SeqGen

Marko Raina

	COLLABORATORS			

	TITLE : SeqGen			
ACTION	*NAME*	*DATE*		*SIGNATURE*
WRITTEN BY	Marko Raina	June 8, 2025		

	REVISION HISTORY		

NUMBER	DATE	DESCRIPTION	NAME

Contents

Chapter 1

SeqGen

1.1 SeqGen manual

```
                       SeqGen 1.0 manual


   General:
            Copyright & disclaimer   Important facts
            Installing SeqGen        Simple!
            Starting the program     Shell and WB startup
            Description              Purpose of this program
            Main window              Gadgets, keyboard, mouse
            The author               If you want to contact me

   Menus:
            Project menu             Files, printing
            Edit menu                The clipboard
            Amino acids menu         Editing tables of amino acids
            Settings menu            Customizing the looks and output
```

```
© Marko Raina 1996
```

1.2 Copyright and Disclaimer

```
This is SeqGen 1.0. The author and all copyrights © Marko Raina 1996. This
archive can be distributed freely, as long as it is not altered and one gets
no economical benefit. In other words, this program is Freeware.

You ARE using this program on your own responsibility. The author takes no
responsibility in case the user loses valuable information, files or his/her
temper (8->) due to bugs or the inherent features of the program.

Despite of the things above, the author would be very pleased if you sent
him detailed information about possible encountered bugs or
incompatibilities. Suggestions for future development are welcome.
```

1.3   Installing SeqGen

Installing SeqGen 1.0 is very easy. Just uncompress the archive with command
"lha e SeqGen.lha". The archive contains a directory called SeqGen. No third
party libraries are needed, but program uses the system's iffparse.library
for clipboard work and asl.library for file requesters. It knows also how to
use the amigaguide and locale libraries.

Contents of the archive:

```
  SeqGen (dir)
      Config (dir)
        def_app.info                    def_seq.info
        NormalAcids.aa                  SeqGen.prefs
      catalogs (dir)
          suomi (dir)
            SeqGen.catalog
      Sequences (dir)
        Little.seq                      Little.seq.info
      develop (dir)
        SeqGen.cd
    Config.info                       PepScan.ilbm
    ReadMe.first                      ReadMe.first.info
    SeqGen                            SeqGen.guide
    SeqGen.guide.info                 SeqGen.info
    SeqGenFinn.guide                  SeqGenFinn.guide.info
    Sequences.info
  SeqGen.info
```

SeqGen should work with OS version 2.04 and above. If you have WB2.1 or
later, the program is fully localized, if not, the language is english. At
the moment there is only Finnish (suomi) catalog file in the archive. For
those who want to create one of their own language the .cd file is also
included. Also you need the CatComp© program or some other localization
tool; see Aminet for them.

If you want to remove SeqGen from your disk, just delete the whole SeqGen
directory with "delete SeqGen all". But, be sure that you don't destroy any
other valuable files on your disk!

1.4   Starting SeqGen

You can start SeqGen either from Shell or Workbench. If no options are
selected, the program tries to open the amino acid table NormalAcids.aa and
the settings file SeqGen.prefs. Both files should lie in the program's
config directory. If these files cannot be loaded, the program uses its
default settings. In this case, the amino acid table will be empty.

1. Starting from Shell

The template for the program is (see AmigaDOS manual for more information):

    SEQUENCE,ACIDS/K,SETTINGS/K,PUBSCREEN/K,BUFFER/N/K

None of the arguments is obligatory. Explanations:

```
    SEQUENCE   - the sequence file to open
    ACIDS      - the acid table to open
    SETTINGS   - the settings file to open
    PUBSCREEN  - the windows will be opened onto this screen
    BUFFER     - the size of the editing buffer (minimum 100)
```

Examples:

```
    SeqGen
        starts the program with default table and settings

    SeqGen file.seq
        loads the sequence file.seq with default settings

    SeqGen file.seq settings abnormal.prefs
        like above, but settings come from file abnormal.prefs
```

