CABIOS

Vol.8. no.4. 1992
Pages 331 – 337

PROFILEGRAPH: an interactive graphical tool for protein sequence analysis

K.Hofmann and W.Stoffel[1]

Abstract

The computer program PROFILEGRAPH, a graphical interactive tool for the analysis of amino acid sequences, is described. The main task of the program is to integrate a variety of sliding-window methods into a single user-friendly shell. The program allows the user to combine any amino acid specific parameter with a selection of several possible types of analysis and to plot the resulting graph in one of several windows on the screen. It is also possible to calculate the moment of the amino acid specific parameter for a given secondary structure and to display both the absolute moment value and the moment angle relative to a reference residue. Also included are several utilities that facilitate visual analysis of protein primary structures like, for example, helical-wheel diagrams. It is possible to adapt the majority of published sliding-window analysis procedures for use with PROFILEGRAPH.

Introduction

One of the most frequently encountered tasks in the computer-assisted analysis of amino acid sequences is a graphical representation of the distribution of some amino acid properties along the primary structure. A well-known example for this type of analysis is the hydropathy plot (Kyte and Doolittle, 1982).

The general procedure for the calculation of a sliding-window profile requires as input (i) the amino acid sequence; (ii) a table of the amino acid specific parameter to be analyzed, holding a numerical value for each different amino acid; and (iii) the width of the window that is used for averaging. In the first step of the calculation, for each residue in the original sequence an abscissa value is given by the position of the residue, while the ordinate is the parameter value associated with the amino acid under consideration. In most cases, a plot at this stage of the analysis would be too noisy to be useful, so the ordinate value of a specified position i is usually not only a function of the amino acid A_i but an average of a window centered around the considered position including also values for the neighboring amino acids A_{i-n}, ..., A_{i-1}, A_i, A_{i+1}, ..., A_{i+n} if the window width is $2n+1$.

Many different sets of amino acid parameters have been proposed for sliding-window analysis of protein sequences, among them different measures for amino acid hydrophobicity (Ponnuswamy *et al.*, 1980; Hopp and Woods, 1981; Kyte and Doolittle, 1982; Eisenberg, 1984; Parker *et al.*, 1986). Other examples are measures for flexibility (Karplus and Schulz, 1985; Bhaskaran and Ponnuswamy, 1988; Ragone *et al.*, 1989), residue volume (Chothia, 1975), surface area (Rose *et al.*, 1985) and propensity for different secondary structure types (Levitt, 1978). A database of all published amino acid parameters has been compiled and is being updated (Nakai *et al.*, 1988).

Most of these amino acid parameters are used with only one recommended window width while in some cases the window width has to be adjusted depending on the type of information to be obtained. As an example, the hydrophobicity parameters when used with a window width of $\sim 17-19$ give information about possible transmembranal domains of proteins, while for locating smaller hydrophobic stretches, as they occur within the core of globular proteins, window widths of ~ 7 are usually applied.

Besides these simple average calculations of a single amino acid parameter, some more complex types of analysis also based on profile calculation have been proposed. Some of these are for determining the distribution of a particular amino acid property along a proposed secondary structure. An example is the calculation of the hydrophobic moment (Eisenberg *et al.*, 1982) as a measure of the excentricity of hydrophobicity if considering an α-helix. Another method in dealing with asymmetrical hydrophobicities is the calculation of the average hydropathy on only one helical face (Jaehnig, 1989). These types of analyses can be generalized to secondary structures other than the α-helix as well as to parameters other than the hydrophobicity.

In visualizing the amphipathy of a helical structure, a type of diagram, known as 'helical wheel', is usually applied (Schiffer and Edmundson, 1967). Here one looks along the helical axis while the amino acid side-chains are plotted around a wheel-like structure with an angle of 100° (the typical inter-residue angle of an α-helix) between two successive residues.

Recent works have pointed out the importance of choosing an appropriate averaging procedure for smoothing the profiles. The classical approach of applying a rectangular window

Institute für Biochemie, Medizinische Fakultät der Universität zu Köln, Joseph Stelzmann Strasse 52, D-5000 Köln 41, Germany
INTERNET: KHOFMANN @ BIOMED.BIOLAN.UNI-KOELN.DE
[1]To whom reprint request should be sent

centered around the considered residue has the disadvantage of a major loss of spatial resolution. An improvement of the situation can be achieved by application of triangular windows (Claverie and Daulmerie, 1991). The advantages of median formation instead of averaging have been discussed and a 'data-sieve' consisting of several cascaded median formations, has been proposed for exact determination of transmembranal domains (Bangham, 1988). While the usefulness of these modifications have mostly been demonstrated for hydrophobicity profiles, it seems likely that other parameters would also benefit from these improvements.

