BioNumerics®

MICROBIAL DATA ANALYSIS SOFTWARE

Sequence Translation Tools plugin

PLUGINS

VERSION 7.6

www.applied-maths.com

Contents

NOTES

SUPPORT BY APPLIED MATHS

While the best efforts have been made in preparing this manuscript, no liability is assumed by the authors with respect to the use of the information provided.

Applied Maths will provide support to research laboratories in developing new and highly specialized applications, as well as to diagnostic laboratories where speed, efficiency and continuity are of primary importance. Our software thanks its current status for a part to the response of many customers worldwide. Please contact us if you have any problems or questions concerning the use of BioNumerics®, or suggestions for improvement, refinement or extension of the software to your specific applications:

Applied Maths NV
Keistraat 120
9830 Sint-Martens-Latem
Belgium
PHONE: +32 9 2222 100
FAX: +32 9 2222 102
E-MAIL: info@applied-maths.com
URL: http://www.applied-maths.com

Applied Maths, Inc.
11940 Jollyville Road, Suite 115N
Austin, Texas 78759
U.S.A.
PHONE: +1 512-482-9700
FAX: +1 512-482-9708
E-MAIL: info-US@applied-maths.com

LIMITATIONS ON USE

The BioNumerics® software, its plugin tools and their accompanying guides are subject to the terms and conditions outlined in the License Agreement. The support, entitlement to upgrades and the right to use the software automatically terminate if the user fails to comply with any of the statements of the License Agreement. No part of this guide may be reproduced by any means without prior written permission of the authors.

BioNumerics® uses following third-party software tools and libraries:

- The Python® 2.7.4 release from the Python Software Foundation (http://www.python.org/).
- A library for XML input and output from the Apache Software Foundation (http://www.apache.org).
- NCBI toolkit version 2.2.10 (http://www.ncbi.nlm.nih.gov/BLAST/).
- The Boost c++ libraries (http://www.boost.org/).
- Samtools for interacting with SAM / BAM files (http://www.htslib.org/download/)
- The 7-Zip command line version (7za.exe) from 7-Zip, copyright 1999-2010 Igor Pavlov. http://www.7-zip.org/
- Velvet for Windows, source code can be downloaded from http://www.applied-maths.com/download/open-source.
- Ray for Windows, source code can be downloaded from http://www.applied-maths.com/download/open-source.
- Mothur for Windows, source code can be downloaded from http://www.applied-maths.com/download/open-source.
- Cairo 2D graphics library version 1.12.14 (http://cairographics.org/).
- Crypto++ Library version 5.5.2 (http://www.cryptopp.com/).
- libSVM library for Support Vector Machines (http://www.csie.ntu.edu.tw/~cjlin/libsvm/).
- SQLite version 3.7.17 (http://www.sqlite.org/).
- Gecko engine version 21 (https://developer.mozilla.org/en-US/docs/Mozilla/Gecko).
- pymzML Python® module for high throughput bioinformatics on mass spectrometry data (https://github.com/pymzml/pymzML).
- Numpy Python® library version 1.8.1 (http://www.numpy.org/).
- BioPython Python® library version 1.64 (http://www.biopython.org/).
- PIL Python library® version 1.1.7 (http://www.pythonware.com/products/pil/).
- The SPAdes genome assembler version 3.7.1 (http://bioinf.spbau.ru/spades).

Chapter 1

Starting and setting up BioNumerics

1.1 Startup program

When BioNumerics is launched from the Windows start panel or when the BioNumerics shortcut ( ) on your computer's desktop is double-clicked, the **Startup program** is run. This program shows the *BioNumerics Startup* window (see Figure 1.1).