2. Starting from Workbench

You can double-click either the icon of the program or of the sequence file.
The tooltypes are the same as in the template of the Shell (of course,
there's no SEQUENCE tooltype). You can modify the tooltypes through
Workbench Icons/Information menu item. See Workbench manual for more
information. The following example shows how to start the program with
different settings file. The ACIDS-tooltype is unselected:

```
    (ACIDS=<name>)
    SETTINGS=abnormal.prefs
    ...
```

1.5   Purpose of SeqGen

SeqGen is a program designed to help editing, saving and printing amino acid
sequences of known proteins. With SeqGen you can easily feed the sequence by
typing the one character symbols of acids with the keyboard. Or, you can
enter the acids by clicking the respective acid button gadget in the window.
Gadgets show short or long symbols, depending on the state of Symbols
gadget. It's also possible to enter the sequence through clipboard (see
example); this is how you can move the sequence from a database to SeqGen.

With the program, you can generate files compatible with the Pepscan
planning program of Commonwealth Serum Laboratories, inc. The CSL program I
have been using is a very old version (v3.20, 1987), and it's made for
MS-DOS computers. It's not very nice to use, particularly if you have to
feed the sequence manually. Problems will also arise if you need to use
offset other than 1 for the component peptides. SeqGen removes these
problems. You choose the part of sequence you want to map, save it with a
specially defined format settings and use the resulted file as a source to
CSL program (for example, BNET synthesis). Using the same settings, you can
also print the sequence with your printer.

1.6 Main window

This is the layout of the main window:

```
     + --------------------------------------+
     | Go to:              Find:             |
     |                                       |
     |                         Symbols       |
     |  ------------         ----------------  |
     | |001:editing |       | @    Short     | |
     | |010:area    |        ----------------  |
     | |            |                         |
     | |            |       acid1  acid2  acid3 |
     | |            |       acid4   ...        |
     | |            |                         |
     |  ------------                          |
     +---------------------------------------+
```

The editing area is naturally for editing the sequence. It lies on the
lefthand side of the window. Editing is possible when the area is active,
that is, when the cursor can be seen. You can make the area active by
point-and-click'ing it with the mouse.

Gadgets

Go to   If you want to move the editing cursor to a certain acid in the
         sequence, type the ordinal number of the acid in the "Go to"
         gadget.

Find   To find a smaller sequence (like a peptide) from the sequence being
         edited, type it into this gadget with or without '-'s. The search
         begins from the cursor position.

Symbols Clicking this will toggle you between short and long amino acid
         symbols. This affects clipboard functions also.

Acids   With these gadget you can enter the acids into the sequence without
         keyboard. Editing area must be active; if you don't see the cursor,
         use mouse to activate the area. If there are no acid buttons, the
         amino acid table is empty. Create or load one from disk.

Keyboard

Acids   When the editing area is active (you see the cursor), you can enter
         an acid by simply typing it's single letter symbol. Note that
         SeqGen is case sensitive: c and C are different.

Cursor   With the cursor (arrow) keys, you can move around the sequence. By
         pressing SHIFT key with the cursor key you can move faster.

Delete   Deleting acids is very easy. Backspace button deletes the acid on
         the left side of the cursor. "Del" button removes the acid under
         the cursor. Bigger areas can be removed by selecting them with a

```
        mouse and clearing the area with Edit/Clear
        function.

Help    Open this manual. The page opened depends on what window (and
        gadget) is active. If you press "Help" while picking up a menu item
        the manual page describing the menu item will be shown.
```

```
Mouse
```

Cursor can be moved by clicking the appropriate acid with the mouse. You can choose part of the sequence by drag selecting: press the left mouse button down and move the mouse. The selected part can be used with the Edit menu functions.

1.7   The author

The author of this program is

```
    Marko Raina
    Ilmarinkatu 23 D 37
    33500 Tampere
    Finland
    Europe

    Email:
    blmara@uta.fi
```

I'll gladly accept comments, bug reports, suggestions etc. I know the program is very simple, but it suits for my present needs. Perhaps an AREXX interface and better, more font sensitive GUI would perhaps be the most urgent improvents... No Enforcer compatibility has been checked. Please inform me in case you notice any Enforcer hits!

This program was developed with an A500/3MB/WB2.1 + A590/20 MB system and Dice C v.2.07.56R. Click here to see the complete thanks page.

