News from IntelliGenetics

Volume 4 Number 2 March/April, 1988

Predicting Transmembrane Alpha Helices

by John Moore

The WINDOW option of the PEP program allows you to use any table of amino acid values to calculate a weighted average over an amino acid interval of any size in a protein sequence.

This example illustrates the use of WINDOW to look for transmembrane alpha helices in a human acetylcholine receptor protein precursor. The amino acid values are those of the buried-helix parameter as described in Rao, M.J.K., and Argos, P. 1986. *Biochimica and Biophysica Acta* 869:197-214. The buried helix parameter is the sum of the hydration potential, free energy of transfer from aqueous helix to nonpolar helix, polarity, bulk conformational preference, and turn conformational preference. You can compare the results shown here to those reported on page 206 of that paper.

MAKING AN AMINO ACID TABLE

First, enter the PEP program and LOAD a peptide file into the program memory. The **ac.pir** file contains the human acetylcholine alpha chain precursor protein, a known transmembrane protein. Next, enter the WINDOW level. The program loads the Hopp and Woods amino acid hydrophilicity values as the default values. Next, enter the Rao and Argos amino acid values by using EDIT-WINDOW-FILE to edit the file. To enter a new set of values rather than edit the old ones, select the CREATE option. Table 1 lists the actual buried helix parameter values.

However, in this example, we make the transmembrane helices easily visible on the plot by adjusting these amino acid values. Reducing each value by 1.05 makes the transmembrane helix candidate segments appear on the positive side of 0 on the plot with noncandidates on the negative side. Table 2 shows the adjusted values. (Since WINDOW does

A	1.360	B	1.000	C	1.270
D	1.110	E	1.250	F	1.570
G	1.090	H	1.680	I	1.440
J	1.000	K	1.090	L	1.470
M	1.420	N	1.330	O	1.000
P	1.540	Q	1.330	R	1.150
S	1.970	T	1.080	U	1.000
V	1.370	W	1.000	X	1.000
Y	1.830	Z	1.000		

Table 1

A	.310	B	.000	C	.220
D	-.940	E	-.800	F	.520
G	.040	H	-.370	I	.390
J	.000	K	-.960	L	.420
M	.370	N	-.720	O	.000
P	-.510	Q	-.720	R	-.900
S	-.080	T	.030	U	.000
V	.320	W	-.050	X	.000
Y	-.220	Z	.000		

Table 2

(continued on page 2)

User Manual To Be Released

by Cindy Cohen

This spring, a new IntelliGenetics Suite *User Manual* will replace the current *Short Course*. Unlike its predecessor, the *User Manual* is organized in terms of activities performed in the lab rather than by program. The examples come from Release 5.1 of the IntelliGenetics Suite, which utilizes GenBank (Release 55), EMBL (Release 14), PIR (Release 14) and Vectorbank (Release 4.1).

The examples are step-by-step guides, with annotations, for using IntelliGenetics programs to perform common operations in the following categories of functions:

- **Sequence entry and editing**
- **Sequence location**
- **Sequence and map display**
- **Sequencing project management**
- **Sequence translation**
- **Sequence composition**
- **Sequence structure**
- **Restriction analysis**
- **Cloning simulation**
- **Sequence comparison and alignment**

The *User Manual* covers all three systems on which the IG Suite is available: Sun , VAX, and the IntelliGenetics DEC 2065. In cases in which there are differences in file names or system commands, the examples note the system-specific names and commands. The manual includes methods for searching and printing files, and for using common system commands of each of the three systems. In addition, the manual discusses basic mechanics of the software, including file structure, data bases, and program commands. It also explains how to display and print program output files.

DISTRIBUTION

Each BIONET Principal Investigator will receive one copy of the manual. Each IntelliGenetics VAX or Sun site will receive three copies. Additional copies of the *User Manual* will be available for $95 apiece. To order, please contact Mary Harrington at the IntelliGenetics Sales Department in writing, or phone (415) 962-7354.

(continued on page 4)

Predicting A Helix

(continued from page 1)

not allow you to choose either the range or the central value of the plot, you can only adjust the amino acid values or the weights to change the display.)

After you edit the values, use LIST to check them. Use SAVE-WINDOW-FILE to save the new set of amino acid values in a file with appropriate comments; sets of amino acid parameters are given the file extension **.win**.

If you have saved a window file in a previous work session, you can LOAD it instead of entering new values or editing old ones.

REAVERAGING

The Rao and Argos method averages over an interval size of 7. The weight for each position in the interval is 1. After the average for each interval is found, the calculation is repeated using the newly calculated values as the amino acid values. Thus the plot of the buried helix parameter values along the sequence is "smoothed" by averaging the current values three times.

