

**GCG<sup>®</sup> Wisconsin Package<sup>™</sup>**

# ***User Release Notes***

**What's New in Version 10.3**

## **Version 10.3**

**Accelrys  
9685 Scranton Road  
San Diego, CA 92121**

Phone: (800) 756-4674  
Fax: (858) 799-5102  
E-mail: [support-us@accelrys.com](mailto:support-us@accelrys.com) or [support-eu@accelrys.com](mailto:support-eu@accelrys.com)  
WWW: <http://www.accelrys.com/bio>



## ***Licenses and Trademarks***

All trademarks or registered trademarks are the property of their respective holders and are used in this document for identification purposes only.

## ***Copyright***

© 2001, Accelrys Inc. A subsidiary of Pharmacoepia. All rights reserved.

Printed in the United States of America

December 31, 2001

---

## **Contents**

<b>New Programs</b> .....	<b>5</b>
<b>Programs Not Available for This Release</b> .....	<b>6</b>
<b>Program Enhancements</b> .....	<b>7</b>
<b>Program Bug Fixes</b> .....	<b>8</b>
<b>SeqLab Bug Fixes</b> .....	<b>11</b>
<b>Package-Wide Bug Fix (Linux Only)</b> .....	<b>13</b>



---

## **New Programs**

The programs listed below are new to Version 10.3 of the Wisconsin Package.

### **Database Searching**

#### **PSIBLAST**

PSIBLAST, or Position-Specific Iterated BLAST, uses the methods described in Altschul, et al. *Nucleic Acids Res.* 25(17): 3389-3402 (1997) and Schaffer, et al. *Nucleic Acids Res.* 29(14): 2994-3005 (2001) to search for similarities between protein query sequences and all the sequences in one or more protein databases.

PSIBLAST uses position-specific scoring matrices (PSSMs) to score matches between query and database sequences, in contrast to BLAST, which uses pre-defined scoring matrices such as BLOSUM62. PSIBLAST may be more sensitive than BLAST, meaning that it may find distantly related sequences not found with a BLAST search.

### **Protein Analysis**

#### **TransMem**

TransMem builds on the method of Sonnhammer et al. (*Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology*, 175-182 (1998)) to predict likely transmembrane helices in one or more input proteins. The method is based upon a Hidden Markov Model (HMM) that has been trained on a set of membrane proteins with helical membrane spanning regions.

---

## **Programs Not Available for This Release**

The programs listed below are not available for Version 10.3 of the Wisconsin Package on the Linux platform.

### **Evolution**

#### **PAUPDisplay**

Because the version of the PAUP programs in Wisconsin Package 10.3 is not compatible with Linux, all programs that call the PAUP package within the Wisconsin Package have been disabled on the Linux platform. We hope to be able to provide these programs on all platforms with the next release of the package.

#### **PAUPSearch**

Because the version of the PAUP programs in Wisconsin Package 10.2 is not compatible with Linux, all programs that call the PAUP package within the Wisconsin Package have been disabled on the Linux platform. We hope to be able to provide these programs on all platforms with the next release of the package.

---

## Program Enhancements

### Database Searching

#### NetFetch

**Enhancement:** The mechanism used by NetFetch to connect to NCBI's Entrez server has been modified such that the parameter **-URL** has no effect. In addition, all input sequences must now be of the same type (i.e., either nucleotide or protein).

### Database Utilities

#### GCGToBLAST

**Enhancement:** GCGToBLAST now uses the new parameter **-PARSEseqid** by default. Databases formatted with this parameter will consist of two additional index files, which permit BLAST and PSIBLAST to correctly display sequence names in any of the multiple alignment output formats. This behavior can be suppressed by specifying **-NOPARSEseqid**.

### Gene Finding and Pattern Recognition

#### FindPatterns

**Enhancement:** A new command line parameter, **-LISTfile**, can now be used to produce list file output, in addition to the normal output file. This parameter replaces **-NAMEs**, which will be removed in a future release. To ensure compatibility with earlier versions, **-LISTfile** has no effect if **-NAMEs** is also specified.

#### Motifs

**Enhancement:** A new command line parameter, **-LISTfile**, can now be used to produce list file output, in addition to the normal output file. This parameter replaces **-NAMEs**, which will be removed in a future release. To ensure compatibility with earlier versions, **-LISTfile** has no effect if **-NAMEs** is also specified.

---

## Program Bug Fixes

### Database Searching

#### NetFetch

**Known Bug:** If you retrieve a sequence larger than 350,000 residues using NetFetch, you get one RSF file containing the entire sequence, which is unusable with other Wisconsin Package programs. Because Breakup is not intended to work on RSF files, you cannot produce GCG sequence files of proper size from the large RSF file.

**Workaround:** When retrieving a long sequence using NetFetch, add the **-RAW** parameter to the command line. This will additionally produce a GenBank sequence file containing the sequence. Now you can run FromGenBank to produce several small single source files of your sequence.