While several programs using sliding-window protein analysis are available, most of them are either very limited in scope, slow and inconvenient, or part of expensive commercial packages. These programs are also often severely limited regarding the format of the sequence files and usually have a rigid grouping of analysis methods.

The main task in designing PROFILEGRAPH was to integrate most of the currently used sliding-window methods into a single user-friendly shell while keeping it an open system that can easily be extended according to the user's need. Most of these extensions can be done by adding further method tables and need no programming.

System and methods

PROFILEGRAPH was written in Turbo Pascal 5.5 and developed on personal computers running MS-DOS 4.0. It has been extensively tested on a variety of computers equipped with different graphical adapters and using different versions of MS-DOS.

PROFILEGRAPH uses the BGI graphical system (Borland International Inc., Scotts Valley, CA) for screen display. For input of protein sequences the READSEQ-routine, kindly provided by Don Gilbert (University of Indiana, 1990), is used.

Since there is no graphical menu system available for programming with Turbo-Pascal that is free of copyrights, a basic user interface using pop-up menus had to be developed. For supporting device-independent vectorial graphical output, the 'Turbo Pascal Graphics description language' (TPGL) has been defined (K.Hofman, unpublished). TPGL files produced by this and other programs can be converted into a plotter-specific format or into a high-resolution bitmap at a later time. Some TPGL-converters are available from the authors; more are in preparation.

Algorithms and capabilities

When PROFILEGRAPH calculates the points to be displayed within a single output frame, it performs the following steps successively: (i) input of the sequence file; (ii) input of an amino acid specific parameter; (iii) assignment of the appropriate parameter value to each sequence position; (iv) execution of a special calculation routine according to the actually chosen 'moment method'; (v) application of an averaging routine according to the actually chosen 'averaging method', (vi) rescaling of the output values according to the chosen 'scaling method'; and (vii) output of the calculated values to the screen using the actual valid 'fill mode'.

Treatment of border positions

During the execution of the moment and averaging method special care has to be taken regarding the start and end positions of the sequence. All windows applied by PROFILEGRAPH are centered around the position under consideration and application of a window of width $(2n + 1)$ leaves the peripheral n residues without a clearly defined environment. Two different ways of treating these border positions are used in PROFILEGRAPH; the decision is made by the value of the global flag 'padding mode'. If this parameter is set to 'false', the peripheral residues are not accounted for in the calculation and are not displayed; if it is set to 'true', an environment for the peripheral residues is simulated by filling in average values of the parameter table in use.

Moment methods

Use of a 'moment method' allows the execution of a special calculation with the amino acid specific parameter to be applied before averaging. Currently available moment methods include calculations based on the assumption of a particular type of secondary structure for the residue under consideration. The periodicity of the assumed secondary structure is specified by the inter-residue angle, which is $\sim 100°$ for an α-helix and $180°$ for a β-sheet. Calculation of moments, generalizing the approach of the hydrophobic moment, involves the application of a window width that is usually to be set to a value of two periods of the secondary structure. This window of $(2m + 1)$ residues, here called the moment window, is independent of the averaging window made of $(2n + 1)$ residues.

A brief discussion of the moment methods provided by PROFILEGRAPH and the respective algorithms follows. Here $V(i)$ denotes the original parameter value at sequence position i, L the sequence length, m the half-width of the moment window and ω the inter-residue angle.

Direct/no moment. No further calculation is applied.