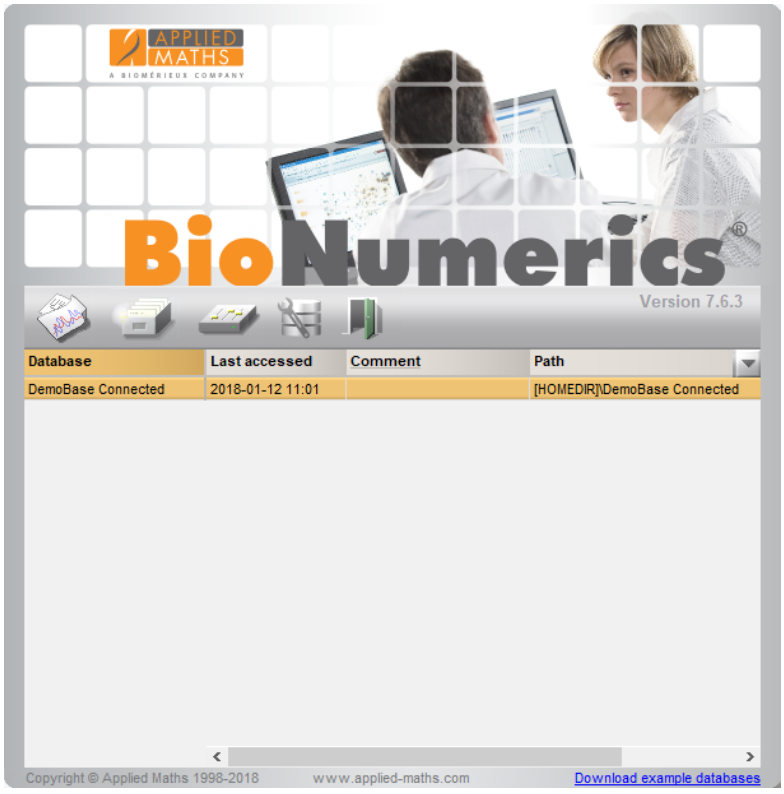

Figure 1.1: The *BioNumerics Startup* window.

A new BioNumerics database is created from the Startup program by pressing the button.

An existing database is opened in BioNumerics with or by simply double-clicking on a database name in the list.

1.2 Installing the Sequence Translation Tools plugin

If a database is opened for the first time, the *Plugins* dialog box will appear by default (see Figure 1.2).

If the database has already been opened previously, the *Plugins* dialog box can be called from the *Main* window by selecting ***File*** > ***Install / remove plugins...*** ( ).

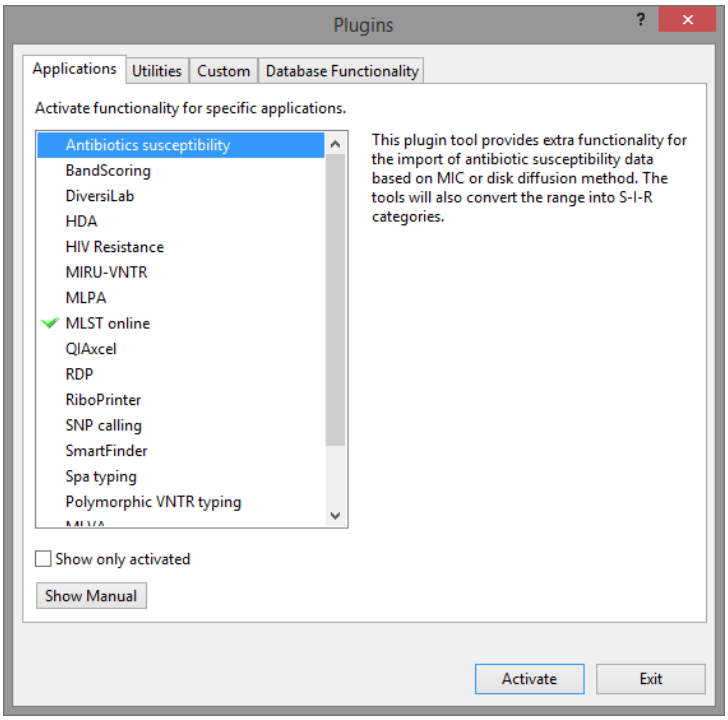

Figure 1.2: The *Plugins* dialog box.

When a particular plugin is selected from the list of plugins, a short description appears in the right panel.

A selected plugin can be installed with the <***Activate***> button. The software will ask for confirmation before installation. Some plugins depend on functionality offered by specific BioNumerics modules. If a required module is missing, the plugin cannot be installed and an error message will be generated.

Once a plugin is installed, it is marked with a green V-sign. It can be removed again with the <***Deactivate***> button.

If the selected plugin is documented, pressing <***Show Manual***> will open its manual in the *Help* window.

2.1 To install the *Sequence Translation Tools plugin* in your database, select the *Utilities tab* and select the *Sequence translation tools plugin* from the list of plugins.