1.8   Pepscan

H.Geysen and his colleagues developed a new method, "Pepscan", for synthetizing small peptides^1. The peptides are short polymers or chains consisting of several amino acids. The idea was that identical peptides are built onto a small plastic pin, which can be used several times during, for example, serological tests. The pins can be bought for example in sets of 96, which are mounted in a small plastic rack. One of the most important applications of Pepscan is linear epitope mapping.

Here's a picture describing the PepScan method shown by Display (< WB3.0) or MultiView (WB3.0 or above) For some reason, the first link works only if you are reading this directly (outside the SeqGen program).

^{1} Geysen et al: Use of peptide synthesis to probe viral antigens for
epitopes to a resolution of a single amino acid. Proc Natl Acad Sci 81:
3998-4002, 1984)

1.9 Epitope

Epitope is the part of an antigen which specifically induces the
immunoresponse. Often, it's a certain area in some structural or functional
protein of a virus (bacterium, fungus,...). Usually there are many epitopes
in an antigen, and there's several antigens in a microbe.

1.10 Epitope mapping

Epitope mapping means searching for the epitopes of an antigen. Pepscan
method is able to find B-cell epitopes, i.e. the parts of protein which
trigger antibody formation. Due to short length of peptides (8-10 acids),
only linear epitopes are found. Non-continuous epitopes which are formed by
protein bending to three-dimensional conformation would need longer
sequences.

An example

Let's say the amino acid sequence of a viral protein is A-C-D-E-F-G-H, where
the letters are the one-letter symbols of amino acids. If it's assumed that
the epitope is four acids long, 4 different 4-acid peptides with an offset
of 1 are synthetized:

```
    Pin1: A-C-D-E
    Pin2:   C-D-E-F
    Pin3:     D-E-F-G
    Pin4:       E-F-G-H
```

After the synthesis, the pins are incubated with patient serum (blood
samples) collected from individuals who have had the virus disease. These
patients have developed antibodies against the virus particles, and some of
the antibodies bind to the mentioned protein, particularly to the epitope
area. Binding is detected with secondary antibodies and the antigenic
sections of the protein are found. These short sections can be synthesized
and used, for example, to diagnose the disease.

1.11 Antigen

An antigen is any structure considered hostile by our immunological defence
system and therefore is attacked by the system. Antigens can be parts of
bacteria, viruses, fungi and even parts of our own tissues.

1.12   Amino acids

Amino acids are small molecules which have at least one carboxyl group
(-COOH) and one amino group (-NH2). Due to the special chemical features of
amino acids they can easily polymerize to short chains (peptides) and even
very long chains called proteins. Amino acids can be said to be the basis of
life.

An international standard defines one-character and 3-character
abbreviations for amino acids, so the structures of peptides and proteins
can be presented compactly. There are hundreds of known amino acids, but
only 20 of them are components of natural proteins. The file
config/NormalAcids.aa in the archive contains the definitions of these
acids:

```
    Short      3-char      Whole name
      A        Ala         Alanine
      C        Cys         Cysteine
      D        Asp         Aspartic acid
      E        Glu         Glutamic acid
      F        Phe         Phenylalanine
      G        Gly         Glycine
      H        His         Histidine
      I        Ile         Isoleucine
      K        Lys         Lysine
      L        Leu         Leucine
      M        Met         Methionine
      N        Asn         Asparagine
      P        Pro         Proline
      Q        Gln         Glutamine
      R        Arg         Arginine
      S        Ser         Serine
      T        Thr         Threonine
      V        Val         Valine
      W        Trp         Tryptophan
      Y        Tyr         Tyrosine
```

1.13   An example

Here is explained how to create a peptide file with 94 peptides. The
peptides have been "cut" from the start of the protein, they measure 10
amino acids and the offset between the acids is 3 acids.

1. Entering the sequence

The sequence was acquired from a sequence database and it was transferred
with CrossDos from PC to Amiga. Header information and position numbers were
removed with a text editor and the remaining sequence was copied into the
clipboard. The sequence was pasted into the SeqGen editing buffer.