Unnormalized moment. The absolute value of the vectorial sum of the single parameter values is calculated as $M_{2m+1,\omega}(i)$ and given by:

$$M_{2m+1,\omega}(i) = \frac{\left\{\left[\sum_{j=i-m}^{i+m} V(j)\,\sin(\omega(j-1))\right]^2 + \left[\sum_{j=i-m}^{i+m} V(j)\,\cos(\omega(j-i))\right]^2\right\}^{1/2}}{2m+1} \quad (1)$$

Normalized moment. Since the moment is meant as a measure of the distribution asymmetry of the specified parameter, it should be independent of the absolute values of the parameter itself. While the original moment calculation (Eisenberg *et al.*, 1984) is suitable only for normalized scales like the OMH-scale (Sweet and Eisenberg, 1983), a normalization of the table during the calculation makes this method generally applicable. In PROFILEGRAPH normalization of the parameter values is based on amino acid probabilities derived from the current sequence. The normalized moment $N_{2m+1,\omega}(i)$ is given by:

$$N_{2m+1,\omega}(i) = \frac{\left\{\left[\sum_{j=i-m}^{i+m} V(j)\,\sin(\omega(j-1))\right]^2 + \left[\sum_{j=i-m}^{i+m} V(j)\,\cos(\omega(j-i))\right]^2\right\}^{1/2}}{(2m+1)\left\{\sum_{i=1}^{L} V^2(i) - \frac{\left(\sum_{i=1}^{L} V(i)\right)^2}{L}\right\}^{1/2}} \quad (2)$$

Moment angle. In certain cases, not only the absolute value of the vectorial moment but also the angle relative to a direction of reference can be of interest. The angle $\Phi_{2m+1,\omega}(i)$ relative to the first residue is given by:

$$\Phi_{2m+1,\omega}(i) = \frac{\sum_{j=i-m}^{i+m} V(j)\,\sin(\omega(j-i))}{\sum_{j=i-m}^{i+m} V(j)\,\cos(\omega(j-i))} \quad (3)$$

Since the angle representation is periodic in nature, i.e. 0° = 360°, there is a problem in graphical representation of moment angles fluctuating around 0°, causing inappropriate discontinuities. Plotting the angles over an extended range of 0°−440° avoids some of the difficulties; this method is called the 'smooth moment angle'.

Face of α-helix. Another way of accounting for asymmetrical distribution of parameters given an assumed secondary structure is to consider only residues positioned on one face of the α-helix or the β-sheet. At the moment only the two methods proposed for calculating one-faced hydrophobicities (Jaehnig, 1989) are implemented in PROFILEGRAPH. The method 'face of α-helix' uses a fixed moment window of 19, representing the minimal requirements for helices spanning a lipid bilayer. $J_{19,\alpha}(i)$ is given by:

$$J_{19,\alpha}(i) = [h(i \pm 8) + h(i \pm 7) + 0.25 * h(i \pm 5) + h(i \pm 4) + 0.75*h(i \pm 3) + 0.5*h(i \pm 1) + h(i)]/10$$

Face of β-sheet. This method is equivalent to that used for the α-helix. The minimal requirement for a β-sheet spanning a lipid bilayer is 9 residues, so $J_{9,\beta}(i)$ is given by:

$$J_{9,\beta}(i) = \frac{[h(i \pm 4) + h(i \pm 2) + h(i)]}{5} \quad (5)$$

Averaging methods

After execution of one of the calculations described above it is usually desired to apply an averaging function to the data in order to separate the interesting features from the underlying noise.

PROFILEGRAPH can use several methods for averaging. In the following description $A_{2n+1}(i)$ denotes the averaged result at position i using an averaging window of half-width n, while $V(i)$ is the original value before averaging but after execution of the specified moment method.

Mean over window. This is the standard type of averaging in sliding-window analyses. A rectangular window is applied, resulting in

$$A_{2n+1}(i) = \frac{\sum_{j=-n}^{n} V(i+j)}{2n+1} \quad (6)$$

Hat over window. An improvement over rectangular windows, resulting in higher spatial resolution, has recently been described (Claverie and Daulmerie, 1991). Here a triangular window of the width $2n+1$ is applied, resulting in

$$A_{2n+1}(i) = \frac{\sum_{j=-n}^{n} V(i+j)\left(1 - \frac{|j|}{n+1}\right)}{n+1} \quad (7)$$

Median over window. The application of median-based averaging procedures and their advantages in pronouncing certain features have been discussed (Bangham, 1988). For calculating median values, PROFILEGRAPH uses an algorithm that takes benefit from the non-randomness of the original sequence of values. It involves a simple test for median-properties of specified data-point starting at the center of the window and then moving towards the borders until the true median has been found (Figure 1).

$$A_{2n+1}(i) = \mathrm{MED}_{i-n}^{i+n}(V) \quad (8)$$

The speed of this algorithm is of crucial importance for the next averaging method, which also involves median calculations.