2.2 Press the <***Activate***> button, confirm the installation of the plugin and close the *Plugins* dialog box.

2.3 Close and reopen the database to activate the features of the *Sequence Translation Tools plugin*.

The *Sequence translation tools plugin* installs menu items in the *Main* window and *Comparison* window.

Chapter 2

Sequence translation tools functionality

2.1 Translate selected sequences

The **Translate selected sequences tool** translates nucleic acid sequences into amino acid sequences for a selected set of entries.

1.1 In the *Main* window, select one or more entries linked to a nucleotide sequence experiment (use the **Ctrl-** and **Shift**-keys to select entries in the database).

1.2 The *Translate the selected sequences* dialog box is called with ***Analysis*** > ***Sequence types*** > ***Translate selected sequences*** (see Figure 2.1).

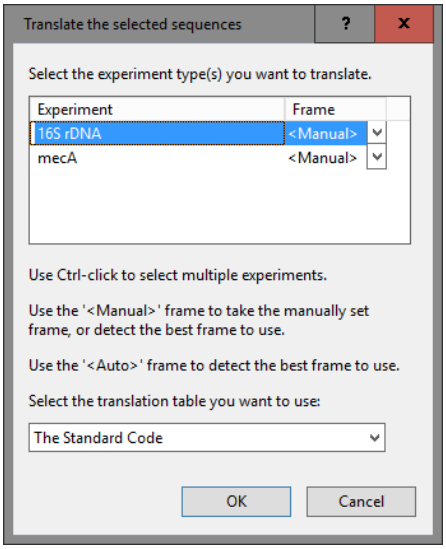

Figure 2.1: The *Translate the selected sequences* dialog box.

All nucleotide sequence types defined in the database are displayed in the upper panel of the *Translate the selected sequences* dialog box. One or more sequence types can be selected using the **Ctrl-** key.

A translation frame can be specified for each sequence type separately using the pull-down menu in the ***Frame*** column: ***Frame 1***, ***Frame 2*** or ***Frame 3***. With the ***Manual*** option the frame that was manually set (see 2.2) is used if no better frame is detected. Using the ***Auto*** frame option the best frame to use is screened by the plugin.

From the list under ***Select the translation table***, the different translation tables, corresponding to variants of

the standard genetic code, can be selected. By default, ***The Standard Code*** is selected.

1.3 To translate the sequences press the <***OK***> button.

A new sequence experiment is added to the database with the name of the corresponding nucleic acid sequence type, plus the suffix "_TRANSL" and contains the translated sequences. The sequences are translated based on the specified reading frame.

1.4 Click on the colored dot in the *Experiment presence* panel corresponding to the translated sequence type to open the *Sequence editor* window (see Figure 2.2).