2. Defining the format

```
CSL program file format for file of sequences is:
    * Number of different peptides
    * Peptide1 (single character symbols, no separators)
    * Number of pep1's to synthetize
    * Peptide2
    * Number of pep2's to synthetize
    * ...
    * Peptide "number of peptides"
    * Number to synthetize

So, the format settings selected were:
    * Header         : 94\n
    * Format         : %s\n\1\n
    * Footer         :
    * Peptide Length : 10
    * Offset         : 3

    "Short symbols" and "Output only" were selected and "Use Separator"
    unselected.

    * Number of peptides      : 94
    * starting from (acid #) : 1
```

3. Creating the peptide file

The Project menu item Save in format... was selected and an appropriate name
(you know, 8+3 chars...) was given. The file was moved back to a PC disk and
the final CSL peptide synthesis plan was created using the generated file as
a source for BNET synthesis.

1.14 Thanks!

Thanks to many persons for tools that made this program possible:

```
    * Jan van den Baard      GadToolsBox 2.0c – GUI editor
    * Matthew Dillon         DICE 2.07.56R – C compiler
    * Edd Dumbill            Heddley 1.1 – Amigaguide editor
    * Dietmar Eilert         GoldEd 3.13 – text editor
    * David Zvekic           FixHeddley 1.2 – Heddley fixer
```

1.15 File types

SeqGen generates three types of files: sequence, amino acid and settings
files. Here is some information about them.

1. Sequence file

This file contains the sequence saved through Project/Save menu item. The
file contains also format settings and the name of amino acid table used
when Save item was chosen. File name has normally postfix ".seq", for

example, "ß-galactosidase.seq".

2. Amino acid file

This is a list of short and respective long symbols of amino acids. In the
SeqGen archive the file "NormalAcids.aa" contains the 20 amino acids found
in natural proteins. File name has normally postfix ".aa". When started,
SeqGen tries to load file "config/Normalacids.aa", where "config" directory
lies in the program directory.

3. Settings file

This file contains information about your editing settings, format settings
and window size and position preferences. File name has normally postfix
".prefs". When started, SeqGen tries to load file "config/SeqGen.prefs",
where config directory lies in the program directory. Window settings are
loaded only when starting SeqGen.

1.16   Antibodies

Antibodies are B-lymphocyte-derived small proteins, which are synthetized
against and bind to the causative agent of infection (bacterium, virus,
fungus). By binding they help other parts of the immune system to destroy
the microbe.

Secondary antibodies are antibodies against other antibodies (!), which are
made for laboratory use. They can be coupled with some marker, for example a
colour-producing enzyme, and used to detect human antibodies.

1.17   Project menu

Project menu:

     Open...
     Clear

     Save
     Save as...

     Save formatted...
     Print formatted

     Information...
     Help

     Iconify

     Quit

1.18 Project/Open

Opens a sequence file saved by SeqGen.

1.19 Project/Clear

Clears the sequence. If you haven't saved it, asks first.

1.20 Project/Save

Saves the current sequence.

1.21 Project/Save as

Asks the name of the sequence file and saves it.

1.22 Project/Save formatted

Saves the sequence (or a part of it) according to the format settings.

1.23 Project/Print formatted

Prints the sequence (or a part of it) according to the format settings. You
can stop printing by clicking "Stop" in the requester shown.

1.24 Project/Information

Gives information about the program, the current sequence and the amino acid
table.

1.25 Project/Help

Opens this manual (actually, file "SeqGen.guide" in program directory).

1.26 Project/Iconify

Iconifies the window, i.e. closes the window and creates an appicon onto
Workbench. You can reopen the window by double-clicking the appicon. The
image of the appicon is taken from Config/def_app.info, so you can replace
it with a better looking icon if you like.

1.27   Project/Quit

Quits the program.

1.28   Edit menu

Edit menu:

     Cut
     Copy
     Paste

     Select all

     Clear

1.29   Edit/Cut

Moves the area selected with the mouse or Select all -function from the
sequence into clipboard.

1.30   Edit/Copy

Copies the area selected with the mouse or Select all function from the
sequence into clipboard. The area is not removed from the sequence.

1.31   Edit/Paste

Pastes the contents of clipboard into the position of the cursor. The
contents are interpreted according to the present short/long symbols and
amino acid table settings. White space characters like carriage return, ' '
and '-' are ignored. If there's a name of an acid which is not found in the
current amino acid table SeqGen will show an error message.