Data-sieve. This averaging method consists of a cascade of median formations starting with small median windows and

```
FUNCTION MEDIAN ( data, mid_interval, halfwidth) ;

 left_border:=mid_interval - halfwidth;
 right_border:=mid_interval + halfwidth;
 i:=mid_interval;                                          (1)
 step:=0;                                                  (2)
 repeat                                                    (3)
  if (step mod 2 = 0) then i:=i+step else i:=i-step;
  step:=step+1;
  l_count:=0; g_count:=0;                                  (4)
  j:=left_border;
  suggestion:=data[i];
  repeat                                                   (5)
    if data[j] > suggestion then g_count:=g_count+1;       (6)
    else if data[j] < suggestion then l_count:=l_count+1;
    j:=j+1;
  until (g_count>halfwidth) or (l_count>halfwidth) or (j>right_border);
 until ((g_count≤halfwidth) and (l_count≤halfwidth)) or
(step>2*halfwidth);
MEDIAN := suggestion;                                      (7)
```

Fig. 1. Core routine for median calculation in PROFILEGRAPH. This algorithm performs best if the median is near the center of the interval. (**1**) 'i' points to the data value that is tested as the median. Start with the central point. (**2**) 'step' is the difference between the old median and the new one. (**3**) The outer repeat-loop tests all data values of the interval as median, starting in the middle and moving towards the borders. (**4**) 'l_count' is the number of data-points less than the suggested median, 'g_count' is the nubmer of greater data-points. (**5**) The inner repeat-loop performs the median-test. (**6**) Values greater or less the suggested value are counted. (**7**) If the inner loop is done and neither 'l_count' nor 'g_count' are greater than half of the window, the median has finally been found.

extending up to a specified mesh-size. A data-sieve of mesh size s consists of a cascade of medians over windows $2n+1$ with n ranging from 1 to s.

$$A_s(i) = \mathrm{MED}_{i-s}^{i+s}\,(\ldots\,(\mathrm{MED}_{i-3}^{i+3}\,(\mathrm{MED}_{i-1}^{i+1}))) \qquad (9)$$

The data-sieve tends to eliminate all features that are represented by less than s data points.

Fourier smoothing. This technique involves a Fourier transformation of the original data, followed by a low-pass filtering for noise-reduction and an inverse Fourier transformation. A value that influences the strength of low-pass filtering is formally equivalent to the width of the averaging window. The characteristics of this method are similar to that of triangular window smoothing.

Scaling and graphical output

The data obtained from the averaging procedure are then scaled to fit into a screen section. Depending on the user's choice, the data are either scaled to fill their respective screen section entirely, giving maximal vertical resolution or by using a standard scaling factor that is contained within the parameter table. The latter mode of scaling allows consistent comparisons between two or more sequences. The graphical output on screen

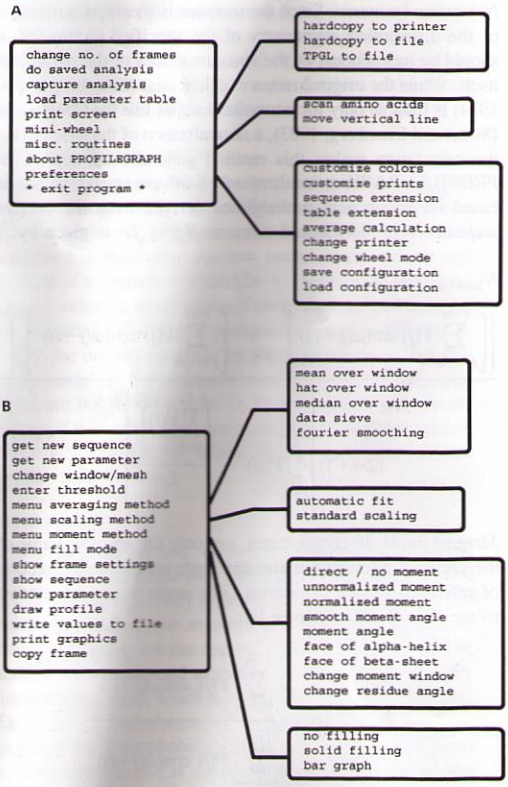

Fig. 2. Overview of the first two levels of PROFILEGRAPH's menu system. (**A**) The global menu manages features valid for the entire program, such as preferences, hardcopy output and routines that work on the whole screen. (**B**) The frame menu manages features valid for one screen section only. Among these are choice of modes for moment, averaging, scaling and display, as well as input of sequence, parameters and window-width.

or another device can be obtained either as a line-drawing, as a filled polygon or in a bar-graph style.

