substitutions rates.

- `seqneighjoin` — New function. Neighbor-joining method for phylogenetic tree reconstruction.

Phylogenetic Tree Methods

- `getcanonical` — New method. Calculate the canonical form of a phylogenetic tree.

- `getnewwickstr`— New method. Create Newick formatted string.

- `reroot` — New method. Change the root of a phylogenetic tree.

- `subtree` — New method. Extract a subtree.

- `weights` — New method. Calculate weights for a phylogenetic tree.

Microarray Functions

- `probesetplot` — New function. Plots values for Affymetrix CHP file probe set.

Statistics Functions

- `rankfeatures` — Renamed function. The previous name was `sqtlfeatures`.

Major Bug Fixes

The Bioinformatics Toolbox Version 2.1 includes the following important bug fixes made since Version 2.0.1.

All Know Affyread Problems Corrected

In previous versions, there was a bug that worked with old libraries, but caused `affyread` to fail when reading CHP files with the new libraries. This bug is fixed.

Known Software Problems

The Bioinformatics Toolbox Version 2.1 includes the following known software and documentation problems.

Mouse Wheel with Seqtool

The mouse wheel does not work correctly with `seqtool`.

Updating Affymetrix Runtime Libraries

If you have a previously installed version of the Bioinformatics Toolbox (version 2.0.1 and before), you need to update the Affymetrix runtime library.

1 Run the installer `GdacFilesRuntimeInsatll-v4.exe` from the directory

```
<matlabroot>/toolbox/bioinfo/microarray/lib/
```

Updating Affymetrix Chip Libraries

To update Affymetrix libraries with CDF files, see the Affymetrix Web site:

```
http://www.affymetrix.com/supprort/technical/libraryfilesmain.affx
```

Enable Java Figures on Macintosh

In Bioinformatics Toolbox 2.1 the functions `seqlogo`, `seqshowwords`, `seqshoworfs`, and `showalignment` use Java based figures. Currently on the Macintosh, Java figures are not enabled by default. If you use these functions on a Macintosh, you should start MATLAB with

```
matlab -useJavaFigures
```

Some functionality does not work with modal windows. For example running a Bioinformatics Toolbox demo opens a blank quest dialog box. A work around is to minimize and maximize the quest dialog box to force X-windows to refresh it.

Bioinformatics Toolbox 2.0.1 Release Notes

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 2.0.1 (CD Release R14SP2).

Updated RBASE table

RBASE is the enzyme table that the function `restrict` uses to locate sequence patterns.

Expanded Bioperl Demo

Example of calling MATLAB from Perl scripts now includes several examples of passing various types of data (both directly and by variant variable) back and forth between Perl and a MATLAB Automation Server. To view the demo, type `bioperldemo`.

Major Bug Fixes

The Bioinformatics Toolbox Version 2.0.1 includes the following important bug fixes made since Version 2.0.

Zoom mode in msviewer is persistent

Before, the zoom mode was deselected after selecting part of an MS spectra. Now, the zoom mode remains selected.

New Tutorial Demonstrations

With the new tutorial demonstrations, you can

- Type the name of the demonstration in the MATLAB command window, and run the demonstrations step-by-step through the code.

- Select a demonstration from the Demo tab in the Help browser and view a printed format, view the code, or run the demonstration.

- Get a list of demonstrations and functions by typing `help bioinfo`, and then selecting a demonstration to run.

Known Software Problems

The Bioinformatics Toolbox Version 2.0.1 includes the following known software and documentation problems.

Enable Java Figures on Macintosh

In Bioinformatics Toolbox 2.0.1 the functions `seqlogo`, `seqshowwords`, `seqshoworfs`, and `showalignment` use Java based figures. Currently on the Macintosh, Java figures are not enabled by default. If you use these functions on a Macintosh, you should start MATLAB with

```
matlab -useJavaFigures
```

Some functionality does not work with modal windows. For example running a Bioinformatics Toolbox demo opens a blank quest dialog box. A work around is to minimize and maximize the quest dialog box to force X-windows to refresh it.

Possible Perl error message with bioperldemo

The demonstration `bioperldemo` was originally implemented with Bioperl version `1.2.3`. If you don't have a complete installation of Perl on your system, using the current release of Bioperl version `1.4` can cause an error message . This happens when `MW_BLAST.pl` in `bioperldemo` is used to generate fresh BLAST results rather then using the sample results included with the Bioinformatics Toolbox.