WINDOW in PEP does not directly perform the reaveraging. However, reaveraging is equivalent to using the original amino acid values on a wider interval with different weights. The original averaging is performed over an interval of seven amino acids. In reaveraging, each new value for an amino acid depends on the seven amino acids nearest it. The contributions of each nearby amino acid to the value of the central amino acid (denoted by "|") can be followed by drawing a graph:

```
First Average

Second Average

Third Average . . . . . .
```

For three averagings, the amino acid value nine amino acids away from the first central amino acid contributes something to the smoothed average of the first central amino acid. Therefore the interval size needed is 19, so use the WINDOW-SIZE option to change the interval.

ASSIGNING WEIGHTS

How do you assign weights to the 19 positions in the interval? All you have to do is add the number of contributions from each original amino acid value in the new interval. The farthest away position contributes once, the next farthest away contributes three times, etc. Therefore, the relative (raw) weights of the positions are: 1, 3, 6, 10, 15, 21, 28, 33, 36, 37, 36, 33, 28, 21, 15, 10, 6, 3, and 1. This is the number of times each position is used in calculating the value for the central amino acid. The sum of these numbers is 343. The next graph shows how these weights were obtained. You can simulate any smoothing by representing it in a graph and then adding the contributions.

```
(7 times)
(6 times)
(6 times)
(5 times)
(5 times)
(4 times)
(4 times)
(3 times)
(3 times)
(2 times)
(2 times)
 (1 time)
.(1 time)
```

FINDING THE AVERAGE

The WINDOW program finds the average by dividing the sum of the weights times the values by 19, the width of the window. However, the total number of values contributing to the smoothed average is 343 (the number in the interval, 7, times the number of repeated intervals, 49). In order to get an average value that has the same size as the buried helix parameter and so get a true average, you must divide each of the relative weights by 343/19. This restores the average value to that of the original buried-helix parameter. Thus, the weights we must use are: 0.055, 0.166, 0.332, 0.554, 0.781, 1.163, 1.551, 1.828, 1.994, 2.040, 1.994, 1.828, 1.551, 1.163, 0.781, 0.554, 0.332, 0.166, and 0.055. Use the WEIGHT option to assign these weights to the 19 positions in the interval.

NOTE: When you LIST values to check them, the program displays the amino acid values as decimals but it displays the weights to the nearest integer in order to fit the interval on one line; the program actually uses the decimal weights you entered.

To find the average, use arithmetic averaging. For an odd interval centered at an amino acid, the program takes the sum of the products of the weight for that position times the value of the amino acid at that position in the interval, and then divides the sum by the size of the interval, in this case, 19. The program assigns that average to the central amino acid.

NOTE: WINDOW also lets you use geometric averaging. Averages can be over odd or even intervals. (Refer to the documentation in the Reference Manual for more information about the method used.) After you select the averaging method, RUN the analysis.

PLOTTING THE RESULTS

The program displays the name of the protein, the values of the amino acids, the maximum and minimum values, and a graph of the results along the sequence. The center of values is always 0, and the range is automatically adjusted to contain all the plotted values. A "+" or a "-" after a residue number indicates a charged residue. (See the plot on page 3.)

According to Rao and Argos, a segment that is a candidate for a transmembrane helix must have a series of 12 or more amino acids with values of 1.05 or higher; at least one of these must be greater than 1.13. The plot shows a possible transmembrane helix (a signal

(continued on page 3)

Predicting A Helix

(continued from page 2)

sequence) at the beginning of the sequence, three possible helices starting at position 233, and one at the end.

The five transmembrane helices in the protein stand out clearly, just as they do in the Rao and Argos paper. The 1.05 level is equivalent to 0 on this plot, and the 1.13 level is equivalent to +0.08. NOTE: The values on the plot are given in tenths.