Implementation

The current implementation of PROFILEGRAPH for MS-DOS computers uses its own graphical user interface. A version for use with MS Windows 3.0 is in preparation.

System requirements

PROFILEGRAPH v. 1.3 can run on any MS-DOS computer equipped with a hard disk and sufficient memory. Exact memory requirements cannot be determined because the program uses largely dynamical memory allocation so the

amount of memory used depends on the graphics adapter. The current version v. 1.3 can use 640 K of memory; there is currently no EMS or XMS support.

Since PROFILEGRAPH uses Borland's BGI graphical device driver system, most standard graphics adapters are supported. A color-based system with at least moderately high resolution, like EGA or VGA, is recommended because the program makes extensive use of color codings. Color preferences can be chosen from a palette and saved in a configuration file for subsequent use, thus allowing monochrome, grayscale and color systems to use sensible color assignments. To work with the pop-up type menu system of PROFILEGRAPH, use of a pointing device like a mouse is highly recommended.

The speed of PROFILEGRAPH is sufficient for interactive modification of the parameters used. Calculation times for a single diagram display for a 300 residue protein on an Intel 80386SX-based computer range from <1 s for small rectangular windows to ~ 5 s for data-sieves with mesh $= 9$ or for moment calculations.

Menu-system

PROFILEGRAPH uses a graphical shell based on hierarchical organized pop-up menus that can be used either with a mouse or the keyboard. An overview of the first two menu levels is given in Figure 2. Two different root menus exist: the global menu comprises items and submenus valid for the execution of the whole program, while the frame menu affects only its associated screen region.

One essential feature of the global menu is the choice of program preferences like colors, default filenames, printer type, type of helical wheel plot, etc. The management of standard analysis files and graphical output to printers or files are also parts of the global menu. Standard analysis files are a means to store a set of predefined analyses in a disk file in order to apply the saved operations to other proteins afterwards. Exclusive use of preassembled standard analysis files is the easiest way to work with PROFILEGRAPH, and a set of example analyses is distributed with the program. Some interactive analysis routines are included in the global menu. It is possible to define a screen region with the mouse in order to obtain a helical-wheel diagram. There are also utilities facilitating the comparison of features in distant screen regions and for finding the amino acid associated with a specified feature in one of the curve diagrams.

Central parts of the frame menu are the choice of the sequence and the analysis parameters of the associated screen region. Utilities for display of the chosen sequence and parameter values are added to the frame menu.

Input and output compatibilities

The READSEQ routine used for sequence input supports several common sequence file formats, among them are the IG, Genbank, NBRF, EMBL, GCG, DNAstrider, Fitch, Pearson and NCR format. The PROFILEGRAPH configuration files and the parameter tables are ASCII files and can be written and modified by any standard editor.

Graphics output can be obtained either as a screen hardcopy or as a TPGL file. Supported printers for screen hardcopy are EPSON compatible 9-pin or 24-pin printers and HP-PCL using printers such as the Hewlett-Packard Deskjet or Laserjet series. TPGL files are ASCII files and can be converted into other vectorial or bitmap graphics formats by special TPGL-converters.

PROFILEGRAPH can create numerical output files in several user-defined formats for export to standard spreadsheet programs. Supported output formats include delimited as well as fixed-spaced ASCII files.

Availability

The current version of PROFILEGRAPH, v. 1.3, is available from the authors (KHOFMANN@BIOMED.BIOLAN.UNI-KOELN.DE) or from NETSERV@EMBL-HEIDELBERG.DE as a single file in compressed format. It is supplied with source code and documentation for the program and the menu system, and can be modified and recompiled using Turbo Pascal 5.5 or later. The program and all sources are completely public and may be freely redistributed, modified and used.