Warning message with Bioperl 1.4

In Bioperl version `1.4` there is a new function, `GuessSeqFormat.pm` that includes a `warnings.pm` dependency. A message tells you that Perl can't find `warnings.pm` in your Perl path. You can correct this error by commenting out the warnings dependency in `GuessSeqFormat.pm` or downloading the file `warnings.pm` to your Bioperl installation directory.

Bioinformatics Toolbox 2.0 Release Notes

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 2.0 (Web Release R14SP1+) .

Mass Spectrometry Data

New set of functions designed for preprocessing and classification of raw mass spectrometry data from SELDI-TOF and MALDI-TOF spectrometers.

- msresample — Resample with antialias filtering.

- msbackadj — Correct a baseline by estimation.

- msalign — Align a spectrum to a set of candidate peaks.
- msheatmap — Draw a heat map image for a set of spectra and check alignments.
- msnorm — Normalize a set of spectra.
- mslowess — Non-parametric smoothing using Lowess method.
- mssgolay — Least-squares polynomial smoothing.
- msviewer — Plot a spectrum or a set of spectra.

Graph Visualization

New set of functions to view relationships between data with interactive maps.

- biograph — Create a biograph object.
- dolayout — Calculate node and edge positions.
- getnodesbyid — Get handles to nodes.
- getedgesbynodeid — Get handles to edges.
- view — Render a graph in its viewer.
- getancestors — Find ancestors.
- getdescendants — Find descendants.

- getrelatives — Find neighbors.

Statistical Learning

New set of functions to classify data and identify features in the data.

- classpert — Evaluate the performance of a classifier.
- crossvalind — Cross-validation index generation.
- knnclassify — K-Nearest neighbor classifier.
- knnimpute — Impute missing data using the nearest neighbor method.
- randfeatures — Randomized subset feature selection.
- sqtlfeatures — Sequential forward feature selection. This function will be renamed to rankfeatures in version 2.1.
- svmclassify — Classify using a support vector machine classifier.
- svmtrain — Train a support vector machine classifier.

Sequence Analysis

New functions for analysis and visualization of multiple sequences.

- seqconsensus — Computes the consensus sequence for a set of sequences.
- seqlogo — Displays sequence logos for DNA and protein sequences.

- seqprofile — Computes the sequence profile of a multiple alignment.

Protein Analysis

- pdbplot — Plots 3D backbone structure of proteins in a PDB file.

Microarray Analysis

- quantilenorm — Quantile normalization.

New set of functions for working with Affymetrix GeneChip data sets.

- probelibraryinfo — Get library information for a probe.

- probesetlink — Show probe set information from NetAffx.

- probesetlookup — Get gene information for a probe set.

- probesetplot — Plot probe set values.

- probesetvalues — Get probe set values from CEL and CDF information.

- manorm — Normalization by scaling and centering replaces the functions
  `mamadnorm` and `mameannorm`.

Note If you use `mamadnorm` or `mameannorm` in any of your personal m-files,
Please update your files with the new function `manorm`. These functions are
now obsolete and may be removed from future releases of the Bioinformatics
Toolbox. .

Enhanced Features

Bioinformatics Toolbox 2.0 has a number enhancements added to functions
in previous releases.

Web Connectivity

- getgenbank — Now returns CDS information for a gene in a structure
  allowing direct access to the transcribed sequence.

Sequence Analysis

- palindromes— Now allows for gaps in the palindrome.

- seqshoworfs, seqshowords, showalignment — Now displays the results in a
  Figure window. [This may cause problems on the Mac]

In Bioinformatics Toolbox 2.0 the functions `seqlogo`, `seqshowwords`,
`seqshoworfs`, and `showalignment` use Java based figures. Currently on the
Macintosh, Java figures are not enabled by default. If you use these functions
on a Macintosh, you should start MATLAB with

```
matlab -useJavaFigures
```

Microarray Data Analysis

- affyread — Output structures have changed slightly. Some redundant fields have been removed from CDF and CHP structure. GIN database files are now supported. Version 4 of the Affymetrix GDAC File Access Runtime Libraries is provided.

- geosoftread — Now supports Gene Expression Omnibus Database records (GDS files).

- maimage — Now supports Affymetrix CEL data.