### Gene Finding and Pattern Recognition

#### CodonPreference

**Problem:** The codon preference statistics for each window determined by CodonPreference were calculated using one too many codon preference statistic values in the sum. That is, instead of using the formula

$$P = \exp(\text{sum over window } \ln(p)) / \text{window}$$

P was calculated using

$$P = \exp(\text{sum over window}+1 \ln(p)) / \text{window}.$$

As a result, if you asked for the window preference statistics with **-PWINDOW** =3, the sum over the codon preference statistics of the 3 codons in a window plus the next codon were used instead.

**Update:** CodonPreference now calculates the correct codon preference statistics for each window.

#### FindPatterns

**Problem:** In RSF output, matches longer than 350 were always reported as features that were 350 positions in length.

**Update:** Matches are now correctly reported in RSF output.

## Importing and Exporting

### BreakUp

**Problem:** The description for BreakUp in the Program Manual identifies several file formats as possible input files, which are not actually allowed. The program accepts single sequence files and plain ASCII files, but does not take as input files RSF, MSF, GCG list format or database references.

**Workaround:** For example, you might encounter large sequences in RSF format, which have been retrieved via NetFetch. For a workaround on how to produce valid GCG files for these sequences, see the workaround for NetFetch under Database Searching above.

## Pairwise Comparison

### FrameAlign

**Problem:** If you ran FrameAlign on the Irix or Solaris platform with the parameters **-GLobal** and **-ENDWeight** set, the results would contain incorrect alignments and quality scores.

**Update:** You can now use both **-GLobal** and **-ENDWeight** and the program will produce the correct results on all supported platforms.

## Primer Selection

### Prime

**Problem:** If you ran Prime and selected a range of the input sequence to be analyzed, the resulting RSF file would contain only the selected range of the input sequence, but the feature positions correlated to the entire input sequence. Therefore, the features listed in the RSF file would not match onto the sequence in the RSF file.

**Update:** The RSF file now contains the entire input sequence and the feature positions are correct.

## Utilities

### PrositateToGCG

**Problem:** If you ran PrositeToGCG on the Solaris platform, the output files would be incomplete and have bad file names.

**Update:** PrositeToGCG now produces a complete set of output files with valid names.

## Translation

### Translate

**Known Bug:** When run interactively, Translate ignores any **-BEGIN** and **-END** ranges specified on the command line.

**Workaround:** Suppress interactive mode by including the **-Default** parameter on the command.

---

## SeqLab Bug Fixes

**Problem:** Deletion of tildes from the 5' end of a sequence in the SeqLab editor caused the edited sequence in the trace viewer to shift relative to the trace and the raw sequence. Reinserting tildes did not shift the edited sequence back into register.

**Update:** SeqLab now maintains proper registration between the original base calls and the trace peaks in the presence of alignment offsets.

### Editor Mode (LINUX Only)

**Known Bug:** Using the Control key in combination with the left mouse button to select individual entries or several non-contiguous sequence residue stretches does not work correctly.

**Workaround:** Use the Control key in combination with the right mouse button, or use the Select by Name or Select Range buttons under the Edit menu to select individual entries or several non-contiguous sequence residue stretches.

**Known Bug:** When switching among Insert, Check, and Overstrike modes, typing in symbols from the keyboard does not work, even with the correct protection set.

**Workaround:** After switching to a different mode, click the Protect button to check that the protections are set correctly, then click the OK button. The selected editing mode should work.

### Programs Run Through SeqLab

#### MEME

**Problem:** If you selected the “Exactly one occurrence of each motif in each sequence” option when analyzing a protein sequence, the **-TWOSTrands** parameter was automatically added to the command line. Then, when trying to run the program, it would fail with the log message “You must use default DNA alphabet if using complementary strands!”

**Update:** Now when you select “Exactly one occurrence of each motif in each sequence” with a protein sequence, MEME runs correctly, without adding the **-TWOSTrands** parameter.

#### ProfileMake

**Problem:** If you ran ProfileMake and selected the “linear weighting” option, the program would run using the default “exponential weighting” option instead.

**Update:** Now the “linear weighting” option works correctly.

## **StemLoop**

**Problem:** The main window for StemLoop would select “see the stems” by default, and would also provide an option to “see the stem coordinates.” Both of these options would display the output of the StemLoop run in the Job Manager window.

**Update:** The StemLoop window now contains the options “file the stems” and “file the stems as points for DotPlot,” which produce files containing the results in the Output Manager.

---

## **Package-Wide Bug Fix (Linux Only)**

**Known bug:** If you run one of the multi-threaded programs that report CPU times in the summary (such as the FastA family and FrameSearch), those times will be incorrect.

**Workaround:** You can still determine the correct process time by using the “time” system switch. For example, use the following command line to start FrameSearch and get the correct process time:

```
$time framesearch
```

After entering this command, a line displays before the next system prompt. The first value in this line provides the process time in seconds.