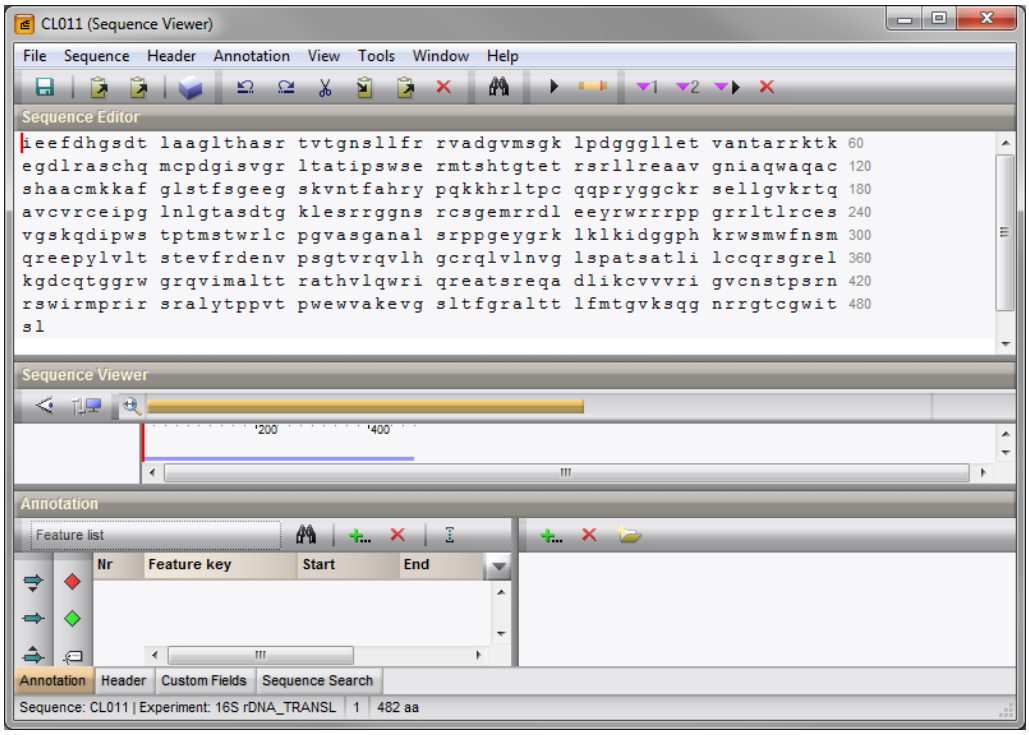

Figure 2.2: The *Sequence editor* window.

1.5 Select ***File*** > ***Exit*** to close the *Sequence editor* window.

1.6 Hold the **Shift**-button and click on the colored dot in the *Database entries* panel to open the sequence card (see Figure 2.3).

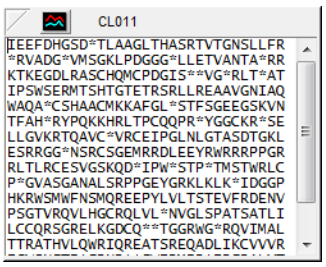

Figure 2.3: The Sequence card.

1.7 Click on the triangle in the top left corner of the experiment card to close it.

2.2 Set reading frame

With the **Set reading frame tool**, the reading frame can be set for a particular entry.

2.1 Double-click on an entry to open the *Entry* window and select ***Sequence*** > ***Set frames***. This opens the *Sequence frames* dialog box (see Figure 2.4).

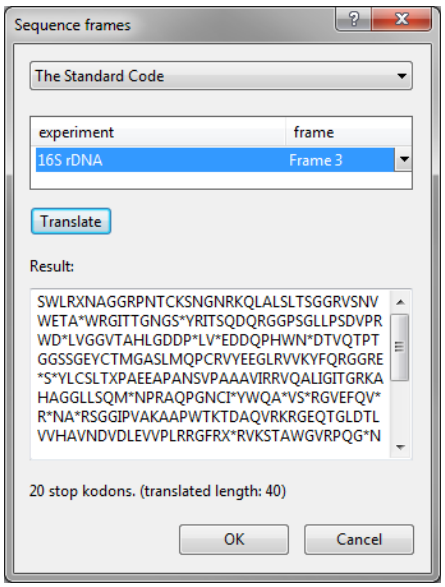

Figure 2.4: The *Sequence frames* dialog box.

In the *Sequence frames* dialog box a translation table can be specified for the selected entry.

A translation frame can be specified for each sequence type separately: select the ***Experiment*** in the list and set the frame (***Frame 1***, ***Frame 2***, ***Frame 3***) using the pull-down menu in the ***Frame*** column.

To see a preview of the translation in the panel below press the <***Translate***> button. The number of stop codons in the translated sequence is reported, together with the number of nucleotides that appear before the first occurrence of a stop codon.

2.2 After having specified the translation settings, close the *Entry* window and run ***Analysis*** > ***Sequence types*** > ***Translate selected sequences*** again to apply the changes.

2.3 Map protein alignment on DNA sequences

The **Map protein alignment on DNA sequences tool** allows an amino acid alignment to be imposed on its corresponding nucleic acid sequences. Its proper function requires having previously used the **Translate selected sequences** tool (see 2.1).

3.1 In the *Main* window select one or more entries linked to an amino acid sequence experiment (use the **Ctrl-** and **Shift**-keys to select entries in the database).

3.2 Create a new comparison by highlighting the *Comparisons* panel in the *Main* window and selecting ***Edit*** > ***Create new object...*** ( ).

3.3 Display the nucleic acid and amino acid sequences in the *Experiment data* panel by clicking on the eye button ( ) next to the experiments in the *Experiments* panel.

3.4 Align the amino acid sequences and select ***Sequence*** > ***Map protein alignment on DNA sequences***.

The corresponding nucleic acid sequences are aligned in the *Experiment data* panel.