- maboxplot — Now supports Affymetrix CHP data.

Affymetrix and Gene Chip are registered trademarks of Affymetrix Inc.

New Tutorial Demonstrations

Bioinformatics Toolbox 2.0 includes several new demonstrations. You can

- Type the name of the demonstration in the MATLAB command window, and run the demonstrations step-by-step through the code.

- Select a demonstration from the Demo tab in the Help browser and view a printed format, view the code, or run the demonstration.

- Get a list of demonstrations and functions by typing help bioinfo, and then selecting a demonstration to run.

affydemo — How to work with Affymetrix Gene Chip data.

biovbscript — How to connect to MATLAB from a Web page using VBScript.

rasmolddedemo — How to connect to RASMOL using DDE.

Three new demos for preprocessing raw mass spectrometry data and searching for patterns within a data set.

1 mspreprodemo — Pre-processing raw mass spectrometry data.

2 `cancerdetectdemo` — Identifying significant features and classifying protein profile.

3 `biodistcompdemo` — Batch processing of spectra using distributed computing.

Known Software Problems

There are two known problems when working with Perl and the Bioinformatic Toolbox 2.0.

Possible Perl error message with bioperldemo

The demonstration `bioperldemo` was originally implemented with Bioperl version 1.2.3. If you don't have a complete installation of Perl on your system, using the current release of Bioperl version 1.4 can cause an error message . This happens when `MW_BLAST.pl` in `bioperldemo` is used to generate fresh BLAST results rather then using the sample results included with the Bioinformatics Toolbox.

Warning message with Bioperl 1.4

In Bioperl version 1.4 there is a new function, `GuessSeqFormat.pm` that includes a `warnings.pm` dependency. A message tells you that Perl can't find `warnings.pm` in your Perl path. You can correct this error by commenting out the warnings dependency in `GuessSeqFormat.pm` or downloading the file `warnings.pm` and copying it in your Bioperl installation directory.

Bioinformatics Toolbox 1.1.1 Release Notes

- "New Features" on page 6-2
- "Major Bug Fixes" on page 6-2

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 1.1.1 (CD Release R14SP1).

There are no new features in this release.

Major Bug Fixes

The Bioinformatics Toolbox 1.1.1 includes several bug fixes made after the version 1.1 release.

- "Change to the data structure returned by affyread for CDF files." on page 6-2
- "Changes to aminolookup.m varargin" on page 6-2
- "Possible Perl error message with Bioperldemo." on page 6-3
- "hmmalign changes to match HMMER scores" on page 6-3

Change to the data structure returned by affyread for CDF files.

When loading CDF files, the data structure returned by the function `affyread` has changed. The new structure no longer contains redundant fields. In earlier versions of the toolbox, the X and Y coordinate information for certain probes was not read. This is now corrected and all coordinate information will be read.

Changes to aminolookup.m varargin

To create a consistent format between the functions `aminolookup` and `baselookup`, 'letter' was changes to 'code' for the 1-letter abbreviation, and 'code' was changed to 'abbreviation' for the 3-letter abbreviation for `aminolookup`.

Possible Perl error message with Bioperldemo.

Bioperldemo was originally implemented with Bioperl v1.2.3. Using the current release of Bioperl (v1.4) can cause an error message if you don't have a complete installation of Perl on your system. This happens when MW_BLAST.pl in Bioperldemo is used to generate fresh BLAST results rather than using the sample results included with the Bioinformatics Toolbox. The error message is:

```
??? Error using ==> perl System error: Can't locate warnings.pm in @INC
```

In Bioperl v1.4 there is a new function GuessSeqFormat.pm which includes a 'warnings.pm' dependency. This message is simply telling you that Perl can't find 'warnings.pm' in your Perl path. You can correct this error by commenting out the warnings dependency in GuessSeqFormat, or downloading 'warnings.pm' and dropping it in your Bioperl installation directory.

hmmalign changes to match HMMER scores

Initializing the matrix now uses log odds ratio probability instead of straight probability.

All hmmprof related functions were updated to accept models with NullX and LoopX transition probabilities.

- Null transition probabilities can be loaded from pfam (pfamhmmread) and used to score the model appropriately. A new flag, in hmmprofalign, selects the use of null transitions. By default the NullX probabilities are not considered.