1.32   Edit/Select all

Selects the whole sequence.

1.33   Edit/Clear

Clears the part of the sequence selected with the mouse or Select all
function. Be careful; this cannot be undone.

1.34   Amino acids menu

Amino acids menu:

        Edit...

        Open...
        Save...

1.35   Amino acids/Edit

Opens the amino acid table editing window.

Here you can create and edit the "amino acid table" needed when you edit and
print the sequence. You must define a "long" and a "short" symbol for each
acid. Long versions can be up to 5 characters, the short symbol is a single
character. It's reasonable to choose the symbols according to the present
standard. By pressing the "Help" key you can read this manual page.

Example: long short
         Ala   A
         Tyr   Y
         ...

The listview contains the long symbols of the amino acids. If you select an
acid with the mouse, the small "Short" gadget shows the short symbol.

Gadget explanations

Add   When you want to add a new acid, click this button. Edit the long
        symbol in the gadget under the listview. By pressing Return or TAB
        you move to the "Short" gadget, where you can edit the short
        symbol.

Delete Removes the current acid from the table.

Up   Moves the current acid upwards in the acid table.

Down   Moves the current acid downwards in the acid table.

Short   Here you can enter the short symbol of the current acid. All
        printable character except '-' can be used as a short symbol. The
        hyphen is reserved as a separator of acids.

Use   When you're done, you accept the changes by clicking this button.

Cancel If you want to cancel you changes, you should click this button.

1.36   Aminoacids/Open

Loads amino acid table file from disk.

1.37   Amino acids/Save

Saves amino acids table file into disk.

1.38   Settings menu

Settings menu:

        General...
        Format...

        Load..
        Save...

         Create icons?

1.39   Settings/General

Opens the general settings window.

These settings affect the format of the editing area, To read this manual
page, you can press "Help" key when the general settings window is active.

Gadgets

Acids/line   Maximum number of acids to print on one line of the
                editing area. The minimum is one; no more
                acids are printed than fits there!

Number length   How many digits are used for the ordinal number of an acid
                in the editing area. Maximum is ten digits.

Use separator   If selected, SeqGen will separate acids with a hyphen ('-')
                in the editing area. This concerns also Edit/Cut
                and Edit/Copy function.

1.40   Settings/Format

Opens window for format settings

These settings affect only format functions. An example how to use the
format system is presented elsewhere. To see this manual page while program
is running, press "Help" key.

Gadgets

Header   This line is saved or printed as-is in the beginning of

formatted save or formatted print. You can include several
lines separated with symbol '\n' representing line feed.
If you want to get the backslash '\', you have to write '\'.
For example, you write 'Header\n' to get a line 'Header'.
Please don't use the printf-style %s,%d,%u etc., because
these can really jam the machine.

Format   Here you define the printf-style format string, which is
            used to save/print every peptide. You can use normal
            characters, '\n' for carriage return, '\' for backslash,
            and the following control codes (To get a line like
             "020: PEPT" you give format "%03.lu: %s\n").

   code   prints
   %%   %
   %s   the peptide - once and only once!
   %-lu   the number of the first acid of the peptide (left justified)
   %lu   "   "   " (right justified)
   %4.lu   as above, but the width of the number field is 4
   %04.lu   as above, but the number field is filled with 0's

Footer   This is saved/printed after the peptides. See "Header"
            for more info.

Peptide l. The length of the save/print peptide.

Offset   The distance between the first acids of sequential peptides.
            If you want the classical one-by-one peptide set, choose 1.

Short symb. If selected, prints short amino acids.

Use   If selected, prints separating '-' between amino acids.
separator

Output only If this is chosen only the number of peptides defined in the
            gadgets under this gadget is saved/printed. The starting point
            is also selected with a number gadget under this gadget.

1.41 Settings/Load

Loads the settings file from disk.

1.42 Settings/Save

Saves the current settings into disk file.

1.43 Settings/Create Icons

If this item is checked, SeqGen saves a project icon when saving a sequence.
The image is taken from Config/def_seq.info, and you can replace it with a
better looking icon if you like.