2.4 Invert-complement selected sequences

The **Invert-complement tool** creates the invert-complement sequences for all selected entries and saves the resulting sequences in a sequence type experiment in the BioNumerics database.

First, a new nucleic acid sequence type needs to be created for the storage of the invert-complement sequences:

4.1 In the *Main* window, highlight the *Experiment types* panel and select ***Edit*** > ***Create new object...*** ( ). Highlight "Sequence type" and press <***OK***>.

4.2 Enter a name for the sequence type, press <***Next***>, make sure ***Nucleic acid sequences*** is checked and press <***Finish***>.

4.3 In the *Main* window select one or more entries linked to a nucleotide sequence experiment (use the **Ctrl-** and **Shift**-keys to select entries in the database).

4.4 To launch the Invert-complement tool, select ***Analysis*** > ***Sequence types*** > ***Invert + complement selected sequences***. The *Invert and complement the selected sequences* dialog box is displayed (see Figure 2.5).

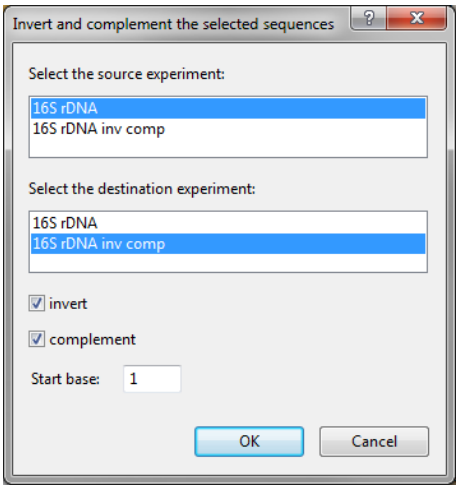

Figure 2.5: The *Invert and complement the selected sequences* dialog box.

All nucleic acid sequence types that are present in the BioNumerics database are listed.

Select the ***Source sequence type*** from the upper list and select the ***Destination sequence type*** from the lower list.

The plugin can create the ***Invert*** and/or ***Complement*** of the selected sequences.

By default, all nucleotides of the resulting sequence will be stored in the destination sequence experiment (***Start base*** = 1). Increase this number if you wish to remove nucleotides at the beginning of the sequences.

4.5 Press the <***OK***> button.

The resulting sequences are saved in the destination sequence type.

2.5 Translate and export sequences

The **Export translated sequences in FASTA format** translates nucleic acid sequences into amino acid sequences for a selected set of entries and exports the translated sequences in FASTA format.

5.1 In the *Database entries* panel of the *Main* window, select the entries to export. A single entry can be selected by holding the **Ctrl**-key and left-clicking (**CTRL+click**). Check boxes for selected entries are indicated as ☑. In order to select a group of entries, hold the **Shift**-key and click on another entry. All the entries in the database can be selected using the keyboard shortcut **Ctrl+A** or with ***Edit*** > ***Select all***.

5.2 Select ***File*** > ***Export...*** to call the *Export* dialog box (see Figure 2.6).

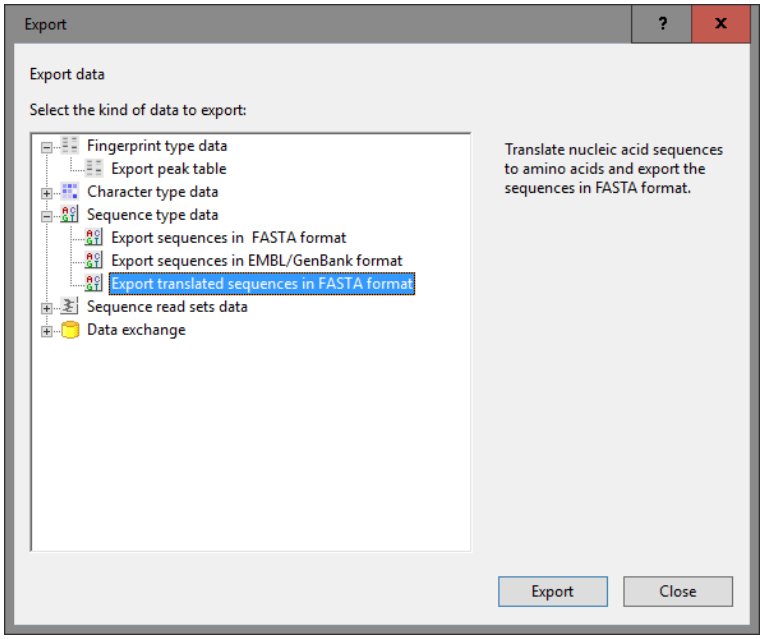

Figure 2.6: The *Export* dialog box.