- Loop transition probabilities are loaded from the Pfam database, however they are not considered for scoring during alignment. Currently, hmmprofalign does not perform tandem repeat alignments.

A new flag was added to hmmprofalign to activate/deactivate the scoring of the flanks. By default flanks are not scored, however, in order to reproduce HMMER 2.2g results, flanking scores need to be included when this flag is selected.

A third output was added to hmmprofalign to easily recover the domain indexes of the alignment, the output follows the same idea as the output

indexed in `hmmpromerge`. In both cases, Null pointer are now defined by NaN instead of Zero.

Bioinformatics Toolbox 1.1 Release Notes

New Features

This section summarizes the new features and enhancements introduced in the Bioinformatics Toolbox Version 1.1 (CD Release R14).

- "Phylogenetic Analysis Functions" on page 7-2
- "Phylogenetic Tree Object and Methods" on page 7-3
- "Hidden Markov Model (HMM) Profiles" on page 7-3
- "aminolookup, baselookup" on page 7-4
- "blastncbi, blastread, getblast" on page 7-4
- "clustergram" on page 7-4
- "imageneread" on page 7-4
- "mapcaplot" on page 7-4
- "isoelectric" on page 7-4
- "seqdisp" on page 7-4
- "seqdotplot" on page 7-4
- "seqmatch" on page 7-5
- "New Demos" on page 7-5

Phylogenetic Analysis Functions

New functions for phylogenetic tree creation and analysis.

- phytreeread — Read a Newick formatted tree file into the MATLAB workspace and return a phytree object with data from the file. Data in the file uses the Newick (New Hampshire) format for describing trees.

- phytreewrite — Copy the contents of a phytree object from the MATLAB workspace to a file.

- phytreetool — Interactive GUI that allows you to view, edit, and explore phylogenetic tree data. This GUI allows branch pruning, reordering, renaming, and distance exploring. It can also open or save Newick formatted files.

- seqlinkage — Construct a phylogenetic tree from pairwise distances.

- seqpdist — Calculate the pairwise distance between biological sequences.

Phylogenetic Tree Object and Methods

New object for manipulating phylogenetic tree data.

- phytree — Function to create a phytree object.

- get— Get property values from a phytree object

- getbyname — Get node names from a phytree object.

- pdist — Calculate the patristic distances between pairs of leaf nodes.

- plot — Draw a phylogenetic tree object in a MATLAB figure window as a phylogram, cladogram, or radial tree.

- prune — Remove nodes from a phylogenetic tree.

- select — Select branches and leaves from a phylogenetic tree using a specified criteria.

- view — Opens a phylogenetic tree in a phytreetool window.

Hidden Markov Model (HMM) Profiles

Updated Hidden Markov Model profile functions

- The model structure, that HMM functions use, now includes loop and null transition probabilities. You can read null and loop probabilities from PFAM files using pfamhmmread and from PFAM web databases using gethmmprof.

- When the function hmmprofstruct builds an HMM model, the loop and null transition probabilities default to pre-defined values. If necessary, you can later modify the probabilities using the same function.

- hmmprofalign includes two new properties to control the scoring of flanking states and null transition probabilities. In addition, a third output argument with indices pointing to the respective symbols of the query sequence was added.

aminolookup, baselookup

Updated functions (`aminolookup`, `baselookup`) to get IUB/UPAC character codes, integer codes, and names for nucleotides and amino acids.

blastncbi, blastread, getblast

New functions (`blastncbi`, `blastread`, `getblast`) BLAST sequences and view results from within MATLAB.

clustergram

Updated function (`clustergram`) to do two way bi-clustering.

imageneread

New function (`imageneread`) to read microarray data from an ImaGene Results file. Other functions read microarray data from Affymetrix GeneChip files (`affyread`) and GenePix Results (GPR) files (`gpread`).

mapcaplot

New function (`mapcaplot`) to create a Principal Component plot of expression profile data

isoelectric

New function (`isoelectric`) for protein analysis. Estimate the isoelectric point (the pH at which the protein has a net charge of zero) for an amino acid sequence and estimate the charge for a given pH.

seqdisp

New function (`seqdisp`) formats sequence output for easy viewing.

seqdotplot

Enhanced function (`seqdotplot`) now returns a second output (the matrix of matches as a sparse matrix).

seqmatch

New function (`seqmatch`) to find matches for every string in a library.