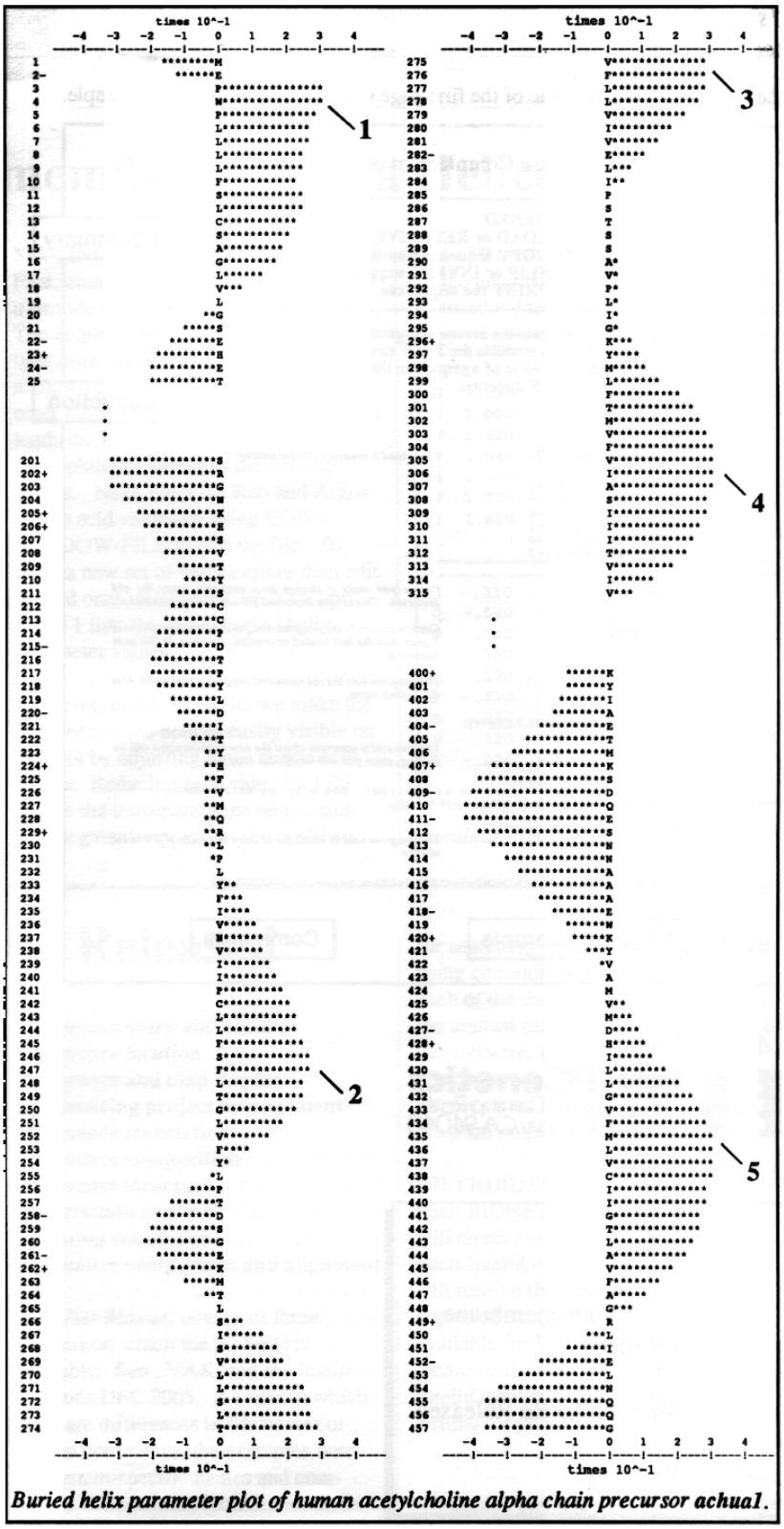
CREATING A NEW DEFAULT

When you have finished the RUN, use QUIT to leave the WINDOW level and to return to the PEP prompt. When you exit WINDOW, you are given the option to save the Rao and Argos values for weights and window size in an output parameter file. If you save the values, they will become the default values for the next time you use WINDOW. This is useful if you want to run this analysis regularly instead of the default Hopp and Woods analysis. You can also add comments to the file.

OTHER APPLICATIONS

You can use the WINDOW option of the PEP program to calculate a weighted average over any interval of amino acids in a protein sequence. You can use these techniques to calculate the physical and chemical parameters of amino acids, including hydropathicity (Kyte and Doolittle), hydrophilicity (Hopp and Woods), hydrophobicity (Rose and Rose or Eisenberg), free energy, secondary structure propensities (Chou and Fasman), hydration potential, polarity, flexibility (Karplus and Schultz), and isoelectric points. WINDOW can "smooth" the values or show the raw unaveraged values. You can also adjust amino acid values or weights to create a plot which emphasizes certain features or to show a cutoff point. ♦

NEWS FROM INTELLIGENETICS, edited by Cindy Cohen, is a customer newsletter published by IntelliGenetics, Inc. Please address all correspondence to 700 East El Camino Real, Mountain View, CA 94040, or phone (415) 962-7346.



Buried helix parameter plot of human acetylcholine alpha chain precursor achua1.