5.3 Select ***Export translated sequences in FASTA format*** under ***Sequence type data*** and press the <***Export***> button to call the *Export translated sequences* dialog box (see Figure 2.7).

In the *Export translated sequences* dialog box all nucleotide sequence types defined in the database are displayed in the upper panel. One or more sequence types can be selected using the **Ctrl-** key.

A translation frame can be specified for each sequence type separately using the pull-down menu in the ***Frame*** column: ***Frame 1***, ***Frame 2*** or ***Frame 3***. With the ***Manual*** option the frame that was manually set (see 2.2) is used if no better frame is detected. Using the ***Auto*** frame option the best frame to use is screened by the plugin.

From the list under ***Select the translation table***, the different translation tables, corresponding to variants of the standard genetic code, can be selected. By default, ***The Standard Code*** is selected.

A new or existing ***Output file*** can be selected with the <***Browse***> button.

5.4 Select the sequence type(s) from the list, set the reading frame(s), specify the translation table and output file and press <***OK***>.

Pressing <***OK***> translates the sequences of the selected entries and sequence type(s) based on the specified reading frame, and exports the translated sequences in FASTA format to the selected file (see Figure 2.8 for an example).

Each sequence begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater than (">") symbol. The description line contains

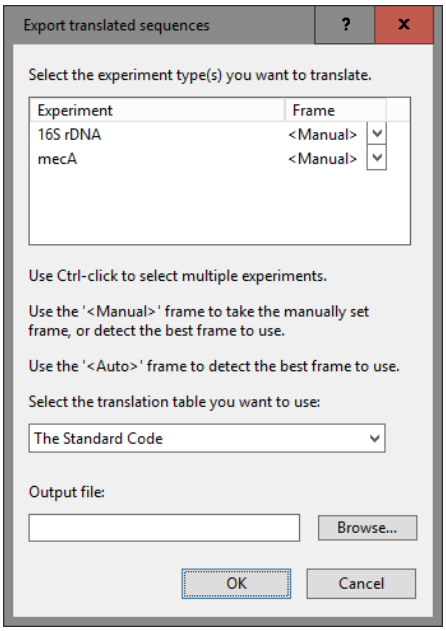

Figure 2.7: The *Export translated sequences* dialog box.

the entry ***Key***, the name of the ***Sequence type*** and the reading ***Frame*** that was used for translation, separated by a pipe (”|”) symbol.

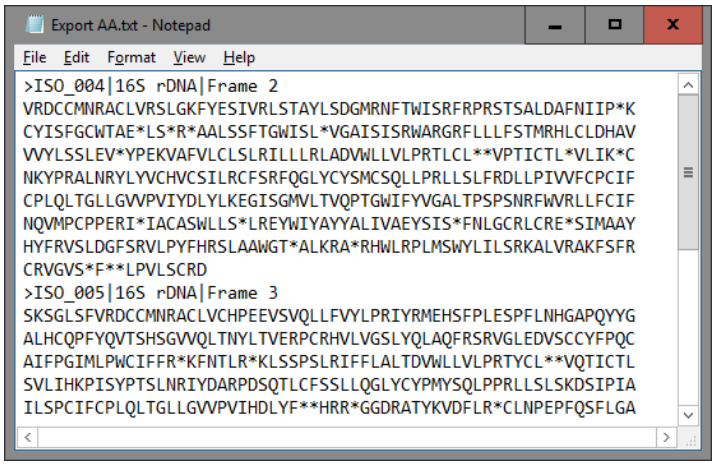

Figure 2.8: Translated sequences.

APPLIED MATHS

A BIOMÉRIEUX COMPANY

Please contact us for any additional information you might require, we will gladly help you!

Headquarters
Keistraat 120 • 9830 Sint-Martens-Latem • Belgium
+32 922 22 100 info@applied-maths.com

USA and Canada
11940 Jollyville Rd., Suite 115N • Austin, TX 78750 USA
+1 512 482 9700 info-us@applied-maths.com