Discussion

The program PROFILEGRAPH is meant as a tool to help the biochemist in the interpretation of protein primary structure. The main goal of the program is to include the diversity of published sliding-window methods in one consistent and clearly defined shell. In doing this, the user has control over all parameters influencing the resulting plot, it is possible to combine any amino acid specific parameter with any moment method, any averaging method and any window width. This flexibility leads to the danger of inappropriate combinations. As an example, it is senseless to calculate moments of β-sheet propensities or to do averaging on values obtained from one-sided helix calculations. A high degree of flexibility in the use of a program is often paid for by a reduced simplicity in program operation. To abbreviate the tedious procedure of combining the desired parameters for every protein to be analyzed, the possibility of grouping frequently used parameter sets into analysis files has been introduced. These analysis files can be understood as a kind of 'macro-instruction'; they can also be used to simulate the output of other sliding-window sequence analysis programs. The result of the application of an example analysis grouping is shown in Figure 3 and serves to demonstrate some of PROFILEGRAPH's main features. The underlying amino acid sequence is that of the human brain myelin proteolipid protein (Stoffel *et al.*, 1985); the analysis applied examines the hydrophobic properties of the integral membrane protein in order to predict the membrane-spanning segments.

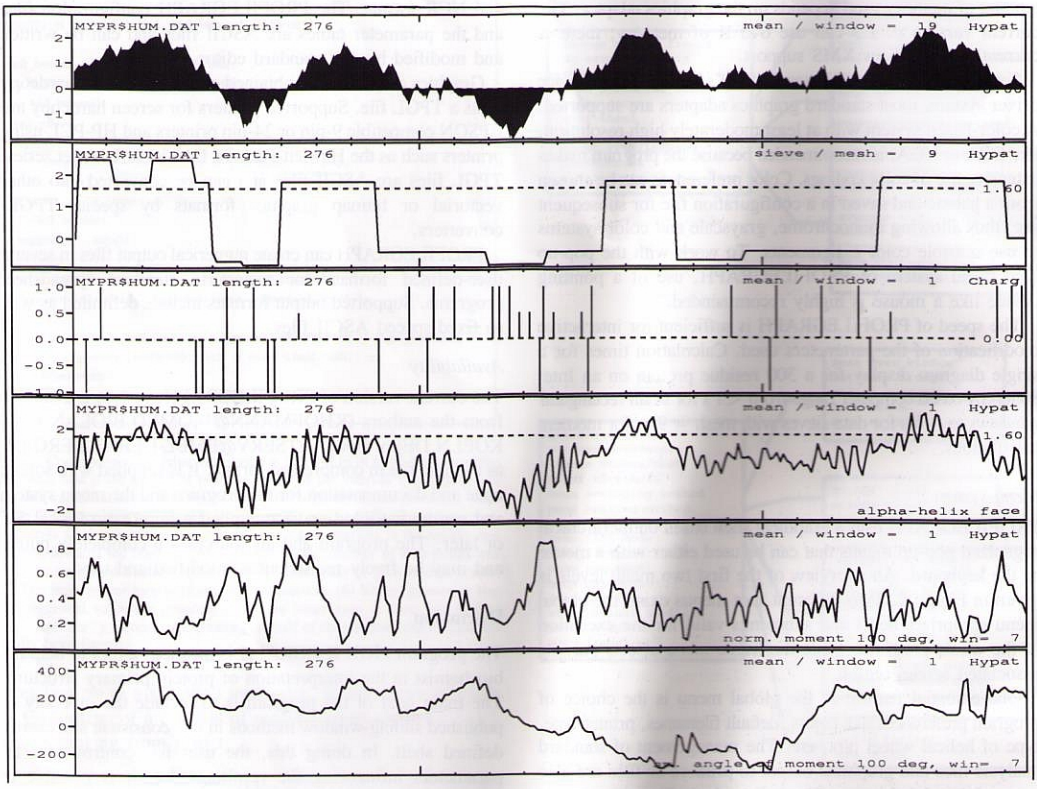

Fig. 3. Screen copy of a sample analysis with PROFILEGRAPH. The screen is divided into six frames, which in this example are all based on the sequence of the human CNS myelin proteolipid protein. Each frame contains in its upper left corner a short description of the underlying sequence, in its upper right corner the averaging method and the parameter in use are displayed. The lower right corner of each frame describes the moment method used. A dotted horizontal line marks a threshold value that can be specified.