New Demos

- **Bicluster demo** — Demonstrates some of the options of the clustergram function.

- **Bioperl demo** — Illustrates the interoperability between MATLAB and Bioperl - passing arguments from MATLAB to Perl scripts and pulling BLAST search data back to MATLAB

- **Phytree demo for Hominidae species**— A phylogenetic tree is constructed from mtDNA sequences for the Hominidae taxa (also known as pongidae). This family embraces the gorillas, chimpanzees, orangutans and the humans.

- **Phytree demo for HIV/SIV** — Analyzes the reconstruction of phylogenetic trees from infected HIV/SIV organisms.

Major Bug Fixes

The Bioinformatics Toolbox 1.1 includes several bug fixes made after the version 1.0 release.

Corrected Emission Probability

For Hidden Markov Model profiles, the emission probability for the first symbol in the profile is now correct. Previously the emission probability did not include null transition probabilities. Ignoring the null transition probabilities caused a small difference in the answer.

Platform Limitations for HP and IBM

HP-UX Help Browser Limitations

When opening the MATLAB Help browser or a function that users the browser, to display results, a number of exceptions are written to the

MATLAB Command Window. These exculpations do not affect viewing information in the browser window.

However, in the left pane, the text labels for the tabs are missing. These tabs are, from left t right, **Contents**, **Index**, **Search**, **Demos**, and **Favorites**.

To view documentation for the Bioinformatics Toolbox, open a Web browser (for example, IE or Netscape), and enter the address.

```
http://www.mathworks.com/access/helpdesk/help/toolbox/bioinfo/
```

Bioinformatics Toolbox 1.0 Release Notes

- "Introduction" on page 8-2
- "Features" on page 8-2

Introduction

The Bioinformatics Toolbox Version 1.0 (Web Release R13 SP1+) extends MATLAB with basic sequence analysis and gene expression analysis functions. The Bioinformatics Toolbox is a collection of tools built on the MATLAB numeric computing environment. The toolbox supports a wide range of common sequence analysis and expression analysis tasks, from accessing web-based databases, to sequence alignment, to microarray normalization and visualization.

The Bioinformatics Toolbox is dependent upon many functions from the Statistics Toolbox including some functions only available in the latest version of the Statistics Toolbox, 4.1. We recommend that you install the latest version of the Statistics Toolbox before running the Bioinformatics Toolbox.

Features

This section introduces the features for the Bioinformatics Toolbox 1.0. The Bioinformatics Toolbox has more than 100 functions implemented using M-files. For a complete list of functions, in the MATLAB Command window., type

```
help bioinfo
```

Data I/O
The toolbox provides functions to directly access many standard Web-based databases such as GenBank, EMBL, PIR, and PDB. There are also functions to read many standard file formats, including FASTA and PDB. For microarray data, there are functions to read Affymetrix, GenePix, SPOT format data, and a function to access data directly from the NCBI Gene Expression Omnibus Web site.

Sequence Alignment
The toolbox has functions for pairwise sequence alignment and for hidden Markov model-based sequence profile alignment, including efficient MATLAB implementations of the Needleman-Wunsch and Smith-Waterman algorithms. In addition to the alignment functions there are several tools for visualizing

sequence alignments. The toolbox provides many standard scoring matrices, including the PAM and BLOSUM families.

Sequence Utilities and Statistics

The toolbox contains many functions for working with sequences. There are functions for converting DNA sequences to RNA or amino acid sequences; there are functions that report various statistics about sequences, and functions to search for patterns within the sequence; there are functions for creating random sequences, and there are functions to perform in-silico digestion of sequences with restriction enzymes and proteases.

Microarray Normalization and Visualization

The toolbox contains a number of functions for normalizing microarray data including lowess normalization, global mean normalization, and MAD normalization. The toolbox provides several functions for visualizing microarray data, including spatial heat maps, box plots, loglog, and I-R plots. The toolbox also uses functions from the Statistics Toolbox to perform cluster analysis and to visualize the results.

Protein Structure Analysis

In addition to standard sequence analysis functions, there is also a graphical user interface (GUI), `proteinplot`, for visualizing properties of protein sequences.

Tutorial Demonstrations

There are also several tutorial examples that demonstrate how to use the functions in the toolbox. These tutorials would be a good place to start using the toolbox.