In this example, the screen is divided into six frames, each containing the same protein sequence. All frames treat the sequence borders by padding with average values, and the scaling mode is set to optimal fit. In the first frame, the Kyte−Doolittle hydropathy is plotted using a rectangular window of 19 residues, displayed as a filled polygon train. The second frame shows the same parameter, but after application of a data-sieve with mesh size = 9. A comparison of these two frames demonstrates that by data-sieving, the borders between hydrophobic and hydrophilic segments are much more pronounced but not necessarily correct. The third frame shows charged amino acids, displayed in a 'bar-graph' style. Amino acids with positive charges point upwards, negative charges point down. It becomes evident that in the example protein there are four extended regions essentially free of charged amino acids that correspond with the hydrophobic domains of the Kyte−Doolittle plot. The fourth frame displays the hydropathy again, here calculated for one face of a 19 residue α-helix. A threshold value of 1.6 is marked by the dotted line. From a comparison with the hydropathy plot of the first frame it can be seen that there are no additional segments able to form α-helices with one side more hydrophobic than the threshold value. The fifth frame shows the hydrophobic moment of an α-helix of seven residues with no further averaging, while the sixth frame shows the hydrophobic moment angle under the same conditions. Evidently the proteolipid protein does not possess extended amphipathic helical regions. Several short stretches of ~10 amino acids—located at positions 1−10, 60−70, 95−105, 130−140 and at the extreme C terminus of the protein—show relatively strong amphipathic character if these are assumed to be α-helical. A subsequent helical-wheel analysis of the interesting regions after selecting them on screen allows further insight into the causes of the amphipathy. This example shows how PROFILEGRAPH can be helpful in performing a

hydrophobicity analysis. We hope that because of the flexibility and easy extendability of the program it will also be useful for other types of analysis.

References

Bangham,J.A. (1988) Data-sieving hydrophobicity plots. *Anal. Biochem.*, **174**, 142–145.

Bhaskaran,R. and Ponnuswamy,P.K. (1988) Positional flexibilities of amino acid residues in globular proteins. *Int. J. Pept. Prot. Res.*, **32**, 242–255.

Chothia,C. (1975) Structural invariants in protein folding. *Nature*, **254**, 304–308.

Claverie,J.M. and Daulmerie,C. (1991) Smoothing profiles with sliding windows: better to wear a hat! *Comput. Applic. Biosci.*, **7**, 113–115.

Eisenberg,D. (1984) Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.*, **53**, 595–623.

Eisenberg,D., Schwarz,E., Komaromy,M. and Wall,R. (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.*, **179**, 125–142.

Hopp,T.P. and Woods,K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA*, **78**, 3824–3828.

Jaehnig,F. (1989) Structure predictions for membrane proteins. In Fasman,G.D. (ed.), *Prediction of Protein Structure and the Principles of Protein Conformation*. Plenum Press, New York, pp. 707–717.

Karplus,P.A. and Schulz,G.E. (1985) Prediction of chain flexibility in proteins. *Naturwissenschaften*, **72**, 212–213.

Kyte,J. and Dootlittle,R.F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.*, **157**, 105–132.

Levitt,M. (1978) Conformational preferences of amino acids in globular proteins. *Biochemistry*, **17**, 4277–4285.

Nakai,K., Kidera,A. and Kanehisa,M. (1988) Cluster analysis of amino acid indices for prediction of protein structure and function. *Prot. Engng.*, **2**, 93–100.

Parker,J.M.R., Guo,D. and Hodges,R.S. (1986) New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, **25**, 5425–5431.

Ponnuswamy,P.K., Prabhakaran,M. and Manavalan,P. (1980) Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. *Biochim. Biophys. Acta*, **623**, 301–316.

Ragone,R., Facchiano,F., Facchiano,A., Facchiano,A.M. and Colonna,G. (1989) Flexibility plot of proteins. *Prot. Engng.*, **2**, 497–504.

Rose,G.D., Gierasch,L.M. and Smith,J.A. (1985) Turns in peptides and proteins. *Adv. Prot. Chem.*, **37**, 1–109.

Schiffer,M. and Edmundson,A.B. (1967) Use of helical wheels to represent the structures of protein and to identify segments with helical potential. *Biophys. J.*, **7**, 121–135.

Stoffel,W., Giersiefen,H., Hillen,H., Schroeder,W. and Tunggal,B. (1985) Amino acid sequence of human and bovine brain myelin proteolipid protein (lipophilin) is completely conserved. *Biol. Chem. Hoppe-Seyler*, **366**, 627–635.

Sweet,R.M. and Eisenberg,D. (1983) Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *J. Mol. Biol.*, **171**, 479–488.

Received on August 12, 1991; accepted on January 3, 1992